Notiziario Tecnico

Archivio

1/2002

TIM

NOTIZIARIO TECNICO TELECOM ITALIA

Notiziario Tecnico Telecom Italia
Anno 11 - n. 1 - Giugno 2002

Editore
Telecom Italia S.p.A.

Direttore responsabile
Rocco Casale

Comitato di direzione
Claudio Carrelli, Gianfranco Ciccarella, Gabriele Falciasecca, Claudio Gentile, Andrea Granelli, Franco Pattini, Stefano Pileri, Aldo Roveri, Roberto Saracco, Mauro Sentinelli, Carlo Giacomo Someda, Giuseppe Tilia, Francesco Vatalaro

Segreteria tecnica
Andrea Baiocchi

Segreteria di redazione
Francesca Romana Belgiovine

Consulenza Redazionale
Adriano Santelli

Progetto grafico e copertina
Modo Comunicazione S.r.l.

Grafica e impaginazione
Carlo Guerra,
Modo Comunicazione S.r.l.

Fotografie
Guido Bruno, Compaq, Graphic Design, Lockheed Martin Space Systems, Paolo Impiglia, Nikon Instruments, Sirti, Telecom Italia, Tesmec / Alpitel, TILAB.

Stampa
C.S.C. Grafica S.r.l.
Via G. G. Arrivabene, 40
00159 Roma

Registrazione
Periodico iscritto al n. 00322/92 del Registro della Stampa presso il Tribunale di Roma in data 20/05/92

Direzione e Redazione
Via di Val Cannuta, 250 - 00166 Roma
tel. +39+06+3688-3801
fax +39+06+6633035
e-mail: notiziario.redazione@telecomitalia.it

Il personale di Telecom Italia può disporre della rivista in formato elettronico all'indirizzo 10.10.18.169/Notiziario/Tecnico

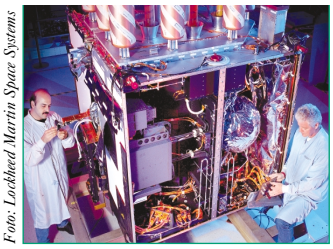


Foto: Lockheed Martin Space Systems

Realizzazione di un satellite GPS (Global Positioning System)

SOMMARIO

Aspettando Godot

Innumerevoli sono stati, da parte della critica, i tentativi d'interpretazione del famoso capolavoro di Samuel Beckett. L'opera narra l'attesa di Vladimir ed Estragon, due viandanti che si trovano in un luogo indefinito (sappiamo soltanto che accanto a loro c'è un salice piangente, simbolo di tutto e di niente....) ad aspettare un certo Godot, che dovrebbe offrirgli un lavoro.

È interessante notare come l'opera, definita il *dramma della non comunicazione*, si presti oggi a rappresentare la situazione del mercato delle comunicazioni, o meglio delle telecomunicazioni e dell'ICT nel suo complesso. Che cosa sta succedendo nel comparto ICT, o meglio, che cosa stiamo aspettando dopo la vivace turbolenza degli ultimi due anni?

I due viandanti potrebbero immedesimarsi nella larga banda e nel mobile, con il salice a rappresentare il processo di regolamentazione.

Il percorso evolutivo delle Telecomunicazioni passa inevitabilmente attraverso l'intreccio della tecnologia, del mercato e delle regole. E nel cerchio che racchiude queste tre componenti fondamentali s'inserisce il ruolo dell'azionista e del suo imprevedibile, ma pur condizionante, comportamento.

Mai come in questo momento l'intreccio tra le tre variabili è stato così forte, e mai come ora ciascun settore tende a far ricadere le responsabilità della crisi soprattutto sugli altri due. Le responsabilità sono ribaltate, nell'attesa di .... un magico Godot, che ci porti verso un futuro più roseo.

La prospettiva di continue nuove tecnologie disorienta il consumatore, sostengono i conservatori; no, è l'eccesso di regolamentazione che disturba il mercato, dichiarano per contro gli innovatori; il mercato è disorientato dalla potenzialità delle nuove tecnologie e dalle politiche di pricing che esse comportano, ribadiscono gli analisti.

Siamo in presenza comunque di almeno tre punti fermi: anzitutto il mercato del mobile è vicino alla saturazione, e la sua crescita deve essere necessariamente rivolta verso nuovi servizi, più che verso nuovi clienti.

Il mercato della larga banda, sia fisso, che mobile, è in secondo luogo, inequivocabilmente, nella fase infantile, e deve essere stimolato per un suo rapido decollo.

In assenza di una domanda sostenuta, infine, l'azione della regolamentazione può addirittura risultare frenante, piuttosto che trainante.

A quest'analisi va anche aggiunto che il processo di evoluzione tecnologica non può essere visto in maniera dogmatica, ma che le nuove tecnologie devono necessariamente integrarsi con l'esistente in un processo continuo di sviluppo.

È storia ben nota come la radio non abbia sostituito i giornali, o come la televisione non abbia a sua volta soppiantato la radio. Tuttavia il gusto per l'attesa di un'altra dirompente novità è sempre presente, sia pure silenziosamente, e sembra oggi dominare più che mai lo scenario delle telecomunicazioni.

Affermazioni come *"il PC è morto"*, *"protocollo IP e fibra dappertutto"*, *"la vita comincia a 100 Mbit/s"*, *"Wi-Fi verso UMTS"*, *"ultra wide-band verso blue-tooth"*, riempiono le cronache quotidiane delle telecomunicazioni, che sembrano essere alla ricerca di un prossimo evento straordinario, della prossima "panacea", come a suo tempo lo furono l'avvento del numerico, del mobile, o di Internet.

È inutile ormai recriminare sulle esorbitanti cifre pagate per le licenze UMTS o sulle avventurose azioni di *mergers and acquisitions* intraprese con dubbio successo a livello nazionale o internazionale.

Sono gli stessi analisti, che non più di due anni fa spingevano vigorosamente in tal senso, che ora invitano ad un sostanziale ridimensionamento. È vero, infatti, che il settore è oggi in fase di consolidamento, ma è altrettanto vero che la necessità di comunicare, specialmente con le nuove prospettive che sono annunciate e/o offerte, presenta ancora enormi potenzialità ed è suscettibile di un enorme sviluppo.

La competizione si sta affermando anche in Europa; i regolatori non nascondono, tuttavia, la loro perplessità per un ruolo ancora determinante ricoperto dagli incumbent.

L'attenzione è oggi soprattutto rivolta all'*ADSL*, come precursore della larga banda sulla rete fissa, ed alle wireless Lan, che accrescono le prospettive di mobilità, lasciando anche intravedere ulteriori promettenti sviluppi nelle applicazioni di domotica e di infomobilità.

Analizzare la situazione nei singoli Paesi del mondo risulterebbe troppo complesso; un caso interessante, anche se certamente emblematico, è, però, rappresentato dalla Germania. A seguito di una crescita a due cifre riscontrata nel fatturato del comparto ICT nello scorso anno, il rapporto dell'Authority apre in prima pagina con la seguente frase: *"Die Lage.. ist besser als die Stimmung"* e cioè la "situazione" è migliore della "sensazione" (seguono quindi i dati di crescita in Germania e la loro analisi di dettaglio). Per contro il valore delle azioni di DT, così come di numerosi altri operatori, è ben lontano dai picchi raggiunti al momento della bolla Internet.

Se guardiamo più da vicino la situazione del mercato tedesco, si nota come il traffico *dial-up Internet* abbia già superato il traffico telefonico locale, e come la diffusione dell'ADSL abbia già abbondantemente oltrepassato la soglia dei due milioni di clienti. La diffusione dell'ISDN ha già raggiunto i 20 milioni di accessi base equivalenti, ed il successo dell'unbundling del local loop è testimoniato da oltre 700mila accessi. Si tratta, certamente, di grandi numeri, ma è interessante anche sottolineare come questi valori, per la sola Germania, siano ben superiori alla somma dei valori corrispondenti di tutti gli altri quattordici Stati dell'Unione messi insieme.

Sarebbe troppo facile estrapolare tali dati, sostenendo che questa sia "la via" da seguire: che L'ADSL sia l'unica soluzione vincente, o che l'unbundling risolva tutti i problemi di competizione. La realtà in altri Paesi è profondamente diversa, come è testimoniato, ad esempio, dalla nuova spinta verso la fibra negli Stati Uniti, accompagnata peraltro da un sostanziale fallimento dell'unbundling del local loop, inventato, tra l'altro, proprio negli USA circa 20 anni fa.

Mobilità e larga banda rappresentano certamente il futuro, ma non devono restare in attesa come i due viandanti dell'opera di Beckett. Sono personalmente convinto che non si possa correre il rischio di aspettare il mitico Godot, ma che sia opportuno affrontare la situazione attuale pur con le sue effettive difficoltà, ma anche con altrettante chiare prospettive di una rapida ripresa. L'attuale incertezza dei mercati non deve frenare gli entusiasmi, ma dovrà piuttosto costituire uno stimolo all'innovazione.

Il sistema è ancora forte e motivato e l'integrazione, da un lato dei progressi dell'Information Technology, e dall'altro delle più avanzate tecnologie di telecomunicazione, permetterà senza dubbio di superare le criticità del presente, per imboccare la porta di un futuro sviluppo.

È assolutamente necessario concentrarsi e collaborare per la creazione di nuovi servizi,resi possibili anche dalle nuove tecnologie, ma soprattutto occorre comprendere e soddisfare le richieste della clientela, anche se ancora inespresse, e, al tempo stesso, insistere nella ricerca delle più appropriate politiche di pricing.

Le tecnologie di accesso, sia wireline sia wireless, e le problematiche legate ai contenuti (tra cui le tipologie, gli aspetti di proprietà intellettuale, i criteri di tariffazione, ...) costituiscono oggi i filoni principali da battere ed è su questi obiettivi che vanno concentrati gli sforzi di ricerca e sviluppo.

La collaborazione tra gli operatori sarà sempre più fondamentale, e costituirà sicuramente una chiave vincente per la ripresa della crescita a due cifre nel prossimo futuro. Essa dovrà essere indirizzata non solo a garantire una continua e completa interoperabilità dei servizi, ma anche a minimizzare i rischi e a condividere risorse umane e finanziarie, per un comune sforzo all'insegna dell'innovazione e del progresso.

Claudio Carrelli

Direttore EURESCOM

Sicurezza per le aziende in rete

Maurizio Dècina
Vittorio Trecordi

Il tema della sicurezza è destinato a rivestire un ruolo crescente via via che le piattaforme ICT abbassano le barriere per agevolare le relazioni di business. L'applicazione della tecnologia nei settori business critical non può prescindere dal governo della sicurezza per ragioni che, pur essendo evidenti, sono state finora largamente sottovalutate. Il lavoro vuole offrire una panoramica sullo stato dell'arte della sicurezza per le aziende in rete, privilegiando la completezza rispetto all'approfondimento di singoli aspetti.

1. Introduzione

La pervasiva adozione delle piattaforme tecnologiche legate al trattamento dell'informazione e all'esecuzione delle transazioni, a sostegno dell'operatività e dello sviluppo delle relazioni con il mondo esterno, ha determinato uno scenario in cui, ai benefici di un'efficiente ed efficace esecuzione delle operazioni di business, si accompagna una valenza di segno opposto: la riduzione delle barriere di accesso alle informazioni e ai servizi in rete e, più in generale, all'utilizzo di potenti strumenti di trattamento dei dati sensibili, accresce, infatti, l'esposizione e, quindi, la vulnerabilità rispetto a eventi, accidentali o causati da azioni eseguite con finalità illecite, che compromettono la sicurezza delle aziende.

In questo articolo sono esaminate le problematiche che riguardano la sicurezza delle aziende in rete, in relazione alle azioni di attacco eseguite da soggetti animati dal proposito di violarla. Sarà omesso, invece, lo sviluppo delle tematiche, riferite generalmente ai problemi della sicurezza, della protezione dagli effetti di errori o di eventi accidentali - inclusi i guasti o i disastri - e del loro impatto sull'affidabilità dei sistemi e sulla normale operatività delle aziende *(business continuity)*.

Le minacce possono avere origine all'interno o all'esterno del dominio di un'organizzazione (azienda o amministrazione pubblica).

Indagini recenti mostrano una lieve prevalenza di azioni illecite provenienti dall'esterno rispetto a quelle che nascono dall'interno. La rete Internet, offrendo un'agile via di contatto tra le centinaia di milioni di utenti a essa collegati, amplia enormemente il fronte di esposizione esterno. Posto che sia bene identificato il perimetro del dominio e che siano ben individuati i punti di interfacciamento con l'esterno, le azioni volte a elevare il livello di protezione o a rilevare attacchi si possono concentrare su pochi punti focali.

Sul versante interno alle organizzazioni, le zone di attenzione della sicurezza sono più sfumate e diffuse. Dipendenti infedeli, ex-dipendenti e collaboratori che hanno accesso al dominio aziendale rappresentano una minaccia particolarmente insidiosa.

Se è vero che all'interno del dominio si ha la titolarità di imporre regole e di verificarne l'applicazione, l'adozione di misure di sicurezza non deve essere di eccessivo intralcio all'operatività, ma piuttosto deve costituire il risultato di un delicato bilanciamento tra usabilità e sicurezza, che in ogni caso porta a disporre barriere di protezione più vulnerabili rispetto a quelle impiegate al confine del dominio.

I soggetti che perpetrano le violazioni, ossia coloro che attuano gli attacchi, sono spinti da moventi di varia natura: la frode e il furto di informazioni, ad esempio per finalità di spionaggio industriale; il danneggiamento per scopi dimostrativi, politici, vandalici o concorrenziali; l'accesso indebito a sistemi e a informazioni.

La figura dell'*hacker* viene solitamente indicata come attore delle azioni di attacco senza distinguere tra le varie sfumature che caratterizzano i diversi attaccanti. L'*hacker* è un soggetto che trova la propria

soddisfazione nella sfida della violazione dei sistemi di protezione più sofisticati. Figure di questo tipo trovano più spazio nell'immaginario collettivo che non nella realtà, nella quale sono, invece, sempre più frequenti gli attacchi intenzionalmente eseguiti da specialisti animati da interessi materiali: spionaggio industriale (ad esempio, l'asportazione indebita di copie delle banche dati aziendali, attuata da dipendenti infedeli o da impiegati che cambiano azienda) o le frodi (ad esempio, l'intercettazione dei numeri di carta di credito usati per i pagamenti *on-line*).

Sempre più spesso si verificano situazioni di attacco che prendono spunto da moventi di carattere politico. Si pensi, ad esempio, ai fenomeni di attacco alle multinazionali e agli organismi istituzionali in corrispondenza degli incontri del G8, o a moventi di carattere industriale, come gli attacchi volti a impossessarsi del codice sorgente delle applicazioni di Microsoft.

L'impiego delle piattaforme tecnologiche comporta la creazione di identità cibernetiche, caratterizzate da identificativi e da credenziali che abilitano la partecipazione ai flussi telematici. La necessità di legare strettamente l'identità reale di un soggetto alla sua identità cibernetica rappresenta un elemento essenziale per raggiungere due obiettivi di rilievo: associare a ogni utente un profilo che esplicita le *autorizzazioni* a svolgere determinate azioni (accesso a risorse e a servizi), e, al tempo stesso, permettere di attribuire in modo certo all'utente le azioni attuate con l'impiego degli strumenti telematici *(accountability)*.

L'operazione di *autenticazione* di un utente a un sistema telematico ha lo scopo di verificare la corrispondenza tra l'identità dell'utente e la corrispondente identità cibernetica, sulla base delle credenziali opposte al sistema. La personificazione di un utente, utilizzandone in modo illecito l'identità telematica, è una via praticata quando si desideri violare le regole di autorizzazione, raggiungendo così la condizione di avvalersi dei permessi propri di un utente legittimo.

L'applicazione di misure volte a sostenere le garanzie di protezione dagli effetti delle azioni illecite è una priorità riconosciuta dai manager di azienda, dagli amministratori pubblici e dall'opinione pubblica. I tragici eventi del settembre dello scorso anno hanno aumentato la consapevolezza della criticità della sicurezza, nei riguardi degli effetti sia privati che pubblici, accrescendo significativamente il volume di risorse tecnologiche, umane e organizzative, indirizzate ad aspetti legati alla sicurezza.

I governi stanno destinando investimenti rilevanti ai temi della sicurezza pubblica. I vertici delle aziende hanno incluso tra le priorità delle organizzazioni lo sviluppo di una strategia e di un piano sistematico per garantire l'operatività del business e la sicurezza delle persone e dei valori aziendali, raggiungendo un profilo di rischio residuo compatibile con la *mission* aziendale.

L'accresciuta sensibilità induce ad affrontare, con uno spirito più genuino e motivato, le tematiche di sicurezza che in passato sono state fronteggiate con un approccio frammentato e occasionale, pervenendo così a un assetto fragile. La frammentarietà porta tipicamente a raggiungere condizioni di vulnerabilità riconducibili alla fragilità degli anelli deboli della catena di un sistema di accesso alle informazioni (il livello di sicurezza di un sistema *end-to-end* è commisurato al grado di sicurezza della tratta più vulnerabile della catena) o ai servizi o alla vulnerabilità dei versanti non adeguatamente presidiati di un dominio (il livello di sicurezza di un sistema è adeguato, infatti, al grado di sicurezza del punto più debole del sistema stesso).

Si sta affermando anche la percezione che la sicurezza non sia legata solo alle tecnologie, che cambiano lo scenario in cui essa si applica e che offrono, d'altra parte, strumenti per attuare interventi di protezione, ma che essa sia piuttosto un processo da eseguire con continuità nel tempo. La salvaguardia dei requisiti di sicurezza va perseguita attraverso l'applicazione di interventi di carattere organizzativo, fisico e logico che non possono essere affidati a interventi *una tantum*, ma che debbono seguire una pratica programmata, gestita e continua [1, 2, 3, 4, 5].

2. Incidenti di sicurezza

Un'azione di attacco che raggiunga i propri scopi, sfruttando le vulnerabilità di un sistema di protezione, produce un incidente di sicurezza, ossia configura una situazione in cui un attaccante, utilizzando strumenti che gli consentono di sfruttare le vulnerabilità del sistema, svolge azioni volte a compromettere una risorsa obiettivo, conseguendo il risultato di poter eseguire un'azione indebita. Le vulnerabilità e il loro sfruttamento seguono un ciclo di vita, quale quello rappresentato in figura 1.

Le nuove vulnerabilità sono scoperte da intrusori con notevoli capacità tecniche, che, dopo aver sfrut-

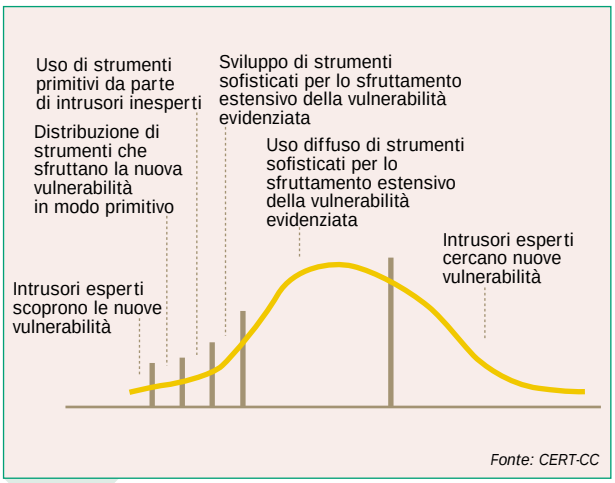


Figura 1 *Ciclo di vita di una vulnerabilità.*

tato il varco nei sistemi, normalmente rendono disponibili in rete Internet strumenti grezzi che consentono anche a intrusori meno esperti di sferrare attacchi. Sono poi sviluppati e resi disponibili in rete strumenti che consentono di verificare, con un test sui

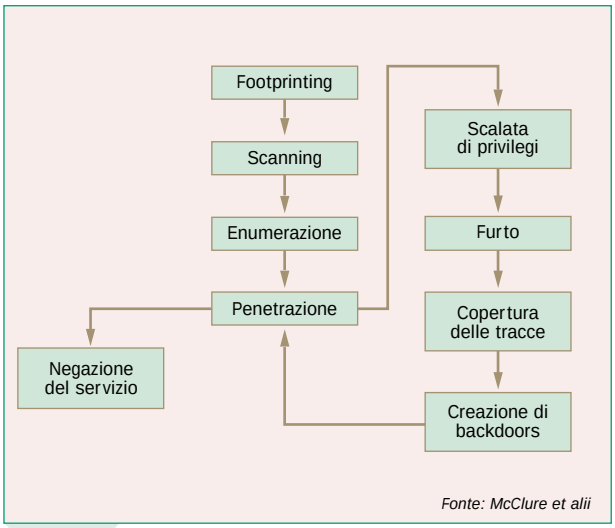


Figura 2 *Anatomia di un attacco.*

sistemi, la presenza della vulnerabilità, eventualmente assieme alle funzioni che la sfruttano, qualora essa sia presente. Alla diffusione generale degli strumenti di verifica e sfruttamento delle vulnerabilità segue generalmente l'attuazione massiccia delle azioni correttive degli amministratori di sistema e quindi il decadimento dello sfruttamento della vulnerabilità. In questa fase gli intrusori esperti si concentrano sulla ricerca di nuove vulnerabilità.

Un attacco alla sicurezza di un sistema in rete, finalizzato ad accedere a macchine protette da un sistema di difesa perimetrale, segue generalmente uno schema ben definito (figura 2), la cui conoscenza è un elemento utile per chi deve preparare la strategia di difesa [6]. La prima fase, denominata *footprinting* nel gergo degli addetti ai lavori, ha l'obiettivo di recuperare informazioni sulla rete obiettivo, usando strumenti e fonti di informazioni pubbliche, come *DNS (Domain Naming System)* e motori di ricerca. L'informazione più rilevante, che normalmente è ricavata in questa fase, è la conoscenza del blocco di indirizzi impiegato dalla rete obiettivo.

Segue una fase di *scanning*, o di scansione, in cui l'aspirante intrusore cerca di scoprire quali sistemi operativi sono in uso e quali servizi sono erogati (ad esempio posta elettronica o WWW). In questa fase si impiegano strumenti automatici che sviluppano una scansione esaustiva delle possibilità fornendo le informazioni cercate in tempi ridotti (lo strumento più utilizzato in questa fase si chiama *Nmap*).

La fase successiva, detta di *enumerazione*, cerca di recuperare informazioni maggiormente dettagliate in termini di risorse condivise, utenti configurati e privilegi di accesso alle risorse, programmi installati e relative versioni. Anche in questo caso si usano strumenti disponibili per i diversi sistemi operativi (ad esempio *finger* per i sistemi Unix).

Sulla base delle informazioni raccolte si cerca di sfruttare le vulnerabilità note per i sistemi installati allo scopo di *rendere il servizio indisponibile* agli utenti legittimi (attacco di *denial of service*), o per penetrare abusivamente nei sistemi, appropriandosi di privilegi che spetterebbero solo a utenti legittimi e sottraendo informazioni riservate o eseguendo transazioni indebite. L'intrusore si preoccupa quindi di coprire le tracce dell'attacco, cancellando i *file* usati strumentalmente e modificando i parametri che potrebbero insospettire gli amministratori delle macchine compromesse.

Non è infrequente il caso in cui l'incursore installi un programma di *backdoor*, che può servire per facilitare intrusioni successive, aggirando alcune difese superate con fatica all'atto della prima intrusione.

I programmi che fungono da *backdoor* sono mimetizzati in modo da renderli difficilmente rivelabili con semplici indagini.

Il rilevamento di un attacco di sicurezza sollecita l'attivazione di una procedura di risposta all'incidente.

La figura 3 mostra in maniera sintetica ma con molta efficacia, le diverse fasi del fenomeno ora descritto, fornendo oltretutto esemplificazioni di ciascun elemento che interessa un incidente di sicurezza [6]. Le frecce trasversali forniscono una rappresentazione del fatto che sul tema della sicurezza si fronteggiano due categorie di attori: i *black-hat* incarnano le figure degli esperti di sicurezza che mettono le proprie competenze al servizio dell'attuazione di azioni illecite. I *white-hat* sono invece gli esperti di sicurezza, schierati sul fronte della difesa.

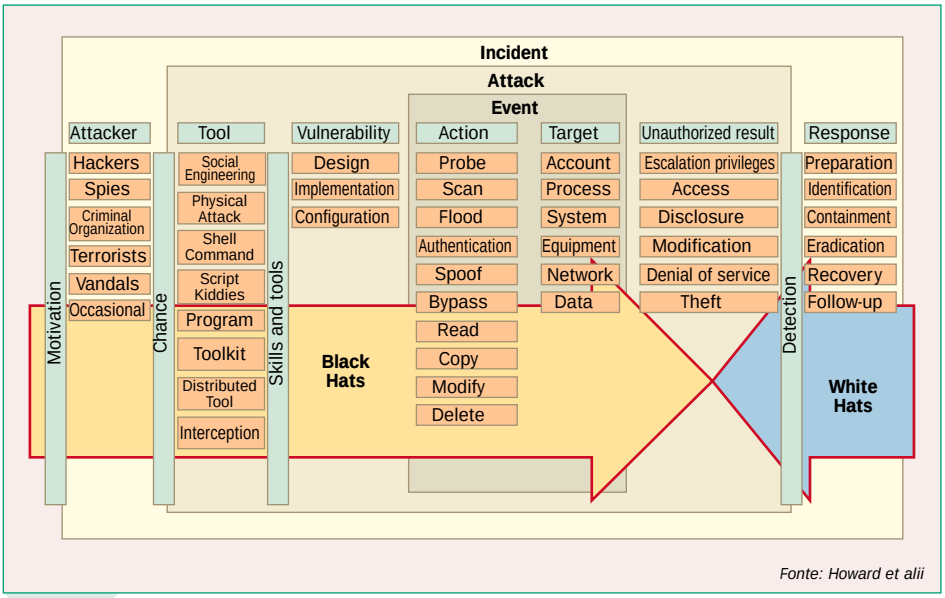


Figura 3 *Caratterizzazione di un incidente di sicurezza.*

Gli attaccanti possono essere ricondotti a una tra le seguenti categorie: gli *hacker*, le spie, le organizzazioni criminali, i terroristi, i vandali, i dipendenti (o gli ex-dipendenti) infedeli o gli attaccanti occasionali.

Tra gli strumenti impiegati per sferrare un attacco, nella figura sono rappresentati:

- il *social engineering* (ossia le pratiche sociali volte a impadronirsi di informazioni utili);
- l'attacco alle risorse fisiche (la violazione delle barriere di accesso ai locali o la manomissione degli apparati);
- gli strumenti atti a fornire la possibilità di aprire una console da cui eseguire comandi illeciti sui sistemi;
- i programmi riconducibili alla categoria degli strumenti per verificare l'esistenza di vulnerabilità e per sfruttarle;
- gli strumenti di intercettazione dei flussi di dati.

Le vulnerabilità sono classificate in base alla loro origine: quelle imputabili a una falla nel progetto del sistema e quelle sorte nella fase di implementazione o in quella di configurazione di un sistema (a causa, eventualmente, di negligenze o di errori).

Le azioni elementari di attacco comprendono: il *probing* e lo *scan* (le azioni volte a identificare la natura dei sistemi e l'esistenza di vulnerabilità); il *flood* (le azioni volte a sovraccaricare i sistemi con flussi di traffico o con transazioni per comprometterne il regolare funzionamento); l'autenticazione indebita; lo *spoofing* (ossia la generazione di traffico e di transazioni che simulano comportamenti particolari allo scopo di ingannare i sistemi); il *bypass* (ossia l'aggiramento dei sistemi di difesa) e, infine, tutte le azioni di *trattamento dei dati* a scopi illeciti (lettura, copiatura, modifica e distruzione).

Gli obiettivi verso cui sono indirizzate le azioni elementari di attacco sono le risorse sensibili: l'*account* di un utente, i processi, i sistemi, i dispositivi, le reti e i dati.

Tra i possibili effetti di un attacco riuscito sono, invece, compresi: l'*escalation* dei privilegi di un utente, l'accesso a risorse pregiate, l'accesso a informazioni riservate, la modifica di informazioni, la compromissione della possibilità di erogare un servizio e il furto.

Le attività di risposta all'incidente comprendono: le attività preparatorie (ad esempio la predisposizione di ruoli, di responsabilità e di procedure di risposta codificate); l'identificazione (il riconoscimento che determinati segnali corrispondono a un attacco); il contenimento (le azioni volte a confinare l'attaccante a una porzione del sistema, impedendo che si propaghi altrove); l'estirpazione (le azioni volte a sradicare l'attaccante dalle risorse compromesse); il *recovery* (ossia il ripristino delle condizioni normali); il *follow-up* (ossia le azioni volte a risalire all'autore dell'attacco attraverso l'analisi delle tracce).

La rete Internet offre uno strumento straordinario di diffusione delle informazioni e di per sé non ha una finalità positiva o negativa. La rete Internet mette, però, a disposizione un'enorme banca dati in cui è agevole reperire informazioni e strumenti utilizzabili per perpetrare attacchi alla sicurezza.

Infatti, Internet è anzitutto utilizzata dai fornitori di tecnologia e dai centri di competenza per rispondere agli incidenti, divulgandone le vulnerabilità o gli attacchi che emergono dall'operatività dei sistemi in esercizio. Questi attori intendono, così, raggiungere - fornendo un'informazione tempestiva - tutti gli amministratori di sistema per aggiornarli, per metterli in guardia e per offrire le indicazioni per attuare le contromisure.

Queste stesse conoscenze possono essere sfruttate per fini diametralmente opposti dai *black-hat*, e cioè dai soggetti che sfruttano la conoscenza per attuare i propri attacchi verso i sistemi che non hanno posto rimedio alle vulnerabilità pubblicate. Non sempre, infatti, i rimedi sono applicati con tempestività, poiché le scarse risorse dedicate all'amministrazione dei sistemi sono impegnate nell'operatività o nella correzione degli errori di configurazione [7]. Si è acceso di recente un dibattito tra i maggiori protagonisti del mercato dell'IT con posizioni difformi sulla necessità di provvedere a mutare questa pratica, limitandosi a comunicare le informazioni sulle vulnerabilità di sicurezza solo a soggetti selezionati e in modo riservato (Microsoft è il principale fautore di questa nuova strategia).

La propagazione dell'informazione sulla sicurezza attraverso la rete è anche soggetta a falsi allarmi che, amplificandosi in modo esponenziale con il credito che viene ingenuamente attribuito da una massa di utenti, finiscono per generare traffico in rete e per distogliere l'attenzione degli amministratori dei sistemi.

Questi falsi allarmi sono classificati in due categorie: *hoaxes*, ossia la propagazione in rete di notizie false sulla pericolosità di messaggi di *e-mail* con un determinato oggetto (come, ad esempio, il *good times hoax* nel 1994); *hypes*, ossia la propagazione sui media di notizie false allarmistiche sull'imminente avvento di virus catastrofici (come ad esempio il *Michelangelo hype* del 1992).

La generazione di falsi allarmi è spesso causata da goliardi burloni; ma è stata avanzata, anche, l'ipotesi che in qualche caso potesse esserci l'interesse dei fornitori di software antivirus.

La politica del software *open source* - ossia con un codice sorgente accessibile gratuitamente a chi voglia realizzare sistemi o sviluppare nuove funzionalità - ha implicazioni notevoli dal punto di vista della sicurezza. L'accesso ai sorgenti del software permesso a una vasta comunità di specialisti offre un livello di trasparenza massimo sulle caratteristiche del codice, consentendo di identificare e di eliminare possibili varchi per attacchi di sicurezza.

Il software *open source* persegue una strategia opposta a quella della *security through obscurity* che affida le garanzie di sicurezza all'assenza di conoscenza sulle caratteristiche dei sistemi. Un esempio dell'efficacia di questa politica è rappresentato dall'applicazione server Web, nota come *Apache*, notoriamente molto meno soggetta a attacchi di sicurezza rispetto ai prodotti commerciali con funzionalità analoghe (il caso più eclatante è quello dell'*Internet Information Server* di Microsoft).

L'osservazione delle tendenze in atto ci porta a approfondire due fenomeni interessanti che ci aiutano a valutarne portata e dinamiche.

LA SICUREZZA NON È SOLO UN PROBLEMA TECNOLOGICO

La diffusione pervasiva delle tecnologie informatiche e delle reti rende agevole lo sviluppo di relazioni complesse tra soggetti, che appartengono alla medesima organizzazione e interagiscono in modo stabile, così come tra soggetti che non si conoscono e che interagiscono in modo occasionale. Tutte queste relazioni sono caratterizzate da requisiti di sicurezza (confidenzialità, integrità, non ripudiabilità, controllo degli accessi, ...), commisurati alle esigenze dettate dal contesto applicativo. Le moderne piattaforme tecnologiche sono progettate in modo da indirizzare in larghissima parte le esigenze applicative in termini di sicurezza. Nonostante la tecnologia sia all'altezza delle esigenze, il problema della sicurezza aziendale rimane un punto di attenzione per tutte le organizzazioni che si appoggiano alle piattaforme telematiche.

Per raggiungere gli obiettivi di sicurezza attesi è, infatti, necessario che si verifichino alcune circostanze: la tecnologia deve essere correttamente calata nelle realtà applicative, gli attori che la impiegano debbono operare in base a regole ben definite (politiche e procedure), armati della consapevolezza/preparazione e della volontà di agire in modo diligente ed etico.

Le condizioni necessarie non sono, tuttavia, sufficienti in quanto le imperfezioni dei prodotti, la negligenza e gli errori di configurazione/impiego dei sistemi e le mutazioni continue dell'assetto dei sistemi aprono numerosi varchi, che possono essere sfruttati per compromettere la sicurezza aziendale.

Peraltro, la realizzazione di progetti di sicurezza deve superare ostacoli importanti: innanzitutto i costi imputabili alla sicurezza non sono associabili a ricavi, ma possono essere giustificati da un'azione di riduzione dei rischi riferibili allo sfruttamento di una vulnerabilità (per queste considerazioni il business della sicurezza è piuttosto affine a quello assicurativo); in secondo luogo l'applicazione della sicurezza riduce tipicamente l'usabilità e l'accessibilità dei sistemi.

Affrontare nel modo corretto la sicurezza aziendale comporta l'attivazione di un processo continuo che, partendo dall'analisi del rischio, imposta le adeguate contromisure in termini di protezione degli asset, rilevamento dei tentativi di violazione con monitoraggio continuo e procedure codificate di risposta agli incidenti di sicurezza. Il bagaglio metodologico e l'arsenale degli strumenti per l'applicazione della sicurezza in azienda sono in continua crescita, all'ombra di iniziative di standardizzazione volte a favorire la messa in comune dei risultati più interessanti e a garantire l'integrazione delle componenti di sicurezza.

La complessità delle piattaforme tecnologiche e la pervasività delle azioni che riguardano la garanzia dei requisiti di sicurezza rendono, tuttavia, il raggiungimento di un buon bilancio rischi-benefici un'impresa difficile.

La figura 4 mostra l'andamento nel tempo della complessità delle tecniche di attacco rapportata agli *skill* necessari per metterle in pratica: all'inizio degli anni Ottanta le tecniche di attacco si basavano sulla capacità di identificare una *password* senza l'ausilio di strumenti automatici, ma semplicemente basandosi sulla conoscenza di informazioni che tipicamente sono impiegate per la scelta delle *password* (ad esempio un nome di donna, piuttosto che la *password di default* impostata dal fornitore di un apparato).

Sono apparsi successivamente i primi virus informatici, ossia programmi che sono trasmessi da una macchina all'altra, come una sorta di infezione, originariamente usando in prevalenza come veicolo i supporti di memorizzazione di massa rimovibili (i dischetti). Quando un virus infetta un calcolatore, esso altera il *file system* andando a modificare file che sono frequentemente utilizzati, inserendo l'esecuzione del programma maligno, che è contenuto nel virus, all'insaputa dell'utente.

Una forma differente di virus è denominata *worm* e identifica programmi maligni, in grado di propagarsi attraverso la rete e di installarsi sui computer compromessi, ripartendo per insidiarsi in altre macchine.

La pericolosità dei *worm* è stata chiaramente percepita quando nel 1988 un famoso *worm* mise in crisi l'intera rete Internet, infettando circa seimila macchine e occupando le risorse di calcolo, fino a renderle inutilizzabili. Sono stati quindi distribuiti in rete programmi in grado di effettuare analisi sistematiche dei *file* di *password*, con le *password* che sono cifrate, allo scopo di decriptarle *(cracking)*.

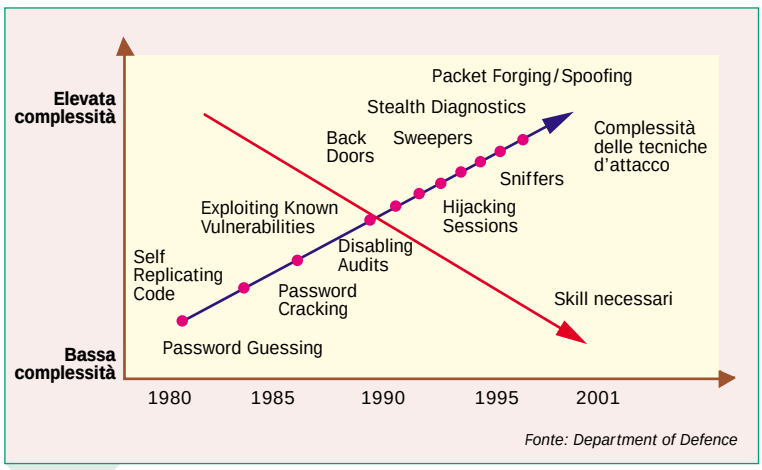


Figura 4 *Evoluzione della complessità delle tecniche di attacco rispetto agli "skill" necessari.*

L'impiego di vulnerabilità note, l'evasione dei meccanismi di audit messi in campo dagli amministratori di sistema e l'introduzione di *back door* (programmi istallati su un sistema protetto per disporre di una via illecita di accesso) sono stati i sistemi maggiormente utilizzati per gli attacchi negli anni a seguire.

Un caso classico di attacco è riportato nel riquadro a pagina 18 e 19.

Più complesso è l'attacco di *session hijacking*, con il

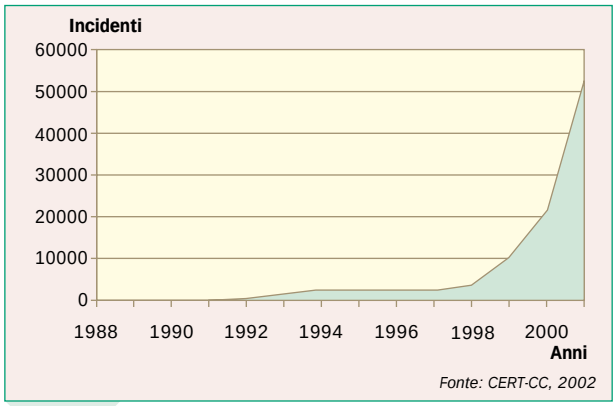


Figura 5 *Incidenti di sicurezza.*

quale un attaccante è in grado di subentrare a un utente su una sessione attiva nel colloquio con il sistema obiettivo dell'attacco.

Negli anni Novanta sono stati resi disponibili programmi in grado di effettuare un *benchmark* sui sistemi in rete per rilevarne eventuali vulnerabilità ad attacchi noti, gli *sweepers* (il più noto tra questi era *SATAN*). Gli autori dei programmi volevano mettere a disposizione strumenti di verifica utili agli amministratori di rete per asseverare il proprio grado di protezione. L'uso di questi strumenti è tuttavia alla portata di attaccanti per utilizzi diametralmente opposti a quelli che hanno mosso gli autori.

Allo stesso tempo sono stati sviluppati e resi disponibili nella rete Internet strumenti in grado di analizzare il traffico in transito su una tratta di rete, estraendo i pacchetti in transito, e consentendo di ricostruire i flussi di dati in transito *(sniffer)*. Uno *sniffer* può essere, infatti, utilizzato per analizzare il traffico in modo da rilevare tentativi di attacco in corso; ma può anche essere impiegato da un malintenzionato per intercettare le *password* in transito.

Le tecniche di attacco più sofisticate impiegano sistemi di generazione di traffico che simula comportamenti leciti ma usati per perseguire finalità illecite *(forging e spoofing)*.

Negli ultimi anni la crescente complessità delle tecniche di attacco è stata accompagnata da una paradossale riduzione degli *skill* necessari a metterle in atto. Questo andamento si spiega con la messa a disposizione, in Internet, di strumenti in grado di attuare tecniche complesse di attacco, con un corredo delle modalità di utilizzo e di documentazione "a prova di principiante".

La figura 5 riporta l'andamento rilevato nell'ultimo decennio del numero di attacchi, mentre la figura 6 presenta l'andamento delle vulnerabilità segnalate nel periodo dal 1995 al 2001, inclusa una proiezione per il 2002. I dati consuntivati dal *CERT/CC (Computer Emergency Response Team/Coordination Center)*, - che è il centro di coordinamento delle risposte agli incidenti di sicurezza - riportano per il primo trimestre del 2002 un numero pari 26.829 incidenti di sicurezza, mentre le vulnerabilità segnalate ammontano 1.065. Dalle curve emerge che il tasso di crescita del numero di vulnerabilità non è che una porzione assai modesta del tasso di crescita degli attacchi. Oltretutto, si verifica un'impennata preoccupante negli ultimi anni rappresentati. A parità di qualità dei sistemi utilizzati in relazione alla sicurezza, sono quindi cresciute in modo cospicuo le azioni di attacco e le capacità di rilevarle (nonché la sensibilità a sentirsi parte di una comunità di difesa, ricorrendo alla segnalazione comunicata al CERT/CC).

In Italia è stato istituito un *OCI (Osservatorio per la Criminalità ICT)* a cura del *FTI (Forum per le Tecnologie dell'Informazione)*, con l'obiettivo di raccogliere informazioni sulla criminalità ICT nel nostro Paese. Il terzo rapporto dell'OCI, pubblicato nel 2001, ha segnalato una serie di elementi interessanti sulla criminalità ICT nel 2000 in Italia, sulla base di un indagine effettuata su duecento aziende e sulla Pubblica Amministrazione.

Dall'indagine risulta che il 73 per cento degli attacchi provengono dall'esterno, contro il 27 per cento dall'interno dell'azienda. Il 78 per cento delle aziende e degli Enti intervistati ha subito nel 2000 una contaminazione da virus, contro il 69 per cento

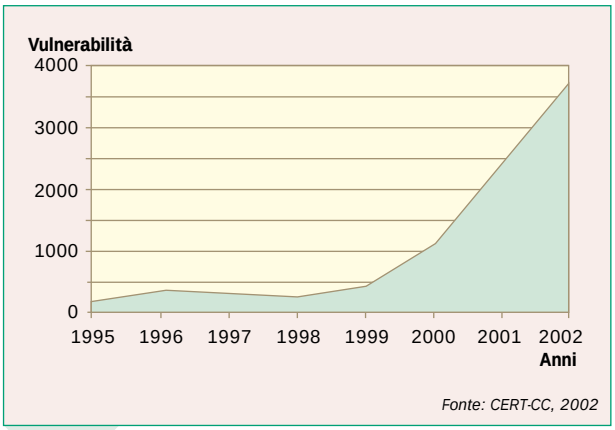


Figura 6 *Vulnerabilità segnalate.*

del 1999. Nel 41 per cento, sono stati sottratti apparati contenenti dati, mentre l'anno precedente la percentuale era del 33 per cento.

Fra i tipi di attacchi, il virus resta quello maggiormente impiegato. Nel 2000 l'attacco è stato portato da virus nel 39 per cento dei casi, mentre nel 1999 la percentuale era del 32,8 per cento. Dati Euros, relativi al 2000, indicano la posta elettronica come il veicolo principale della diffusione dei virus (53 per cento dei casi), seguita dai dischetti (28 per cento), dall'ac-

cesso WWW (15 per cento) e dal CD-Rom (4 per cento). Il furto di apparati dati passa dal 15,6 per cento del 1999 al 22 per cento del 2000. Al terzo posto della classifica dei tipi di attacchi si trova la saturazione delle risorse (14 per cento degli attacchi, contro l'8,3 per cento del 1999).

Le motivazioni che inducono a violare la sicurezza dei sistemi informatici delle aziende sono: il vandalismo (36 per cento), l'azione dimostrativa (22 per cento), il sabotaggio (17 per cento) e la frode informatica (14 per cento).

anno	Virus	Impatto economico nel mondo (miliardi di $)
2001	Nimda	0,635
2001	Code Red(s)	2,62
2001	Sircam	1,15
2000	Love Bug	8,75
1999	Melissa	1,10
1999	Explorer	1,02

Fonte: Computer Economics

Tabella 1 *Perdite economiche imputabili ai virus.*

Le conseguenze per le aziende sono prevalentemente economiche: il 41,3 per cento degli intervistati denuncia una perdita in termini economici; il 14,3 per cento accusa un danno di immagine e il 23,8 per cento afferma che gli attacchi informatici non hanno avuto alcun esito. A renderli vani, contribuisce anche il miglioramento delle condizioni di sicurezza dei sistemi: un miglioramento testimoniato anche dalla riduzione dei tempi necessari per il ripristino del sistema. Nel 2000, nel 68 per cento dei casi è bastata meno di una giornata per far ripartire i sistemi (nel 24 per cento ce ne sono volute tre), mentre nel 1999 solo nel 44 per cento dei casi era possibile risolvere il problema entro ventiquattro ore.

La situazione italiana, secondo gli esperti dell'OCI, è sostanzialmente omogenea con quella americana. In Italia, come negli Stati Uniti, è, inoltre, in crescita la sensibilità nei confronti del problema degli attacchi informatici, visti sempre più come una minaccia alla produttività e alla sicurezza delle aziende.

La rassegna della sicurezza del *CSI (Computer Security Institute)* relativa al 2001 e pubblicata nella primavera del 2002, mostra i dati raccolti da 503 aziende e amministrazioni pubbliche americane di varie dimensioni. Oltre il 90 per cento del campione ha rilevato attacchi di sicurezza negli ultimi dodici mesi; l'80 per cento ha denunciato perdite economiche e il 44 per cento ha quantificato le perdite, per un totale di 455.848mila $.

Le cause di perdita più ingenti sono il furto di informazioni (circa 170 milioni di $) e la frode finanziaria (circa 115 milioni di $). Il 74 per cento del campione ha indicato Internet come il fronte da cui provengono gli attacchi, mentre il 33 per cento delle aziende ha indicato i sistemi interni come fonte degli attacchi.

Il 78 per cento delle aziende intervistate ha denunciato abusi nell'utilizzo del collegamento aziendale a Internet (prevalentemente per accesso a informazione pornografica o per l'uso inappropriato della posta elettronica). Il 40 per cento del campione ha segnalato la violazione dei sistemi di difesa perimetrali, e la stessa percentuale di aziende ha denunciato attacchi volti a impedire la disponibilità del servizio dei propri sistemi. I virus informatici hanno interessato l'85 per cento delle aziende. Il 34 per cento delle aziende del campione hanno denunciato le violazioni all'autorità giudiziaria.

Secondo Computer Economics - azienda specializzata in analisi dei fenomeni che riguardano la sicurezza - le perdite economiche imputabili ai virus informatici sono di tutto rispetto e sono da attribuire soprattutto ai costi per il fermo dei sistemi infetti e per le risorse necessarie al ripristino della normale funzionalità, a valle della bonifica (tabella 1).

È impressionante osservare che il virus Nimda abbia infettato ben 37318 server Web solo negli Stati Uniti e ben 1874 in Italia. Questo virus si è propagato rapidamente grazie alla capacità di diffondersi con quattro diverse modalità, compromettendo in modo serio i sistemi infetti e consentendo all'attaccante di acquisire i privilegi di amministratore. Oltretutto, la bonifica è complessa poiché il virus opera numerose modifiche nel *file system* e nei registri di configurazione (*registry* dei sistemi Windows).

Se allarghiamo l'orizzonte dell'impatto economico legato agli incidenti di sicurezza, la stessa fonte riporta cifre significative e in crescita repentina,

anno	Impatto economico nel mondo (miliardi di $)
2001	13,2
2000	17,1
1999	12,1
1998	6,1
1997	3,3
1996	1,8
1995	0,5

Fonte: Computer Economics

Tabella 2 *Perdite economiche imputabili a incidenti di sicurezza.*

anche se nel 2001 è stata osservata una contrazione dei costi rispetto all'anno precedente (tabella 2).

3. Sicurezza aziendale

La sicurezza aziendale è definita in modo sintetico ed efficacie dalla norma UNI 10459 [8] che prescrive: "... studio, sviluppo e attuazione delle strategie, delle politiche e dei piani operativi volti a prevenire, fronteggiare e a superare eventi in prevalenza di

natura dolosa o colposa, che possono danneggiare le risorse materiali, immateriali organizzative e umane di cui l'azienda dispone o di cui necessita per garantire un'adeguata capacità concorrenziale nel breve, nel medio e nel lungo termine".

In quest'articolo si vuole fornire una visione del percorso metodologico che si intende proporre per introdurre la sicurezza nelle aziende, attraverso un piano orientato verso il raggiungimento di obiettivi ben precisi.

3.1 Fattori ambientali decisivi

L'esperienza ha messo in luce fattori ambientali e di sostegno che concorrono in modo decisivo al successo delle iniziative mirate a introdurre e a gestire la sicurezza in azienda: è, innanzitutto, fondamentale che il vertice aziendale abbia maturato la consapevolezza dell'importanza e della portata del tema della sicurezza aziendale, attribuendogli un ruolo corretto nell'ambito delle strategie di sviluppo dell'azienda.

Per un'efficace attuazione delle strategie di sicurezza aziendali è, infatti, opportuno costituire strutture organizzative specifiche, incaricate del presidio e del monitoraggio dell'attuazione delle strategie e allo stesso tempo, strutture organizzative volte a effettuare gli interventi, con chiare attribuzioni di compiti e di responsabilità *(Information Security Office)*.

Il piano di intervento deve poi essere affiancato da un programma di diffusione della consapevolezza e delle strategie, e deve essere seguito da programmi specifici volti a educare le diverse componenti aziendali alla sensibilità alla sicurezza e all'esecuzione delle procedure sviluppate per un comportamento aderente all'attuazione della sicurezza.

Il coinvolgimento delle risorse umane nella realizzazione dei piani di intervento rappresenta, quindi, un elemento chiave nell'attuazione delle strategie di sicurezza.

3.2 I motori decisionali

I motori decisionali che guidano il posizionamento rispetto al tema della sicurezza hanno diverse matrici e dipendono dalla realtà specifica dell'azienda. La *mission* aziendale, la sua organizzazione e l'impiego delle tecnologie informatiche e di telecomunicazioni influenzano infatti in modo determinante l'approccio delle organizzazioni rispetto alla sicurezza.

3.2.1 Strategie di business, organizzazione e tecnologie

Un'azienda che opera in un certo segmento di business, che è organizzata per processi operativi e che impiega le tecnologie, in base a criteri funzionali allo svolgimento dei propri compiti è, in genere, soggetta a minacce, ossia è sensibile a eventi che possono compromettere l'operatività, l'efficienza o l'immagine.

Si pensi, ad esempio, a una qualunque azienda che svolga attività commerciali su Internet attraverso il proprio sito. Essa è esposta a tutte le minacce che in tale ambiente sono operative e deve, quindi, attuare le misure necessarie a fronteggiare il rischio di attacco e le perdite finanziarie ad esso collegate.

Un attacco provoca, generalmente, perdite che possono essere suddivise in *materiali* e *immateriali*. Le prime sono le perdite secche dovute all'indisponibilità di un anello della catena di produzione del valore per un certo periodo, a seguito, ad esempio, di un attacco di tipo *DoS (Denial of Service)* che provoca il fermo del sito dedicato al commercio elettronico.

Il secondo caso riguarda, invece, le perdite immateriali di immagine e di fiducia nella sicurezza di un sito di commercio elettronico e di un'azienda, che si verifica, ad esempio, a seguito di un attacco destinato al furto delle informazioni personali degli utenti registrati sullo stesso sito.

3.2.2 La spinta legislativa e normativa

Il contesto legislativo nazionale e internazionale impone all'azienda di ottemperare a norme stilate appositamente in materia di sicurezza (vedi riquadro a pagina 14). La maggior parte delle norme riguarda il trattamento di dati personali, per i quali i requisiti necessari per ottenere la piena conformità risultano essere particolarmente stringenti. Non vanno tuttavia sottovalutate le problematiche relative alla proprietà intellettuale e ai crimini informatici.

Le leggi nazionali in materia di tutela della *privacy* assumono un rilievo fondamentale dal punto di vista dell'impegno richiesto alle aziende e dell'attribuzione di responsabilità a soggetti che in azienda rivestono un ruolo manageriale di rilievo (legge del 31 dicembre 1996 n. 675 per la *Tutela delle persone e di altri soggetti rispetto al trattamento dei dati personali* [9] e successivo Decreto del Presidente della Repubblica 28 luglio 1999, n. 318 contenente il *Regolamento recante norme per l'individuazione delle misure di sicurezza minime per il trattamento dei dati personali* a norma dell'articolo 15, comma 2, della legge 31 dicembre 1996, n. 675) [10].

Appartengono al contesto normativo di interesse anche le seguenti norme con valenza di carattere generale: Decreto Legge 518/1992 [11] e Legge 248/2000 [12], relative alla proprietà intellettuale, che si applicano sia al materiale sviluppato all'interno dell'azienda, sia a quello fornito all'azienda stessa in licenza d'uso da terzi; la Legge 547/1993 relativa alla criminalità informatica [13].

A queste direttive si aggiungono le leggi che regolano specifiche attività, come, ad esempio, le norme relative all'esecuzione di attività commerciali con l'impiego degli strumenti telematici nelle varie forme del commercio elettronico o le leggi che regolano il trattamento dei documenti, a valore legale, legati alle norme per la firma digitale emesse dall'AIPA [14, 15, 16].

Di più, la trans-nazionalità di attività, che comportano l'impiego delle infrastrutture telematiche, determina la necessità di norme che trascendono i confini dei territori delle singole nazioni e che sono riconosciute e applicate dal diritto internazionale nelle sedi deputate [17, 18].

Oltre all'ottemperanza alle disposizioni di legge, per attuare la propria strategia di business le aziende si trovano spesso a perseguire obiettivi di conformità

SICUREZZA INTERNA ED ESTERNA

Il tema della sicurezza aziendale presenta diversi risvolti e due fronti principali di intervento: interno ed esterno. Il fronte esterno riguarda la protezione degli asset aziendali dagli attacchi che possono essere perpetrati dal mondo esterno all'azienda. Con l'introduzione delle piattaforme telematiche il perimetro dell'azienda non corrisponde a confini fisici. Le recinzioni, i sistemi di controllo degli accessi e i sistemi di sorveglianza e di allarme servono a realizzare la protezione fisica delle sedi e dei beni aziendali.

I sistemi di difesa perimetrali telematici servono ad attuare la protezione in corrispondenza dei punti di contatto telematico del dominio dell'azienda con l'esterno. In corrispondenza di questi punti sono dispiegate le tecnologie di difesa e di monitoraggio: firewall, sistemi di analisi dei contenuti maligni (sistemi antivirus), sistemi di segnalazione dei tentativi di intrusione, sistemi di autenticazione e di autorizzazione per gestire l'attraversamento selettivo della frontiera telematica agli utenti autorizzati. I sistemi di difesa sono progettati e realizzati in pochissimi punti concentrati di interfaccia del dominio telematico. Questi ingressi sono sottoposti a una sorveglianza continua in modo da mantenere aggiornate le difese e di rilevare i segnali di violazione, attivando le procedure di intervento appropriate.

Il fronte interno della sicurezza riguarda la protezione dagli attacchi che possono provenire dall'interno del dominio dell'azienda.
Innanzitutto, all'interno di un'azienda esistono zone che pongono requisiti di sicurezza altamente differenti ed è opportuno, naturalmente, che le soluzioni di sicurezza siano commisurate alle esigenze.

Inoltre, a ogni soggetto che opera nel dominio dell'azienda corrispondono ruoli e responsabilità che è opportuno vengano circoscritti grazie all'uso degli strumenti tecnologici che consentano di attuare l'autenticazione, l'autorizzazione e la registrazione delle azioni effettuate. A ogni soggetto è associata in modo robusto un'identità digitale, alla quale corrispondono ruoli e responsabilità in linea con la funzione aziendale.

La scalata dei privilegi di accesso o l'abuso dei privilegi legittimi da parte di attori operanti nel dominio dell'azienda, rappresentano violazioni della sicurezza estremamente pericolose in quanto alla portata di utenti che interagiscono con gli *asset* aziendali in modo privilegiato e facilitato.

L'impiego di sistemi di autenticazione, autorizzazione e di *accounting*, caratterizzati dall'integrazione delle modalità di governo dell'interfacciamento degli utenti con una varietà ampia di sistemi e tecnologie presenti in azienda (con l'uso di sistemi di *directory* e *meta-directory*, di certificati digitali e di sistemi *single-sign-on*) rappresentano quindi la linea di intervento essenziale per rafforzare la sicurezza interna.

agli standard di sicurezza internazionali. In numerosi settori le aziende sono tenute ad attuare valutazioni di conformità per le applicazioni software, per gli apparati hardware o per gli algoritmi di crittografia utilizzati e/o sviluppati internamente con i vari standard codificati a livello internazionale, ottenendo eventualmente il rilascio della relativa certificazione (ISO, CC, ITSEC, TCSEC).

3.2.3 La valutazione del rischio

Ogni azienda è tenuta ad agire in modo da attestarsi consapevolmente su un livello di protezione dai rischi di sicurezza [19]. Per perseguire questo obiettivo risulta fondamentale valutare il rischio, ossia individuare gli eventi che possono compromettere la sicurezza aziendale, sia come probabilità che essi si presentino, sia in termini di portata dei danni che possono essere arrecati alla stessa azienda *(business impact analysis)*.

Ogni azienda, in funzione delle proprie peculiari caratteristiche, è esposta a un certo livello di rischio che può essere ridotto realizzando interventi tecnologici e organizzativi volti a ridurre le vulnerabilità e a contrastare le minacce.

Questi interventi comportano costi che devono essere tarati sul valore dei beni da proteggere (asset aziendali, immagine, informazioni sensibili, …) o alle perdite, che potrebbero essere subite dall'azienda, in funzione del danneggiamento dei beni da proteggere. Quest'operazione di bilanciamento dovrebbe portare l'azienda ad attestarsi su un livello di rischio residuo - ossia di esposizione a valle della realizzazione degli interventi - che è ritenuto il compromesso ottimale per l'azienda.

L'analisi del rischio può essere eseguita solo conoscendo in maniera approfondita l'azienda e il settore in cui essa opera e attribuendo un valore specifico ai diversi asset che possono essere compromessi come effetto del successo di un attacco alla sicurezza. Quest'analisi è condotta, quindi, attraverso una raccolta di informazioni che portano a costruire un quadro completo della realtà aziendale dal punto di vista organizzativo e tecnologico.

L'esecuzione dell'analisi del rischio per i sistemi *IT (Information Technology)* è guidata dagli standard internazionali [20, 21, 22] che ne hanno codificato la terminologia e la metodologia: ISO 17799 / BS7799 (figura 7).

Lo standard ISO 17799, derivato dalla specifica britannica BS7799, copre numerosi aspetti legati alla sicurezza: partendo dalla problematica del *Business Continuity Planning*, attraverso le tematiche del controllo degli accessi e dello sviluppo e della manutenzione di sistemi sicuri, la norma si occupa infatti della sicurezza ambientale e fisica, propone i criteri di aderenza alle norme, tratta la sicurezza personale, introduce i criteri per l'organizzazione della sicurezza e i

STANDARD PER LA SICUREZZA INFORMATICA

La necessità di nuovi approcci metodologici alla sicurezza informatica, motivata dal crescente uso della tecnologia ICT, ha spinto allo sviluppo di un gran numero di standard che sono riconducibili a una prospettiva più ampia di quella della sicurezza informatica, detta *"information system assurance"*, ossia di garanzia e fiducia sia nell'efficacia che nella correttezza delle funzioni e dei meccanismi di sicurezza. L'ampliamento di prospettiva è determinato dalla maggiore enfasi sulla sicurezza come tecnologia abilitante per le transazioni elettroniche in cui è necessario garantire un rapporto di fiducia tra le parti e la disponibilità dell'informazione e dei servizi in generale, oltre alle più elementari esigenze di protezione degli *asset*. In questo riquadro sono riportati alcuni degli standard di riferimento di maggior rilievo che attengono all'*information system assurance*, escludendo dalla trattazione gli standard che riguardano le tecnologie crittografiche di base o applicate.

- ***ISO/IEC 13335***

L'ISO 13335 [24] non è uno standard ma una collezione di cinque rapporti tecnici che offrono le linee guida per la gestione della sicurezza IT, note come *GMITS (Guidelines for the Management of IT Security)*. Le parti trattano i seguenti argomenti: i concetti e i modelli per la sicurezza IT, la gestione e la pianificazione, le tecniche per la gestione, la scelta delle soluzioni di protezione e la protezione per le connessioni con l'esterno. I documenti hanno un profilo che enfatizza le definizioni, peraltro non sempre in accordo con gli altri documenti ISO sulla sicurezza, piuttosto che la codifica delle regole di azione.

- ***ISO/IEC 17799***

Lo standard ISO 17799 [20] ratifica a livello di norma internazionale un lavoro di standardizzazione in gran parte sviluppato in seno a un contesto di standardizzazione regionale, il *BSI (British Standard Institution)*, artefice dello standard BS7799 [21, 22]. Il lavoro BSI sul BS7799 ha avuto inizio nel 1990 e la versione più aggiornata risale al 1999.
La norma ISO recepisce per ora solo la prima parte del lavoro BS7799. Infatti, lo standard BS7799 si compone di due parti: BS7799-1, noto come *standard code of practice*, è la guida per rendere sicuro un sistema informativo, BS7799-2, detto *standard specification*, fornisce in termini di requisiti, obiettivi di controllo e *framework* di riferimento, un sistema completo per la gestione della sicurezza, detto *ISMS (Information Security Management System)*.
L'obiettivo dello standard è quello di progettare la sicurezza dei sistemi IT, attuando un insieme di azioni di controllo: politiche, procedure, *best practice*, strutture organizzative e funzioni software. Il contesto di riferimento è l'organizzazione di una realtà aziendale e l'enfasi è posta sugli aspetti organizzativi e sui fattori umani.
Lo standard fornisce indicazioni abbastanza generali di gestione della sicurezza a vantaggio di chi riveste la responsabilità di avviare, applicare e mantenere la sicurezza nell'organizzazione di appartenenza.
Il consenso su questa norma è stato fortemente dibattuto in sede ISO, tant'è vero che Belgio, Canada, Francia, Germania, Italia, Giappone e Stati Uniti hanno votato contro l'adozione.
Lo standard è organizzato in sezioni che trattano i seguenti temi: *BCP (Business Continuity Planning)*; controllo degli accessi; sviluppo e manutenzione dei sistemi; sicurezza fisica e ambientale; criteri di conformità; sicurezza del personale; sicurezza dell'organizzazione; gestione della comunicazione e dell'operatività; classificazione degli *asset* e politiche di sicurezza.
Le organizzazioni possono essere certificate da tempo secondo la norma BS7799, mentre, solo di recente è stato messo a punto l'iter di certificazione ISO 17799.

• ***ISO/IEC 15408***

L'ISO 15408 [25] è uno standard sviluppato dal 1993 da un gruppo di lavoro che comprendeva Stati Uniti, Canada e Ue (non hanno partecipato l'Italia, il Giappone e l'Australia), sotto l'etichetta di *"Common Criteria for IT Security"* ed è confluito nel 1998 nello standard ISO, emesso in versione definitiva nel 1999.
Questo standard prende spunto dallo standard *TCSEC (Trusted Computer System Evaluation Criteria)* [26] del *DoD (Department of Defence)* americano, comunemente noto come *"orange book"*, dedicato ai criteri per valutare e classificare il grado di sicurezza di un sistema o di un dispositivo IT.
Sulla scia del TCSEC, alcuni Paesi europei (tra cui la Francia, la Germania, la Gran Bretagna) hanno sviluppato, sotto l'egida della Comunità Europea, lo standard *ITSEC (Information Technology Security Evaluation Criteria)* [27] orientato alla valutazione di sistemi o prodotti, denominati *TOE (Target Of Evaluation)*.
La versione 1.2 dello standard ITSEC risale al 1991 ed è tutt'ora in uso.
ITSEC è uno strumento di valutazione della sicurezza e non uno strumento di progettazione. La valutazione consiste nella misura dell'*assurance* riscontrata nel sistema esaminato. Definiti i criteri di sicurezza *(Security Target)*, si esamina la presenza delle funzioni o delle contromisure di sicurezza *(Security Enforcing Function)*.
La norma AIPA per la firma digitale fa esplicito riferimento ai criteri ITSEC per la sicurezza dei sistemi e dei prodotti da impiegare nella certificazione.
Lo standard ISO 15408 ha l'obiettivo di conciliare l'approccio nordamericano con quello europeo, attraverso una definizione comune e condivisa dei requisiti funzionali e di affidabilità dei sistemi e dei prodotti IT.
Lo standard ISO usa la nomenclatura ITSEC per quanto riguarda i TOE e gli ST. Vengono introdotti i *PP (Protection Profile)* che possono essere considerati modelli di riferimento a cui si raffrontano i realizzatori o gli acquirenti di sistemi o prodotti IT.
I prodotti che sono proposti per applicazioni in cui la sicurezza ha un rilievo importante sono certificati; per essi, infatti, è stato sviluppato un *PP* che ne classifica il grado di sicurezza.

• ***STANDARD PER L'AUDIT***

Nell'area dell'audit di sicurezza sono presenti diverse linee di standardizzazione che hanno radici principalmente negli Stati Uniti e, in molti casi, sono un'appendice di funzioni di audit più generali.
Nel continente nordamericano operano oggi associazioni e fondazioni dedicate all'auditing dei sistemi informativi, ciascuna con i propri programmi di certificazione.
Le Organizzazioni principali sono:

- ***ISACA (Information Systems Audit and Control Association)*, che gestisce la certificazione *CISA (Certified Information Systems Auditor)* e che ha prodotto lo standard *CoBiT (Control OBjectives for Information and related Technology)*;**
- ***IIA (Institute of Internal Auditors)* che offre la certificazione *CIA (Certified Internal Auditor)* e che prevede precise competenze riguardo alla gestione dei rischi legati all'IT;**
- ***(ISC)2 (International information systems Security certification Consortium)*, supportato da *CSI (Computer Security Institute)*, *ISSA (Information Systems Security Association)*, e da altri Enti e che propone la certificazione *CISSP (Certified Information System Security Professional)*.**

In Italia, l'*AIEA (Associazione Italiana EDP Auditor)* rappresenta in Italia l'ISACA, mentre l'*AIIA (Associazione Italiana Internal Auditors)* rappresenta l'IIA e organizza la certificazione CIA.

concetti di sicurezza applicata ai sistemi telematici, definisce i criteri per la classificazione delle risorse in relazione alla sicurezza e alle politiche riguardanti la sicurezza.

3.2.4 Vulnerability Assessment & Penetration Testing

Una componente essenziale dell'analisi del rischio è lo studio delle vulnerabilità, ossia l'individuazione dei punti deboli in cui l'azienda è esposta a possibili attacchi alla sicurezza. L'analisi delle vulnerabilità di tipo tecnologico sulle infrastrutture informatiche e di telecomunicazioni dell'azienda si avvale tipicamente dell'esecuzione di prove volte a valutare il grado di esposizione a possibili attacchi noti di sicurezza.

Questa operazione chiamata *VAPT (Vulnerability Assessment & Penetration Testing)* comprende due fasi: l'analisi della presenza di vulnerabilità ad attacchi noti, attraverso l'esplorazione esaustiva dei casi conosciuti (avvalendosi generalmente di strumenti commerciali che scandagliano i sistemi in modo da accertare la presenza di varchi praticabili per un eventuale attacco) e, in secondo luogo, lo sfruttamento delle vulnerabilità individuate per violare i sistemi, in modo da determinare le possibili conseguenze del successo di un attacco di sicurezza (questa prassi viene anche denominata *Ethical Hacking*).

Le attività di VAPT possono concorrere al consolidamento dell'analisi del rischio, apportando un contributo sperimentale all'attività.

3.3 Sviluppo della sicurezza

I motori decisionali, che spingono e orientano l'azienda in materia di sicurezza, alimentano lo sviluppo di azioni specifiche volte a posizionarla nel punto di bilancio ottimale tra costi e rischi. Per eseguire queste azioni occorre avviare una fase di progettazione e di pianificazione degli interventi, affiancata da una fase di revisione di quelli già effettuati.

L'attività di analisi del rischio, partendo dalla raccolta delle informazioni sui processi e sull'organizzazione e sul loro intreccio con le piattaforme tecnologiche, produce relazioni di interdipendenza tra le componenti e un'analisi di impatto, che alimenta la valutazione del rischio e la definizione delle contromisure.

Il bilanciamento tra costi e benefici per gli interventi ipotizzati, necessari per attuare le contromisure e l'applicazione di priorità d'intervento, portano a identificare il punto di lavoro in cui l'azienda si attesta su livelli di rischio residui accettabili. La definizione delle politiche di sicurezza, l'identificazione dell'architettura di sicurezza, la scelta delle tecnologie e la definizione delle procedure, accompagnate da attività di diffusione della consapevolezza del rischio e di istruzione del personale, concorrono a determinare l'*enforcement* della sicurezza. Completano il quadro delle linee di intervento il monitoraggio, l'*audit* (ossia l'ispezione periodica per accertare lo stato di attuazione delle pratiche previste dalle politiche) e la gestione degli incidenti di sicurezza.

3.3.1 Politiche di sicurezza

Le politiche di sicurezza vogliono fornire una definizione di come un'organizzazione, intesa come un'unica entità, affronta le problematiche di sicurezza nella loro globalità.

Una politica di sicurezza è costituita in genere da due parti: una sezione generale che descrive l'approccio alle problematiche di sicurezza e una collezione di regole che specificano le azioni permesse e le attività non autorizzate. Le regole sono in genere integrate con procedure, strumenti e meccanismi di intervento.

Le politiche di sicurezza sono rivolte tipicamente a un uditorio costituito dalla direzione e dal management aziendale. Esse infatti definiscono un insieme di decisioni che l'azienda consapevolmente approva. Tali decisioni sono generalmente ad alto livello e sono accompagnate da procedure che specificano come esse devono essere attuate.

Obiettivi della sicurezza aziendale sono: la sicurezza fisica dei dipendenti e la protezione dei beni di loro proprietà; la sicurezza fisica di edifici, apparati e sistemi da incendi, inondazioni, furti e usi non autorizzati; la protezione delle risorse critiche, che comprende, anche, la conformità ad accordi con terze parti e *non-disclosure agreement*; la sicurezza dei sistemi di elaborazione e dell'infrastruttura di rete che include la gestione, la protezione da attacchi, l'uso corretto, la confidenzialità e la proprietà intellettuale delle informazioni conservate, elaborate e trasmesse.

La politica di sicurezza di un'azienda riguarda aspetti organizzativi, fisici e informatici. In particolare le politiche di sicurezza

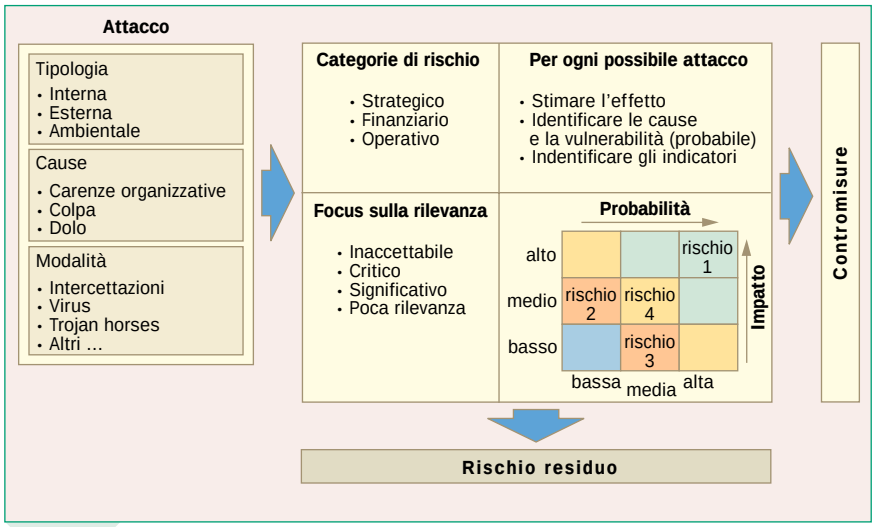


Figura 7 *Analisi del rischio.*

devono specificare alcuni ambiti quali:

- *sicurezza fisica;*
- *sicurezza logica*: *reti e sistemi;*
- *sicurezza organizzativa;*
- *piani di business continuity, disaster recovery e contingency*.

3.3.2 Procedure

Le procedure definiscono le norme particolareggiate di comportamento delle singole componenti aziendali che portano all'attuazione delle politiche di sicurezza definite (figura 8).

A partire dalla definizione delle *procedure di sicurezza ordinarie e straordinarie* (gestione degli incidenti), si procede quindi alla stesura dei *manuali operativi* per descrivere le attività svolte sulle risorse critiche.

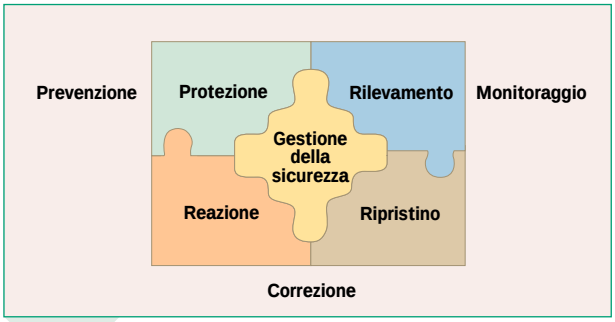


Figura 8 *Fasi degli interventi per la sicurezza di un'azienda.*

Sono di norma previste attività di prova *(testing)* e di sintonizzazione *(tuning)* delle procedure ordinarie e straordinarie.

Le procedure messe a punto per la sicurezza sono impiegate in diverse fasi del ciclo di vita di un sistema distribuito.

Assieme all'utilizzo di misure tecnologiche, che sono elencate nel seguito dell'articolo, gli interventi contemplati dalle procedure si distinguono in preventivi, di monitoraggio e correttivi.

Gli *interventi preventivi* hanno lo scopo di aumentare la difficoltà di attuazione di un attacco alla sicurezza di un sistema distribuito, accrescendo la protezione del sistema.

Gli *interventi di monitoraggio* sono indirizzati a mantenere il sistema sotto controllo, per rilevare tempestivamente e, possibilmente, in anticipo i segnali di una danneggiamento generato da un attacco in corso.

Gli *interventi correttivi* si esplicano nell'applicazione di operazioni di reazione a un attacco rilevato e, eventualmente, al ripristino della funzionalità del sistema o al recupero dei dati.

Le operazioni sviluppate nella fase correttiva comprendono le procedure di *incident handling (*ossia le procedure da applicare a cura dell'*incident response team* in caso di attacco accertato alla sicurezza con sistemi compromessi) e le procedure di *forensics analysis* (ossia di analisi volte a seguire l'operato dell'attaccante al fine di raccogliere le tracce del percorso di attacco, come ausilio per un eventuale iter investigativo teso ad attribuire la responsabilità per l'esecuzione dell'attacco).

La definizione dei processi e delle procedure di gestione degli incidenti di sicurezza deve essere supportata dalla presenza di un'ulteriore coppia di elementi: anzitutto la creazione di un *gruppo deputato alla gestione* di tali incidenti e, in secondo luogo, l'*utilizzo di meccanismi e di strumenti* atti a rilevare, identificare, contenere, contrastare e analizzare eventuali incidenti di sicurezza e a permettere il ripristino dello stato dei sistemi coinvolti.

L'adozione di un approccio sistematico e di una metodologia per la gestione degli incidenti di sicurezza sono passaggi indispensabili per limitare l'impatto della compromissione di sistemi *business critical*. Fattori critici nella gestione di un incidente di sicurezza sono la rapidità della risposta, l'efficienza delle risorse utilizzate e l'efficacia nel contenere e nel limitare i danni provocati. Questi elementi possono essere assicurati solo mediante un'attenta pianificazione, l'uso di meccanismi e di strumenti adeguati e la costituzione di una *task force* dedicata a tale attività.

Numerosi fattori hanno incrementato i rischi di incidenti di sicurezza. Tra questi possono essere citati: l'importanza della riservatezza delle informazioni; il rispetto dell'integrità o disponibilità; la diffusione e l'uso di infrastrutture di rete locali e remote (Extranet, Internet); la diffusione di PC e di sistemi informatici; l'incremento delle vulnerabilità a causa dell'aumento del numero di applicazioni e della loro complessità.

La necessità di un team di gestione degli incidenti di sicurezza è guidato da due elementi fondamentali: anzitutto la diffusione degli strumenti informatici all'interno delle aziende e la dipendenza della continuità aziendale da tali strumenti e, in secondo luogo, la crescente esposizione dei sistemi e delle reti a rischi informatici quali virus, intrusioni e vulnerabilità.

3.3.3 Misure e contromisure

Le contromisure comprendono tutti gli interventi sui sistemi o sull'organizzazione, volti a contrastare attacchi di sicurezza e a raggiungere una migliore garanzia del rispetto dei requisiti di sicurezza. Viene usato normalmente anche il termine *misure di sicurezza*.

I requisiti di sicurezza riguardano tipicamente la confidenzialità, l'integrità, l'autenticità delle informazioni, il non-ripudio delle transazioni, la disponibilità delle informazioni, la continuità del servizio e il controllo degli accessi.

Il grado di sensibilità a questi requisiti e la loro criticità sono calati nei diversi contesti operativi dell'azienda, sia per rispondere a necessità operative o di business sia per motivi di natura legale. Un'attività essenziale legata all'applicazione della sicurezza aziendale riguarda la classificazione degli asset e, in particolare, delle informazioni in relazione alla sensitività dal punto di vista della sicurezza.

Possono essere adottate una serie di misure organizzative e tecnologiche, idonee a rafforzare il grado di rispondenza ai requisiti.

UN CASO DA MANUALE: MICROSOFT IIS UNICODE EXPLOIT / DOS ATTACK

Un attacco classico e molto diffuso è quello che sfrutta la vulnerabilità conosciuta come MS IIS Unicode Exploit, tuttora pericolosa per una lunga serie di server Web Internet Information Server (IIS) 4 e IIS 5 a cui non sono stati applicati gli interventi correttivi.

Questa vulnerabilità si presta a essere utilizzata per poter lanciare programmi ostili su macchine che ospitano server Web.

La maggior parte dei server Web IIS sono configurati con una struttura delle directory simile alla seguente:

C:\Inetpub

\scripts	**directory con permessi di esecuzione contenente gli script**
\wwwroot	**directory contenente il sito Web ospitato dal server**

Un noto attacco verso i server Web consiste nel tentare di attraversare l'albero delle directory, sfruttando la stringa *../..* per salire alla *directory "padre"*, con l'obiettivo di attivare l'esecuzione di un programma ostile, caricato in precedenza dall'attaccante.

Ad esempio digitando nel browser il seguente indirizzo:

http://www.unpatched_server.com/../miofile.txt

si dovrebbe accedere al file *C:\Inetpub\miofile.txt*, situato sotto la *root* del sito Web (ossia in una zona del disco nella quale un navigatore non dovrebbe accedere). Questo accesso è possibile poiché la stringa *".."* indica la *directory parent* sia in sistemi Windows sia in sistemi Unix.

Questo attacco sfrutta una vulnerabilità scoperta diversi anni orsono e perciò non è più attuale in quanto tutti i programmi server Web ricercano ed eliminano le stringa *../..* prima di effettuare il *parsing* della URL (e, cioè, prima di analizzare l'URL per comprenderne il contenuto). Una variante dell'attacco usa la codifica Unicode dei caratteri della URL per eludere i controlli del programma server Web: il codice *%5C* corrisponde al carattere / e quindi una stringa di caratteri *..%5C..%5C* corrisponde nuovamente a *../..*

Anche queste sequenze di caratteri sono oggi verificate dai server IIS durante il *parsing* e quindi l'attacco è prevenuto. È possibile sofisticare ulteriormente l'attacco, in modo da eludere i controlli, applicando la codifica Unicode due volte:

% = %25
5 = %35
C = %43

Da cui: *../.. = ..%5C.. = ..%25%35%43..*

Questa doppia codifica inganna i controlli effettuati dal *parser* e permette a un intrusore di invocare comandi del sistema operativo DOS sulla macchina che ospita il pro-

Qui di seguito sono elencate alcune *misure tecnologiche* (a scopo puramente esemplificativo), sottolineando che queste misure debbono essere sempre affiancate da adeguate *misure organizzative* (procedure) che ne consentono la corretta applicazione. La tecnologia offre, infatti, solo strumenti che debbono essere applicati correttamente, per evitare di vanificarne gli effetti o addirittura di esaltarne il livello di rischio.

Va sottolineato, in particolare, che le misure di protezione possono essere applicate a vari livelli di una piattaforma ICT.

gramma server Web. Questa intrusione è resa possibile per mezzo di URL che, attraverso il prompt di MSDos (cmd.exe richiamabile attraversando opportunamente l'albero delle directory), eseguono comandi singoli sul sistema che ospita il server Web.

In particolare occorre:

"fingere" di trovarsi nella directory scripts/, per ottenere i permessi di esecuzione,
http://www.upatched_server.com/scripts/
risalire di due directory nell'albero
http://www.upatched_server.com/scripts/..%25%35%43..%25%35%43../
e accedere a /winnt/system32,
http://www.upatched_server.com/*scripts/..%25%35%43..%25%35%43../winnt/system32*
dove è situato cmd.exe,
http://www.upatched_server.com/scripts/..%25%35%43..%25%35%43../winnt/system32/cmd.exe

Esso permette di eseguire comandi singoli se gli viene passato lo switch /c.

Per superare questo switch includiamo un carattere ? che funge da delimitatore per i parametri da passare:

http://www.upatched_server.com/scripts/..%25%35%43..%25%35%43../winnt/system32/cmd.exe?/c

È così possibile semplicemente aggiungere il comando da eseguire in coda alla URL costruita, con l'accortezza di usare il carattere + al posto degli spazi nella sintassi del comando (per assecondare ancora il comportamento del *parser*).

Se si desidera, ad esempio, elencare i file contenuti nella directory scripts/ basta invocare l'URL seguente:

http://www.upatched_server.com/scripts/..%25%35%43..%25%35%43../winnt/system32/cmd.exe?/c+dir

Per elencare i file contenuti nella directory C:/ si deve invocare l'URL:

http://www.upatched_server.com/scripts/..%25%35%43..%25%35%43../winnt/system32/cmd.exe?/c+dir+c:/

Un possibile attacco di interruzione del servizio, *DoS (Denial of Service)*, che sfrutta questa semplice vulnerabilità consiste nel far eseguire al server Web molti comandi che determinano l'apertura di connessioni che non sono chiuse. Il comando *comp* confronta, ad esempio, due file alla ricerca di differenze.

***Comp* gira in background in attesa di input e non termina automaticamente. Lanciando diverse istanze si satura il numero massimo di connessioni possibili per il sistema operativo, impedendo alle connessioni lecite di essere instaurate.**

Microsoft ha rilasciato la patch per questa vulnerabilità, reperibile all'indirizzo: *http://www.microsoft.com/technet/security/bulletin/MS01-026.asp*

Per chiarire quest'esigenza basta far riferimento ai numerosi contesti in cui può essere applicata la cifratura allo scopo di proteggere un sistema: si può intervenire cifrando la trasmissione su un *link* di una rete, oppure si può cifrare tutto il traffico IP impiegando il protocollo IPSEC, o, anche, si può introdurre la cifratura tra *client* e *server* utilizzando un protocollo specifico *end-to-end* come *SSL (Secure Socket Layer)*. In alternativa si può introdurre la cifratura a livello applicativo.

La scelta tra le diverse soluzioni, applicate eventualmente anche in combinazione tra loro, è frutto di

un'analisi del bilanciamento *(trade-off)* tra funzionalità e costi in cui tra le funzionalità venga inclusa la necessità di contrastare possibili attacchi di sicurezza.

a) Controllo dell'integrità

L'integrità dell'informazione, ossia la garanzia che l'informazione originale non abbia subito manomissioni, è di norma garantita applicando tecniche che associano all'informazione dei sigilli di controllo che ne attestano l'aderenza all'originale (ad esempio tecniche di *hashing*). L'intervento in campo informatico riproduce la metafora fisica dell'apposizione del sigillo di ceralacca sul bordo di una busta.

b) Controllo della confidenzialità

La confidenzialità delle informazioni, ossia la garanzia che le informazioni siano accessibili solo a chi è autorizzato, è ottenuta impiegando tecniche di cifratura dei dati e di gestione, relative ai permessi di accesso alle risorse. Queste tecniche si applicano ai dati sia negli archivi di memorizzazione che nei canali di comunicazione.

c) Controllo della autenticità

L'autenticità dell'informazione, ossia la garanzia che la paternità dell'informazione sia accertabile in modo incontrovertibile, è perseguita applicando tecniche di firma digitale. La firma digitale si effettua apponendo ai dati un sigillo che ne attesta l'origine.

Il contesto normativo nazionale ha stabilito le condizioni per attribuite alla firma digitale un valore legale. La firma digitale è introdotta con l'uso della tecnologia di crittografia a chiave pubblica o asimmetrica e con le *PKI (Public Key Infrastructure).*

d) Controllo della non ripudiabilità

La non ripudiabilità è la capacità di stabilire in modo certo che una determinata transazione è avvenuta e di associare a essa le entità coinvolte. Il non ripudio di sorgente è, ad esempio, la capacità di impedire che una sorgente di informazione neghi di averla generata o trasmessa. Il non ripudio di destinazione è, invece, la capacità di impedire che il destinatario di un'informazione neghi di averla ricevuta. Tecniche di firma digitale e di registrazione notarile digitale rispondono all'esigenza di garantire questo tipo di requisiti.

e) Filtraggio dei contenuti

I contenuti scambiati attraverso i sistemi telematici possono essere compromessi in modo da veicolare pericolosi attacchi alla sicurezza dei sistemi, normalmente chiamati *virus* per la loro capacità generica di riprodursi e di diffondersi, oltre che per la possibilità che essi hanno di generare danni di varia entità.

Il filtraggio dei contenuti, volto ad analizzarli e a prevenire che possano essere introdotti nei sistemi aziendali in presenza di indizi virali, è una tecnica diffusa, che principalmente fa uso delle tecnologie antivirus per sistemi terminali o per serventi di posta elettronica e siti Web.

f) Garanzia di disponibilità dei dati e della continuità dei servizi

La disponibilità dei dati e la continuità dei servizi rappresentano naturalmente elementi essenziali per l'operatività delle aziende; la perdita irreparabile di dati o di programmi può quindi costituire un danno di portata assai elevata. E, a questo scopo, tecniche che provvedono a proteggere i dati dai tentativi di alterazione o di distruzione, ovvero tecniche di replicazione dei dati (totale o parziale), sono spesso introdotte per far crescere il livello di protezione dei dati da incidenti di sicurezza o da eventi accidentali.

L'indisponibilità dei servizi dei sistemi telematici, siano essi strumentali al flusso di lavoro dell'azienda o siano il prodotto finale dell'attività dell'azienda, causa anch'essa perdite economiche di rilievo. Possono essere effettuati attacchi alla sicurezza volti espressamente a determinare il sovraccarico dei sistemi deputati all'erogazione di servizi in modo da precluderne la normale fruizione. Questi attacchi vanno sotto il nome di *denial of service*. Sono state messe a punto misure specifiche volte a prevenire o a contrastare queste categorie di attacchi di sicurezza.

L'indisponibilità di dati o di servizi può essere causata in maniera dolosa o accidentale; per contrastare gli effetti si applicano tecniche di *business continuity* che irrobustiscono i sistemi e i processi di business in modo da garantire che le discontinuità siano ridotte al minimo. Il *business continuity plan* definisce le procedure dettagliate da attuare in caso di emergenza. Il *disaster recovery* è, invece, il piano di azione da attivare nel caso di eventi ambientali catastrofici che compromettano le funzionalità dei siti presso cui si trovano i sistemi telematici o degli stessi sistemi. Le procedure di recupero della funzionalità dei processi può essere basata sull'impiego di risorse alternative mantenute in *backup caldo* oppure in *backup freddo* in una sede diversa.

g) Configurazione sicura dei sistemi

L'*hardening* è l'intervento di configurazione sicura dei sistemi, sia a livello di sistema operativo che a livello di software applicativo. Esso è volto a eliminare le vulnerabilità note e per le quali esistono opportune correzioni *(patch)* che provvedono a sanare la situazione di esposizione ad attacchi.

Quest'attività necessita di continui aggiornamenti, in quanto deve mantenersi al passo con la diffusione di nuove forme di attacco le cui caratteristiche sono divulgate, insieme ai relativi antidoti, da centri specializzati (quale, ad esempio, il CERT) [23] e dai fornitori dei sistemi.

h) Protezione perimetrale

L'infrastruttura telematica privata di un'azienda è in genere collegata con infrastrutture telematiche esterne, quali ad esempio Internet o la rete di altre aziende. La protezione perimetrale si concentra sull'applicazione di tecniche volte a garantire la sicurezza nei punti di frontiera del perimetro dell'infrastruttura telematica di un'azienda. Nei punti di contatto del perimetro dell'infrastruttura telematica dell'azienda con le infrastrutture esterne si focalizza lo sforzo di protezione. Tra le tecnologie impiegate in

questi punti possono essere ricordati i *firewall* e i sistemi *IDS (Intrusion Detection System)* descritti qui di seguito.

Nei sistemi di difesa perimetrale possono essere impiegati anche sistemi che svolgono una funzione di bersaglio diversivo, i cosiddetti sistemi *honeypot*, che cercano di attirare l'attenzione degli attaccanti e di mantenerli impegnati, prolungando la durata dell'operazione di attacco e guadagnando così tempo per le operazioni di risposta. Sui criteri base del sistema diversivo è stato sviluppato il progetto *Honeynet*, che ha lo scopo di studiare il comportamento degli attaccanti, per poter progettare le difese sulla base della conoscenza della dinamica delle azioni di attacco.

I *firewall* sono sistemi che controllano l'attraversamento del confine tra il mondo esterno e l'infrastruttura telematica dell'azienda, selezionando i flussi che legittimamente attraversano la frontiera e filtrando con varie tecniche tutti gli altri flussi.

I firewall sono anche impiegati per suddividere un dominio di rete in sottodomini, con due obiettivi: il primo provvede a separare i segmenti di rete che hanno criteri di sicurezza distinti e che appartengono a organizzazioni diverse della medesima azienda; il secondo realizza le zone di rete compartimentate in modo che l'accesso a una zona non implichi automaticamente la possibilità di accedere alle altre.

I sistemi firewall sono realizzati, in base alla complessità e ai volumi di traffico, impiegando architetture e tipologie di macchine distinte.

Il principio più elementare dei firewall impiega alcune regole di filtraggio del traffico che si presenta alle sue interfacce, basate sul contenuto dell'intestazione dei pacchetti IP e dei pacchetti a livello di trasporto TCP/UDP. Questa funzione è direttamente disponibile negli apparati convenzionali di interconnessione, i *router*, in cui si applicano delle *ACL (Access Control List)* ottenendo un router filtrante.

I principi più complessi di funzionamento dei *firewall* si avvalgono dell'impiego di *proxy* applicativi e della *stateful inspection*, ossia dell'analisi dei contenuti dei flussi di dati con verifiche che comportano il tracciamento dei dati scambiati da una sessione di comunicazione.

Si realizzano in genere "aree demilitarizzate", le *DMZ (De-Militarized Zone)*, che ospitano le macchine che debbono essere accessibili dall'esterno, separandole in modo netto da quelle che si vuole vengano escluse dall'accesso dall'esterno.

I sistemi firewall adottano varie tecniche di filtraggio basate su criteri sia statici che dinamici di trattamento selettivo dei flussi di dati, e possono intervenire a vari livelli del modello di riferimento *OSI (Open System Interconnection)*.

I sistemi di rilevazione delle intrusioni, *IDS (Intrusion Detection System)*, possono essere considerati il complemento logico delle misure di sicurezza preventive come, ad esempio, i firewall, poiché estendono la gestione della sicurezza includendo il monitoraggio dello stato dei sistemi e delle reti, il riconoscimento *(rilevazione)* e la risposta a eventuali attacchi o intrusioni *(risposta)*.

I sistemi di *intrusion detection* sono sistemi complessi in grado di analizzare i flussi che transitano in una porzione di rete, *NIDS (Network-based IDS)*, o di analizzare il comportamento di un sistema, *HIDS (Host-based IDS)*, per riconoscere sequenze di eventi che possono corrispondere a tentativi di attacco alla sicurezza.

La segnalazione di anomalie o di utilizzi impropri delle risorse telematiche è veicolata sotto forma di allarme ai sistemi di monitoraggio, attivando, così, l'intervento degli operatori addetti alla supervisione dei sistemi di sicurezza perimetrale.

I sistemi IDS devono essere tarati sull'effettiva composizione dell'infrastruttura telematica senza perturbarne le caratteristiche funzionali e deve essere effettuata un'operazione di *tuning*, atta a contenere entro limiti accettabili il fenomeno dei falsi allarmi *(falsi positivi)*, senza precludere la possibilità di intercettare e segnalare tentativi di attacco *(falsi negativi)*.

La rilevazione e la risposta alle intrusioni sono parte di un processo che coinvolge tecnologie, persone e strumenti. L'importanza di quest'affermazione risiede nel fatto che un processo implica un'interazione continua nel tempo di entità diverse e ciò rappresenta il problema più critico dei sistemi di IDS. La maggioranza degli incidenti di sicurezza, documentati e studiati in letteratura nei casi in cui si è riusciti a catturare la sorgente dell'attacco, ha comunque coinvolto tecniche manuali elaborate da personale esperto in sicurezza.

L'identificazione o la rilevazione di una intrusione può avvenire prima, durante o dopo che l'attività fraudolenta sia messa in atto: se l'identificazione avviene prima, allora l'intrusione può essere prevenuta e qualsiasi danno potenziale è evitato. Nel caso in cui l'identificazione avvenga durante la presenza di attività fraudolente, le problematiche riguardano le decisioni o di permettere tale attività, monitorandola e acquisendo informazioni sulla sorgente, oppure di procedere nella generazione di un allarme ovvero di bloccare l'intrusione stessa.

L'identificazione di un'intrusione, dopo che essa sia avvenuta, riguarda naturalmente la valutazione dell'impatto sulle risorse compromesse in termini materiali e immateriali. Una risposta è successiva a un'identificazione tranne i rari casi in cui il sistema di *intrusion detection* utilizza meccanismi di analisi predittiva, e le problematiche sono relative al fatto che la risposta persegua l'obiettivo di terminare l'attacco in corso *(reset)*, di individuare e catturare la sorgente dell'attacco *(counter-intelligence)* o di contrattaccare (azione necessaria in un contesto di *information warfare*).

L'attività fraudolenta può essere definita come un insieme di azioni volte a compromettere risorse di rete (dati, programmi, servizi erogati).

Le funzioni critiche svolte da un IDS sono quelle relative al monitoraggio, alla reportistica e alla risposta, cioè:

- *monitoraggio:* gli IDS esaminano ed elaborano informazioni riguardanti attività di risorse di rete sia in transito sia memorizzate su uno o più sistemi. Il grado di accuratezza e di riservatezza con cui queste informazioni sono analizzate ed elaborate (soprattutto attuando correlazioni tra eventi raccolti da più fonti o dalla stessa fonte in

istanti di tempo diversi), determinano la bontà del sistema nell'individuare correttamente un attacco;
- *reportistica:* gli IDS presentano informazioni, raccolte ed elaborate, sulle risorse di rete monitorate all'interno di un'infrastruttura di gestione della sicurezza;
- *risposta:* scopo ultimo degli IDS è di ridurre i rischi relativi alla sicurezza delle infrastrutture di rete. Con le funzioni di risposta si vogliono avviare attività volte a mitigare il rischio. Le problematiche relative alla tipologia di risposta di IDS riguardano la tempestività, l'accuratezza e la correttezza.

i) Reti private virtuali

Sempre più spesso si evita il ricorso a risorse di rete dedicate alle necessità di una sola organizzazione ma si adottano, piuttosto, tecniche che virtualizzano le risorse di rete in modo da allocare per ogni esigenza una rete privata virtuale.

Una *VPN (Virtual Private Network)* è una rete, ricavata "ritagliando" le risorse da una rete fisica in modo da conseguire i vantaggi della condivisione delle risorse con altre VPN, ma realizzando la suddivisione delle risorse di rete, in modo che vengano garantite le caratteristiche di una comunicazione su rete privata (rendendola possibile solo tra i punti che appartengono alla VPN), la confidenzialità e l'integrità dei dati.

Il protocollo IP offre diverse tecniche per realizzare VPN su reti IP non dedicate. L'uso del protocollo IPSEC consente di realizzare scambi di pacchetti IP cifrati ed eventualmente autenticati.

Il protocolli *MPLS (MultiProtocol Label Switching)* offre una via alternativa alla realizzazione di reti private virtuali in reti IP condivise, senza offrire la cifratura del traffico, ma utilizzando la tecnica di commutazione che crea percorsi virtuali, su cui sono instradati i pacchetti IP, mediante etichette aggiunte all'intestazione del pacchetto.

j) Controllo degli accessi

La regolazione dell'accesso alle risorse e ai servizi di un sistema distribuito rappresenta uno dei temi di maggior interesse nell'ambito della sicurezza. Gli aspetti da affrontare per applicare in modo corretto una politica di controllo degli accessi riguardano: l'*autenticazione*, ossia la verifica dell'identità riguardante un utilizzatore delle risorse del sistema, attraverso l'esame delle credenziali presentate, e l'*autorizzazione*, ossia l'applicazione di una politica di accesso per ogni entità qualificata come utente delle risorse del sistema distribuito.

Qui di seguito sono esaminati gli aspetti di maggior rilievo relativi a questi due controlli.

L'*autenticazione* è un processo complesso che normalmente è affrontato attraverso l'impiego di credenziali basate sulla coppia *userid* e *password*, ossia dell'identificativo dell'utente presso il sistema e della parola segreta per la verifica dell'identità.

L'efficacia di questa forma elementare dipende dall'applicazione di politiche intelligenti di gestione delle password (lunghezza, non inclusione nei dizionari delle lingue utilizzate, tecniche di memorizzazione, tasso di rinnovo delle password, non condivisione, cifratura delle password nella comunicazione e nell'archiviazione, *one-time usage*, …).

La forma elementare enunciata, basata sull'assunto che la parola d'ordine sia mantenuta segreta e sia difficilmente individuabile, può essere rafforzata affiancando al semplice principio del *what you know*, i due principi di potenziamento delle credenziali di sicurezza: il *what you have* e il *what you are*.

Il principio del *what you have* arricchisce le credenziali di sicurezza di un utente fornendogli un *token* (oggetto fisico) che deve essere da questo custodito in modo da impedirne l'accesso a terzi. Il *token* può essere una *memory-card* o una *smart-card* che debbono essere introdotte in un lettore oppure sono card che non richiedono un lettore (sono impiegate ad esempio carte dotate di un visore su cui appare una porzione di parola d'ordine da usare all'atto dell'autenticazione).

Il principio del *what you are* arricchisce le credenziali di sicurezza, facendo uso di elementi biometrici, ossia di elementi legati alle caratteristiche fisiche dell'utente che deve autenticarsi. Le applicazioni biometriche richiedono la presenza di sensori in grado di rilevare le caratteristiche fisiche di interesse, quali ad esempio la forma del viso, lo sfondo dell'iride dell'occhio, la geometria della mano o l'impronta digitale.

I sensori rilevano le caratteristiche biometriche e sono in grado di confrontare queste caratteristiche con le credenziali archiviate per gli utenti registrati e, quindi, di accertare la corrispondenza con l'identità dichiarata.

L'introduzione di elementi che rafforzano la tecnica base della *password* contrassegna le tecniche di autenticazione cosiddetta "forte".

A queste tecniche si affianca la tecnica dell'uso dei certificati digitali, utilizzati per autenticare utenti, programmi clienti e serventi, sfruttando le caratteristiche della crittografia a chiave pubblica.

Questa tecnica presenta la caratteristica interessante di facilitare l'uso combinato dell'autenticazione e della cifratura dei dati per la comunicazione attraverso canali di comunicazione non sicuri *(untrusted)*. Un'utente che dispone di un certificato digitale può firmare con la propria chiave privata una sequenza di *challenge* che gli viene inviata da un sistema cui l'utente desidera autenticarsi: l'utente è autenticato se il sistema è in grado di ricavare la sequenza di *challenge* decifrando il messaggio ricevuto mediante l'utilizzo della chiave pubblica dell'utente, contenuta nel suo certificato.

Una volta accertata l'identità dell'utente, attraverso la verifica delle credenziali di sicurezza, i sistemi ICT applicano le politiche di accesso selettivo alle risorse e ai servizi basate su mappe di corrispondenza tra l'identità dell'utente e la griglia di accesso (risorse e operazioni ammesse). Nei sistemi distribuiti complessi assume un particolare interesse la capacità di applicare l'*autorizzazione* specifica a una moltitudine di sottosistemi che compongono lo stesso sistema, come conseguenza di un'unica autenticazione. Questa caratteristica particolare è denominata *SSO (Single Sign-On)*.

L'uso di tecniche SSO coniuga una maggiore usabilità dei sistemi con l'impiego di sistemi di *AAA*

(Authorization Authentication Accounting) centralizzati e permette di godere di notevoli vantaggi dal punto di vista della sicurezza (registrazione centralizzata degli eventi di autenticazione e centralizzazione della gestione delle politiche sulle credenziali di sicurezza) e della gestione dei profili relativi all'autorizzazione degli utenti.

k) Sicurezza network-based

L'evoluzione delle forme di accesso alla rete e il passaggio dalla commutazione di circuito a quella di pacchetto portano a una maggiore diffusione delle tecniche di connettività, che si basano su una modalità di collegamento alla rete di tipo *always-on*, in cui il collegamento con la rete è sempre attivo a differenza dei sistemi di collegamento *dial-up* in cui il collegamento è espressamente attivato dall'utente con una chiamata verso un nodo di rete. Questa tendenza è accompagnata da un significativo aumento della capacità trasmissiva a disposizione in ogni accesso.

L'accesso xDSL, o quello in fibra, così come l'accesso *wireless GPRS (Generalised Packet Radio System)* testimoniano queste tendenze. La connettività *always-on* e l'impiego della larga banda accrescono in modo significativo il problema della sicurezza perimetrale, interessando da subito i segmenti di mercato *SME (Small Medium Enterprise)* e *SOHO (Small Office Home Office)* e, in prospettiva, il mercato residenziale.

In questi segmenti di mercato, le soluzioni di sicurezza debbono essere alla portata dei modesti budget ICT. Per predisporre soluzioni che possano migliorare la sicurezza in mercati che presentano queste caratteristiche, appaiono particolarmente convenienti le tecnologie che si basano sulla protezione *network-based*, che applicano, cioè, nei nodi di rete le funzioni che normalmente sono svolte nel sistema di difesa perimetrale disposto in corrispondenza della terminazione di rete, al confine della rete del cliente, presso le sue sedi (prassi convenzionalmente indicata con il termine di *CPE-based security*, ossia sicurezza basata sul *Customer Premises Equipment*).

Operazioni di *firewalling*, di analisi di contenuti e di *intrusion detection*, così come di VPN svolte nei nodi di rete, ben si prestano ad offrire sicurezza a un'estesa popolazione di clienti con una modesta capacità di spesa.

La protezione offerta dalla *network-based security* lascia scoperta la tratta che congiunge il rilegamento d'utente al nodo di rete, sulla quale debbono essere eventualmente attivate ulteriori misure di sicurezza.

l) Gestione della sicurezza

L'impiego di una qualunque misura di sicurezza all'interno di un'azienda di notevoli dimensioni risulta del tutto inefficace se non è garantita da una struttura centralizzata di gestione e coordinamento che si occupi delle seguenti attività:

- *pianificazione, organizzazione e controllo* e cioè definizione di attività mirate: a conseguire una presa di coscienza collettiva del problema della sicurezza all'interno dell'azienda (con l'organizzazione, ad esempio, di riunioni periodiche o di corsi di formazione per il personale); a definire, poi, un piano di sicurezza aziendale e a rivederlo periodicamente; a formalizzare ruoli e responsabilità; a definire attività periodiche di analisi e di gestione del rischio;
- *politiche e procedure di sicurezza* che riguardano regole e standard aziendali (codice di comportamento per i propri dipendenti, modalità di attuazione dei processi aziendali e delle singole attività operative, definizione e modalità di impiego delle misure di sicurezza adottate, gestione degli incidenti di sicurezza, gestione dei contratti …) e stesura dei relativi documenti;
- *accordi per l'accesso a reti e a servizi da parte di terzi (consulenti, clienti, partner commerciali)* che concerne: la definizione di un contratto, la formalizzazione delle rispettive responsabilità e del comportamento in caso di incidenti di sicurezza e l'indicazione delle politiche di sicurezza, adottate da entrambe le parti nei confronti del particolare servizio erogato o delle informazioni (relative a dati personali, a dati aziendali riservati, …) trattate attraverso il servizio stesso o a cui venga comunque fornita visibilità all'interno della rete;
- *formazione e gestione del personale* che consiste nelle misure di sicurezza adottate per la gestione del personale e, in particolare, l'indicazione di alcuni punti relativi, ad esempio, a responsabilità, formazione sul problema della sicurezza, controllo delle attività, gestione del personale neo-assunto o dimissionario.

3.3.4 Security audit

Un intervento di *security audit* consiste nell'attuare un'ispezione, asincrona o ripetuta a intervalli regolari di tempo, mirata ad accertare lo stato di sicurezza di un'azienda o di una porzione di un'azienda a cui l'operazione si applica.

Il *security audit* è quindi un'operazione di verifica, che normalmente indica le esposizioni e il portafoglio interventi da attuare, per accrescere la protezione e per ridurre il grado di rischio.

L'esecuzione di un processo formale di *security audit* (detta anche *security review*) deve essere giustificata da vantaggi e benefici per l'azienda soggetta a tale processo.

L'attività di *security audit* permette di valutare le misure di sicurezza adottate dall'azienda e la loro adeguatezza rispetto ai rischi cui l'azienda è esposta. Essa, in particolare, consente di:

- *giustificare i costi sostenuti per la sicurezza:* l'introduzione di nuove soluzioni di sicurezza ha, infatti, un costo che non genera direttamente profitti. Il costo deve essere giustificato in termini finanziari e il processo di *security audit* dovrebbe servire a validarlo e fornire suggerimenti per incrementarne il livello di sicurezza;
- *distribuire informazioni:* il processo di *security audit* permette infatti di informare e notificare i risultati ottenuti a diverse funzioni e livelli aziendali, promuovendo la creazione di relazioni intra-aziendali e svolgendo un ruolo attivo di coordinamento;
- *integrare la sicurezza all'interno della realtà aziendale*, attribuendo responsabilità ai diversi ruoli aziendali;

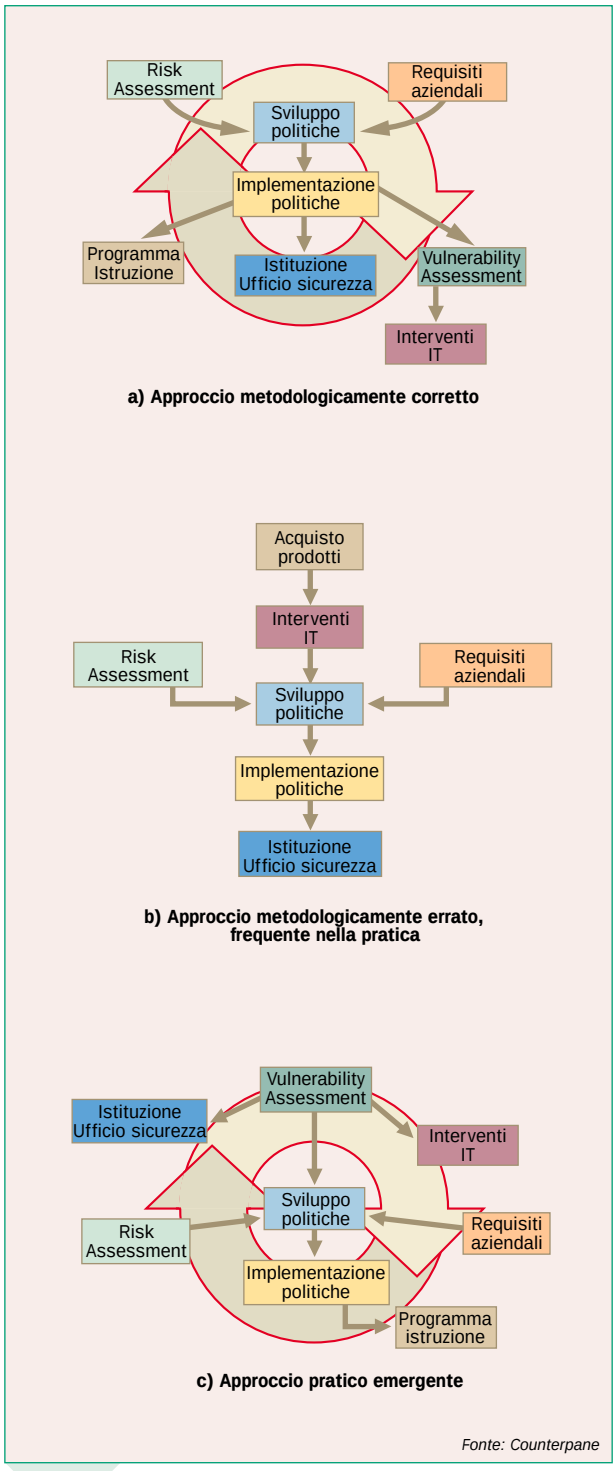


Figura 9 *Tra metodologie e pratica.*

- *incrementare la consapevolezza aziendale:* il processo di *security audit*, coinvolgendo diverse realtà aziendali e personale eterogeneo, promuove, infatti, approfondimenti inerenti la sicurezza e sensibilizza l'intera realtà aziendale;
- *identificazione degli obiettivi di sicurezza* correlati agli impatti potenziali, alle minacce e alle vulnerabilità esistenti. La definizione di obiettivi di sicurezza, non coerenti con le reali esigenze aziendali, potrebbe infatti condurre a spese eccessive o non necessarie ovvero a decisioni errate;
- *definire una security baseline:* le aziende in molti casi devono assicurare livelli preassegnati di sicurezza in quanto sono tenute a rispondere a obblighi legislativi, di regolamentazione interna, o ad accordi con terze parti;
- *accertare la conformità alle norme vigenti in materia di sicurezza:* la normativa nazionale prescrive, infatti, che le aziende adottino provvedimenti in materia di sicurezza e che vi sia evidenza dell'adozione delle misure previste dalla legge, sia da parte dell'azienda sia dalle figure alle quali è attribuito un ruolo indicato nelle norme per la sicurezza.

L'attività di *security audit* ha come obiettivo la verifica delle misure di sicurezza adottate dall'azienda, in termini organizzativi, logici e fisici e l'esame della coerenza di esse con gli obiettivi e gli indirizzi indicati dalle politiche di sicurezza e dai piani operativi aziendali.

Un *security audit* ha, anche, lo scopo di stabilire il grado di aderenza a standard internazionali, a vincoli legislativi, a norme interne o a vincoli contrattuali con terze parti. L'attività di *security audit* dovrebbe quindi essere svolta con cadenza periodica o a seguito di trasformazioni significative a livello organizzativo, tecnologico o strutturale.

3.3.5 Metodologie e approcci pragmatici

La modalità con cui le aziende affrontano il tema della sicurezza è estremamente varia (figura 9), soprattutto in dipendenza dal diverso grado di criticità che si attribuisce alla sicurezza, sia in relazione al *core business* sia rispetto a fatti occasionali che possono focalizzare l'attenzione dell'azienda sul tema della sicurezza.

Un'impostazione metodologicamente corretta presuppone che si affronti la sicurezza considerandola una caratteristica pervasiva del sistema aziendale in continua evoluzione. Partendo dall'analisi e dalla valutazione dei rischi, sotto la guida dei requisiti aziendali, l'approccio metodologico affronta lo sviluppo delle politiche e ne verifica l'attuazione, e comprendono: il programma d'istruzione, l'istituzione dell'ufficio di sicurezza e la verifica della tenuta dei sistemi di difesa.

Da quest'ultima attività si determinano gli interventi correttivi sulla piattaforma tecnologica. Data la continua evoluzione del sistema aziendale sotto i profili organizzativi e tecnologici, è necessario iterare il percorso metodologico delineato, con un impegno proporzionale all'intensità e alla frequenza dei cambiamenti endogeni ed esogeni.

Le aziende spesso affrontano la sicurezza partendo dall'acquisto di tecnologia che ha finalità di protezione degli asset aziendali (principalmente firewall).

L'introduzione di queste componenti tecnologiche ha impatti sull'architettura IT e rende opportuna l'adozione di politiche di sicurezza, sviluppate con il concorso dei risultati di attività di analisi del rischio e dei requisiti aziendali.

L'aggiustamento della componente organizzativa si attua generalmente a posteriori per conferire un ruolo istituzionale ad assetti organizzativi che si sono imposti nei fatti.

Questo percorso è in genere avviato per risolvere qualche situazione critica e si persegue, a questo scopo, solo una finalità tattica, escludendo cioè l'impostazione di attività continuative o ricorrenti.

Tra queste due impostazioni, nella pratica si va affermando un percorso intermedio che prende le mosse dall'attività di verifica della tenuta dei sistemi di difesa, allo scopo di dare corpo alle vulnerabilità tecnologiche e ai rischi ad essi associati. Le risultanze di quest'analisi, che normalmente conducono a verdetti di rischio, che non possono essere più ignorati, offrono lo spunto per dedicare risorse alla sicurezza, non limitandosi ad attuare interventi tecnologici, per mettere riparo alle situazioni più rischiose, ma procedendo all'istituzione dell'ufficio sicurezza e allo sviluppo delle politiche da esso perseguite. Segue, poi, la realizzazione pratica delle politiche, accompagnata dall'attuazione del programma d'istruzione. Il percorso è impostato consapevolmente in modo da ripercorrere le tappe delineate in sintonia con l'evoluzione del sistema aziendale e dell'ambiente in cui esso si colloca.

4. Esternalizzazione della gestione della sicurezza

La sicurezza di un sistema telematico investe in modo pervasivo tutte le componenti del sistema, sia architetturali e tecnologiche, sia organizzative. La sicurezza interessa, altresì, gli aspetti più intimi e di dettaglio delle varie componenti, richiedendo non solo una visione di insieme, ma anche una conoscenza specialistica.

La sicurezza del sistema è, poi, influenzata sia dal tasso di evoluzione delle caratteristiche degli apparati (innovazione tecnologica e mutazioni di configurazione), sia dalla continua scoperta di vulnerabilità che espongono il sistema ai rischi di un attacco.

Queste motivazioni, unite alla scarsa attenzione del mondo della formazione di base ai temi della sicurezza, concorrono a mettere in luce la complessità e la ricchezza delle competenze richieste al personale che professionalmente deve occuparsi della sicurezza dei sistemi telematici a tutti i livelli (analisi, progettazione, realizzazione, gestione).

I professionisti della sicurezza sono, quindi, assai rari e "l'aggiornamento e la manutenzione della loro competenza" ha costi elevati.

La scarsità e l'elevato costo delle risorse per la sicurezza, unite al fatto che generalmente tale aspetto non costituisce il *core business* dell'azienda, stimolano il ricorso all'esternalizzazione *(outsourcing)* delle funzioni esecutive dedicate alla sicurezza.

Questo trasferimento di responsabilità costituisce un'operazione molto delicata vista la sensibilità del tema. Accanto ai ruoli tradizionali della consulenza, della fornitura di apparati e della *system integration*, si va così affermando il ruolo dei cosiddetti *MSP (Managed Security Provider)* e, cioè, aziende in grado di gestire la sicurezza perimetrale (ossia la difesa della rete aziendale dagli attacchi esterni) configurando, aggiornando e monitorando con continuità (h24 x 365gg) il sistema di protezione.

Gli esperti della sicurezza di un MSP, operando da un *Security operation center* attraverso collegamenti di rete, mettono la propria competenza al servizio di molti clienti che si ripartiscono così i costi ad essi relativi.

Tra i servizi correntemente erogati sono compresi: la verifica della vulnerabilità dei sistemi perimetrali agli attacchi noti di sicurezza e la gestione dei *firewall*, dei sistemi antivirali, delle reti private virtuali e dei sistemi di rilevamento intrusioni.

5. Conclusioni

Un'azienda che intenda essere partecipe al contesto di business emergente non può prescindere dall'affrontare il tema della sicurezza in accordo con la propria prospettiva e con le proprie esigenze.

Il dominio del tema della sicurezza si presenta come una sfida complessa e importante, soprattutto per la necessità di interessare in modo profondo le componenti tecnologiche, organizzative, di business e legali di un'azienda.

Interventi estemporanei, frammentari o superficiali non offrono, infatti, garanzie per la sicurezza dell'azienda in rete che richiede viceversa un approccio pervasivo, continuo e competente.

La complessità del tema si presta a essere affrontata da attori fortemente specializzati e in grado di offrire le necessarie garanzie di protezione. La corsa al business della security è iniziata dopo che si è smorzata, anzi si è invertita, la corsa alla *net economy*, principalmente per le distorsioni e per le disillusioni da essa indotte.

Il mercato della sicurezza è certamente in crescita, anche se si è in attesa che esso si consolidi in modo che faccia emergere i protagonisti che stabilmente opereranno nel mercato.

Bibliografia

[1] Tipton, H.F.; Krause, M.: *Information Security Management Handbook* - 4th Edition. Auerbach, 2001.

[2] Stinson, D.R.: *Cryptography Theory and Practice.* CRC Press, Inc., Boca Raton, Florida, 1995, 1996.

[3] Menezes, A.J.; van Oorschot, P.C.; Vanstone, S.A.: *Handbook of applied cryptography.* CRC Press, Inc., Boca Raton, Florida, 1996, 2001.

[4] Schneier, B.: *Applied Cryptography*. John Wiley and Sons, 1994, 1996.
[5] Byrnes, C.; Kutnick, D.: *Securing Business Information: Strategies to Protect the Enterprise and Its Network*. Intel Press, 2001.
[6] McClure, S.; Scambray, J.; Kurtz, G.: *Hacking Exposed: Network Security Secrets & Solutions*. Terza edizione, Osborne McGraw-Hill, 2001.
[7] Howard, J.D.; Longstaff, T.A.: *A Common Language for Computer Security Incidents*. Sandia National Laboratories, Sandia Report: SAND98-8667, 1998.
[8] Arbaugh, W.A.; Fithen, W.L.; McHugh, J.: *Windows of Vulnerability: A Case Study Analysis*. «IEEE Computer», Vol. 33, n. 12, pp. 52- 59, dicembre 2000.
[9] *Funzioni e profilo del professionista della security aziendale*. UNI 10459, 1995.
[10] *Tutela delle persone e di altri soggetti rispetto al trattamento dei dati personali*. Legge n. 675, dicembre 1996.
[11] *Modificazioni ed integrazioni alle norme del codice penale e del codice di procedura penale in tema di criminalità informatica*. Decreto Legislativo n. 547, dicembre 1993 (G.U. n. 305 del 30 dicembre 1993).
[12] *Regolamento recante norme per l'individuazione delle misure minime di sicurezza per il trattamento dei dati personali*. A norma dell'articolo 15, comma 2, della legge 31 dicembre 1996, n. 675, settembre 1999, DPR n. 318 (G.U. n. 216).
[13] *Nuove norme di tutela del diritto d'autore*. Legge 18, agosto 2000, n. 248 (G.U. n. 206 del 4 settembre 2000).
[14] *Attuazione della Direttiva 91/250/CEE relativa alla tutela giuridica dei programmi per elaboratori*. Decreto Legislativo n. 518, dicembre 1992.
[15] *Testo unico delle disposizioni legislative e regolamentari in materia di documentazione amministrativa*. DPR n. 445, dicembre 2000.
[16] *Regolamento contenente i criteri e le modalità per la formazione, l'archiviazione e la trasmissione di documenti con strumenti informatici e telematici a norma dell'articolo*. DPR n. 513, novembre 1997 (G.U. n. 60, serie generale, del 13 marzo 1998).
[17] Articolo 17 del Decreto del Presidente della Repubblica 10 novembre 1997, n. 513: *Utilizzo della firma digitale nelle Pubbliche Amministrazioni*. Circolare AIPA n. 27 del febbraio 2001 (G.U. n. 47 del 26 febbraio 2001).
[18] *Direttiva sul commercio elettronico*. Direttiva del Parlamento Europeo e del Consiglio n.2000/31/CE, giugno 2000 (G.U. CE n. L 178 del 17 luglio 2000).
[19] *Commercio elettronico - Disciplina della vendita di beni tramite mezzo elettronico*. Decreto Legislativo n. 114, 31 marzo 1998, Ministero dell'Industria, del Commercio e dell'Artigianato, Circolare n. 3487/C del 1° giugno 2000, (G. U. n. 174 del 27 luglio 2000).
[20] Peltier, T.R.: *Information Security Risk Analysis*. Auerbach, 2001.
[21] Information Technology*: Code of Practice for Information Security Management*. First Edition 12/2000, ISO/IEC 17799, *www.bspsl.com/17799/*.
[22] Information Security Management - Part 1: *Code of practice for information security management*. BS7799-1: 1999.
[23] Information Security Management - Part 2: *Specification for information security management systems*. BS7799-2: 1999.
[24] *Computer Emergency Response Team / Coordination Center*. *www.cert.org*.
[25] Guidelines for the Management of IT Security (GMITS). ISO/IEC 13335, nelle cinque parti: *Concepts and Models for IT Security*. ISO 13335-1: *1996 Managing and Planning IT Security*. ISO 13335-2: *1997 Techniques for the Management of IT Security*. ISO 13335-3: *1998 Selection of Safeguards*. ISO 13335-4: *2000Safeguards for External Connections*. ISO 13335-5: 2001.
[26] Information technology - Security techniques: *Evaluation criteria for IT security*. ISO IS 15408, 1999 (corrispondente alla norma Common Criteria version 2.1).
[27] *Trusted Computer System Evaluation Criteria* [TCSEC]. DoD 5200.28-STD, dicembre 1985, *www.radium.ncsc.mil/tpep/library/rainbow/*.
[28] *Information Technology Security Evaluation Criteria* [INFOSEC]. ISBN 92 826 3004 8, versione 1.2, Lussemburgo 1991, in *www.itsec.gov.uk/*.

Abbreviazioni

AAA	Authorization Authentication Accounting
ACL	Access Control List
AIEA	Associazione Italiana EDP Auditor

AIIA	Associazione Italiana Internal Auditors
BCP	Business Continuity Planning
BSI	British Standard Institution
CC	Common Criteria
CERT/CC	Computer Emergency Response Team/Coordination Center
CIA	Certified Internal Auditor
CISA	Certified Information Systems Auditor
CISSP	Certified Information System Security Professional
CoBiT	Control OBjectives for Information and related Technology
CPE	Customer Premises Equipment
CSI	Computer Security Institute
DMZ	De-Militarized Zone
DNS	Domain Naming System
DoD	Department of Defense
DoS	Denial of Service
FTI	Forum per le Tecnologie dell'Informazione
G8	Group of Eight
GMITS	Guidelines for the Management of IT Security
GPRS	Generalised Packet Radio System
HIDS	Host-based Intrusion Detection System
IDS	Intrusion Detection System
IIA	Institute of Internal Auditors
IP	Internet Protocol
IPSEC	IP Secure
ISACA	Information Systems Audit and Control Association
(ISC)2	International information systems Security certification Consortium
ISMS	Information Security Management System
ISO	International Standard Organization
ISSA	Information Systems Security Association
ITSEC	Information Technology Security Evaluation Criteria
ITSEM	Information Technology Security Evaluation Manual
MPLS	MultiProtocol Label Switching
MSP	Managed Security Provider
NIDS	Network-based Intrusion Detection System
OCI	Osservatorio per la Criminalità ICT
OSI	Open System Interconnection
PKI	Public Key Infrastructure
PP	Protection Profile
SME	Small Medium Enterprise
SOC	Security Operation Center
SOHO	Small Office Home Office
SSL	Secure Socket Layer
SSO	Single Sign-On
TCP	Transmission Control Protocol
TCSEC	Trusted Computer System Evaluation Criteria
TOE	Target Of Evaluation
UDP	User Datagram Protocol
VAPT	Vulnerability Assessment & Penetration Testing
VPN	Virtual Private Network
WWW	World Wide Web

Maurizio Dècina è professore ordinario al Politecnico di Milano, Facoltà di Ingegneria, dove è titolare del corso di "Reti per Telecomunicazioni". È il direttore scientifico del CEFRIEL, un Centro di Ricerca e Formazione per laureati in tecnologia dell'informazione a cui partecipano il Politecnico di Milano e le maggiori aziende del settore telecomunicazioni e informatica. È stato Presidente della "Communications Society" dell'*IEEE (Institute of Electrical and Electronics Engineers)* per gli anni 1994 e 1995, e direttore della rivista tecnica internazionale "European Transactions on Telecommunications", per gli anni 1991-1997. Il prof. Dècina ha lavorato nell'industria (Telecom Italia, Italtel e AT&T) ed è consulente tecnico scientifico di varie aziende nazionali ed internazionali. Da molti anni collabora con l'*ITU (International Telecommunication Union)* di Ginevra nel settore degli standard e come consulente esperto di cooperazione internazionale. Ha scritto in italiano e in inglese vari libri e un centinaio di pubblicazioni tecniche, che coprono il campo delle telecomunicazioni. Nel 1986 è stato nominato "Fellow" dell'IEEE, mentre nel 1997 e nel 2000 gli sono stati assegnati i premi dell'IEEE "Award in International Communications" e "Third Millennium Medal Award".

Vittorio Trecordi si è laureato nel 1986 al Politecnico di Milano in Ingegneria Elettronica e delle Telecomunicazioni e ha conseguito il Master in Tecnologia dell'Informazione presso il centro CEFRIEL nel 1989. Ha svolto attività di ricerca e sviluppo nel settore delle telecomunicazioni presso Sirti, LABEN e Pirelli. Dal 1991 al 1998 è stato impiegato presso il CEFRIEL ove ha ricoperto l'incarico di direttore tecnico. Nel 1998 è tra i fondatori della ICT Consulting, Società di consulenza focalizzata sul settore delle reti e della sicurezza informatica. Alla fine del 2000 è stato tra i fondatori di «Securmatics», azienda spin-off' di ICT Consulting dedicata ai servizi di gestione remota della sicurezza delle reti dei clienti. È autore di numerose pubblicazioni tecniche su riviste e atti di congressi internazionali. Ha fatto parte del comitato tecnico di conferenze internazionali ed è membro del Comitato Editoriale della rivista "IEEE Network Magazine". Le attività professionali di maggior rilievo includono: la consulenza per la Rete Unitaria della Pubblica Amministrazione, per la rete dei Ministeri delle Finanze, del Lavoro e della Pubblica Istruzione, per la rete del Politecnico di Milano e del Comune di Milano, nonché la partecipazione al Comitato per la Cablatura del Comune di Milano. Da nove anni è professore a contratto presso il distaccamento di Cremona del Politecnico di Milano nell'ambito del corso di Laurea in Ingegneria Informatica ed Automatica.

Realizzazione e visualizzazione remota di rappresentazioni interattive di oggetti e luoghi

GUIDO MARIA CORTELAZZO

Le rappresentazioni interattive sono uno strumento che consente di mostrare oggetti o realtà distribuite nello spazio in modo più coinvolgente ed efficace di foto e filmati. Ciò è dovuto alla loro abilità di imitare le modalità con cui gli individui ispezionano le articolazioni spaziali della geometria meglio di ogni altro mezzo iconografico.

Questo lavoro presenta le caratteristiche delle due tipologie di rappresentazioni interattive attualmente utilizzate: ossia le rappresentazioni a modelli e quelle ad immagini. Il loro potenziale applicativo con l'affermarsi di Internet è amplissimo. Particolare attenzione viene dedicata alle applicazioni ai beni culturali per l'interesse che rivestono per il nostro Paese.

A fronte della banda dei sistemi Internet attuali, è opionione comune che il più grosso ostacolo alla diffusione delle rappresentazioni interattive sia costituito dal protocollo di visualizzazione remota corrente. Questo problema, la cui soluzione richiede competenze principalmente di telecomunicazioni, è messo in evidenza. Possibili soluzioni richiedono di affrontare problemi di compressione non standard; infatti per le rappresentazioni a modelli sono necessari nuovi algoritmi di compressione, dedicati alle viste generate da modelli tridimensionali (3D), e per le rappresentazioni a immagini nuovi algoritmi di compressione dedicati a famiglie di immagini relative allo stesso soggetto (ma non simili come le immagini delle sequenze video). Le soluzioni proposte per superare i problemi di visualizzazione remota delle rappresentazioni interattive sui sistemi Internet attuali, possono essere estese alla visualizzazione remota di rappresentazioni interattive verso dispositivi con risorse limitate di calcolo e memoria di tipo generale, quali ad esempio telefoni cellulari, computers palmari e simili.

1. Introduzione

Per introdurre il concetto di rappresentazione interattiva di oggetti e luoghi (tanto difficile a descriversi a parole quanto immediato a riconoscersi nell'uso diretto), conviene partire dalle riflessioni di Bruno Zevi, nel suo libro degli inizi degli anni Sessanta dal titolo: "Saper vedere l'architettura".

In quell'opera Zevi osserva che la percezione dello spazio è caratterizzata dal fatto che è l'individuo a scegliere autonomamente la sua posizione o la sua traiettoria nello spazio tra le infinite possibili e lamenta i limiti fondamentali degli strumenti iconografici tradizionali per presentare l'architettura, ossia disegni tecnici, foto e filmati. In buona sintesi, i disegni tecnici danno una rappresentazione schematica dello spazio comprensibile solo a chi possiede una specifica preparazione tecnica. Le foto danno viste singole dei luoghi, e i filmati danno una sola delle infinite traiettorie spaziali secondo cui un luogo può essere visitato (la traiettoria seguita dall'operatore). Naturalmente le osservazioni di Zevi sulle limitazioni del senso dello spazio relative alle rappresentazioni tradizionali dell'architettura, si possono applicare anche alle rappresentazioni iconografiche convenzionali di oggetti o di scene tridimensionali.

I computers contemporanei, dietro richiesta dell'utente, sono in grado di presentare e di calcolare in frazioni di secondo immagini relative a viste di oggetti e luoghi. Questa possibilità consente un nuovo tipo di presentazioni iconografiche di oggetti e luoghi, chiamate rappresentazioni interattive o anche visite virtuali[1], caratterizzate da una forte interattività con l'utente.

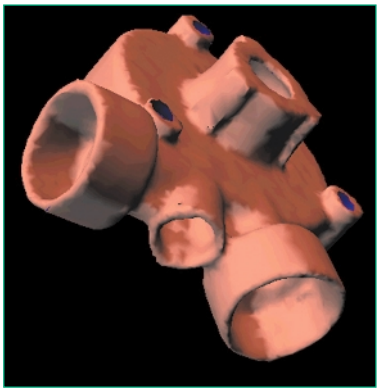
(1) *Per la precisione questi metodi si riferiscono alla realtà virtuale di tipo non-immersivo, caratterizzata dal non richiedere particolari attrezzature per essere fruita (a parte, ovviamente, un Personal Computer).*

Le rappresentazioni interattive non sono certamente equivalenti o sostituibili all'ispezione diretta; tuttavia per rappresentare oggetti o realtà distribuite nello spazio, esse sono più chiare e coinvolgenti di foto e filmati perché sono in grado di imitare meglio di ogni altro strumento iconografico la libertà con cui gli individui seguono e ispezionano le articolazioni dello spazio. La loro efficacia rappresentativa e il loro impatto emotivo possono essere di particolare utilità in molte applicazioni.

Un grande vantaggio di questo metodo di rappresentazione interattiva consiste nella completa libertà di punti di vista, della distanza di osservazione e del tipo di illuminazione nei confronti di ciò che i modelli rappresentano.

È inoltre possibile costruire anche modelli di oggetti e luoghi che non esistono. In effetti è proprio quest'applicazione che ha fatto da propulsore alla modellazione tridimensionale sintetica, campo ormai tradizionale della grafica computerizzata.

a) "Testa di Medusa" di Gian Lorenzo Bernini



b) Pompa

c) Cranio proveniente da una tomba etrusca

Figura 1 *Viste virtuali calcolate da modelli tridimensionali.*

Al momento ci sono due tipologie di rappresentazioni interattive: le rappresentazioni a modelli, dall'inglese *model-based* e quelle a immagini, dall'inglese *image-based* le cui caratteristiche sono richiamate in questo lavoro nei paragrafi 2 e 4. Il paragrafo 3 è dedicato alle applicazioni dei modelli 3D. Particolare attenzione viene dedicata alle applicazioni dei modelli 3D ai beni culturali, visto l'interesse di questo argomento per il nostro territorio. Il paragrafo 5 considera le difficoltà che si incontrano nella fruizione remota di rappresentazioni interattive. Il paragrafo 6 contiene le conclusioni e alcune osservazioni sul ruolo che possono giocare le telecomunicazioni rispetto alle rappresentazioni interattive.

2. Rappresentazioni interattive a modelli

Le visite virtuali "a modelli" ricorrono a modelli matematici tridimensionali delle superfici degli oggetti o dei luoghi per rappresentarli. La figura 1a mostra un esempio di vista calcolata da un modello tridimensionale della "Testa di Medusa" di Gian Lorenzo Bernini. La figura 1b una vista calcolata del modello tridimensionale di una pompa e la figura 1c una vista del modello tridimensionale di un cranio proveniente da una tomba etrusca. Un modello tridimensionale di un "Apollo" (scultura greca del V secolo a.C.) è disponibile nel sito *http://freia.dei.unipd.it/cultural-heritage.htm*.

L'industria del cinema è stata una delle prime ad utilizzare questa tecnologia, prima con i film di fantascienza e oggigiorno anche con i cartoni animati. Gli spot video in cui spesso i grafici pubblicitari ricorrono a modelli 3D per ottenere effetti spettacolari sono nell'esperienza di tutti. La ricostruzione di ambienti che non esistono ha in seguito trovato particolare interesse in archeologia e nella presentazione di beni culturali (vedi paragrafo 3).

Per quanto possa apparire paradossale, la realizzazione di modelli 3D di oggetti o luoghi esistenti pone molti più problemi della realizzazione di oggetti o luoghi che non esistono. In effetti la realizzazione di modelli 3D di oggetti è una disciplina recente, che di fatto utilizza tecniche di misura delle superfici, per la quale gli strumenti tradizionali della grafica computerizzata non bastano, ma devono essere integrati da strumenti e da tecniche di visione elaborativa e di trattamento delle immagini.

La procedura standard per realizzare un modello 3D della superficie di un oggetto richiede quattro fasi. La prima consiste nell'acquisire le coordinate di nuvole di punti dense della superficie dell'oggetto, tipicamente dette immagini 3D (figura 2a). Gli strumenti che svolgono questa operazione sono chiamati *range cameras*. Ve ne sono di varie tecnologie, prestazioni e costi [1]. Caratteristiche importanti sono la portabilità, la precisione e la possibilità di catturare anche informazioni sul colore dei punti acquisiti.

Esistono *range cameras* portatili con capacità di acquisire con precisione di 30 μ per piccoli volumi (10x10x10 cm) e con informazione sul colore. Per avere il modello della superficie di un oggetto a tutto tondo è necessario acquisire una sequenza di immagini 3D che lo ricoprano. È anche fondamentale che le immagini 3D acquisite abbiano regioni di sovrapposizione.

Le regioni di sovrapposizione vengono utilizzate nella seconda fase della modellazione per registrare le immagini 3D in un unico sistema di riferimento (figura 2b). Infatti, durante l'acquisizione ogni immagine 3D nasce in un proprio sistema che ignora i sistemi delle altre immagini 3D. La registrazione delle immagini 3D viene fatta in genere con la supervisione di un operatore [2-7]. Un fronte di ricerca corrente riguarda gli algoritmi per la registrazione automatica delle immagini 3D [8-10].

L'insieme delle immagini 3D registrate tra loro dà luogo a un insieme di nuvole di punti parzialmente sovrapposte che costituisce una rappresentazione ridondante e di aspetto visivo poco intuitivo. A queste difficoltà risponde la terza fase della modellazione 3D, che consiste nell'interpolare in un'unica superficie tipicamente formata da una griglia di triangoli o di poligoni di grado basso le nuvole di punti (figura 2c). Per questa operazione esistono algoritmi che operano in modo automatico [6]-[7]. Se i dati 3D raccolti contengono informazioni di colore, è necessario ricorrere a strategie che ne combinino opportunamente i valori sulle regioni di sovrapposizione senza creare artefatti visivi. Questo problema è oggetto di attiva ricerca. Le tecniche più efficaci ricorrono a stime della riflettanza dei materiali [7].

Per oggetti reali a geometria articolata è di fatto inevitabile che le superfici costruite a partire dai dati presentino lacune. La quarta fase della modellazione (figura 2d) consiste nella chiusura delle lacune tramite opportuni strumenti interattivi. Questa è un'operazione, tipicamente chiamata correzione *(editing)*, di difficile automazione, data la grande varietà di lacune che si possono presentare a fronte della geometria degli oggetti.

La costruzione di modelli 3D, a seconda delle dimensioni degli oggetti e delle precisioni desiderate, che condizionano direttamente i volumi di ripresa e quindi il numero di immagini 3D da acquisire e successivamente da comporre in un'unica superficie, richiede tipicamente quantità di lavoro non banali. La sostanziale semplificazione delle procedure per la realizzazione di modelli tridimensionali, con l'obiettivo finale della completa automazione, costituisce un obiettivo perseguito al momento da tutti i ricercatori che operano in questo settore. Non sono disponibili soluzioni robuste e generali al momento.

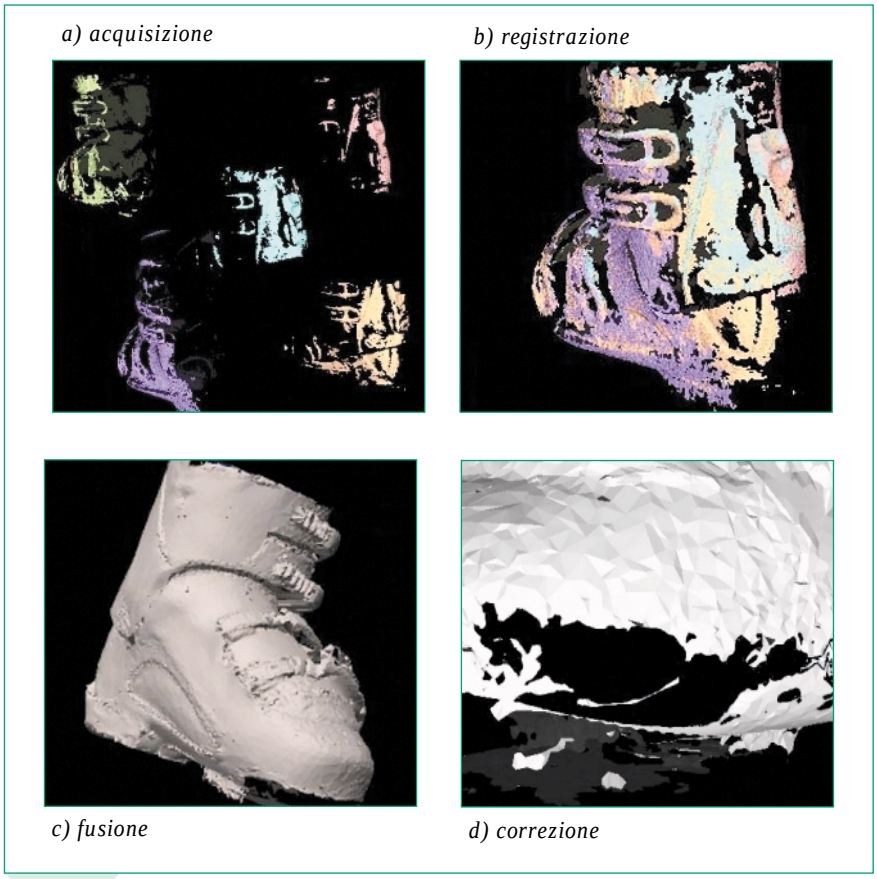


Figura 2 *Procedura per realizzare un modello 3D della superficie di un oggetto.*

Per modelli 3D di precisione non elevata sono molto promettenti i metodi che usano la teoria delle superfici di rivoluzione [11-12], che hanno tra l'altro il vantaggio di richiedere semplici fotocamere come strumenti di acquisizione.

Modelli 3D di oggetti di grandi dimensioni e articolati - quali edifici e monumenti - sono fattibili; al momento tuttavia le difficoltà ed i costi di realizzazione rendono di fatto proponibili i modelli 3D solo per oggetti di dimensioni limitate.

3. Applicazioni dei modelli 3D

I metodi e gli strumenti per la realizzazione di modelli tridimensionali sono tipici di alcune applicazioni dell'industria meccanica, quali la prototipazione rapida (ossia la realizzazione fisica in tempi brevi di pezzi meccanici, progettati al computer) e la metrologia industriale (ossia la misurazione con precisioni micrometriche delle forme di manufatti industriali).

I modelli tridimensionali sono anche usati per la duplicazione fisica di oggetti in luoghi diversi da quello in cui si trovano (applicazione denominata FAX 3D). Nell'industria cinematografica dei cartoni animati sono tipicamente impiegati modelli 3D di pupazzetti realizzati da scultori, che vengono in seguito animati tramite computer. L'industria aerospaziale di punta (ad esempio, la NASA) usa modelli tridimensionali per la documentazione dei pezzi meccanici ed è previsto che questo modo di operare con il tempo diventi comune in questo settore e che si diffonda anche a settori vicini dell'industria meccanica. Un altro ambito in cui vengono impiegati i modelli tridimensionali è l'antropometria, ossia le misure di parti del corpo umano, con applicazioni sul versante storico (figura 1c), medico e in quello merceologico.

Figura 3 *Snapshots del modello della Cappella Ovetari con gli affreschi di Andrea Mantegna distrutti in seguito al bombardamento del 1944.*

Una lista esaustiva degli ambiti applicativi correnti dei modelli tridimensionali è al di fuori dei nostri scopi; basta qui segnalare che le applicazioni sono già molte, in settori anche molto lontani tra loro e che certamente sono destinate a crescere con il perfezionarsi della tecnologia e la diminuzione dei costi di realizzo.

L'applicazione dei modelli 3D ai beni culturali merita un commento particolare per l'interesse che riveste per il nostro territorio. Il realizzare modelli 3D di edifici mai costruiti è particolarmente importante per l'architettura (è piuttosto tipico - emblematico a questo riguardo è il caso del Palladio - che solo una frazione dei progetti di un architetto venga effettivamente costruita). Il realizzare modelli 3D di luoghi non più esistenti è importante per l'archeologia e, in generale, per il patrimonio artistico.

Un esempio è dato dal caso della Cappella Ovetari, annessa alla Chiesa degli Eremitani in Padova. Gli affreschi di Cappella Ovetari furono eseguiti per la maggior parte da Andrea Mantegna ventenne, che operava allora accanto a Nicolò Pizzolo, pittore più anziano di lui e sua guida ufficiale in quel progetto. Gli affreschi di Cappella Ovetari sono stati da sempre considerati un capolavoro della pittura italiana, spesso accostato a quelli di Masaccio e Masolino nella Cappella Brancacci a Firenze, dato che in entrambi i casi si registra la vicinanza di un allievo e di un maestro, con dimostrazioni inconfondibili della superiorità del primo rispetto al secondo. Purtroppo una bomba sventrò Cappella Ovetari l'11 marzo del 1944 e distrusse completamente la maggior parte degli affreschi del Mantegna. Per un fortunato caso una troupe di fotografi aveva ripreso a colori gli affreschi la settimana precedente il bombardamento. Queste foto costituiscono la documentazione iconografica di un libro di Giuseppe Fiocco [13].

A partire dalle foto di questo libro nel 1996, in una tesi di laurea volta a sperimentare le possibilità della modellazione 3D, è stato realizzato un modello 3D della Cappella Ovetari con gli affreschi del Mantegna prima del bombardamento. L'ispezione di questo modello 3D di cui la figura 3 mostra alcuni *snapshots*, è il miglior modo oggi disponibile per apprezzare la relazione tra gli spazi della Cappella e gli affreschi del Mantegna descritta con ammirazione da tutti gli storici che poterono vederli intatti.

Per la scultura i modelli 3D probabilmente forniscono la forma di documentazione più pregnante di informazione che si conosca: basta ricordare che un modello tridimensionale consente di ricavare lette-

ralmente infinite immagini o filmati video e consente anche di duplicare fisicamente la scultura (figura 4). Aspetto rilevante di queste tecniche è che esse consentono la duplicazione fisica senza contatto con gli oggetti.

Infine i modelli tridimensionali dei beni culturali hanno il potenziale di essere fruiti tramite Internet o CD-ROM. Questa possibilità una volta che i problemi di compressione dei dati, relativi alle rappresentazioni interattive [14] siano stati adeguatamente risolti, si presta a essere messa a frutto concretamente per scopi di promozione, presentazione e gestione su rete, con risvolti applicativi ed economici di tutto rispetto.

Figura 4 *La figura entro il cerchio mostra la duplicazione, in materiale plastico in scala 1:10, della "Madonna con Bambino" di Giovanni Pisano (Cappella degli Scrovegni, Padova) accanto alla rappresentazione interattiva del modello da cui è stata ricavata, mostrata a monitor.*

4. Rappresentazioni interattive ad immagini

Il modello 3D di una scena è ottenuto tramite procedimenti di stereopsi[2] computazionale che usano come dati di ingresso immagini della stessa scena. Questa osservazione ha suggerito l'idea di utilizzare direttamente come rappresentazioni interattive, opportune collezioni di immagini possibilmente coadiuvate da dati di profondità (ottenuti elaborando le stesse immagini) [15-18] anziché i modelli 3D.

Queste rappresentazioni vengono denominate interattive a immagini. È stato dimostrato che opportuni insiemi di immagini sono equivalenti a un modello 3D per quanto riguarda la possibilità di generare a partire da essi, nuove viste della scena, ossia per quanto riguarda la rappresentazione interattiva della scena [19-20]. Inoltre, se si accettano limitazioni alla libertà di ispezione, le collezioni di immagini utili per rappresentazioni interattive possono essere drasticamente semplificate. Di particolare interesse sono alcuni tipi di rappresentazioni a immagini di tipo semplificato che mantengono forte fotorealismo e capacità di coinvolgimento e la cui realizzazione è molto meno complessa delle rappresentazioni a modelli [21].

Un caso estremo di rappresentazioni a immagini di tipo ridotto - è divenuto molto popolare - è costituito dalle visite virtuali a *panoramiche* [23-25], le quali rappresentano la scena tramite collezioni di viste panoramiche a 360 gradi; un esempio di immagine panoramica è illustrato in figura 5. Per alcuni esempi di realizzazioni con questa tecnologia si rinvia a *http://freia.dei.unipd.it/RESEARCHI-b-repres.htm#*.

Le rappresentazioni interattive a panoramiche si sono rivelate molto efficaci per la rappresentazione di ambienti, perché riescono a fornirne rappresentazioni molto vivide senza i costi e il lavoro richiesto dai modelli 3D. Inoltre è possibile costruire le immagini panoramiche in modo automatico a partire da semplici immagini riprese in modo opportuno [24].

Per quanto riguarda le rappresentazioni a immagini, quelle di tipo ridotto che non usano informazione ausiliaria relativa a dati di profondità ma solo un insieme di immagini riprese in modo da fornire cinture di viste attorno agli oggetti, hanno incontrato particolare favore a causa della loro spettacolarità. Alcuni esempi sono disponibili in *http://freia.dei.unipd.it/RESEARCHI-b-repres.htm#*.

La realizzazione automatica di questa tipologia di rappresentazioni interattive richiede di cimentarsi con il difficile problema della segmentazione automatica di un oggetto dallo "sfondo".

Soluzioni robuste e generali a questo problema non sono note, per cui di fatto queste operazioni sulle immagini richiedono la supervisione di un operatore. Procedure per automatizzare il più possibile la realizzazione delle rappresentazioni a immagini sono oggetto di attiva ricerca.

5. Visualizzazione remota di rappresentazioni interattive

È noto che modelli 3D di qualità fotorealistica richiedono notevoli quantità di dati per rappresentare sia le caratteristiche di forma *(struttura)* che di

(2) Stereopsi è la localizzazione relativa degli oggetti visivi in profondità. Si può realizzare solo nella visione binoculare poiché è basata su un processo fisiologico derivato dall'organizzazione del sistema video sensoriale (www.ortottica.it).

Figura 5 *"Panoramica" che mostra la "Basilica" di Gio' Ponti dalla visita virtuale di Palazzo Bo, realizzata per il sito web dell'Università di Padova (http://www.dei.unipd.it/conferences/3DPVT/visita/index.htm).*

colore *(tessitura)*[3] degli oggetti. A titolo di esempio il modello della scultura mostrata in figura 1a è dell'ordine dei cento Megabyte. Il paradigma attuale di visualizzazione a distanza di modelli 3D prevede la trasmissione dell'intero modello (possibilmente in forma compressa) al computer dell'utilizzatore *(client)* seguito dalla visualizzazione del modello effettuata localmente dal client stesso. L'operazione di visualizzazione *(rendering)* consiste nella generazione delle viste *(immagini)* corrispodenti alla posizione dell'utente. Questa operazione, poiché deve essere effettuata in tempo reale, è estremamente impegnativa rispetto alle risorse di calcolo e oggigiorno viene generalmente effettuata con l'ausilio di apposite schede hardware dette *acceleratori grafici*.

Questo modo di procedere presenta alcuni problemi intrinseci: a) la visualizzazione può iniziare sul client solo dopo che il modello è stato ricevuto, quindi con un ritardo che dipende dalle dimensioni del modello 3D e dalla banda del canale disponibile; b) se il modello è complesso la capacità di calcolo del client può non essere adeguata a fornire una navigazione fluida; c) con l'arrivo del modello sul client viene perso il controllo della proprietà intellettuale del modello 3D.

Si noti, in particolare, che dal punto di vista del progetto di un sistema di telecomunicazioni per la visualizzazione remota questa impostazione porta a un problema intrinsecamente mal posto.

Non è infatti possibile con questo paradigma di visualizzazione remota determinare le risorse (banda, potenza) necessarie a garantire una prefissata qualità di servizio (definita da vari parametri, quali, ad esempio, ritardo di visualizzazione inferiore a 30 secondi) perché le prestazioni del sistema richiedono informazioni sulle quantità dei dati da trasmettere che non sono mai disponibili (le dimensioni dei modelli possono variare dai Megabyte ai Terabyte a seconda della complessità della scena rappresentata e della qualità della rappresentazione).

In prospettiva con l' aumento della banda dei sistemi di trasmissione questo paradigma di visualizzazione remota diventerà via via più utile; ovviamente la sua utilità è estremamente limitata per la visualizzazione remota di modelli 3D di qualità fotorealistica (dell'ordine almeno di svariate decine di Megabyte) su canali a banda stretta, quali ad esempio i modem a 56 kbit/s oggigiorno comuni per i collegamenti Internet da casa.

Al fine di ovviare a questi problemi è stato recentemente proposto un paradigma di visualizzazione remota di superfici 3D [25-26] che sostanzialmente sposta il *rendering* dal client al server e di fatto trasforma l'ispezione remota di un oggetto in un problema legato alla trasmissione di viste (ossia, di immagini 2D) dell'oggetto da posizioni specificate dall'utente.

Conviene osservare come in questo caso, non essendo più necessario trasmettere l'intero modello, ma dovendo semplicemente trasmettere immagini da esso generate, non sia più presente il malcondizionamento intrinseco del problema di progetto dei sistemi di telecomunicazione sopra accennato, ossia la dipendenza delle prestazioni del sistema dalle dimensioni imprecisabili dei dati.

I dati trasmessi con quest'impostazione sono sempre e solo immagini di dimensione nota, più alcuni dati di ritorno del client, per i quali è semplice definire dei limiti superiori di velocità di cifra *(bit-rate)*; e, quindi, il dimensionare i sistemi di telecomunicazioni per prefissate qualità di servizio non solo diventa un problema ben posto, ma anche un problema per il quale esistono soluzioni ben collaudate.

Inoltre, con questa impostazione, anzitutto il ritardo iniziale si riduce da quello richiesto per la trasmissione di un intero modello 3D a quello richiesto per la trasmissione di un'immagine; la potenza di calcolo del client diventa poi ininfluente per la visualizzazione, dato che tutto quello che si

(3) *Si accetti per semplicità l'uso di questo anglicismo nel ruolo della parola inglese "texture".*

richiede al client è la capacità di visualizzare immagini e, infine, la proprietà intellettuale del modello 3D è garantita, dato che il modello 3D non lascia mai il server (vengono trasmesse solo immagini del modello).

L'efficacia pratica di questo metodo di visualizzazione remota è direttamente legata all'efficienza degli schemi di compressione delle immagini trasmesse.

Vale la pena osservare che, poiché le immagini trasmesse in questo schema di visualizzazione remota sono viste di modelli 3D matematicamente definiti, esse possono essere compresse secondo metodi molto più efficienti (ancorché totalmente diversi) da quelli utilizzati nella codifica video. Quali metodi di compressione siano i più adatti per questo ambito applicativo è oggi terreno aperto di ricerca.

Uno schema di compressione da utilizzare entro un sistema di visualizzazione remota con le caratteristiche indicate in [26], consiste nel trasmettere oltre alla vista corrente anche l'informazione dello *z-buffer* opportunamente compressa. Lo z-buffer è da utilizzarsi al client per il calcolo della vista successiva (si noti che gli acceleratori grafici, ormai comuni anche sui PC, fanno di default ricorso allo z-buffer per il calcolo delle viste).

Un secondo schema [26] attualmente in fase di verifica consiste nell'applicazione di trasformazioni affini di tipo 2D alle immagini dei triangoli delle superfici 3D al fine di ottenere la nuova vista a partire dalla vista precedente. Questo schema si basa sul fatto noto che immagini di oggetti planari (come i triangoli piani che costituiscono le approssimazioni delle superfici 3D), inquadrate da punti di vista diversi, sono legate da trasformazioni proiettive 2D. Queste trasformazioni sono determinabili analiticamente quando è nota la posizione del piano, una sua vista e la posizione della nuova vista.

Altri schemi presumibilmente verranno suggeriti dal confronto tra le prestazioni della tecnica che usa l'informazione dello z-buffer e quella che usa le trasformazioni proiettive 2D.

Infine, è utile osservare che gli schemi di visualizzazione che spostano il *rendering* al *server* sopra proposti per la visualizzazione remota di modelli 3D via Internet, dato che alleggeriscono gli oneri computazionali del client al massimo, sono intrinsecamente adatti alla visualizzazione remota di modelli 3D su dispositivi con risorse di calcolo e memoria limitate di tipo generale. Pertanto se oggigiorno i PC collegati alla rete Internet ne sono percepiti come l'utenza privilegiata, in prospettiva, quando i PC avranno sensibilmente aumentato le loro risorse, a beneficiarne potrebbero essere telefoni cellulari, computers palmari, audioguide museali e simili.

Per quanto riguarda la visualizzazine remota delle rappresentazioni a immagini di oggetti, è necessario osservare che nella compressione delle immagini sono state tradizionalmente considerate solo due situazioni: la compressione dell'immagine singola e la trasmissione della sequenza di immagini.

Non sono disponibili al momento algoritmi efficienti dedicati alla situazione che si incontra nelle rappresentazioni interattive a immagini, ossia la situazione di insiemi di immagini, non così simili come quelle delle sequenze video, ma pure con caratteri di similarità dati dal fatto che si riferiscono allo stesso soggetto.

In questo caso l'efficienza di codifica deve provenire dall'uso dell'informazione comune a tutte le immagini.

Purtroppo questa informazione non è legata a semplici traslazioni tra immagini vicine, come nel caso delle sequenze video, e non è semplice da estrarre dalle immagini.

Nell'ambito della standardizzazione corrente la situazione di un insieme di più immagini può rientrare nei casi gestiti da *motion JPEG* e dal recente standard di *motion JPEG2000*; questi standards sono stati tuttavia sostanzialmente concepiti per applicazioni video di tipo professionale (in cui ad esempio è importante poter manipolare le immagini senza degradarne la qualità) e non usano informazione *inter-frame*. La loro efficienza di codifica è pertanto molto bassa e essi non sono adeguati alle esigenze delle rappresentazioni interattive.

Non esistono al momento algoritmi efficienti per la compressione di dati relativi alla rappresentazione a immagini di oggetti.

6. Conclusioni

Quest'articolo presenta le caratteristiche delle rappresentazione a modelli e ad immagini, i due metodi oggi utilizzati per le rappresentazioni interattive, con l'intento di dare il senso sia dei problemi tecnici di questo settore (senza entrare nel merito di alcuno di essi), che del potenziale applicativo di questi strumenti.

Tra le osservazioni finali è utile includere il fatto che i modelli tridimensionali vengono sostanzialmente realizzati tramite metodi di misura delle superfici; i dati relativi ai modelli possono perciò avere uno straordinario valore di documentazione o di misura degli oggetti rappresentati. Una chiara manifestazione di questa caratteristica è che i modelli possono anche consentire la duplicazione fisica degli oggetti per via automatica.

Inoltre è opportuno ricordare che le due metodologie di rappresentazione possono essere anche utilmente ibridate. Un caso emblematico è dato da un museo di sculture in cui l'ambiente può venire rappresentato "a immagini" per ragioni di semplicità e le sculture "a modelli" per motivi di spettacolarità e accuratezza di documentazione (a titolo di esempio si veda il prototipo di visita virtuale realizzato per la collezione Mantova-Benavides del Dipartimento di Scienze dell'Antichità dell'Università di Padova nel sito *http://freia.dei.unipd.it/RESEARCHI-b-repres.htm#*).

È anche bene spezzare una lancia in favore delle applicazioni ai beni culturali perché, se da un lato la tecnologia delle rappresentazioni interattive può essere di grande utilità per la documentazione e la presentazione interattiva e su rete dei beni culturali, quest'applicazione specifica a sua volta può svolgere

un ruolo del tutto particolare per l'evoluzione tecnica delle rappresentazioni interattive. Infatti i beni culturali per le loro dimensioni, per la loro articolazione che ne rende difficile la modellazione (si pensi alla semplicità degli oggetti meccanici rispetto ai panneggi o ai volti delle sculture), per la precisione con cui è opportuno realizzarli e per il fatto che il patrimonio culturale molto spesso non è trasportabile (sviluppare metodi e strumenti in grado di funzionare anche fuori dal laboratorio, in condizioni operative difficili, come non è infrequente incontrare in musei o chiese, è un problema nel problema) costituiscono una grande sfida tecnologica e un vero e proprio banco di prova per la modellazione tridimensionale.

Risolvere i problemi posti dalla modellazione tridimensionale dei beni culturali farebbe compiere un sostanziale passo in avanti a questa disciplina, con immediate ricadute alla modellazione di altre tipologie di oggetti di tipo industriale e commerciale.

La diffusa consapevolezza di questo ruolo di banco di prova, esercitato dalla modellazione di oggetti del patrimonio culturale, è indicata anche dai progetti pilota di alcune grandi ditte rivolti ai beni culturali; può essere ricordato in proposito il progetto "Pietà Rondanini" di IBM, ed il "Michelangelo Project" di Interval [7].

Le rappresentazioni interattive consentono di presentare oggetti e luoghi in modo molto più efficace e coinvolgente di foto, filmati e, in genere, di altri strumenti iconografici tradizionali.

Questo fatto apre a questa tipologia di dati un amplissimo bacino applicativo. Da più parti si ritiene che grazie alle applicazioni Web, il loro uso nel medio termine sarà confrontabile con quello delle immagini statiche e dei filmati video.

Esistono tuttavia ancora molte difficoltà nella realizzazione e fruizione remota delle rappresentazioni interattive, in parte adducibili alla multidisciplinarietà della materia, che richiedono nozioni di optoelettronica, di visione elaborativa, di grafica computerizzata, di trattamento delle immagini e di compressione e trasmissione di dati e, in parte, alla sua relativa novità.

Può infine essere utile ricordare che è opinione diffusa che il più grande ostacolo contro l'affermazione delle rappresentazioni interattive sia costituito dai limiti dei paradigmi di visualizzazione remota attualmente in uso. La soluzione di questi problemi richiede competenze principalmente legate alle telecomunicazioni. A questo riguardo è confortante registrare l'interesse crescente dei ricercatori per la trasmissione di rappresentazioni interattive, come indicato indirettamente dal moltiplicarsi di conferenze internazionali dedicate esplicitamente a questi argomenti. A titolo di esempio si cita il *First International Symposium on 3D Data Processing Visualization and Transmission*, per i cui dettagli sui contenuti può essere consultato il sito *http://www.dei.unipd.it/conferences/3DPVT/index.htm*.

È anche opportuno osservare come le soluzioni proposte per ovviare ai limiti della visualizzazione remota sui sistemi Internet odierni, siano intrinsecamente adatte a gestire la visualizzazione remota di rappresentazioni interattive verso dispositivi con risorse limitate di calcolo e memoria di tipo generale, quali telefoni cellulari e computers palmari.

Bibliografia

[1] Rioux, M.; Roth, G. (Eds.): «Proc. Of First Int. Conference on Recent Advances in 3-D Digital Imaging and Modeling (3DIM97)», Ottawa, Canada, maggio 1997.

[2] Besl, P.J.; McKay, N.D.: *A Method for Registration of 3-D Shapes*. «IEEE Trans. on Pattern Analysis and machine Intelligence», febbraio 1992, Vol. PAMI-14(2): 239-259.

[3] Bergevin, R.; Laurendeau, D.; Poussart, D.: *Registering Range Views of Multipart Objects*. «Computer Vision and Image Understanding», gennaio 1995, Vol. 61, pp. 1-16.

[4] Pulli, K.: *Multiview Registration for Large Data Sets*. «Proc. of Second International Conference on 3D Digital Imaging and Modeling (3DIM99)», Ottawa, Canada, ottobre 1999, pp. 160-168.

[5] Curless, B.; Levoy, M.: *A Volumetric Method for Building Complex Models from Range Images*. «Proc. of the 23rd Int. Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'96)», New Orleans, Stati Uniti, agosto 1996, pp. 303-12.

[6] Ikeuchi, K.; Wheeler, D.; Sato, Y.: *Consensus surfaces for modeling 3D objects from multiple range images*. «Proc. of International Conference on Computer Vision (ICCV 98)», gennaio 1998, pp. 917-924.

[7] Levoy, M.; Pulli, K.; Curless, B.; Rusinkiewicz, S.; Koller, D.; Pereira, L.; Ginzton, M.; Anderson, S.; Davis, J.; Ginsberg, J.; Shade, J.; Fulk, D.: *The Digital Michelangelo Project: 3D Scanning of Large Statues*. «Proc. of the 27rd Int. Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'2000)», New Orleans, Stati Uniti, luglio 2000.

[8] Cortelazzo, G.M.; Doretto, G.; Lucchese, L.; Totaro, S.: *Frequency domain methods for the registration of 3D Data*. «Proc. IEEE International Symposium on Circuits and Systems», Monterey, Canada, giugno 1998, pp. 518-521.

[9] Bernardini, R.; Cortelazzo, G.M.: *A post-processing technique for noise removal of range data*. «IEEE Trans. on Circuits and Systems for Video Technology», marzo 2000, pp. 201-206.

[10] Andreeetto, M.; Bernardini, R.; Cortelazzo, G.M.; Lucchese, L.: *Towards automatic modelling of 3D cultural heritage*. «Proc. of the IEEE

International Conference on Image Processing», Salonicco, Grecia, settembre 2001.
[11] Hartley, R.I.; Zisserman, A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
[12] Cipolla, R.; Giblin, P.: *Visual Motion of Curves and Surfaces*. Cambridge University Press, 1999.
[13] Fiocco, G.: *Mantegna, la Cappella Ovetari nella Chiesa degli Eremitani*. Silvania Editoriale d'Arte (1947, prima edizione e 1978, seconda edizione).
[14] Ardito, M.; Cortelazzo, G.; Martelli, M.; Mian, G.A.: *Subband coding of museal images*. «Proc. of SPIE Internatinal Symposium on Fiber Optic Networks and and Video Compression», Berlino, Germania, aprile 1993, pp. 354-363.
[15] Mc Millan, L.; Bishop, G.: *Plenoptic modeling: An image-based rendering system*. «Computer Graphics Proceedings», Annual Conference Series, 1995, pp. 39-46.
[16] Levoy , M.; Hanrahan, P.: *Light field rendering*. «Proc. of the 23rd Int. Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'96)», agosto 1996, New Orleans, Stati Uniti, pp. 31-42.
[17] Laveau, S.; Faugeras, O.D.: *3-D Scene Representation as a Collection of Images*. «Proc. of 12th International Conf. on Pattern Recognition», Gerusalemme, Israele, 1994, pp. 689-691.
[18] Ullman, S.; Basri, R.: *Recognition by linear combinations of models*. «IEEE Trans. on Pattern Analysis and machine Intelligence», 1991, 13(10): 992-1006.
[19] Avidan, S.; Evgeniou, T.; Shasha, A.; Poggio, T.: *Image-based View Synthesis by combining trilinear Tensors and learning techniques*. «ACM Symposium on Virtual Reality Software and Technology», 1997.
[20] Avidan, S.; Shashua, A.: *Novel View Synthesis in Tensor Space*. 1996, CIS Report 9602, Rechnion. 1997, «Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition», pp. 1034-1040.
[21] Chen, S.E.: *Quick-Time VR - An Image-Based Approach to Virtual Environment Navigation*. «Proc. of the 22rd Int. Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'95)», Los Angeles, Stati Uniti, 1995, pp. 29-38.
[22] Cossi, M.; Cortelazzo, G.M.; Frezza, R.: *Reconstruction of structure and texture of planar environments by dynamic vision techniques*. «Proc. of the European Signal Processing Conference», Trieste, settembre 1996, pp. 1681-1684.
[23] Cortelazzo, G.M.; Sacco, S.; Marcotto, F.: *Photorealistic Virtual Visits of Hystorical Architecture Via Image-Based Rendering*. «Proc. of International Workshop on Sinthetic-Natural Hybrid Coding and Three Dimensional Imaging», Rodi, Grecia, 5-9 settembre 1997, pp. 33-36.
[24] Cortelazzo, G.M.; Lucchese, L.: *A new method of image mosaicking and its application to cultural heritage representation*. «Computer Graphics Forum» (Guest, ed.), P. Brunet and R. Scopigno, Blackwell, 1999, Vol. 18, pp. C265-C275.
[25] Bernardini, R.;Cortelazzo, G.M.: *An efficient network protocol for virtual worlds browsing*. «The 12th Tyrrhenian International Workshop on Digital Communications», (Portoferraio, Italy), CNIT, Springer-Verlag, Berlino, Germania, settembre 2000, pp. 67-78.
[26] Bernardini, R.; Cortelazzo, G.M.: *Accessing multi-user virtual worlds over IP*. «Proc. Vision Modeling and Visualization 2001», Stoccarda, Germania, ottobre 2001, pp. 407-413.

Guido Maria Cortelazzo nel 1975 si laurea in Ingegneria Elettronica presso l'Universita' di Padova. Nel 1980 riceve un Master e nel 1984 un Ph. D., entrambi in Electrical Engineering, presso l'Università dell'Illinois ad Urbana-Champaign. Dal 1983 al 1986 lavora presso M/A-COM Linkabit, Inc. di San Diego, California, industria leader nel settore delle telecomunicazioni via satellite. Nel 1987 diventa professore associato al Dipartimento di Elettronica e Informatica dell'Università di Padova (DEI-UP) dove ora è professore ordinario. Al DEI-UP nell'autunno 1986 è tra gli iniziatori del "Laboratorio di Elaborazione Immagini". Nel 1994 avvia il "Laboratorio di Tecnologie e Telecomunicazioni Multimediali" che tuttora dirige. Nel 1997 è Visiting Researcher al California Institute of Technology (Pasadena, California). Il "Laboratorio di Tecnologie e Telecomunicazioni Multimediali", oltre all'attività di didattica e di ricerca, nel corso degli anni ha svolto molti servizi no-profit con neolaureati che hanno portato alla realizzazione di vari CD-ROM relativi a visite virtuali di musei ed edifici storici e ad un corso multimediale di "Fondamenti di visione, fotometria e colorimetria" (Scuola RAI, 1998), di cui Cortelazzo è autore insieme a Riccardo Bernardini. È autore di circa 50 pubblicazioni su riviste internazionali. I suoi attuali interessi di ricerca riguardano l'automatizzazione della realizzazione delle rappresentazioni interattive della scena e la loro trasmissione e visualizzazione remota.

MPLS: dall'idea originale alle attuali applicazioni nelle reti IP

FEDERICO M. RENON
GIANNI ROSSI
PAOLO SALAMANDRA

La crescita esponenziale della domanda di servizi dati, applicazioni multimediali e, in particolare, lo sviluppo preponderante, a livello mondiale, di servizi disponibili sulla rete Internet sono tali da richiedere un continuo progresso delle reti IP e un'evoluzione delle sue funzionalità.

Sin dall'inizio degli anni novanta l'IETF si è posta l'obiettivo di riuscire a adattare l'architettura protocollare delle reti IP a un contesto, in cui il traffico presente è altamente variegato e in cui il classico modello di servizio best-effort non consentiva di rispondere alle esigenze degli utenti.

Uno dei risultati di quest'attività, ha portato alla definizione dell'architettura MultiProtocol Label Switching, che impiega un'ennesima variante del classico principio della commutazione di pacchetto. Essa nasce a partire dagli sforzi intrapresi dall'IETF (MPLS working group) nella seconda metà degli anni Novanta, con l'intento di proporre una nuova soluzione in ambito delle reti multiservizio.

Concepita inizialmente come strumento per ottenere un'efficiente ed efficace integrazione della tecnologia ATM con i meccanismi del protocollo IP per migliorarne le prestazioni MPLS è divenuta, in realtà, un'architettura che consente di realizzare alcuni nuovi servizi IP attraverso l'introduzione di ulteriori meccanismi di segnalazione e d'inoltro dei pacchetti.

L'architettura MPLS è più articolata di quella di una rete IP e maggiormente complessa da gestire e da amministrare, nonostante lo sforzo compiuto dall'IETF, nella definizione degli standard, sia stato quello di limitare la complessità che ha caratterizzato lo sviluppo del piano di controllo ATM.

Oggi MPLS rappresenta una tecnologia matura, in molti casi già introdotta in campo dagli operatori, necessaria per realizzare i servizi VPN IP (Virtual Private Network IP) e di Traffic Engineering.

1. Introduzione

La tradizionale architettura delle reti IP *(connectionless-oriented*, paradigma *best-effort)*, ha contribuito da un lato alla straordinaria crescita e diffusione di queste reti, ma è stata anche vincolata da limiti riconosciuti che hanno stimolato la ricerca di nuove soluzioni di interconnessione in rete, in modo da rendere Internet capace di diversificare e di incrementare ulteriormente le tipologie delle applicazioni e dei servizi offerti *(video on-demand, real-time, Voice over IP)*.

Questo articolo presenta una nuova architettura proposta dalla comunità tecnico-industriale: Organismi di Ricerca e delle Università, *IETF (Internet Engineering Task Force)*, costruttori ed è stata chiamata *MPLS (MultiProtocol Label Switching)*. Nel testo si è cercato di mettere in luce cosa effettivamente sia l'MPLS e cosa effettivamente essa permette di ottenere, in termini di valore aggiunto in una rete IP.

Partendo da questi presupposti, in questo articolo, si sottolinea, dapprima, come in origine MPLS si sia sviluppato quale risposta alle difficoltà di integrazione *IP/ATM (Asynchronous Transfer Mode)* [1], vista in passato come la soluzione ideale per risolvere i problemi legati alle prestazioni presenti nei router tradizionali sfruttando le capacità di trasporto di una rete ATM.

Sono poi presentate le successive evoluzioni di questo paradigma, che hanno portato all'attuale architettura MPLS e ai relativi utilizzi da parte degli

ISP (Internet Service Provider). Sono, in particolare, considerate le caratteristiche funzionali di una rete che realizzi MPLS, mettendone in luce anzitutto gli aspetti salienti che consentono di farne comprendere le principali strutture operative; e, successivamente, si passa a descrivere l'attuale stato dell'arte della architettura MPLS e gli aspetti, invece, che tuttora sono in fase di sviluppo e di evoluzione.

Sono, infine, illustrate le applicazioni oggi maggiormente diffuse, per le quali il livello della sperimentazione è ormai maturo e la messa in campo di fatto già realizzata da numerosi costruttori negli apparati e dai Provider nelle proprie reti. In particolare, sono messi anche in evidenza lo stato di avanzamento dell'introduzione di MPLS nelle reti di Telecom Italia e la tipologia di servizi che conseguentemente possono essere offerti ai clienti.

2. Da dove nasce MPLS

Nella prima metà degli anni Novanta si era diffusa la convinzione che il modo per riuscire a far fronte al crescente sviluppo della rete Internet, in termini di diversità dei servizi gestibili e della banda offerta agli utilizzatori, fosse quello di coniugare nella rete le tecniche basate su IP e ATM, sfruttando, in particolare, le soluzioni ATM come dorsali a livello geografico per gli ISP.

In questo contesto, il primo problema che allora si era posto, riguardava la possibilità di riuscire a mappare l'architettura IP su una rete ATM.

La prima soluzione proposta era, perciò, di tipo *overlay*, ossia con livelli separati e sovrapposti, in cui, di fatto, i protocolli d'instradamento a livello IP e ATM agivano indipendentemente, con una netta distinzione tra la rete IP e quella ATM (figura 1). La rete ATM è, infatti, utilizzata per collegare i router IP e occorreva, quindi, ingegnerizzare l'impiego della banda disponibile.

Nell'ambito del modello overlay sono state poi sviluppate diverse soluzioni, sia proposte dai costruttori, sia inserite nei piani di standardizzazione della ATM Forum e dell'IETF.

In particolare possono essere ricordate:

- *IP over ATM* [1]: questo modello definisce sia i criteri di incapsulamento dei datagrammi IP quando sono trasportati attraverso una rete ATM mediante lo strato di adattamento *AAL5 (ATM Adaptation Layer 5)*, sia un protocollo per associare gli indirizzi IP ai corrispondenti indirizzi ATM (ATMARP), le cui caratteristiche fossero estese anche al mappaggio di indirizzi multicast;
- *LAN Emulation* [2]: il modello stabilisce le procedure per rendere una rete ATM simile - in termini di proprietà legate al servizio multiaccesso e broadcast - a un ambiente LAN, emulando le funzioni relative al livello *MAC (Medium Access Control);*
- *MPOA (MultiProtocol over ATM)* [3]: la specifica è stata normalizzata come uno strumento idoneo a realizzare una soluzione di rete caratterizzata dalla presenza di differenti tecnologie, sia a livello di rete (IP, IPX, ...) sia di trasporto (Ethernet, Token Ring, LANE, ...);
- *IPLPDN (IP over Large Public Data Network)* e successivamente *ROLC (Routing over Large Clouds)*: definisce il *NHRP (Next Hop Resolution Protocol)* [4] per host e router non appartenenti alla medesima sottorete IP, in modo da permettere di stabilire circuiti virtuali diretti attraverso una rete ATM, a condizione, però, che essi appartengano alla stessa rete ATM.

Nell'applicazione tipica del modello overlay, i router IP comunicano tra loro attraverso un insieme di *PVC (Permanent Virtual Circuit)* ATM, che funzionano, perciò, come un insieme di circuiti logici che garantiscono la connettività tra i nodi terminali *(edge)*.

I router non sono in grado di rilevare la topologia fisica della rete ma conoscono solo i PVC, che, di conseguenza, appaiono come semplici collegamenti punto-punto. Su ciascun PVC è poi attivato un protocollo d'instradamento che consente ai router di stabilire le adiacenze *(peer relationships)*.

Alla classica tabella di routing IP, presente su ciascun router, si aggiunge la corrispondenza tra next-hop e identificativo del *VPI/VCI (Virtual Path Identifier / Virtual Circuit Identifier)* del PVC ATM di collegamento tra il router di origine e quello di destinazione del pacchetto.

È da notare che in questo caso, se N è il numero dei router di dorsale, per ottenere una soluzione funzionante in maniera corretta ed efficace, sarebbe necessario magliare completamente la rete e, quindi, configurare un numero di PVC proporzionale al qua-

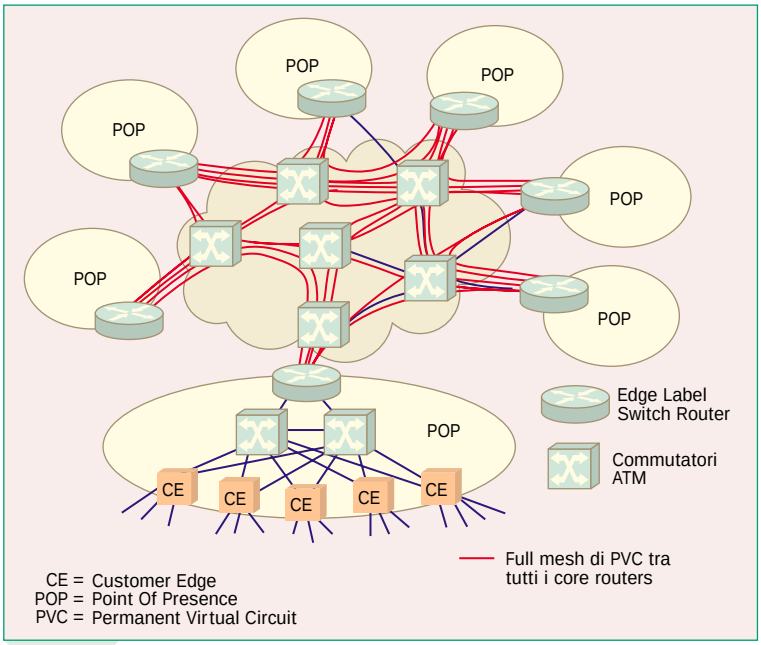


Figura 1 *Integrazione IP/ATM, modello overlay.*

drato del numero N dei nodi (più eventuali PVC di riserva).

Il numero di adiacenze da gestire è pure decisamente elevato e la quantità delle informazioni d'instradamento che viaggiano è dell'ordine della quarta potenza del numero di nodi N, il che potrebbe portare al sovraccarico del protocollo d'instradamento IP in caso di fuori servizio contemporaneo di molti PVC (ad esempio nel caso di un guasto di un nodo ATM), e, quindi, a un crollo drammatico delle prestazioni degli apparati.

Una dorsale IP basata su ATM presenta, dunque, alcune limitazioni di rilievo: la necessità di mantenere la gestione delle due reti ATM e IP sovrapposte e i problemi di crescita modulare e, cioè, di scalabilità *(N-squared problem)* causati dall'impiego dei protocolli d'instradamento IP - in genere di tipo *OSPF (Open Shortest Path First)* - su una magliatura di PVC.

Per questi motivi, al modello *overlay* si contrappose, in un secondo tempo, il *modello integrato*, introdotto con lo scopo di eliminare le difficoltà di indirizzamento e le ridondanze delle caratteristiche funzionali presenti nelle reti IP e ATM per permettere l'inoltro delle informazioni. Già nel corso del biennio 1996/1997 diversi costruttori proponevano soluzioni, sia pur proprietarie, che rispondevano a quest'obiettivo. Tra questi possono essere ricordati: Toshiba con *CSR (Cell Switch Router)* [5]; Ipsilon (ora Nokia) con *IP Switching* [6]; Cisco con *Tag Switching* [7]; IBM con *ARIS* (*Aggregate Route-based IP Switching)* [8].

Nella maggioranza dei casi, con queste soluzioni si intendeva agire sul software di controllo di un router IP per integrarlo con l'hardware di un commutatore ATM. La componente di controllo realizza in questo caso instradamenti basati su un protocollo IP (OSPF, BGPv4, ...), eliminando i problemi di scalabilità e limitando, notevolmente, il numero delle adiacenze da mantenere e il carico conseguente sul protocollo d'instradamento.

Per quanto concerne, invece, la componente di invio dei pacchetti *(forwarding)*, i commutatori IP/ATM utilizzano hardware ATM convenzionale e la tecnica di commutazione di etichetta tipica del livello 2. Si aggiunge, così, un'ulteriore funzione, per la componente di controllo, relativa all'allocazione e alla distribuzione delle etichette.

Il limite principale di queste tecnologie è, però, rappresentato dal fatto che le varie soluzioni proposte dai costruttori non sono tra loro interoperabili e quasi tutte richiedono, come tecnologia di trasporto, l'ATM non essendo infatti in grado di operare su infrastrutture differenti, quali ad esempio quella SDH o quella *PPP (Point to Point Protocol)*.

Proprio a partire da questo contesto, nel 1997 l'IETF ha costituito l'*MPLS Working Group*, assegnandogli l'obiettivo di armonizzare e di integrare le precedenti proposte, in modo da produrre uno standard multivendor impiegabile su qualsiasi tecnologia di trasporto.

L'idea architetturale di base riguarda l'associazione a tutti i datagrammi IP di una breve etichetta *(label)* di lunghezza fissa, con cui gli apparati di rete possono effettuare un instradamento veloce basato sulla commutazione dell'etichetta stessa *(label swapping)*.

La tecnologia risultante è di fatto così in grado di "appoggiarsi" a qualsiasi protocollo di trasporto e di utilizzare qualsiasi protocollo di rete, con il vantaggio di consentire a un ISP di offrire nuovi servizi che non possono essere forniti efficacemente tramite il convenzionale instradamento IP.

Gli studi sull'MPLS erano indirizzati all'inizio anche verso l'obiettivo di trasformare i commutatori ATM in router con elevate prestazioni. I recenti progressi nella tecnologia del silicio consentono, però, di effettuare - mediante nuovi componenti dedicati *ASIC (Application Specific Integrated Circuit)* - consultazioni di tabelle d'instradamento *IP (route lookup)* con prestazioni paragonabili a quelle dell'hardware dedicato, sviluppato per ATM. Gli sviluppi si sono, perciò essenzialmente orientati verso un'architettura di rete MPLS con sistemi SDH, piuttosto che verso MPLS su ATM.

Questa scelta è, anche, coerente con le tendenze architetturali per la rete nel suo complesso che vedono, in prospettiva, la semplificazione dei livelli di rete e la convergenza verso soluzioni che consentano di instradare direttamente IP sui portanti ottici.

3. Architettura MPLS

Il concetto fondamentale di MPLS [9] è, dunque, quello di associare un'etichetta a ciascun pacchetto che attraversa la rete, seguendo a livello architetturale di nodo il criterio della separazione delle due componenti dell'instradamento: *la decisione d'instradamento*, gestita dai protocolli IP, e l'effettiva *attuazione (forwarding) dello smistamento dei flussi di pacchetti*, gestita tramite la commutazione di etichetta.

La componente di decisione - che comprende l'insieme dei moduli demandati all'allocazione e alla distribuzione delle etichette tra nodi adiacenti - e l'intelligenza di livello 3 *(IP addressing, IP routing)*, è del tutto indipendente da quella di attuazione (inoltro dei pacchetti secondo il paradigma *label switching*). L'assenza di vincoli permette di realizzare differenti protocolli su qualsiasi mezzo *(multiprotocol)* e di evitare, come chiarito nel paragrafo 2, configurazioni completamente magliate di percorsi, gli *LSP (Label Switched Path)*, fra i router della dorsale.

3.1 Commutazione di etichetta: la componente di forwarding

Nel modello di *layer 3 forwarding* tradizionale [10], ciascun router di rete consulta la propria IP *FT (Forwarding Table)* e seleziona, così, il nodo successivo verso cui inviare i pacchetti *(next hop)* sulla base dell'indirizzo IP di destinazione contenuto nell'intestazione di livello 3.

La scelta del *next hop* è data dalla combinazione di due funzioni: la prima suddivide l'intero insieme dei possibili pacchetti IP in sottoinsiemi denominati *FECs (Forwarding Equivalence Classes);* la seconda

associa ciascuna FEC a un determinato indirizzo IP di *next hop*.

In questo modo tutte le destinazioni all'interno della rete sono raggiungibili mediante almeno un percorso e, eventualmente, si rendono disponibili percorsi multipli in caso di load balancing. Questo processo è conosciuto in letteratura come *hop by hop forwarding*.

Con la tecnologia MPLS, l'analisi dell'intestazione IP e l'assegnazione conseguente di un pacchetto a una determinata FEC - che può essere effettuata sulla base di numerose informazioni quali *IP precedence*, indirizzo di sorgente e di destinazione, tipo di applicazione - è eseguita una sola volta in corrispondenza dell'*E-LSR (Edge-Label Switch Router)*, posto nei punti di ingresso della rete.

La FEC alla quale il pacchetto è assegnato è codificata con un'etichetta di lunghezza fissa che è

Elementi funzionali e terminologia delle Reti MPLS

Nella struttura tipica di una rete MPLS (figura A) gli elementi base che possono essere riconosciuti sono i seguenti:

- ***Dominio MPLS:*** **porzione di rete costituita da apparati che riconoscono e che sono in grado di dialogare con la rete MPLS;**
- ***Dorsale di rete MPLS:*** **porzione interna del dominio MPLS in cui l'inoltro dei pacchetti avviene unicamente attraverso la commutazione di etichetta MPLS;**
- ***Dominio cliente:*** **insieme di siti della rete di un Cliente connessi al *backbone* MPLS;**
- ***Edge LSR (Edge Label Switch Router):*** **router posti alla terminazione (frontiera) della rete, utilizzati per assegnare e per togliere le etichette ai datagrammi e per eseguire l'operazione conseguente di inoltro verso il dominio MPLS;**
- ***LSR (Label Switch Router):*** **dispositivi collocati in genere all'interno del dominio MPLS capaci di inoltrare i pacchetti unicamente sulla base del contenuto informativo di un'etichetta *(paradigma Label Switching)*;**
- ***LSP (Label Switched Path):*** **percorso attraverso uno o più LSRs seguito da un pacchetto appartenente a una certo flusso di dati;**
- ***LDP (Label Distribution Protocol):*** **protocollo utilizzato, insieme con quelli d'instradamento;**
- ***IP classici,*** **per definire e per distribuire le etichette;**
- ***Router Ru e Rd:*** **Ru è detto *upstream LSR* se un pacchetto, rispetto al processo di distribuzione delle etichette, è inoltrato da un router Ru ad uno Rd; analogamente Rd è chiamato *downstream LSR*.**

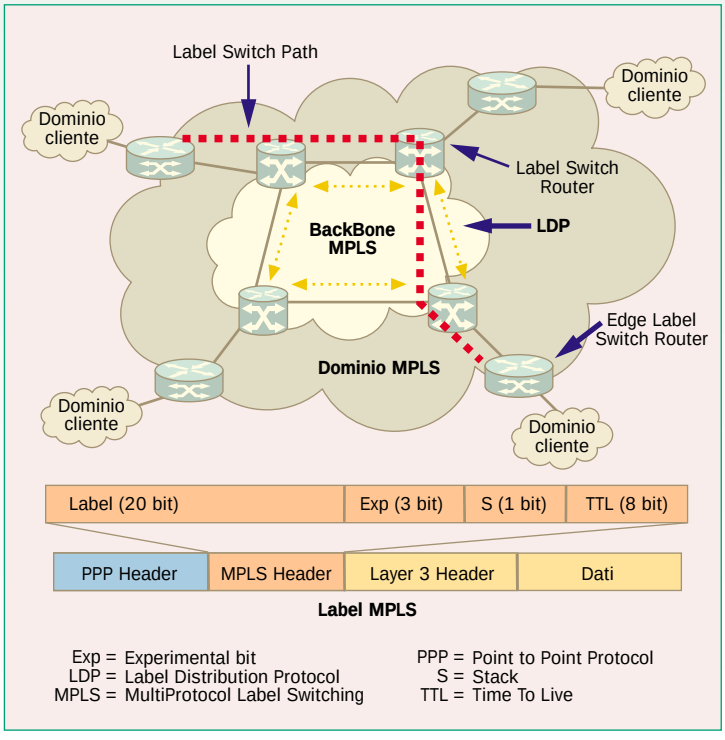


Figura A *Architettura di rete e struttura dell'intestazione MPLS.*

anteposta all'intero pacchetto. Il pacchetto "esteso" è, quindi, inviato verso il *next hop*, che questa volta realizza lo smistamento unicamente in base alle informazioni contenute nell'etichetta, piuttosto che sull'analisi delle informazioni dell'intestazione di livello 3 [11].

Nel punto di uscita dalla rete MPLS, il corrispondente E-LSR rimuove l'etichetta e consegna il pacchetto IP al sito del Cliente finale. In questo modo l'intero processo di trasporto MPLS in rete rimane del tutto trasparente per i siti posti presso le sedi dei clienti.

Il riquadro a pagina 40 illustra gli elementi funzionali e la terminologia delle architetture MPLS; esso mostra anche la struttura di una generica trama MPLS consegnata al livello 2 sottostante per il trasporto.

Il *payload* è costituito da un pacchetto IP preceduto da una sequenza di intestazioni MPLS, che consentono di stabilire, come sarà chiarito in seguito, anche gerarchie su più livelli d'instradamento.

Ogni intestazione MPLS è composta da 32 bit così ripartiti:

- *Label* (20 bit): rappresenta l'etichetta utilizzata per l'instradamento del pacchetto IP;
- *Exp* (3 bit): è il campo oggi definito come sperimentale. Un possibile utilizzo dei bit riguarda il mappaggio di eventuali *classi di servizio IP* su una rete MPLS che consente alcune differenti procedure di trattamento dei pacchetti, chiamati *PHB (Per Hop Behavior);*
- *Stack* (1 bit): è il campo che indica l'eventuale presenza di più etichette messe in sequenza (ovvero in *stack*) per consentire, come sarà chiarito più avanti, lo smistamento in reti realizzate su più livelli MPLS;
- *TTL* (8 bit): è il campo *Time-to-Live*, definito per rilevare ed eliminare *frame* MPLS che per qualche motivo circolino in rete per tempi eccessivi (riquadro a pagina 44). Il suo valore è fissato all'inizio del cammino ed è diminuito di un'unità ogni volta che si attraversa un nodo. Quando il valore di TTL raggiunge lo zero il frame MPLS è scartato.

La *label* di 20 bit sopra specificata ha un valore locale nell'interfaccia utilizzata per l'inoltro dei pacchetti e sintetizza diverse informazioni riguardanti il pacchetto cui essa si riferisce:

- destinazione;
- precedenza;
- appartenenza a *VPN (Virtual Private Network);*
- *QoS (Quality of Service);*
- informazioni di *TE (Traffic Engineering).*

Tra i valori assegnabili alcuni sono riservati:

- valore 0 - *IPv4 Explicit NULL*. Questo valore indica che l'etichetta non contiene effettive informazioni d'instradamento. Il pacchetto deve, quindi, essere inoltrato seguendo le informazioni contenute nell'intestazione di livello 3, che in questo caso è del tipo IPv4;
- valore 2 - *IPv6 Explicit NULL:* analogamente al caso precedente, l'etichetta non contiene vere e proprie informazioni d'instradamento e il pacchetto deve essere inoltrato seguendo le informazioni contenute nell'intestazione di livello 3, che in questo caso è del tipo IPv6;
- valore 1 - *Router Alert*. Questo valore è utilizzato per informare il nodo che il pacchetto può richiedere altre operazioni oltre all'inoltro;
- valore 3 - *Implicit NULL*. Questo valore è impiegato nel protocollo *LDP (Label Distribution Protocol)* per la distribuzione delle etichette tra nodi.

Un pacchetto nel caso più generale, quando, ad esempio, si debbano attraversare *aree multibackbone*, vale a dire, aree gestite da differenti ISP, non trasporta una sola etichetta, ma una serie di etichette, organizzate in sequenza (a *stack*) di tipo *LIFO (Last In First Out).*

L'analisi dello *stack* in ogni nodo MPLS, *LSR (Label Switching Router),* avviene, allora, in maniera indipendente dal livello della gerarchia e sempre guardando l'etichetta posta in cima, senza considerare che altre etichette possono essere state inserite in precedenza "sotto di essa". Può essere rilevato che un pacchetto, a cui non sia stata ancora associata un'etichetta, abbia lo *stack* vuoto, mentre se un pacchetto possiede *m* etichette, quella posta in cima allo *stack* è definita di livello (o di gerarchia) *m*.

Quando un pacchetto è ricevuto, viene analizzata l'etichetta di livello più elevato nello *stack* e da questa lettura, sempre operando sulla base della tabella *NHLFE (Next Hop Label Forwarding Entry),* si ricavano le informazioni necessarie per agire correttamente nell'inoltro del pacchetto stesso, in pratica il *next-hop* da utilizzare, e le successive azioni da eseguire sullo stack, tra cui, ad esempio:

- sostituire l'etichetta posta in cima allo *stack* con una nuova;
- leggere lo *stack* delle etichette;
- inserire nuove etichette.

Per inoltrare, invece, un pacchetto privo di etichetta, in un LSR si analizza direttamente l'intestazione di livello 3 e, sempre dalla tabella NHLFE, si valuta dove instradare il pacchetto e quale etichetta inserire.

La figura 2 mostra un esempio di come avvenga

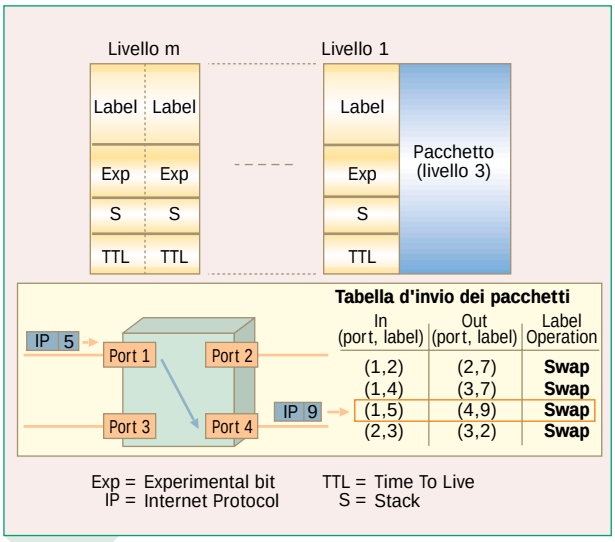


Figura 2 *Stack di etichette MPLS e paradigma di commutazione di etichetta.*

PERCHÈ MPLS

L'architettura di MultiProtocol Label Switching (MPLS), è nata alcuni anni fa per superare i limiti nelle prestazioni dei tradizionali router IP e nell'integrazione fra le reti IP e le reti ATM ma successivamente è stata sviluppata in modo innovativo e indipendente, allontanandosi dai primi obiettivi con essa perseguiti.

Nelle reti IP classiche, il piano di controllo (che decide dove instradare i pacchetti) ed il piano di attuazione (che effettua lo smistamento vero e proprio) agiscono entrambi sulla base dell'indirizzo IP di destinazione finale del pacchetto. Nelle architetture MPLS, invece, il controllo è reso indipendente dall'inoltro dei pacchetti attraverso l'introduzione di una nuova etichetta, la label che dà il nome alla stessa soluzione MPLS.

Nelle reti MPLS il piano di controllo analizza, infatti, la destinazione del pacchetto e, sulla base di logiche, eventualmente diverse e più elaborate di quelle degli instradamenti IP "tradizionali", definisce e inserisce le etichette. Da queste possono poi essere ricavate tutte le informazioni necessarie per inoltrare correttamente i pacchetti.

Questa caratteristica conferisce all'MPLS proprietà di scalabilità e di buona indipendenza dal livello di trasporto utilizzato, e consente di estenderne l'impiego ai sistemi ottici evoluti. Grazie, anche, alla maturità raggiunta dalle realizzazioni industriali, l'impiego di MPLS dà oggi, quindi, agli operatori la possibilità di introdurre nuove funzionalità e servizi nella rete IP.

Una nuova evoluzione sarà resa possibile nel prossimo futuro, quando sarà concluso l'iter per l'introduzione del trasporto diretto su reti completamente ottiche. Con queste reti sarà, infatti, impiegata una nuova architettura MPLS (chiamata *Generalised MPLS*), oggi già in fase di definizione.

una generica propagazione di un pacchetto IP, a cui sia stata applicata l'etichetta MPLS, quando attraversa un nodo MPLS. Secondo quanto riportato nella tabella d'instradamento del nodo rappresentato, un pacchetto proveniente dalla porta 1, con etichetta di ingresso pari a 5, deve essere inviato alla porta di uscita 4, dopo che gli sia stata associata un'etichetta di uscita pari a 9.

Quando anche l'ultima etichetta posta nello *stack* è letta, l'inoltro del pacchetto avviene unicamente sulla base dell'intestazione IP.

3.2 Commutazione di etichetta in MPLS: la componente di controllo

Le funzioni d'instradamento sono suddivise in due componenti: attuazione e controllo. Per effettuare il controllo devono essere inserite nuove prestazioni, legate alla distribuzione delle informazioni d'instradamento tra nodi LSR e alle procedure *(algoritmi)* che gli stessi nodi eseguono per costruire e per aggiornare le tabelle d'instradamento utilizzate, in modo da assegnare e da modificare le etichette.

L'architettura MPLS non fissa peraltro un unico criterio di realizzazione dei meccanismi di segnalazione necessari per la distribuzione e per l'allocazione delle etichette fra gli LSR.

Possono essere, ad esempio, utilizzati a questo scopo alcuni protocolli già esistenti, come il *BGP (Border Gateway Protocol)* [12], inserendo, all'interno dei pacchetti d'instradamento veri e propri, messaggi detti di *piggyback*, che contengono le informazioni relative alle etichette MPLS d'instradamento.

L'IETF ha, però, definito anche un nuovo protocollo, l'*LDP (Label Distribution Protocol)* [13] [14], con l'obiettivo di fissare l'insieme delle procedure attraverso le quali un LSR informa un altro LSR sulle etichette create e sulle associazioni tra percorsi d'instradamento ed etichette.

Due LSR che stabiliscano una comunicazione mediante LDP sono detti *label distribution peers*, rispetto alle informazioni scambiate; e si parla anche di *label distribution adjacency* tra i due LSR.

Le operazioni caratteristiche per l'allocazione e per la distribuzione delle etichette MPLS sono tre:

- *Downstream Label Allocation;*
- *Downstream Label Allocation on Demand;*
- *Upstream Allocation*.

Per un corretto funzionamento del meccanismo è necessario attivare nella rete MPLS un protocollo di routing come quello *IGP (Interior Gateway Protocol)*, che governa il popolamento delle tabelle d'instradamento dei singoli LSR.

Downstream Label Allocation

Nel *Downstream Label Allocation*, un LSR, nel momento in cui un particolare prefisso, il *FEC (Forwarding Equivalence Class)*, è stato appreso tramite messaggi che provengono dal protocollo di routing IGP, associa un'etichetta al prefisso e ad un percorso d'instradamento, la inserisce nella sua tabella d'instradamento, stabilisce un riferimento a essa nel proprio elenco di etichette valide, la *LIB (Label Information Base)*, e comunica, poi, agli LSR adiacenti la relazione tra etichetta d'ingresso e percorso d'instradamento.

Quando un LSR riceve, dal nodo successivo su un dato percorso d'instradamento, l'informazione che consente di stabilire su quel percorso un'associazione tra FEC ed etichetta (permette di effettuare cioè il cosiddetto *label-binding*), l'LSR pone l'etichetta tra quelle d'uscita della LIB che si riferiscono allo stesso percorso. In caso contrario, si limita ad associare un'etichetta a ciascun percorso disponibile.

Downstream Label Allocation on Demand

Nel *Downstream Label Allocation on Demand* un LSR identifica per ciascun percorso d'instradamento

un nodo subito "a valle" *(next hop)*.

Invia, poi, una richiesta (via LDP) per associare un'etichetta a quel percorso. Quando il nodo *next hop* riceve la richiesta, crea un'etichetta e la memorizza nel proprio archivio di etichette valide (nella propria LIB), producendo un'azione successiva che dipende dal modo di funzionamento che può essere di tipo *indipendente* o *ordinato*.

Nella *modalità indipendente* l'LSR interrogato restituisce immediatamente a quello "a monte", che ha effettuato la richiesta, la corrispondenza tra l'etichetta in ingresso e il percorso d'instradamento. Con questa informazione di associazione l'LSR "a monte" crea un elemento nella sua tabella d'instradamento, specificando il valore di etichetta ricevuto come identificativo in uscita, che consenta di raggiungere la destinazione fissata.

Il router LSR "a valle" ripete il processo, inviando una richiesta per la stessa destinazione al nodo successivo nel percorso.

Con questo comportamento, però, l'LSR "a valle" potrebbe non essere in grado di inoltrare un pacchetto in arrivo con un'etichetta per la successiva destinazione, poiché non è necessariamente detto che abbia già un'etichetta di uscita per quel percorso.

Nella *modalità ordinata* è invece avviato un processo mediante il quale l'LSR posto "a valle" nel percorso invia una nuova richiesta al suo *next hop*, anziché restituire la corrispondenza tra etichetta e percorso all'LSR "a monte", finché non venga raggiunto l'LSR di destinazione.

Quest'ultimo LSR invia l'informazione di *label binding* a quello precedente, avviando così un processo a ritroso per la propagazione dell'associazione su tutto il percorso individuato dal protocollo di routing IGP.

Con questa seconda procedura è risolto il problema relativo all'impossibilità temporanea di procedere con lo smistamento del pacchetto, in quanto ogni LSR invia verso "monte" il proprio *label binding solo quando ha a disposizione un'etichetta in uscit*a verso la destinazione interessata.

Upstream Label Allocation

Con la procedura di *Upstream Label Allocation*, un LSR alloca alcune etichette per ciascun percorso, contenuto nella propria tabella d'instradamento e raggiungibili da una delle sue interfacce. Aggiorna poi la propria LIB ponendo l'etichetta tra quelle in uscita e informa il nodo successivo su quel percorso dell'avvenuta associazione.

Il nodo di *next hop*, dopo aver ricevuto questa informazione, mette questa etichetta tra quelle in ingresso nella propria LIB.

Dopo aver inserito sia l'etichetta d'ingresso sia quella di uscita, l'LSR può inoltrare i pacchetti sul percorso individuato, utilizzando l'algoritmo di commutazione di etichetta.

Ogni volta che un LSR crea una nuova associazione tra un percorso e un'etichetta, aggiorna sia la tabella d'instradamento sia la LIB.

Quest'operazione permette di associare etichette anche ai pacchetti a cui non era stata assegnata in precedenza alcuna etichetta e, quindi, ai pacchetti che arrivano in ingresso alla rete MPLS dall'esterno.

Nell'ambito del processo di distribuzione e di allocazione delle etichette può anche presentarsi la situazione in cui un LSR riceva informazioni di *label binding* differenti provenienti da diversi LSR oltre che dal proprio nodo di *next hop*.

In questo caso l'LSR può mantenere in memoria le etichette provenienti dagli altri nodi, l'*LRM (Liberal Retention Mode)*, e può eventualmente utilizzarle se necessario (ad esempio nel caso dell'interruzione di una connessione), oppure può eliminarle, se si comporta con una procedura conservativa, il *CRM (Conservative Retention Mode)*, alleggerendo così il proprio carico elaborativo, ma riducendo allo stesso tempo la propria capacità di adattarsi ai mutamenti della rete.

3.3 Label Switched Path

Il cammino seguito da un pacchetto nel backbone MPLS prende il nome di *LSP (Label Switched Path)* e, genericamente, può essere definito di *livello m* per un particolare datagramma se si tratta di una sequenza di router $R_1 \ldots R_n$ con le proprietà di seguito elencate:

- inizia con un LSR *(LSP Ingress)* che inserisce nel pacchetto un'etichetta di *gerarchia m* (come descritto al precedente punto 3.1);
- tutti gli LSR intermedi nel LSP prendono le decisioni di *label switching* basandosi solo sull'etichetta di *livello m*;
- termina *(LSP Egress)* quando viene deciso di effettuare lo smistamento, basandosi su un'etichetta di livello differente (pari a m-k, con $k > 0$), o quando la decisione dello smistamento non è basata sulla procedura di *label switching*.

L'operazione di eliminazione dell'etichetta di livello *m* può essere eseguita dal *LSP Egress*, ma, in genere, risulta più efficiente se essa è eseguita dal penultimo LSR di un LSP. A livello architetturale questo comportamento risulta, infatti, perfettamente appropriato in quanto l'etichetta di gerarchia *m* ha la funzione di instradare il pacchetto sino a R_n, e, quando R_{n-1} ha deciso di indirizzarlo correttamente, non è più necessario il trasporto dell'etichetta.

L'utilizzo di questa tecnica, che prende il nome di *Penultimate Hop Popping*, evita, di fatto, la necessità di far eseguire per due volte dall'*LSP Egress* l'operazione di decisione d'instradamento: dapprima sulla base dell'etichetta di livello *m*, e poi dall'esame della parte restante del datagramma in modo da consentire l'instradamento verso la destinazione finale.

Per quanto concerne il modo per selezionare un LSP per una particolare FEC, il protocollo MPLS supporta due possibili meccanismi di *route selection:*

- *Hop by Hop Routing;*
- *Explicit Routing.*

Nel caso di un *Hop by Hop Routing* ciascun nodo sceglie il proprio *next-hop* in maniera indipendente dagli altri, sulla base delle informazioni contenute nella propria tabella d'instradamento, popolata ad esempio dalle rotte distribuite attraverso il proto-

Loop control nelle reti MPLS

Il valore di *TTL (Time To Live)*, associato a una trama MPLS in ingresso ad un LSP è il valore del campo TTL dell'etichetta MPLS, estraendola dallo *stack* al momento della ricezione della trama; questo valore è ricavato da quello corrispondente del campo TTL dell'intestazione IP relativa al pacchetto in maniera indipendente dalle etichette MPLS che sono state inserite o prelevate dallo stack.

Il valore di uscita del *Time To Live* può, invece, essere:

- **minore di uno, rispetto al valore di ingresso e in questo caso il pacchetto è inoltrato verso la sua destinazione;**
- **zero, e in questo caso il pacchetto è scartato se non è giunto alla destinazione.**

Una situazione critica si presenta, allora, quando l'etichetta MPLS è inserita nell'intestazione dello strato di collegamento (ad esempio, MPLS su ATM o su *Frame Relay)* e i pacchetti sono inoltrati come in un commutatore di livello 2 che non presenta alcun campo TTL e che non permette, quindi, di ridurre il TTL del pacchetto in corrispondenza di ciascuno dei "salti" LSR attraversati (in questo caso l'LSP prende il nome di *non-TTL LSP segment*).

Quando un pacchetto esce da un *non-TTL LSP segment*, dovrebbe avere un campo TTL che riflette il numero complessivo di LSR attraversati. Nel caso di un pacchetto *unicast* questa prestazione può essere ottenuta, fissando nell'*LSR ingress* il numero di router che compongono l'LSP e abilitandolo a diminuire il TTL di questo valore, prima che il pacchetto venga inoltrato, attraverso il *non-TTL LSP segment*.

Se dovesse accadere che il valore così ridotto fosse minore di zero, allora l'*LSR ingress* non dovrebbe eseguire l'instradamento del pacchetto sulla base dell'etichetta MPLS ma dovrebbe inoltrarlo secondo le regole dell'instradamento IP tradizionale.

In ogni caso, non potendo applicare in generale questo tipo di meccanismo, ogni

collo *OSPF (Open Shortest Path First)*.

In un *Explicit Route* LSP, invece, ogni LSR non esegue la scelta del *next hop* in maniera indipendente. Un singolo LSR, tipicamente l'LSP Ingress o l'LSP Egress, specifica in modo completo *(strictly)*, o quasi *(loosely)*, l'intero LSP.

Questo meccanismo può essere utile per molte ragioni, prima fra tutte, la possibilità di utilizzare MPLS a scopi di corretto bilanciamento del traffico sulle varie direttrici interne alla rete MPLS, in base, cioè, al *TE (Traffic Engineering)*.

4. Stato dell'arte e sviluppi in corso

Dal 1997 ad oggi il lavoro di standardizzazione dell'IETF su MPLS [15], [16] ha prodotto significativi risultati con la definizione di numerose *RFC (Request For Comment)*. Sono ancora aperti, tuttavia, alcuni Internet Draft, a conferma che lo stato dell'arte su MPLS è tuttora in costante e crescente sviluppo (la sintesi della normativa IETF relativa alle architetture e ai protocolli MPLS è illustrata nel riquadro a pagina 46).

A livello di sperimentazione, uno degli obiettivi più attraenti dell'MPLS riguarda oggi la possibilità di utilizzare un meccanismo equivalente *MPλS (MultiProtocol Lambda Switching)* [17], per riuscire a portare IP direttamente sulle reti ottiche, e quindi a realizzare i sistemi di commutazione ottica attraverso i meccanismi d'instradamento IP riducendo, in particolare, il numero dei livelli tipici delle attuali reti per dati (da IP/ATM/SDH/WDM a IP/WDM) via via che si estenda l'impiego della rete *WDM (Wavelength Division Multiplexing)*.

In sostanza, MPλS si propone di combinare i vantaggi che MPLS introduce in termini di *Traffic Engineering* (a livello, quindi, di piano di controllo) con le tecnologie emergenti di commutazione fotonica e ottica, per realizzare reti capaci di fornire in tempo reale servizi di trasporto attraverso canali ottici.

In questo modo si dovrebbe permettere l'utilizzo di una semantica uniforme per tutte le operazioni di controllo e di gestione di reti ibride costituite da permutatori ottici e da sistemi SDH/SONET; da router IP/MPLS; da commutatori ATM e Frame Relay.

Per realizzare MPλS sarà necessario introdurre

volta si attraversi un *non-TTL LSP segment*, è necessario prevedere procedure alternative di *loop control* [35].

Una possibile soluzione potrebbe essere quella di utilizzare un'allocazione di buffer controllata limitando la quantità di memoria riservata a ogni singolo circuito virtuale dei commutatori ATM, ottenendo come effetto quello di contenere i problemi causati dal loop in ambiente MPLS.

Questa tecnica dovrebbe anche consentire di tenere sotto controllo (e di evitare) i problemi causati dai *loop transitori*, ossia la perdita di pacchetti nel tempo impiegato dagli algoritmi d'instradamento a recuperare la convergenza sullo stesso instradamento.

Possono essere, però, impiegate anche altre due tecniche di *loop control*. La prima è basata sull'idea chiamata *path vector*. Un *path vector* è una lista di LSR che sono stati attraversati da un messaggio di *Label Request* o *Label Mapping*. Ad esempio, un messaggio di *Label Request* inviato da un LSR non ATM verso uno ATM contiene un *path vector* con l'indirizzo IP dell'LSR richiedente. L'LSR ATM aggiunge il proprio indirizzo al path vector prima di inoltrare una *Label Request* per questo flusso di dati al successivo LSR. Se si generasse un *loop* nell'instradamento dei pacchetti, un LSR vedrebbe nel *path vector* il proprio indirizzo IP e potrebbe segnalare il problema, avviando procedure per il recupero della convergenza.

Un secondo meccanismo di protezione è basato sul cosiddetto *colored thread*. È possibile applicarlo a qualsiasi tipologia di LSR, ma è necessario mantenere traccia della sequenza in cui vengono generati gli LSP. L'idea su cui si basa è piuttosto semplice e consiste nell'associare la generazione dell'intero LSP a un colore (un campo dell'etichetta MPLS) in modo tale che, se un LSR lungo il percorso vede passare per due volte lo stesso colore, riconosce immediatamente la presenza di un *loop* ed interrompe il processo di generazione dell'LSP fino a che i protocolli d'instradamento non risolvono il problema.

Si tratta di una tecnica che è in grado, non solo di mitigare, ma anche di risolvere in maniera efficiente i problemi di *loop*, così come il metodo prima descritto del *path vector*; al tempo stesso, essa permette di ridurre notevolmente il sovraccarico di informazioni da trasmettere e da immagazzinare in ogni router.

alcune modifiche su MPLS, sia in termini di protocolli d'instradamento, sia di protocolli di segnalazione, per adattarlo alle caratteristiche peculiari dei permutatori ottici.

Questi cambiamenti dovrebbero sinteticamente riguardare:

- l'introduzione di un nuovo protocollo, l'*LMP (Link Management Protocol)* [18], per la gestione delle reti ottiche, che, in particolare, controllerà la connettività fra canali adiacenti e che permetterà di rilevare le condizioni di guasto e di fuori servizio;
- l'impiego di un protocollo, che adatti gli annunci OSPF-versione 2 [19] e IS-IS versione-3 [20] riguardanti la disponibilità di risorse nella rete ottica (numero di lunghezze d'onda disponibili, banda sulle singole lunghezze d'onda, …);
- un'estensione dei protocolli dedicati alla prenotazione delle risorse, RSVP-TE [21], per permettere di definire a priori percorsi espliciti attraverso la dorsale della rete ottica.

Il passo successivo nell'evoluzione dell'MPLS dovrebbe condurre a definire la commutazione generalizzata di etichetta *GMPLS (Generalized MPLS)* [22], che dovrebbe permettere di utilizzare MPLS come meccanismo di controllo per configurare LSP, non solo costituiti da router IP, ma realizzati anche attraverso apparati di tecnologie differenti, come i commutatori ottici ed i multiplatori TDM e quelli ADM dei sistemi SDH/SONET.

Anche il GMPLS richiederà il protocollo per la gestione delle reti ottiche LMP e le estensioni di OSPF e IS-IS precedentemente descritte, per valutare la disponibilità di risorse in questa rete mentre gli altri aspetti maggiormente significativi riguarderanno:

- *Link Bundling:* e cioè il raggruppamento di più collegamenti fisici indipendenti in una singola connessione a livello logico;
- *Link Hierarchy:* e cioè la definizione di pile di etichette in grado di gestire, a livello logico e fisico, tutti i possibili percorsi di rete;
- *Unnumbered Links:* e cioè la capacità di configurare cammini d'instradamento senza che le interfacce costituenti, sia logiche sia fisiche, posseggano un indirizzo IP pubblico;
- *Constraint Based Routing:* e cioè il protocollo per supportare le funzionalità di Traffic Engineering.

Normativa del MPLS

In questa scheda sono presentati i documenti prodotti dall'*IETF (Internet Engineering Task Force)* relativi all'architettura MPLS, raggruppati in base agli argomenti funzionali trattati in modo da facilitarne l'inquadramento (si veda in proposito[15]).

Architettura e strutture basilari:

- **RFC 3031[11]: *Multiprotocol Label Switching Architecture*, definisce specificatamente la struttura dell'architettura di MPLS, così come è stata delineata nel paragrafo 3;**
- **RFC 3032: *Label Stack Encoding*, specifica le procedure di *encoding* da parte di un LSR per inviare pacchetti con stack di etichette attraverso vari *datalink* (PPP, LAN, SONET, ...);**
- **RFC 3063 [35]: *MPLS Loop Prevention Mechanism*, specifica la maniera per eseguire il loop control basata sul meccanismo di *colored thread*.**

Trasporto dei pacchetti MPLS:

- **RFC 3034: *Use of Label Switching on Frame Relay Networks Specification*, normalizza le procedure di realizzazione di MPLS su Frame Relay.**
- **RFC 3035: *MPLS using LDP and ATM VC Switching*, definisce specificatamente le caratteristiche di MPLS su ATM;**

Meccanismi base di segnalazione:

- **RFC 3036: *LDP Specification*, stabilisce le caratteristiche del protocollo LDP;**
- **RFC 3037: *LDP Applicability*, specifica il campo di applicabilità delle precedenti procedure;**
- **RFC 3107: *Carryng Label Information in BGPv4*, indica come può essere utilizzato il protocollo BGP per distribuire insieme a un determinata rotta l'etichetta associata *(messaggi piggyback)*.**

MPLS TE (Traffic Engineering):

RFC 2702 *(Requirements for Traffic Engineering Over MPLS)* stabilisce le specifiche per la realizzazione delle funzionalità di TE in una rete MPLS e, a partire da questa norma, sono in via di completamento numerose bozze di specifiche *(Internet*

5. Applicazioni

L'introduzione dell'architettura e delle funzionalità MPLS in una rete IP può consentire di:

a) gestire le funzioni di Traffic Engineering per un impiego ottimale delle risorse di rete da parte degli ISP;
b) realizzare *VPN (Virtual Private Network)* IP, ossia realizzare infrastrutture di Intranet e di Extranet, gestite dagli ISP per conto dei siti clienti, che spesso sono reti Internet estese;
c) consentire la predisposizione di nuove *CoS (Classes of Service)* con relativa QoS, nell'ambito della fornitura di servizi differenziati, attraverso la realizzazione del modello *Diffserv* congiuntamente a meccanismi di Traffic Engineering;
d) garantire un rapido reinstradamento *(fast rerouting)* per migliorare l'affidabilità, la robustezza e la qualità dei servizi offerti dalle reti IP.

Nei paragrafi successivi saranno esaminate tutte le applicazioni prima citate, che ad oggi sono quelle maggiormente diffuse e ritenute più significative per l'offerta di nuovi servizi da parte di Telecom Italia.

5.1 MPLS Traffic Engineering

Le funzioni di Traffic Engineering consentono a

***draft)* relative ad alcuni aspetti di questi problemi:**

- **estensioni al protocollo RSVP per il supporto di tunnel LSP e relative possibili applicazioni (Internet draft-ietf-mpls-rsvp-tunnel-applicability-02.txt: *Applicability Statement for Extensions to RSVP for LSP-Tunnels*), definisce le estensioni da apportare al protocollo RSVP per realizzare le funzioni di TE;**
- **definizione dello standard CR-LDP per instaurare e modificare i tunnel [36];**
- **estensioni di RSVP e CR-LDP per il ripristino degli LSP in presenza di condizioni di guasto [36];**
- **definizione di meccanismi di recupero degli errori dovuti ad avarie, sia a livello hardware sia software, per migliorare l'affidabilità di una dorsale di rete MPLS;**
- **meccanismi di *Fast Rerouting* (estensione delle caratteristiche della segnalazione).**

MPLS e DiffServ:

Sono aperti alcuni *Internet draft* [36] per la definizione del modello in cui MPLS è realizzato, in un contesto architetturale IP con classi differenziate con la procedura Diffserv.

MPLS Multicast:

Internet draft-ietf-mpls-multicast-06.txt: *Framework for IP Multicast in MPLS*. La norma dà indicazioni per il supporto di comunicazioni IP multicast su MPLS.

MPLS e VPN:

- **RFC 2764: *A Framework for IP Based Virtual Private Networks*. La specifica definisce l'architettura di VPN basate su MPLS, intorno alla quale sono attivi numerosi ambiti di ricerca inerenti le estensioni del MP-BGP per *IPv6, Multicast, label stack encapsulation, VPN ottiche;***
- **RFC 2283[37]: *Multiprotocol Extensions for BGP-4*, fissa le caratteristiche delle estensioni al protocollo di routing esterno BGP.**

Aspetti di gestione:

Ci si riferisce in particolare al *Management Information Base* [36] per MPLS, LDP e MPLS-TE.

un ISP di instradare un certo flusso di traffico lungo un percorso differente da quello individuato dai normali protocolli d'instradamento, in modo da utilizzare, qualora sia necessario, un percorso fisico meno congestionato.

Con il TE si vuole, infatti, evitare che alcuni collegamenti di una rete siano sovraccarichi mentre altri, nello stesso istante, siano inutilizzati, determinando così un'elevata inefficienza della rete.

Si è osservato che, in generale, le congestioni di rete si verificano in due situazioni:

- quando le risorse di rete, come la banda trasmissiva, siano insufficienti a smaltire il traffico a esse offerto;
- quando i flussi di traffico siano trasportati in maniera non efficiente dalle singole risorse di rete.

La capacità di controllo offerta dagli attuali *IGRP (Interior Gateway Routing Protocol)* è inadeguata a risolvere questi problemi. Questi protocolli, essenzialmente di tipo *SPF (Shortest Path First)*, calcolano, infatti, un determinato cammino sulla base della topologia della rete, cercando di rendere minimo il costo a seconda della metrica adottata. Questo sistema di controllo non offre alcun valore aggiunto in termini di TE in quanto esso trascura totalmente, in fase decisionale, le caratteristiche del traffico da trasportare che, se note a priori (ad esempio, classi

VANTAGGI E PUNTI DI ATTENZIONE NELLE REALIZZAZIONI ATTUALI

L'architettura MPLS risulta più articolata di quella di una rete IP "tradizionale", più flessibile ma anche maggiormente complessa da gestire e da amministrare.

Sono stati, infatti, introdotti in rete nuovi protocolli di segnalazione e di instradamento, che integrano quelli esistenti per la realizzazione di servizi più evoluti. Si possono avere spazi di indirizzamento privati sovrapposti per i servizi legati alle reti private virtuali (VPN); sempre in ottica di servizio, si possono distinguere livelli di qualità di servizio (QoS) all'interno delle singole VPN, differenziandoli in modo controllato per dare supporto efficace ad applicazioni real-time; si possono definire e realizzare percorsi di instradamento espliciti protetti, per applicazioni di gestione del traffico in rete *(Traffic Engineering, Fast Rerouting)*, tramite le quali è possibile ridurre i tempi di reazione ai guasti di una rete IP ai valori tipici (dell'ordine di 50 ms) delle reti di trasporto ottiche.

Queste applicazioni sono ritenute al momento di maggior interesse e sono offerte dalle reti di Telecom Italia.

Per gestire poi in modo efficace questa maggiore complessità, occorrerà disporre di opportuni strumenti di gestione e di amministrazione della rete, che dovranno essere progressivamente introdotti dagli operatori all'aumentare del numero di clienti e di applicazioni che utilizzano questa nuova architettura e le funzionalità da essa offerte.

di servizio) possono essere utilizzate nella scelta e nella definizione di metriche alternative per i protocolli d'instradamento.

Prima di adottare l'MPLS-TE, per modificare un percorso calcolato attraverso un protocollo tradizionale di routing potevano essere impiegati differenti meccanismi d'instradamento:

- *a livello IP*, poteva essere modificata la metrica del protocollo, o poteva essere applicato un meccanismo di *Equal Cost Multipath* [23]. In entrambi i casi si tratta di soluzioni in qualche misura "palliative": con il primo sistema si modifica, in sostanza, il percorso scelto, per cui a fronte di un nuovo mutamento dello stato della rete si ritorna alla situazione di partenza, e per di più non è, né conveniente, né semplice operare sulla metrica del protocollo. Con il secondo meccanismo, pur distribuendo per una data destinazione il traffico sulle varie connessioni disponibili, nelle attuali realizzazioni non si tiene conto della banda disponibile sulle stesse connessioni;
- *a livello 2*, può essere adottato un modello del tipo *overlay*, cioè una topologia virtuale costruita sulla topologia fisica della rete. Possono essere utilizzati ad esempio commutatori ATM con circuiti virtuali permanentemente configurati per distribuire il carico uniformemente *(load balancing)*. I problemi che si pongono in questo caso sono legati sia al costo elevato di gestione di due reti (anziché di una) sia alla scarsa modularità della soluzione.

Lo sviluppo delle funzionalità di *Traffic Engineering* su MPLS consente, invece, di superare questi problemi attraverso un'integrazione delle tecnologie di livello 2 e 3, in quanto esso:

- elimina la necessità di configurare manualmente gli apparati di rete per fissare percorsi predefiniti, in quanto utilizza protocolli di segnalazione;
- il calcolo degli LSP è fatto sulla base delle risorse disponibili nella rete in quel momento e tenendo presente le caratteristiche specifiche richieste dal particolare flusso di dati che si deve trasportare attraverso la rete;
- possiede un meccanismo adattativo rispetto ai mutamenti della rete a livello topologico, dovuti a guasti o all'inserimento di nuovi nodi.

Per stabilire gli LSP da utilizzare nell'ottica TE sono utilizzati meccanismi di controllo e di segnalazione differenti rispetto a quelli tipici dell'architettura base di MPLS. Queste differenze sono sintetizzate nella tabella 1.

In una rete MPLS-TE gli LSP sono configurati mediante percorsi espliciti, gli *ER (Explicit Route)*, calcolati e definiti nei router all'ingresso del dominio MPLS, che poi utilizzano un protocollo di segnalazione per coordinare la distribuzione delle etichette nei nodi lungo il percorso, riservare la banda, modificare le risorse e, eventualmente, per indicare la *CoS (Class of Service)* del traffico entrante.

Architettura	Meccanismo di instradamento	Meccanismo di segnalazione
MPLS Base VPN	OSPF, BGP	LDP, External BGP
MPLS-TE	CSPF	RSVP-TE, CR-LDP

BGP = Border Gateway Protocol
CR = Constraint-based Routing
CSPF = Constraint Shortest Path First
LDP = Label Distribution Protocol
MPLS = MultiProtocol Label Switching
OSPF = Open Shortest Path First
RSVP = Resourse reSerVation Protocol
TE = Traffic Engineering

Tabella 1 *Meccanismi di instradamento e di segnalazione utilizzati nelle reti MPLS.*

I protocolli che possono essere utilizzati a questo proposito sono l'*RSVP-TE (Resource ReserVation Protocol-Traffic Engineering)* [24] e quello *CR-LDP (Constraint-based Routing LDP)* [26].

Il protocollo *RSVP* [27] è stato studiato originariamente per riservare alcune risorse della rete a un dato flusso di pacchetti IP. La sua estensione RSVP-TE è, invece, funzionale al supporto delle *ER (Explicit Route)* e può essere utilizzata per la distribuzione delle etichette MPLS in un tunnel TE.

RSVP-TE opera a livello 3, utilizzando datagrammi IP - o eventualmente *UDP (User Datagram Protocol)* - per comunicare con gli LSR adiacenti, non richiedendo di mantenere sessioni *TCP (Transmission Control Protocol)*. Quando un LSR ingress (testa del percorso) individua la necessità di inizializzare un nuovo LSP sino a un *LSR egress (coda del percorso)* esegue un calcolo del cammino da seguire, ossia determina un'ER sulla base sia dei parametri richiesti per la determinata sessione sia delle politiche di gestione e amministrative adottate per la rete.

A partire dal calcolo effettuato, l'LSR ingress costruisce un messaggio di *path* che contiene la successione dei nodi da attraversare (ER) insieme ai parametri del traffico da trasportare e quindi lo invia in un datagramma IP.

Quando un generico LSR lo riceve inoltra la richiesta verso il nodo successivo, presente nell'ER, e il processo prosegue finché il pacchetto giunge al router di coda del tunnel, dove, a partire dai parametri in esso indicati, è calcolata la banda richiesta per il flusso di traffico in questione in modo da allocarla qualora disponibile.

Viene, poi, selezionata un'etichetta per il nuovo LSP che è distribuita a ritroso lungo il cammino, mediante un messaggio di *resv*. Quando un LSR del percorso riceve un messaggio di *resv*, esso determina la corrispondenza rispetto alla richiesta originale di *path* (mediante l'identificativo di LSP contenuto sia nel messaggio di *path* che in quello di *resv*); valuta poi le risorse da riservare, aggiorna la propria *FT (Forwarding Table)* e si alloca un'etichetta per l'LSP.

Questa etichetta è comunicata al router precedente in un nuovo messaggio di *resv*, e il processo è ripetuto finché non è raggiunto il nodo alla testa del percorso che completa la procedura di instaurazione del tunnel TE.

Il *protocollo CR-LDP* è, invece, un'estensione del protocollo LDP e ha caratteristiche atte al supporto e alla costruzione di LSP espliciti. Come LDP, esso utilizza alcune sessioni TCP tra i router di un dominio MPLS, durante le quali sono scambiati messaggi che consentono di assegnare le etichette, ottenendo, così, un meccanismo affidabile di distribuzione delle informazioni di controllo.

In questo caso, *LSR ingress* quando individua la necessità di creare un nuovo LSP, invia al primo LSR del percorso un messaggio di *Label_Request* contenente l'*explicit route* e i parametri del traffico che sarà inviato. Quando un nodo riceve lungo il cammino il messaggio di *Label_Request* inoltra la richiesta all'LSR a valle, riservando le risorse indicate al nuovo LSP.

L'*LSR egress* alloca le risorse necessarie e assegna un'etichetta al percorso appena costituito, che è distribuito, a ritroso, mediante un messaggio di *Label_Mapping* contenente anche le caratteristiche relative alle risorse allocate.

Quando un LSR lo riceve controlla, sulla base dell'identificativo di LSP, la corrispondenza con il precedente messaggio di *Label_Request*, ed effettua la prenotazione di risorse con la definizione di un'etichetta per l'LSP. Il meccanismo prosegue sino a quando il messaggio di *Label_Mapping* non abbia raggiunto l'*LSR ingress* che completa la procedura di segnalazione per la costruzione del percorso MPLS-TE.

Il CR-LDP permette di realizzare una ER-LSP in due modi diversi: la prima, *strict*, e l'altra, *loose*.

Nella *modalità strict*, utilizzata solitamente da Enti che hanno la gestione della rete, la ER-LSP è completamente specificata dal nodo posto alla testa del tunnel. Si tratta in sostanza di un particolare cammino costruito manualmente da un operatore di un centro di gestione, che ha il completo controllo della rete.

Nella *modalità loose*, l'operatore non deve specificare ogni singolo nodo costituente il cammino, bensì, può selezionare un gruppo di nodi che dovranno costituire l'LSP lasciando alcuni gradi di libertà ai protocolli d'instradamento.

La definizione di una ER nella modalità *loose* comporta che l'*LSR ingress* sia in grado di eseguire le seguenti operazioni:

- memorizzare le informazioni provenienti dai protocolli interni IGP;
- registrare le informazioni di TE;
- calcolare il cammino fisico, ossia l'insieme dei nodi costituenti l'LSP;
- rappresentare il percorso mediante un ER e passare questo attributo al CR-LDP per la procedura di segnalazione.

Nelle prime due fasi, nel router di testa dell'LSP, il protocollo Extended IGP [25] contribuisce a distribuire le informazioni topologiche e di TE nella tabella d'instradamento e nella base di dati TE, utilizzando come estensioni IGP anzitutto la massima banda riservabile su di un collegamento (con otto differenti livelli di priorità impostabili) e, in secondo luogo la banda riservabile residua (ancora con otto differenti livelli di priorità) e, infine, eventuali gruppi amministrativi di collegamenti.

In particolare, quest'ultimo aspetto è realizzato associando a ogni collegamento un colore (o eventualmente più di uno), in quanto i colori sono rappresentati mediante vettori i cui bit individuano un singolo elemento. Il calcolo *on-line* del nuovo LSP è realizzato dall'algoritmo *CSPF (Constrained Shortest Path First)* che utilizza come dati di ingresso - oltre quelli contenuti nella base di dati del TE - anche una serie di informazioni, stabilite dallo stesso utente per definire il percorso, che vanno dai requisiti di banda alle limitazioni sugli hop del cammino, ai vincoli sui gruppi amministrativi (colori), alle priorità di inizializzazione e agli eventuali ER.

Il processo di selezione del percorso *constrained-based* fruisce anche delle tabelle d'instradamento IP e di informazioni di segnalazione. Questa procedura permette una notevole flessibilità a livello locale per soddisfare i vincoli imposti lungo il tunnel TE.

Una volta che l'algoritmo CSPF ha determinato l'ER, questa informazione è passata al CR-LDP (o RSVP-TE) ed è instaurato l'LSP come *downstream on demand*.

Confrontando le due tecniche di TE basate su MPLS, può essere rilevato come la soluzione più efficace dovrebbe essere in realtà un ibrido che utilizza, in parte la *modalità loose* per la definizione dei *path*, per assicurarsi rapidità di adattamento ai cambiamenti di rete e una regolazione fine tra differenti fasi di riottimizzazione, e, in parte, quella *strict*, per

ottimizzare periodicamente la rete in maniera globale e centralizzata.

Le differenze principali fra i protocolli CR-LDP e RSVP-TE, sono, invece, legate all'affidabilità del protocollo di trasporto utilizzato per la distribuzione delle etichette e ai criteri per la definizione delle ER. Da queste due differenze fondamentali discendono poi altre difformità di minore importanza.

Per quanto riguarda il protocollo di trasporto, mentre RSVP-TE utilizza UDP - o addirittura può non utilizzare alcun protocollo di trasporto inviando i propri messaggi direttamente all'interno di pacchetti IP - CR-LDP utilizza TCP che risulta essere molto più affidabile. In questo modo risulta essere più semplice ed efficace la gestione di malfunzionamenti della rete, nonché più veloce il reinstradamento dell'ER.

I costruttori non sono pervenuti finora a una scelta comune unica: alcuni infatti, come Cisco, hanno nei propri apparati RSVP-TE, altri, come Nortel, utilizzano CR-LDP.

5.2 Reti Private Virtuali MPLS/IP

Una Rete Privata Virtuale, o *VPN (Virtual Private Network),* è costituita da un insieme di siti di Clienti la cui connettività è basata su una struttura condivisa, dotata delle stesse politiche amministrative applicabili a una struttura privata (ad esempio piano di indirizzamento individuale, traffico limitato ai siti dei clienti).

Sia dal punto di vista dell'instradamento, che da quello della riservatezza, la rete può, infatti, essere classificata come privata, nel senso che per una VPN tutte le altre sono "trasparenti", vale a dire non può essere utilizzata da utenti esterni, e che l'instradamento e il piano di indirizzamento interno - eventualmente privato - sono completamente indipendenti dall'instradamento e dal piano di indirizzamento di tutte le altre reti.

La rete è, invece, virtuale nel senso che l'utilizzo del mezzo fisico è in realtà condiviso fra più utilizzatori *(VPN customer)* ciascuno dei quali desidera disporre di una propria VPN, mentre il fornitore è tipicamente una terza parte che può essere definito come *VPN service provider*.

Nel caso più generale, una VPN è costituita da una serie di siti interconnessi tra loro, a cui è possibile applicare diversi criteri di connessione anche se operano all'interno di una medesima struttura amministrativa. Si può anche fare in modo che un sito, appartenente a una certa VPN, comunichi solo con un sottoinsieme di altri siti appartenenti a un'altra VPN. Nel primo caso si può parlare di *Intranet VPN*, mentre nel secondo di *Extranet VPN*. In virtù di questa definizione risulta chiaro che un determinato sito può appartenere a una o più VPN.

Sino a qualche tempo addietro la maggior parte delle tecniche utilizzate per la realizzazione di VPN era basata sul modello *overlay* [9], nel quale ciascun sito ha uno o più router connessi agli altri siti - o eventualmente a un loro sottoinsieme - mediante collegamenti punto-punto (tipicamente realizzati con tecnologia *Frame Relay* o ATM, in ogni caso a livello 2).

Soluzioni di questo tipo presentano, però, notevoli inconvenienti sia in termini di scalabilità - richiedono una magliatura completa o parziale di collegamenti punto-punto, con una dimensione dell'ordine del quadrato del numero dei router in rete, e cioé $O(N^2)$, sia anche di capacità gestionale del Cliente in termini d'instradamento IP, QoS, ATM o *Frame Relay*.

Il modello MPLS VPN/IP basato sul concetto del "peer" [9] consente, invece, di superare la maggior parte delle precedenti limitazioni e, in particolare, consente ai *VPN service provider* di fornire VPN su larga scala, permettendo, al contempo, di offrire il servizio agli utenti senza richiedere un'esperienza a livello d'instradamento IP in quanto si riduce notevolmente il numero complessivo delle connessioni di livello 2 necessarie.

In tal modo risulta decisamente semplificata la fornitura di servizi VPN rispetto al numero di siti connessi e si riesce a ottenere una connettività *any-to-any* altamente scalabile per Intranet estese e per Extranet che offrano nuovi servizi a valore aggiunto.

Le MPLS VPN consentono, anche, di raggiungere livelli di sicurezza e di riservatezza delle informazioni affidabili e di offrire più classi di servizio, sia all'interno della stessa VPN che fra più VPN.

Nel seguito del paragrafo sono descritte le caratteristiche peculiari più significative dell'architettura del modello MPLS VPN/IP e sono quindi meglio chiarite le precedenti affermazioni.

La struttura tipica di un'area MPLS VPN/IP è rappresentata nella figura 3 dove gli elementi fondamentali sono:

- *PE (Provider Edge router):* sono i router utilizzati dai clienti per accedere all'area MPLS;
- *CE (Customer Edge router):* sono i router che raggruppano l'insieme dei siti del Cliente e sono collegati direttamente con i PE router per accedere alla rete MPLS pur ignorando completamente la struttura del dominio MPLS;
- *P (Provider router):* sono i router dell'area interna di una MPLS VPN.

In una MPLS VPN/IP il router del Cliente, il *CE (Customer Edge)* è connesso al router di accesso del provider *(PE)* con una logica di interconnessione indipendente dal particolare tipo di tecnologia utilizzata per il livello fisico e per il collegamento.

Il meccanismo di controllo della connettività fra i siti è realizzato mediante una *constrained distribution of routing information*, con uno schema, cioè, di distribuzione delle informazioni d'instradamento che può essere decomposto in cinque passi:

a) le informazioni d'instradamento sono, dapprima, avviate dal Cliente (nodo CE) al Service Provider al nodo *PE (Provider Edge)* attraverso uno dei quattro criteri riportati di seguito: instradamento statico; RIPv2; BGPv4; OSPF;
b) sul nodo PE le informazioni d'instradamento sono esportate nel protocollo *BGP (Border Gateway Protocol)* del Service Provider;
c) le informazioni d'instradamento sono poi propagate, tramite *I-BGP (Internal BGP)*, tra i nodi PE a cui sono attestati i siti di una stessa VPN;
d) le informazioni precedenti d'instradamento sono importate dal protocollo di routing BGP nel nodo

PE di uscita (fase esattamente complementare a quella del punto b prima indicata);

e) le informazioni d'instradamento sono propagate dal Service Provider (nodo PE) al client (nodo CE) (fase complementare a quella del punto a).

In una struttura di questo tipo, i router definiti come PE ricevono e memorizzano le informazioni d'instradamento relative solo alle VPN direttamente connesse. Il numero delle informazioni d'instradamento mantenute sul PE è, quindi, direttamente proporzionale al numero di VPN direttamente connesse a ciascuno di essi.

Inoltre, ogni nodo CE mantiene informazioni sui *peer* di tipo PE ai quali è direttamente connesso, trascurando tutti gli altri siti di Clienti della VPN di appartenenza.

È proprio grazie a tale meccanismo che questa soluzione risulta decisamente più scalabile rispetto a quella *overlay*, non solo per quanto riguarda il numero di connessioni risparmiate (è, infatti, evitata la realizzazione di una magliatura completa fra i CE), ma anche perché, per aggiungere o per eliminare un sito da una VPN, è semplicemente necessario aggiornare il database del PE al quale un proprio CE risulta direttamente connesso (indipendentemente dal numero totale di siti presenti in rete).

I router PE non possiedono un'unica tabella d'instradamento per gestire questo meccanismo, ma ne gestiscono differenti, ognuna delle quali - relativa a una particolare VPN - prende il nome di *VRF (VPN Routing Forwarding table)* e, come le normali tabelle d'instradamento IP, si compone di due sottotabelle: la *VRF IP routing table*, nella quale sono contenute le informazione d'instradamento verso le destinazioni appartenenti alla VPN, e la *VRF IP forwarding table*, nella quale sono comprese le informazioni relative alla commutazione dei pacchetti da un'interfaccia entrante a una uscente dal router. L'associazione di un sito a una determinata VPN avviene, quindi, in base all'interfaccia logica di attestazione del CE al PE.

È importante sottolineare che, mentre non è necessario stabilire una relazione uno-a-uno fra siti di utente e VPN, un sito può però essere associato solo a una VRF, che contiene tutte le possibili rotte disponibili al sito e poste all'interno della VPN al quale esso appartiene. Può essere definita una sola istanza VRF su ciascuna interfaccia, mentre è consentito associare la stessa VRF a più interfacce (questo è il caso in cui più siti della stessa VPN siano connessi allo stesso PE su interfacce distinte).

Sulla base delle informazioni contenute nella *VRF IP routing table* e nella *VRF forwarding table*, i pacchetti sono consegnati alla loro destinazione attraverso un meccanismo di inoltro MPLS, che consente di superare il problema relativo all'utilizzo di cammini espressi in termini di *VPN IP address:* a questo scopo è, infatti, disaccoppiata l'informazione utilizzata per l'instradamento dei pacchetti (etichetta MPLS) da quella contenuta nell'intestazione IP.

Sono, in pratica, elaborati e stabiliti alcuni cammini sulla base delle informazioni d'instradamento private contenute nelle singole VRF ed i pacchetti sono successivamente inoltrati lungo questi percorsi mediante MPLS.

Dal punto di vista delle caratteristiche peculiari di un MPLS, un router PE non è altro che un *LSR (Label Switch Router)* di tipo *Egress* che esegue l'operazione di assegnazione delle etichette ai pacchetti che ne sono sprovvisti e di eliminazione, sempre delle etichette, a quelli diretti verso i router CE.

In realtà, per migliorare la modularità e l'espandibilità della rete, un PE quando riceve un pacchetto non etichettato da uno dei CE direttamente connessi, gli assegna una coppia di etichette gerarchiche. Quella di primo livello *(interna)* è associata a un percorso verso il PE di destinazione, e di conseguenza garantisce il corretto instradamento da un PE di ingresso a uno di uscita, mentre quella di secondo livello *(esterna)* controlla l'inoltro dei pacchetti MPLS sino al penultimo PE, prima cioè del PE di destinazione (vedi figura 4).

L'utilizzo di questa tecnica permette ai router dei *provider* P della dorsale MPLS di non memorizzare informazioni riguardanti direttamente l'instradamento interno alle singole VPN, ma di registrare solo quelle relative all'inoltro dei pacchetti mediante MPLS (ossia mediante il LIB) verso i router a essi direttamente connessi.

Le etichette di primo livello sono tipicamente distribuite mediante sessioni *BGP (Border Gateway Protocol)* insieme con i percorsi VPN/IP; quelle di

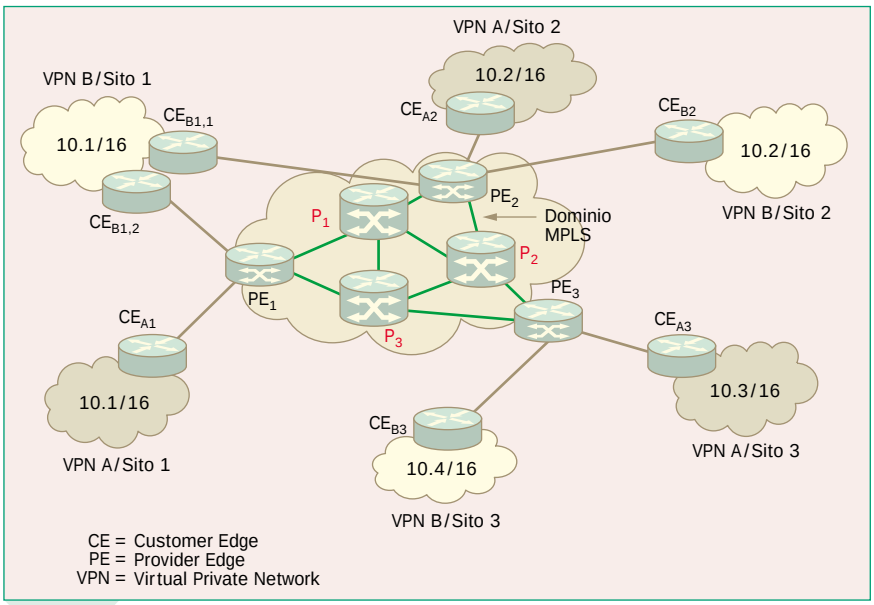


Figura 3 *Architettura di una VPN MPLS.*

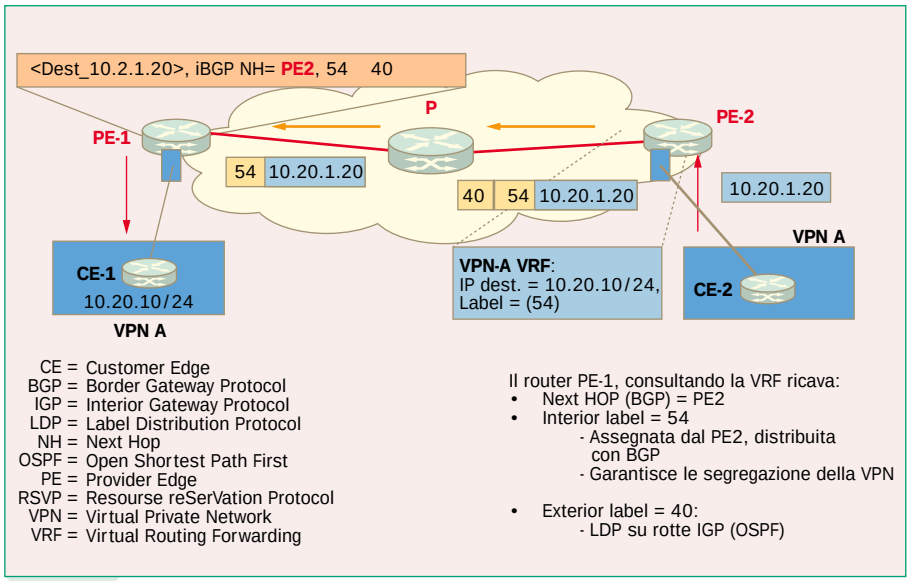


Figura 4 *Esemplificazione del funzionamento delle VPN MPLS con doppia gerarchia di etichette.*

secondo livello mediante *LDP (Label Distribution Protocol)* (come mostrato nella figura 4).

Per avere un'idea concreta del guadagno che si ottiene in termini di scalabilità utilizzando una tecnica MPLS gerarchica basta considerare l'esempio di un service provider che possiede duecento router (PE e P) e che gestisce 10mila VPN, ciascuna mediamente con 100 route. Senza l'ausilio della tecnica MPLS gerarchica, ogni provider (P) router dovrebbe mantenere 10mila X 100 = 1 milione di rotte, mentre con MPLS sono sufficienti per ogni P router, duecento rotte verso tutti gli altri router della dorsale.

Le informazioni d'instradamento sono, dunque, mantenute nella VRF che ne garantisce la sicurezza, prevenendo che vadano indebitamente fuori della VRF informazioni che devono rimanere all'interno, e che pacchetti esterni siano instradati a un router interno alla VPN.

Gli stessi meccanismi associati alle VRF consentono, anche, di poter ripetere indirizzi di utenti (tipicamente privati) all'interno di VPN differenti. Non è, infatti, necessario che un utente, per partecipare a due VPN, debba avere due indirizzi IP differenti, dal momento che sono esclusi a priori eventuali conflitti: anche se uno stesso indirizzo è presente in due tabelle d'instradamento VRF, essendo queste tabelle del tutto indipendenti tra loro, non si può verificare alcuna ambiguità.

Per la gestione degli spazi di indirizzamento sovrapposti è definita una nuova famiglia di indirizzi IP estesi (per mezzo del meccanismo del *route distinguisher*), mentre per la propagazione delle relative informazioni d'instradamento tra i vari router terminali PE sono utilizzate le estensioni del *Multi Protocol BGP (MP-BGP)* [28] (si veda il riquadro a pagina 54).

5.3 MPLS e la differenziazione dei servizi (Diffserv)

La realizzazione di VPN attraverso MPLS permette, già intrinsecamente, ai *service provider* che le offrono di poter differenziare, almeno parzialmente, i servizi offerti garantendo differenti livelli di QoS per le diverse classi di traffico presenti in rete. È infatti possibile realizzare meccanismi che consentano di distinguere la QoS all'interno delle singole VPN, separando, ad esempio, il traffico *VoIP (Voice-over-IP)* che dovrebbe ricevere un trattamento che assicuri un fissato ritardo massimo di trasmissione, da quello dell'*e-commerce* che dovrebbe, invece, ricevere una banda minima garantita (senza vincoli espressi sul ritardo di consegna).

La tecnica MPLS dovrebbe però essere in grado di gestire anche un modello architetturale del tipo *DiffServ* [29] per il controllo della QoS. Sono aperti a questo scopo alcuni *Internet Draft* che ne prescrivono le caratteristiche [30]. In questo caso i pacchetti entranti nella rete sarebbero raggruppati in classi, ciascuna della quali è contraddistinta da una determinata tipologia di servizio offerto. I pacchetti di traffico VoIP possono essere, ad esempio, inseriti nella classe a priorità più elevata, mentre quelli HTTP e-commerce in una classe "gold" e così via. Per differenziare ciascuna classe all'interno di ogni singolo router, ognuna di esse è associata a un determinato colore (una particolare sequenza di bit del campo MPLS Experimental dell'etichetta MPLS) il che consente di rendere il modello assai scalabile e garantisce che anche nel nucleo della rete vengano rispettati i vincoli sulla banda e il ritardo per il traffico trasmesso. L'associazione è realizzata quando un pacchetto entra nella rete ed è marcato in base ai criteri di classificazione applicati.

I router di frontiera possono anche eseguire il controllo sul traffico effettuando *shaping e/o policing*. Essi, ad esempio, effettuano la cancellazione dei pacchetti che eccedono la capacità concordata o eventuali operazioni di *re-marking*.

Ciascun nodo della dorsale applica poi differenti criteri di classificazione del traffico, sia per ciò che concerne la gestione delle code, sia per l'eliminazione di alcuni pacchetti, a seconda di come questi siano stati marcati.

Gli approcci utilizzati per poter marcare il traffico MPLS al fine di realizzare un modello DiffServ sono due. Il problema che si pone in questo caso è, infatti, quello di associare alla trama MPLS le informazioni relative alla classe di servizio cui appartiene un flusso di dati.

Una prima soluzione, definita *EXP Infrared-LSP* (E-LSP) [30], consiste nell'utilizzare un unico LSP per tutte le classi di servizio trasportate e "colorare" le varie trame MPLS utilizzando il campo EXP dell'intestazione MPLS. Questo meccanismo consente di poter definire otto differenti classi di servizio, mentre attraverso il campo *TOS (Type Of Service)* del-

l'intestazione di livello IP in realtà possono esserne definite sino a 64 classi, mediante l'impiego di otto bit. Questa discrepanza è motivata dal fatto che l'etichetta MPLS è stata definita prima della standardizzazione del campo TOS e la sua brevità si giustifica con l'intento di non appesantirla troppo.

È, però, opinione comune che otto classi di servizio potrebbero essere più che sufficienti in futuro per diversificare tutto il traffico presente in rete. Oggi in quasi tutte le realizzazioni del modello DiffServ sono definite una o, al più, due classi di servizio, oltre la tradizionale *best effort*.

In alternativa è stata definita dall'IETF una seconda procedura [30] per il trasporto dell'informazione *DiffServ* nella trama MPLS.

In questo caso l'etichetta associata a ciascun pacchetto MPLS contiene la parte di informazioni relativa al DiffServ marking, che specifica come trattare all'interno di una coda lo stesso pacchetto, mentre ogni singolo tunnel, fra gli stessi nodi di ingresso e di uscita, è associato a una sola classe di servizio.

Questo metodo prende il nome di *L-LSP (Label Inferred-LSP)*, in quanto le informazioni di QoS sono associate direttamente a ogni singolo LSP.

Per quanto concerne, invece, i criteri per gestire le code dei singoli router è stata studiata, anche se è ancora in fase di standardizzazione, un'estensione al protocollo RSVP-TE[31], necessaria per definire le procedure per la distribuzione di messaggi di segnalazione sulla base dei quali dovrebbero essere gestite le code.

Si vuole, così, definire un nuovo campo *(Object E-LSP)*, da trasportare nei messaggi RSVP-TE di path e di resv, che dovrebbe consentire di attivare all'interno dei router meccanismi di gestione preconfigurati, attraverso i quali i pacchetti entranti sono poi elaborati, controllando a livello globale la QoS offerta dalla rete.

5.4 Il ripristino veloce di MPLS (Fast Rerouting)

Un'importante applicazione di MPLS riguarda la possibilità di reagire a condizioni di guasto della rete, quali, ad esempio, fuori servizio di un collegamento, di un nodo, oppure di entrambe queste parti della rete, con tempi di ripristino molto bassi (dell'ordine di 50 ms), tipici dei meccanismi di protezione delle reti SDH.

La tecnica in questione, che prende il nome di *Fast Rerouting* [32] [33], è ottenuta attraverso meccanismi definiti all'interno dell'architettura MPLS-TE che consentono di mantenere per il traffico interessato adeguati livelli di qualità del servizio.

Più in particolare, grazie al *Fast Rerouting*, si può anche predisporre l'opzione di utilizzare i percorsi di protezione solo per il traffico con priorità più elevata, lasciando che il *best-effort* continui a essere gestito dai protocolli tradizionali di routing e garantendo l'immissione dei pacchetti in un *Fast Reroute Path* mediante l'impiego di una classificazione di tipo *Diffserv*.

Di seguito è fornita una breve descrizione di come è possibile realizzare tempi di ripristino assai bassi rispetto a quelli ottenibili in una rete IP classica, o in una in cui l'architettura MPLS è realizzata ma senza funzioni di *Traffic Engineering*.

In generale, in una rete IP il reinstradamento del traffico lungo un percorso che abbia presentato una o più condizioni di avaria si realizza dopo un tempo che dipende, sia dall'intervallo necessario al riconoscimento del guasto, sia dal tempo necessario per l'instaurazione e l'utilizzo del nuovo cammino.

In particolare, questa seconda componente dipende dai tempi necessari all'invalidazione delle rotte contenute nelle tabelle d'instradamento - che utilizzano lo specifico percorso - e al tempo necessario per il calcolo delle nuove tabelle che indirizzano i pacchetti verso eventuali percorsi differenti. Il processo di diffusione delle informazioni di guasto e di ricalcolo delle tabelle corrette per gli instradamenti, conseguenti al malfunzionamento, è distribuito ed è legato al protocollo utilizzato (*OSPF* tipicamente, ma anche *BGP*). Esso dipende, in generale, dalla complessità della rete, ma in ogni caso può assumere valori dell'ordine delle decine di secondi.

In una rete MPLS con funzionalità TE di *Fast Rerouting* i tempi necessari per l'instaurazione dei tunnel alternativi e per il ripristino da una condizione di guasto a una nuova possibile sono minimi, in quanto il percorso di riserva è pre-calcolato e pre-allocato direttamente, durante la fase di instaurazione dell'LSP primario. In aggiunta in questo caso sono utilizzati meccanismi di rilevazione dell'anomalia molto rapidi.

Per quanto riguarda le condizioni di guasto, sono stati definiti tre diversi meccanismi di protezione:

a) *Link Protection*, in grado di reagire a un malfunzionamento su di una connessione.
b) *Node Protection*, che consentono di rispondere a un malfunzionamento su di un nodo. Essi si differenziano dai meccanismi di Link Protection per due aspetti: il primo dipendente dal fatto che, in presenza della condizione di fuori servizio su di un router, gli LSP di protezione sono originati e terminati su LSR che sono fra loro distanti di due salti. Per la caduta di una connessione, invece, i router di origine e di destinazione del percorso alternativo sono distanti un solo salto. Nel caso di caduta di un router devono anche essere messi in atto ulteriori meccanismi di rilevazione del fuori servizio (ad esempio sviluppi del protocollo *hello*), diversamente dalla rilevazione del guasto di una connessione esistono meccanismi già a livello di strato di collegamento.
c) *Path Protection*, in grado di proteggere un intero percorso in seguito a un'anomalia che si presenti su di esso.

I motivi che giustificano i miglioramenti nelle prestazioni, introdotti dal *Fast Rerouting*, sono legati al fatto che il calcolo dei percorsi alternativi non è distribuito e, in secondo luogo, all'instradamento dei pacchetti che non è basato sulla destinazione finale. Nel caso di *Link Protection* se si verifica, ad esempio, la caduta di un collegamento fra due LSR, il percorso di protezione è determinato (e quindi controllato) solo dal router a monte del guasto, in contrasto con gli schemi tradizionali in cui l'instradamento lungo il cammino alternativo è gestito, anche, da tutti gli altri router sino alla destinazione *(hop by hop forwarding)*.

L'attivazione del percorso secondario è poi notificata mediante RSVP (o protocolli IGP) all'LSR ingress

Il MultiProtocol-BGP

Il MP-BGP (definito nella RFC 2283[37]) consente di annunciare univocamente le rotte IPv4 dei clienti, in un ambiente in cui gli indirizzi IP non sono unici, utilizzando un apposito formato denominato VPN-IPv4.

Come rappresentato in figura A, un indirizzo VPN-IPv4 è costituito da un campo *RD (Route Distinguisher)* di lunghezza pari a 64 bit e da un indirizzo IP tradizionale.

Il formato del RD comprende tre sottocampi:

- ***Type:* ha la lunghezza di due byte e determina la lunghezza degli altri due campi e la semantica del campo *Autonomous System*;**
- ***AS (Autonomous System):* contiene un numero identificativo dell'Authority preposta a fissare il valore dell'*Assigned Number* per una determinata applicazione;**
- ***Assigned Number:* è direttamente fissato dal VPN service provider (in genere uno per ogni VPN servita dal medesimo provider, anche se in realtà può essere pure utilizzato per identificare un particolare sito del cliente).**

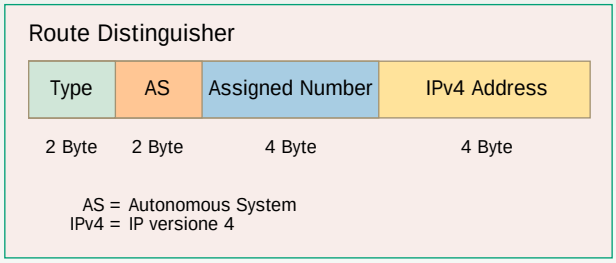


Figura A *Formato Route Distinguisher.*

Impiegando un formato come quello mostrato in figura per l'RD, si è certi dell'unicità totale degli indirizzi VPN-IPv4, in quanto il campo AS è unico rispetto all'insieme dei provider, mentre quello relativo all'*Assigned Number*, per definizione, è unico all'interno dell'ambito relativo al singolo provider.

dell'LSP primario che, a sua volta, avvia le nuove procedure per costituire un altro tunnel di protezione.

Le prestazioni attuali, presenti sui router Cisco, prevedono la gestione del solo meccanismo *MPLS-TE Fast Rerouting Link Protection*.

Nei laboratori di TILAB è invece ora in corso la valutazione sperimentale dei meccanismi di *Node* e di *Path Protection*.

6. L'introduzione dell'MPLS nelle reti IP di Telecom Italia

Dal punto di vista degli sviluppi industriali, può essere ricordato che la quasi totalità dei costruttori [34] *(Cisco, Nortel, Juniper, Unisphere, Lucent)* ha realizzato numerosi apparati in grado di fornire le funzioni richieste in una rete MPLS.

Per la presenza sul mercato di questi nuovi MPLS router e per le effettive potenzialità di una rete MPLS, numerosi operatori hanno introdotto nelle dorsali delle loro reti questa nuova architettura. Fra gli esempi più significativi al riguardo, possono essere citati *Teleglobe* che ha già inserito la rete MPLS nella sua dorsale, sfruttandone le funzioni disponibili per offrire ai propri clienti un servizio di VPN, classi di servizio differenziate e *TE*. *Equant* ha messo a punto una rete analoga a quella di Teleglobe, e poi ancora *Infonet*, *GlobalOne*, *At&T*, e *Global Crossing*. Molti altri gestori sono al momento nella fase di attivazione, come ad esempio Telefonica in Spagna e France Télécom in Francia, che in realtà già offre un servizio VPN IP basato su un nucleo di rete MPLS fornito da Global One.

Anche Telecom Italia si è mossa in questa direzione a partire dall'inizio del 2001, a valle dell'esperienza, realizzata nei laboratori di TILAB, che ha permesso di rilevare come la tecnologia MPLS fosse sufficientemente matura per essere introdotta su larga scala in servizio.

In particolare, a giugno dello scorso anno è stato completato il processo di introduzione di MPLS nella dorsale IP delle reti di Telecom Italia, che ha portato a rendere disponibili le funzioni del tipo P *(provider)* su tutti i router delle dorsali e quelle di tipo PE negli apparati terminali *(di frontiera)*. Inserire le funzionalità P-MPLS ha comportato la necessità di far diventare LSR tutti i router del nucleo della rete, in modo da abilitare ciascuno di essi a rilanciare il traffico IP attraverso la commutazione di etichette MPLS.

Con una struttura di questo tipo il *service provider* è in grado di assegnare il valore di RD in maniera autonoma, tenendo conto del fatto che questi valori hanno un significato solo locale, all'interno dell'area MPLS servita dal Provider, mentre non è nota la struttura del *VPN IP address* a livello BGP e di cliente.

Il valore di *route distinguisher* è configurato su ciascun router *PE (Provider Edge router)* per ciascuna VRF. È importante osservare che questa informazione è trasportata nei messaggi dei protocolli d'instradamento e non nell'intestazione IP utilizzata per inoltrare i pacchetti attraverso le etichette MPLS.

Per ciò che concerne la distribuzione delle informazioni di routing, il principio base, nel contesto dell'MPLS VPN IP, è che questa sia controllata mediante l'impiego di *VPN route target communities* rese disponibili mediante le *BGP extended communities*.

Esse sono realizzate nella maniera seguente:

- **le *route apprese* da un CE sono iniettate nel processo BGP e ad esse è associata una sequenza di *VPN route target extended community attributes*. Tipicamente la lista dei *route target attributes* è assegnata attraverso l'impiego di una *export list* associata alla VRF dalla quale la *route* è stata appresa;**
- **le *route importate* o *esportate* da una VRF sono gestite mediante l'impiego di *import/export list* che definisce i valori dei *route target community attributes*. Se la *import list* per una data VRF contiene ad esempio le route *target communities* A, B e C, ogni *VPN route*, caratterizzata da una delle precedenti *route target extended communities* (A, B, C) è importata nella VRF. L'insieme delle *route* importate in una VRF può essere definito attraverso l'uso di una mappa dell'instradamento *(route-map)* utilizzando gli standard BGP *communities* o altre informazioni portate dal BGP come criteri per la selezione delle route.**

Va precisato che, sebbene le *route distinguisher* e le *route target* abbiano lo stesso formato, le prime sono utilizzate per distinguere indirizzi VPN IPv4 non unici, mentre le seconde consentono di identificare in quale VRF debba essere "esportata" o "importata" una determinata *route*.

Una dorsale di rete IP con questa architettura permette a Telecom Italia di offrire ai propri clienti un servizio di VPN IP con qualità di servizio garantita all'interno della singola VPN.

Con le stesse predisposizioni, è stato anche possibile fornire servizi esterni, come accesso alla rete pubblica *(Internet)*, colloquio affidabile fra differenti VPN *(Extranet)*, accesso ad aree legate alla fornitura di servizi a valore aggiunto *(E-mail, Web caching, Web hosting, DNS, content delivery, database, …)*.

Telecom Italia è anche in grado oggi di offrire ai propri Clienti, grazie alla presenza di MPLS nella propria dorsale IP, un secondo tipo di servizi a qualità differenziata e con elevata affidabilità per il trasporto di differenti tipi di traffico *(real-time, Voice over IP, telefonico)*. In questo contesto sono certamente significative le funzionalità di *TE* e *Fast Rerouting* e gli sviluppi in corso di definizione in ambito IETF e ora in fase di sperimentazione nei laboratori di TILAB.

7. Conclusioni

Come si è sottolineato nell'articolo, l'architettura del MultiProtocol Label Switching era nata come risposta a un problema che agli inizi degli anni Novanta si presentava quanto mai critico ed era legato ai limiti delle prestazioni dei tradizionali router IP che non riuscivano più a sostenere e, quindi, a gestire, il ritmo di sviluppo della rete Internet.

Di fatto, però, questo problema nell'ultimo quinquennio è stato superato grazie agli sviluppi della tecnologia, in particolare della microelettronica attraverso le ASIC dedicate, che permettono di eseguire l'instradamento dei pacchetti con elevate velocità (10 Gbit/s). Questi progressi nei sistemi riguardano oggi i *Gigarouter*, presenti nelle reti degli ISP e potrebbero portare alla realizzazione di *Terarouter* in grado di risolvere per molto tempo eventuali problemi legati alle prestazioni degli apparati di *internetworking*.

Il processo di definizione e di standardizzazione di MPLS è stato piuttosto laborioso e travagliato e per un certo periodo è stata messa anche in dubbio l'effettiva utilità dell'architettura.

Finora, tuttavia, secondo l'esperienza di TILAB, MPLS rappresenta una tecnologia sufficientemente matura per essere utilizzata in campo. Questa opinione è confermata dalla scelta fatta da numerosi ISP che già la impiegano per migliorare la gestione del piano di

controllo e per aumentare il numero di tipi di servizi affidabili di una rete IP multiservizio.

A livello sistemistico e architetturale l'elemento fondamentale che consente di raggiungere questi miglioramenti è rappresentato dalla separazione del piano di controllo della rete tradizionale, da quello di inoltro gestito attraverso il paradigma del *label switching*.

Da un lato questa scelta permette di offrire nuovi servizi agli utenti, le cui prime, ma non ultime, predisposizioni in esercizio sono rappresentate dalle VPN IP e dalle funzioni di *Traffic Engineering* (con informazioni di banda, colori sui *link*, *Fast Rerouting*). La scelta rientra nell'ottica di un progressivo aumento della complessità legata all'architettura della classica rete IP senza garanzie sulla qualità offerta *(best effort)* che dovrebbe tendere verso un nuovo modello di rete, molto più funzionale e in grado di offrire un'ampia varietà di tipologie di servizio sempre più controllabili e affidabili in risposta alle esigenze degli utenti continuamente in crescita.

Il pedaggio da pagare per ottenere questi risultati consiste nella realizzazione di una rete decisamente più articolata a livello architetturale e, quindi, maggiormente complessa da gestire e da amministrare.

Abbreviazioni

AAL	ATM Adaptation Layer
AS	Assigned Number
ATM	Asynchronous Transfer Mode
ARIS:	Aggregate Route-Based Switching
BB	BackBone
BGPv4	Border Gateway Protocol version 4
CE	Customer Edge
CoS	Class of Service
CR-LDP	Constraint-based Routing - LDP
E-LSP	Exp. inferred Label Switched Path
E-LSR	Edge-Label Switch Router
ER	Explicit Route
ERO	Explicit Route Object
FEC	Forwarding Equivalence Class
FT	Forwarding Table
GMPLS	Generalized MultiProtocol Label Switching
IETF	Internet Engineering Task Force
IGP	Interior Gateway Protocol
IGRP	Interior Gateway Routing Protocol
IP	Internet Protocol
IPLPDN	IP over Large Public Data Network
ISP	Internet Service Provider
LDP	Label Distribution Protocol
LIB	Label Information Base
LIFO	Last In First Out
LMP	Link Management Protocol
LSP	Label Switched Path
L-LSP	Label inferred Label Switched Path
LSR	Label Switch Router
MAC	Medium Access Control
MP-BGP	MultiProtocol BGP
MPLS	MultiProtocol Label Switching
MPλS	MultiProtocol Lambda Switching
MPOA	MultiProtocol Over ATM
NHLFE	Next Hop Label Forwarding Entry
NHRP	Next Hop Resolution Protocol
OMP	Optimized MultiPath
OSPF	Open Shortest Path First
P	Provider (Router)
PE	Provider Edge (Router)
PHB	Per Hop Behaviour
PPP	Point to Point Protocol
PVC	Permanent Virtual Circuit
QoS	Quality of Service
RD	Route Distinguisher
RFC	Request For Comment
ROLC	Routing Over Large Clouds
RSVP	Resourse reSerVation Protocol
TCP	Transmission Control Protocol
TE	Traffic Engineering
TOS	Type Of Service
UDP	User Datagram Protocol
VPI/VCI	Virtual Path Identifier / Virtual Circuit Identifier
VPN	Virtual Private Network
VRF	Virtual Routing Forwarding
WDM	Wavelength Division Multiplexing

Bibliografia

[1] Laubach, M.; Halpern, J.: *Classical IP and ARP Over ATM*. RFC 2225, aprile 1998.

[2] *LAN Emulation over ATM*. ATM Forum, Version 1.0 ftp://ftp.atmforum.com/pub/specs/af-lane-0021.000.ps, gennaio 1995.

[3] Benham, D.; Swallow, G.: *Multiprotocol Over ATM: Specification Completed*. MPOA Working Group, The ATM Forum, settembre 1997.

[4] Lucani, J.; Katz, D.; Piscitello, D.; Cole, B.; Doraswamy, N.: *NBMA Next Hop Resolution Protocol*. RFC 2332, aprile 1998.

[5] Nagami, K. et al.: *Toshiba's Router Architecture Extensions for ATM: Overview*. RFC 2332, aprile 1998.

[6] Lin, S.; McKeown, N.: *A simulation Study of IP switching*. Proceedings of ACM SIGCOMM 97. Cannes, Francia, settembre 1997.

[7] Rekhter, Y.; Davie, B.; Katz, D.; Rosen, E.; Swallow, G: *Cisco Systems Tag Switching Architecture Overview*. RFC, 2105, febbraio 1997.

[8] Feldman, N.; Viswanathan, A.: *ARIS Protocol Specification*. IBM Technical Report TR 29.2368, marzo 1998.

[9] Davie, B.; Rekhter, Y.: *MPLS Technology and Applications*. Morgan Kaufman Ed., aprile 2000.

[10] Comer, D.: *Internetworking with TCP/IP: Principles, Protocols, Architecture*. Prentice Hall, 1995.

[11] Rosen, E.; Viswanathan, A.; Callon, R.: *Multiprotocol Label Switching Architecture*. RFC 3031, gennaio 2001.

[12] Rekhter, Y.; Rosen, E.: *Carrying Label Information in BGP-4*. RFC 3107, maggio 2001.

[13] Andersson, L.; Doolan, P.; Feldman, N.; Fredette, A.; Thomas, B.: *LDP Specification*. RFC 3036, gennaio 2001.

[14] Thomas, B.; Gray, E.: *LDP Applicability.* RFC 3037, gennaio 2001.

[15] *http://www.ietf.org/html.charters/mpls-charter.html*

[16] *http://www.mplsforum.org*

[17] Awduche, D. et al.: *MultiProtocol Lambda Switching: Combining MPLS Traffic Engineering Control With Optical Crossconnects.* Internet draft-awduche-mpls-te-optical-03.txt, aprile 2001.

[18] Lang, P.; Mitra, K. et al.: *Link Management Protocol.* Internet draft-ietf-ccamp-lmp-02.txt, maggio 2001.

[19] Kompella, K.; Rekhter, Y. et al.: *OSPF Extensions in Support of Generalized MPLS.* Internet draft-ietf-ccamp-ospf-gmpls-extensions-00.txt, maggio 2001.

[20] Kompella, K.; Rekhter, Y. et al.: *IS-IS Extensions in Support of Generalized MPLS.* Internet draft-ietf-isis-gmpls-extensions-04.txt, marzo 2001.

[21] Ashwood-Smith, P. et al.: *Generalized MPLS - Signaling Functional Description.* Internet draft-ietf-mpls-generalized-signaling-06.txt, ottobre 2001.

[22] Ashwood-Smith, P. et al.: *Generalized Multi-Protocol Label Switching (GMPLS) Architecture.* Internet draft-ietf-ccamp-gmpls-architecture-00.txt, giugno 2001.

[23] Moy, J.: *OSPF Version 2.* RFC 1583, marzo 1994.

[24] Awduche, D.; Hannan, A.; Xiao, X.: *Applicability Statement for Extensions to RSVP for LSP-Tunnels00.* Internet draft-ietf-mpls-rsvp-tunnel-applicability-02.txt, aprile 2001.

[25] Giacalone, S.: *Network Engineering Extensions (NEXT) for OSPFv3.* Internet draft-giacalone-te-optical-next-02.txt, marzo 2001.

[26] Ash, J.; Lee, Y. et al.: *LSP Modification Using CR-LDP.* Internet draft-ietf-mpls-crlsp-modify-03.txt, marzo 2001.

[27] Mankin, A.; Baker, F.; Braden, B.; Bradner, S.; O'Dell, M.; Romanow, A.; Weinrib, A.; Zhang, L.: *Resource ReSerVation Protocol (RSVP) Version1 Applicability Statement Some Guidelines on Deployment.* RFC 2208, settembre 1997.

[28] Rosen, C. et al.: *BGP/MPLS VPNs.* Internet draft-ietf-ppvpn-rfc2547bis-00.txt, luglio 2001.

[29] Nichols, K.; Karpenter, B.: *Definition of Differentiated Services Per Domain Behavior and Rules for their Specification.* RFC 3086, aprile 2001.

[30] Le Faucheur, F.; Wu, L. et. al.: *MPLS Support of Differentiated Services.* Internet draft-ietf-mpls-diff-ext-09.txt, aprile 2001.

[31] Ganti, S.; Seddigh, N.; Nandy, B.: *MPLS Support of Differentiated Services using E-LSP.* Internet draft-ganti-mpls-diffserv-elsp-00.txt, aprile 2001.

[32] Gan, D.; Pan, P.; Ayyangar, A.; Kompella, K.: *A Method for MPLS LSP Fast-Reroute Using RSVP Detours.* Internet draft-gan-fast-reroute-00.txt, aprile 2001.

[33] Atlas, A.; Villamizar, C.; Litvany, C.: *MPLS RSVP-TE Interoperability for Local Protection/Fast Reroute.* Internet draft-atlas-rsvp-local-protect-interop-01.txt, luglio 2001.

[34] *http://www.mplsrc.com/vendor.shtml*

[35] Ohba, Y.; Katsube, Y.; Rosen, E.; Doolan, P.: *MPLS Loop Prevention Mechanism.* RFC 3063, febbraio 2001.

[36] *http://www.mplsrc.com/standards.shtml*

[37] Bates, T.; Chandra, R.; Katz, D.; Rekhter, Y.: *Multiprotocol Extensions for BGP-4.* RFC 2283, febbraio 1998.

Federico M. Renon si è laureato in Ingegneria Elettronica a Pavia nel 1986. Dopo una breve esperienza di lavoro in Aeritalia (ora Alenia Spazio) come responsabile dello sviluppo di sistemi di controllo per il satellite SAX, è passato allo CSELT (oggi TILAB) dove opera dal 1990. Ha lavorato nel campo delle reti dati ad alte prestazioni (reti metro DQDB, geografiche ATM e IP), nel contesto di normativa internazionale IEEE, ETSI e ITU; e in ambito di sviluppo e realizzazione nazionale ed internazionale (rete pilota ATM europea ed italiana, reti ATMosfera e Interbusiness). È attualmente responsabile dell'Area di Competenza "Value Added Networking", dove sta indirizzando le attività di ricerca su soluzioni innovative di rete e servizio in ambito wireline e wireless, quali MPLS e GMPLS, VPN IP, autenticazione e profilatura utente, integrazione voce/video/dati su IP, Content Networking.

Gianni Rossi si è laureato in Ingegneria Elettronica presso l'Università degli studi di Genova nel 1993 e ha conseguito il Master in Telecomunicazioni "Politecnico di Torino - Scuola Superiore G. Reiss Romoli" nel 1996. Dopo alcune esperienze lavorative presso il D.I.S.T. di Genova e la Marconi di Genova, dal 1995 opera allo CSELT (oggi TILAB). Attualmente è responsabile tecnico delle attività di TILAB per lo sviluppo della rete e dei servizi del Backbone IP di Telecom Italia. Ha lavorato nel campo dell'analisi sistemistica e della sperimentazione di piattaforme di networking IP (Gigabit Router, Multi-Layer Switch, Content Networking) e nella progettazione e nel collaudo di soluzioni di networking IP per reti geografiche e corporate (con esperienze sulle architetture MPLS VPN, MPLS-TE, sull'introduzione di servizi con qualità differenziata per il trasporto di applicazioni voce su IP, sull'ottimizzazione del routing BGP e OSPF e sulle tecniche di integrazione IP–ATM). Ha partecipato ad attività di normativa internazionale IETF; è autore di pubblicazioni e ha contribuito alla realizzazione di eventi e conferenze scientifiche.

Paolo Salamandra si è laureato in Ingegneria Elettronica, con la specializzazione in Telecomunicazioni, presso l'Università degli Studi di Perugia nel gennaio del 2001 con una tesi nell'ambito della qualità del servizio in reti IP. Nello stesso anno è stato assunto in TILAB dove si è occupato della qualificazione e il testing di apparati di accesso ADSL. Dall'inizio del 2002 partecipa alle attività per lo sviluppo della rete e dei servizi del Backbone IP di Telecom Italia interessandosi, in particolare, del collaudo dell'architettura MPLS-TE e dell'introduzione di servizi con qualità differenziata per il trasporto di applicazioni voce su IP.

Come velocizzare le applicazioni IP su GPRS

I nuovi sistemi di accelerazione

GIORGIO BRUNO
FABIO MAZZOLI
ALDO VANNELLI

Con il diffondersi dell'uso del GPRS per l'accesso dati in mobilità e in mancanza di applicazioni IP specifiche nate per tale contesto, la maggior parte degli utenti ha semplicemente continuato a usare le applicazioni IP standard esistenti.
Queste applicazioni, nate in un contesto di reti cablate, sono state pensate e ottimizzate in base alle caratteristiche di queste reti.
Per permettere alle stesse applicazioni di operare efficientemente anche sulla rete GPRS, che ha caratteristiche sostanzialmente diverse dalle reti cablate, è stata messa a punto una serie di prodotti, chiamati generalmente acceleratori o compressori, che si occupano di riadattare i protocolli IP in modo da rendere massima l'efficienza sulle reti wireless.
Grazie all'utilizzo degli acceleratori, le prestazioni delle applicazioni IP su GPRS diventano del tutto confrontabili con quelle ottenibili sulla rete fissa.
In questo articolo sono riassunti i fondamenti teorici su cui si basano gli acceleratori, le architetture oggi proposte dai costruttori e i risultati che da essi possono essere ottenuti quando siano introdotti in rete.

1. Introduzione

Con l'introduzione da parte dei maggiori operatori del settore mobile di un servizio *GPRS (General Packet Radio System)* stabile, efficiente e diffuso a livello nazionale, e con la crescente disponibilità di terminali GPRS di costo medio e che possono facilmente essere connessi ai computer portatili, una rilevante percentuale di utenti mobili cominciano a impiegare il GPRS per utilizzare da remoto le tipiche applicazioni *IP (Internet Protocol)*: posta elettronica e navigazione su Web.

Le applicazioni IP oggi disponibili, nate tutte nel contesto delle reti cablate, non offrono il massimo delle loro potenzialità quando utilizzate sulla rete GPRS, a causa delle diversità intrinseche della rete mobile rispetto a quella fissa.

Per migliorare ulteriormente le prestazioni ottenibili con le applicazioni IP su GPRS, è stata approntata una serie di prodotti ad-hoc, chiamati *acceleratori* o *compressori*, che riadattano le caratteristiche dei protocolli IP in modo da rendere massima l'efficienza sulla rete GPRS.

I sistemi di accelerazione oggi disponibili sul mercato svolgono tre tipi di funzione:

- ottimizzazione del protocollo di livello 4 (trasporto), il *TCP (Transmission Control Protocol)*;
- ottimizzazione dei protocolli di livello 7 (navigazione su Web , posta elettronica);
- compressione dei dati: testo *HTML (HyperText Mark-up Language)*; immagini *JPEG* (*Joint Photographic Experts Group)* e *GIF* (*Grafic Interchange Format*).

CHE COSA SONO GLI ACCELERATORI GPRS

Gli acceleratori GPRS sono soluzioni hardware e software che operano a livello IP, e che provvedono a modificare i protocolli sia di livello applicativo sia di livello trasmissivo per ottimizzarne le prestazioni sulla rete GPRS.

I prodotti più evoluti oggi disponibili sul mercato operano secondo tre gruppi di tecniche distinte:

- ottimizzazione del protocollo di livello 4 (trasporto TCP);
- ottimizzazione dei protocolli di livello 7 (navigazione su Web, posta elettronica);
- compressione dei dati (testo HTML, immagini JPEG e GIF).

Da un punto di vista architetturale, sono disponibili due varianti di acceleratori: *server-only* e *client-server*, che offrono un diverso bilanciamento tra prestazioni e semplicità di installazione.

Le soluzioni *server-only* operano entro i limiti imposti dai protocolli standard, e hanno un livello di prestazioni più basso.

Le soluzioni *client-server* sostituiscono i protocolli standard con quelli proprietari studiati e ottimizzati ad-hoc per funzionare in maniera ottimale su reti GPRS.

In entrambi i casi, l'obiettivo degli acceleratori è quello di migliorare le prestazioni delle applicazioni IP in ambito GPRS.

Nel seguito dell'articolo saranno descritti i principi di funzionamento di tutti i protocolli menzionati e le tecniche utilizzate dagli acceleratori per migliorarne l'efficienza su GPRS.

2. Il protocollo di trasporto TCP

2.1 Caratteristiche generali

Il TCP è un protocollo di livello 4 *(trasporto)* che possiede i seguenti attributi:

- *connesso:* prima di poter iniziare lo scambio di dati, trasmettitore e ricevitore devono esplicitamente instaurare una connessione. Questa viene stabilita attraverso il meccanismo del *3-way handshake*[1], che impiega un *RTT (Round Trip Time)* per essere eseguito;
- *punto-punto:* una sessione TCP coinvolge sempre solo due entità;
- *affidabile:* il TCP garantisce con una probabilità estremamente elevata che i dati consegnati al livello superiore della pila protocollare siano integri e completi, facendosi carico di correggere eventuali errori di trasmissione. A questo scopo raggruppa i dati in blocchi, chiamati segmenti, e per ogni segmento spedito attende di ricevere una conferma esplicita, denominata *ACK (ACKnowledge)*;
- *full duplex:* sulla stessa connessione TCP i dati possono essere inviati contemporaneamente in entrambe le direzioni;
- *multiplexing:* possono presentarsi più connessioni TCP indipendenti l'una dall'altra tra la stessa coppia di entità. Le connessioni sono identificate usando le porte logiche. Il protocollo HTTP utilizza, ad esempio, il numero di porta 80.

2.2 Controllo di flusso: i meccanismi a finestre

Come si è visto, il TCP garantisce l'affidabilità, attraverso la conferma esplicita (ACK), della ricezione di ogni singolo segmento.

Per sfruttare efficacemente la banda disponibile, nel TCP è stato introdotto un meccanismo a finestre che funziona secondo il seguente principio:

- sono mandati in rete un certo numero di segmenti senza attendere l'ACK. Il numero è definito dalla dimensione della finestra;
- è tenuta traccia del primo segmento per il quale non è stato ancora ricevuto il riscontro;
- nel momento in cui si riceve l'ACK del primo segmento non riscontrato, la finestra è spostata in avanti di un segmento, e un ulteriore nuovo segmento può essere spedito.

La quantità di dati in volo è variata istantaneamente dal TCP mediante l'uso di due variabili di stato, che agiscono sul funzionamento del suo meccanismo a finestre: la *CWND (Congestion WiNDow)* e la *RWND (Receiver WiNDow)*.

La *CWND*, mantenuta all'interno dal trasmettitore, indica il numero massimo di byte in volo prima di dovere attendere un ACK dal ricevitore. La CWND consente al trasmettitore di variare la velocità di invio dei blocchi per adattarla alla capacità trasmissiva della rete, ed è quindi usata per realizzare un meccanismo di controllo della congestione *(Congestion Control)*.

La *RWND* è mandata dal ricevitore insieme a ogni ACK, e indica quanti byte esso è ancora in grado di accettare.

Essa consente al ricevitore di far diminuire la velocità di invio dei dati al trasmettitore, in modo da adattarla alle capacità ricettive che esso presenta. Viene usata per realizzare un meccanismo di controllo di flusso (*Flow Control)*.

In ogni istante, il trasmettitore mantiene in volo

(1) *Il 3-way handshake è il meccanismo che viene usato per stabilire una connessione TCP tra due entità: il trasmettitore invia un segmento con il bit SYN impostato a 1; il ricevitore risponde con un segmento con i bit SYN e ACK impostati a 1, e, infine, il trasmettitore conferma mandando un ulteriore segmento con il bit ACK impostato a 1 (i bit SYN e ACK sono entrambi contenuti nell'intestazione dei segmenti TCP). Solo da questo momento la connessione è stabilita e può essere impiegata per trasmettervi dati in modo affidabile.*

al massimo un numero di byte pari al valore minore tra CWND e RWND.

2.3 Controllo di flusso: le politiche di slow start e di congestion avoidance

Il TCP gestisce la dimensione istantanea della *congestion window* attraverso due criteri differenti, utilizzati in momenti differenti della sessione TCP: l'avvio lento *(slow start)* e l'aggiramento della congestione *(congestion avoidance)*.

Lo *slow start* è usato all'inizio della sessione TCP e serve ad aumentare progressivamente la portata *(throughput)* del trasmettitore fino a saturare la capacità del canale.

Il funzionamento può essere suddiviso in tre fasi successive:

- la dimensione della CWND viene inizializzata con un valore pari a un segmento;
- sono inviati un numero di segmenti non riscontrati pari alla dimensione della CWND;
- per ogni ACK ricevuto, la CWND è aumentata di un'unità.

Contrariamente a quanto suggerito dal nome, lo *slow start* (avvio lento) genera una portata *(throughput)* con crescita assai rapida (di tipo esponenziale) in funzione del tempo. Il termine "slow" si riferisce piuttosto al fatto che la CNWD parte sempre dal valore minimo di un segmento.

A causa dell'alto valore dell'RTT tipico del GPRS - che ritarda in misura rilevante gli ACKnowledge di ritorno - si ha un marcato sottoutilizzo della banda nella fase iniziale dello *slow start*.

Quando la dimensione della *congestion window* CWND supera una certa soglia, chiamata convenzionalmente *SSTHRESH (Slow Start THRESHold)*, lo *slow start* è sostituito dal regime di *congestion avoidance*, che aumenta progressivamente la finestra di congestione in modo quasi lineare.

Il regime di *congestion avoidance* è utilizzato dal trasmettitore quando reputa che il suo throughput sia prossimo alla capacità massima della rete.

Il valore di SSTHRESH, mantenuto all'interno del trasmettitore, è inizializzato quando è instaurata la connessione TCP e generalmente è posto a un valore pari alla finestra di ricezione RWND del ricevitore.

Nella figura 1 è riportato l'andamento tipico combinato di *slow start* e di *congestion avoidance*, nel caso in cui il valore di SSTHRESH sia inferiore alla RWND iniziale del ricevitore e in assenza di errori.

2.4 Rilevamento e recupero degli errori: il timeout

Il TCP utilizza un meccanismo basato sul *timeout* per rilevare la presenza di errori in rete: ad ogni segmento spedito è associato un timer, che è bloccato alla ricezione dell'*ACKnowledge* ad esso relativo.

Nel caso in cui l'ACK non arrivi entro un tempo massimo prefissato, chiamato *RTO (Retransmit TimeOut)*, il trasmettitore considera perso il segmento.

Poiché il TCP è nato per reti cablate, caratterizzate da un tasso di errore basso, esso interpreta sempre gli errori come sintomo di congestione della rete, e abbassa, perciò, la velocità di trasmissione dopo che si è verificato un errore.

Per il recupero degli errori si seguono le seguenti fasi:

- è ritrasmesso il segmento andato perso;
- si pone SSTHRESH uguale a CWND/2;
- è forzato il regime di *slow start* ponendo CWND = 1.

Nel caso di una rete GPRS, il recupero dagli errori risulta naturalmente oneroso, in quanto comporta l'instaurarsi di una nuova fase di *slow start*.

Il calcolo dell'*RTO* (*Retransmit TimeOut)* è eseguito basandosi sul valore dell'RTT medio corrente, che è aggiornato per ogni ACK ricevuto correttamente.

I dettagli del meccanismo dipendono dal sistema operativo. Una delle politiche utilizzate è la seguente:

- per ogni ACK ricevuto, è misurato il valore dell'RTT istantaneo, pari al tempo intercorso tra la spedizione del segmento e la ricezione del corrispondente ACK;
- l'RTT medio corrente è ricalcolato come media pesata tra il precedente RTT medio corrente e l'RTT istantaneo (generalmente è attribuito un maggior peso all'RTT istantaneo);
- l'RTO è posto pari a β*RTT (dove β è un margine, tipicamente posto pari a 2).

Sulle reti GPRS il calcolo del timeout può portare a un falso rilevamento di errori, a causa delle forti variazioni di throughput istantaneo (che si riflettono a loro volta in forti variazioni del valore dell'RTT istantaneo) dovute alla condivisione della banda radio tra tutti gli utenti che operano in un contesto GPRS attivo all'interno della stessa cella radio.

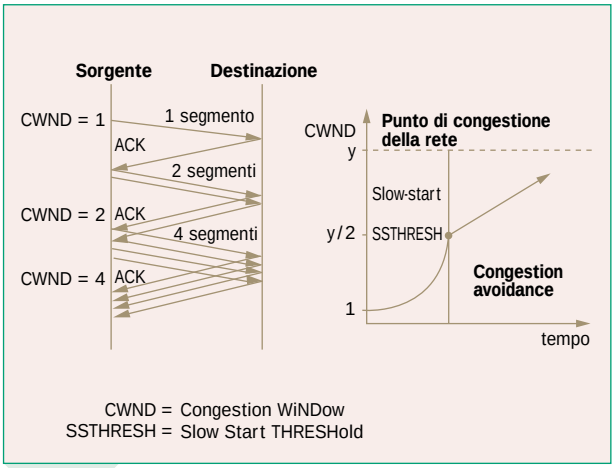


Figura 1 *Schema di funzionamento dello slow start e del congestion avoidance.*

2.5 Inefficienze del TCP su GPRS

Alla luce di tutti i meccanismi del TCP esaminati nei paragrafi precedenti, possono essere indicate le principali cause di riduzione dell'efficienza

ANDAMENTO DI UNA SESSIONE TIPICA DEL TCP

Nella figura A è riportato l'andamento di una tipica sessione TCP in cui si verifichi un errore non recuperabile.

L'asse delle ascisse rappresenta il tempo, mentre quello delle ordinate è relativo alla dimensione istantanea della *congestion window* (in segmenti), che equivale al numero di segmenti in volo.

Sull'asse delle ordinate è anche riportato il valore istantaneo di SSTHRESH: essa è una variabile interna che il TCP impiega per scegliere se aumentare la *congestion window* secondo la politica di *slow start* oppure di *congestion control*.

Il valore di SSTHRESH è aggiornato ogni volta che si verifica un errore.
Nel grafico possono essere distinte cinque fasi:

1) lo slow start iniziale, con crescita esponenziale della CWND;

2) il passaggio da *slow start* a *congestion avoidance* nel superamento della soglia SSTHRESH, punto in cui la crescita della CWND diventa quasi lineare;

3) il crollo del *throughput* in corrispondenza di un errore irrecuperabile *(time-out)*;

4) una nuova fase di *slow start* e la reimpostazione della SSTHRESH con un valore pari a CWND/2, attuate per il recupero dell'errore;

5) una nuova fase di *congestion avoidance*, che questa volta è raggiunta in tempi più brevi, poiché la soglia SSTHRESH risulta inferiore a quella iniziale.

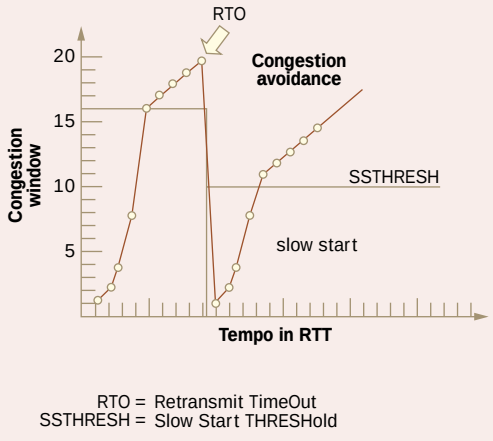


Figura A *Andamento tipico di una sessione TCP con errori.*

del protocollo quando esso è impiegato nella rete GPRS.

Le caratteristiche della rete GPRS, che risultano di difficile gestione da parte del TCP, sono i *ritardi elevati* e la variabilità di *throughput istantaneo*.

Gli effetti che tali caratteristiche di rete generano sono i seguenti:

- lo *slow start*, a causa dell'alto valore dell'RTT medio, impiega diversi secondi prima di saturare il throughput della rete;
- le sensibili fluttuazioni istantanee sia dell'RTT sia del throughput sono scambiate dal TCP per congestione sulla rete, e conseguentemente sono attivati i meccanismi di *congestion control* ed *error recovery*, che limitano l'efficienza nell'impiego della banda del canale;
- la mancanza di una stima esplicita della capacità del canale fa sì che in alcuni casi si arrivi all'*overflow* dei *buffer* di rete, e che si abbia perciò la perdita di segmenti e l'attivazione delle procedure di *error recovery*.

Nei due paragrafi seguenti sono fornite alcune indicazioni di come possono essere migliorate le prestazioni del TCP su una connessione GPRS.

Le tecniche qui descritte sono quelle correntemente impiegate sugli acceleratori per GPRS.

2.6 Tecniche di ottimizzazione TCP

a) cambio del numero iniziale di segmenti per slow start

Per giungere rapidamente a saturare la capacità del canale è possibile modificare il numero di segmenti trasmessi inizialmente dallo *slow start*.

La strategia seguita è abbastanza semplice: poiché il TCP non può ricevere nessun ACK prima che sia passato almeno un RTT, e poiché quindi in questo intervallo di tempo non può essere stimato il throughput effettivo della rete, occorre immettere in rete almeno un numero di byte pari alla quantità che essa è in grado di smaltire entro un RTT.

Il numero di byte che occorre inviare inizialmente in rete, senza attendere conferme, è quindi

STRUTTURA DI UN URL HTTP

Un *URL (Uniform Resource Locator)* è un formato standard che permette di identificare le risorse presenti sulla rete.

Un *URL* contiene le seguenti informazioni:

- **il protocollo con cui si accede alla risorsa;**
- **il nome del server su cui risiede la risorsa;**
- **il nome della risorsa, comprensivo di un eventuale percorso relativo alla radice del *filesystem* del Web server.**

Nel caso specifico dell'HTTP, le risorse sono tipicamente dei *file* presenti sul filesystem del Web server.

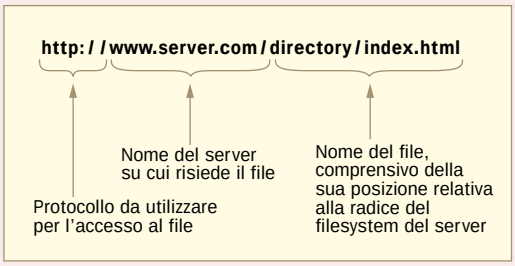


Figura A *Struttura di un Uniform Resource Locator HTTP.*

Nella figura A è riportata la struttura di un *URL HTTP.*

pari almeno al prodotto throughput per RTT.

Nel caso della rete GPRS, può essere assunto con buona approssimazione RTT pari a 1 secondo e throughput pari a 44 kbit/s, e quindi possono essere inviati inizialmente 44 kbit, ovvero circa 5 kbyte di dati.

Poiché lo *slow start* opera in termini di segmenti, e non di byte, occorre, naturalmente, convertire il numero di byte in numero di segmenti.

b) stima esplicita del throughput istantaneo

Come si è visto nei paragrafi precedenti, il meccanismo standard di controllo di congestione del TCP non stima il throughput istantaneo del canale, ma incrementa semplicemente la velocità di trasmissione finché non rileva la perdita di segmenti, e quindi finché non si ha un *overflow di buffer*[2] in rete.

Purtroppo l'*overflow dei buffer* genera la perdita (e quindi la necessità di ritrasmissione) di gruppi interi di segmenti TCP, prima che il meccanismo di recupero degli errori abbia il tempo di intervenire. Si ha quindi un andamento del throughput "a dente di sega" (vedi riquadro a pagina 61).

Senza rompere la semantica del TCP, è invece possibile utilizzare un meccanismo di stima esplicita del throughput istantaneo della rete che operi tramite un algoritmo come quello qui descritto:

- si inviano i segmenti con un certa cadenza, e si misura l'RTT di ogni segmento tramite gli ACK;
- se l'RTT aumenta (segno che il livello dei buffer aumenta, perché si sta trasmettendo sopra capacità), si diminuisce la velocità di trasmissione;
- se invece il valore dell'RTT diminuisce, si aumenta la velocità di trasmissione.

Esistono diverse realizzazioni di questo algoritmo, dipendenti dalla tecnologia e dal sistema operativo.

3. Il protocollo per la navigazione su Web/ HTTP

3.1 Descrizione generale

L'*HTTP (Hypertext Transport Protocol)* è il protocollo standard usato per la navigazione sul Web.

Esso è un protocollo *stateless*, privo quindi del concetto di "sessione", ed è stato definito con l'obiettivo, tra gli altri, di essere inseribile con facilità su macchine con limitate capacità di elaborazione.

L'HTTP ha un funzionamento del tipo richiesta-risposta: il *client* (chiamato anche *Web browser*, o semplicemente *browser*) apre una connessione TCP e chiede al server *(GET request)* un singolo oggetto (tipicamente un *file*). Il server invia l'oggetto richiesto *(GET response)* e chiude la sessione TCP.

Ogni oggetto presente sulla rete è identificato univocamente da un *URL (Uniform Resource Locator)*.

Il protocollo HTTP serve quindi unicamente per trasferire singoli oggetti, identificati da un URL, tra un client e un server.

La struttura di un URL HTTP è indicata nel riquadro sopra riportato.

3.2 Il formato di descrizione della pagina: HTML

Per permettere di visualizzare i contenuti formattati graficamente sul browser dell'utente, è adottato un formato di file denominato *HTML (Hypertext Mark-up Language)*, trasportato tramite il protocollo HTTP sopra descritto.

Un HTML è un *file* di testo, codificato con semplici caratteri ASCII, che contiene la descri-

(2) I buffer che vanno in overflow sono quelli posti ai confini tra il tratto di rete veloce (cablata) ed il tratto di rete lenta (radio).

ESEMPI DI HTTP GET REQUEST E GET RESPONSE

Nelle figure A e B sono riportati due esempi di GET HTTP: la *GET request* mandata da un browser per richiedere una certa pagina HTML e la relativa *GET response* restituita dal Web server.

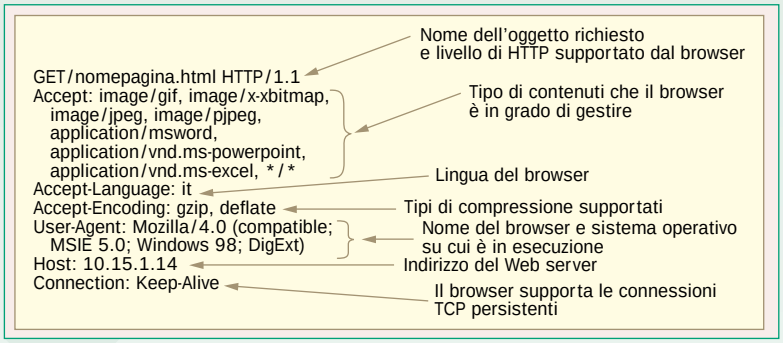


Figura A *Esempio di GET request inviata dal browser.*

Può essere anzitutto osservata la grande dimensione di entrambe le intestazioni (dell'ordine delle centinaia di byte): essa è dovuta sia alla grande quantità di informazioni contenute in entrambe le intestazioni, sia al fatto che le informazioni sono codificate tramite stringhe di caratteri mnemoniche facilmente interpretabili dall'uomo, ma estremamente ridondanti.

Nel caso in cui l'oggetto richiesto sia piccolo, l'*overhead* introdotto dalle intestazioni risulta paragonabile alle dimensioni dell'oggetto stesso, e genera quindi uno spreco di banda non trascurabile.

Si noti come tutte le richieste GET inviate da un browser verso lo stesso server differiscano tra loro per un unico campo: il nome dell'oggetto richiesto presente nella prima riga.

La conseguenza è che gran parte dell'informazione trasportata da tutte le *GET request* successive alla prima oltre che essere codificata in maniera inefficiente è anche ridondante.

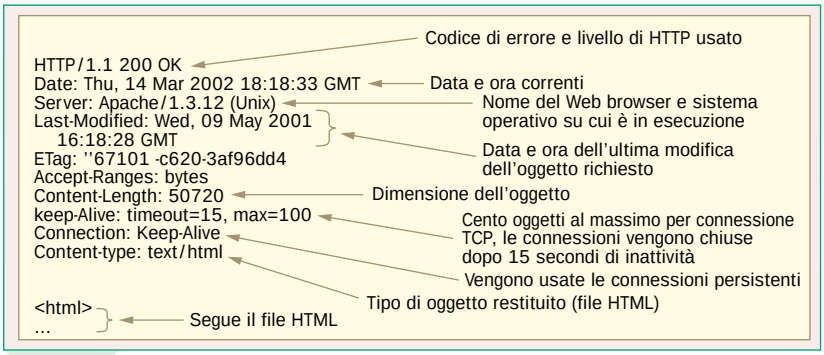


Figura B *Esempio di GET response restituita dal browser.*

I meccanismi GETALL, che accorpano tutte le GET successive alla prima in una sola GETALL, eliminano anche questo tipo di ridondanza e permettono quindi di ottenere un notevole risparmio di banda.

zione di come debba essere visualizzata la pagina Web sul browser dell'utente.

Al suo interno sono presenti numerose informazioni quali: il titolo visualizzato nella barra di stato del browser; lo schema grafico della pagina; il testo contenuto nella pagina; i riferimenti a tutte le immagini presenti (sotto forma di altri URL) e la loro posizione; i link ad altre pagine HTML alle quali può accedere l'utente.

Come si può notare, nel file HTML non sono incluse le immagini, ma sono presenti solo dei puntatori, codificati sottoforma di URL, che indicano la loro posizione fisica sul Web server. È compito del browser, tramite un'operazione detta di *parsing*, analizzare il file HTML e scaricare le immagini dal server per poterle poi presentare sullo schermo.

3.3 Caricamento di una pagina HTML

Il meccanismo complessivo con cui il browser carica e visualizza una pagina Web è il seguente:

- l'utente scrive un URL (cioè l'indirizzo della pagina desiderata) nel browser;
- il browser apre una connessione TCP[3] verso il server, e invia una richiesta del tipo *GET /nomepagina.html;*

(3) In realtà con lo standard HTTP è anche possibile riutilizzare la stessa connessione TCP per richiedere diversi oggetti. Questa modalità è conosciuta col nome di Connection Keep-Alive ed è usata dalla maggior parte dei browser e dei Web server moderni.

- il server legge il file HTML richiesto dal *filesystem*, lo invia al client e chiude la connessione TCP;
- il browser analizza il file HTML ricevuto *(parsing)*, e nel caso al suo interno siano presenti riferimenti ad altri oggetti (per esempio a immagini) li richiede al server con ulteriori aperture di sessioni TCP e *GET request;*
- il browser mano mano che riceve gli elementi referenziati nella pagina HTML li visualizza sul video.

Per evitare di sovraccaricare i server, il browser non richiede contemporaneamente tutti gli oggetti referenziati, ma ne richiede al massimo un certo numero (in genere due), e via via che li riceve richiede quelli successivi.

Il numero di negoziazioni *GET request/response* è, quindi, proporzionale al numero di oggetti presenti all'interno della pagina HTML.

Alcuni esempi di HTTP GET request/response sono riportati nel riquadro a pagina 63.

3.4 La compressione dell'HTTP

All'interno dello standard HTTP sono stati previsti, fin dall'inizio, due formati di compressione *lossless*[4] dei dati, che quando applicati ai tipici file HTML portano a fattori di compressione dell'ordine di 2-5.

Nel caso in cui il client dichiari nell'intestazione della richiesta che manda al server che è in grado di accettare contenuti compressi, il server può, in via opzionale, restituire al client un flusso di dati compresso: il browser si occupa poi di decomprimere i dati per riottenere il file originale.

Tutti i browser più recenti sono in grado di accettare formati compressi; anche i server Web oggi maggiormente impiegati sono in grado di eseguire la compressione dei dati, nativamente oppure mediante l'aggiunta di appositi *plug-in*.

Al momento, tuttavia, la stragrande maggioranza dei siti esistenti non ha abilitato la prestazione di compressione sui propri server.

3.5 Tecniche di ottimizzazione dell'HTTP

Le tecniche di ottimizzazione dell'HTTP operano prevalentemente su quattro aspetti:

a) riduzione del numero di negoziazioni richiesta/risposta *(GET request/response)* tra *client* e *server*. La tecnica risulta essere particolarmente efficace sulla rete GPRS, sia perché il numero di negoziazioni è molto elevato (una tipica pagina HTML contiene dai quaranta ai sessanta oggetti), sia perché ognuna di esse richiede approssimativamente[5] un RTT per essere eseguita, e l'RTT tipico delle reti GPRS attuali è dell'ordine del secondo.

La tecnica è generalmente realizzata sostituendo

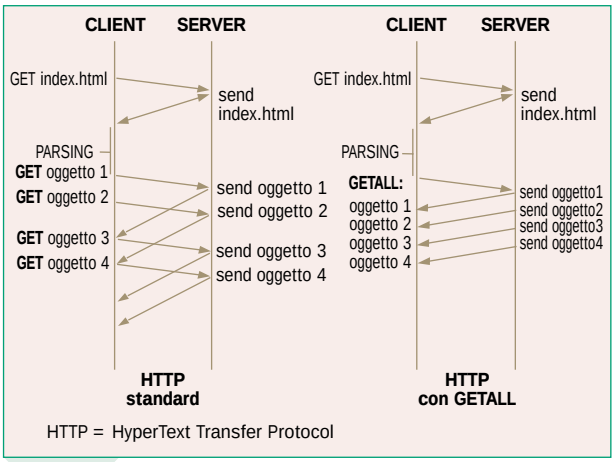


Figura 2 *Confronto tra la tecnica di richiesta HTTP standard e la tecnica cumulativa GETALL.*

le richieste per i singoli oggetti con un'unica richiesta cumulativa, chiamata in letteratura *GETALL* (GET di tutti gli oggetti della pagina), che contiene al proprio interno la lista di tutti gli oggetti voluti. Simmetricamente, lato server, tutti gli oggetti richiesti sono restituiti predisponendoli in serie su un unico flusso TCP.

L'uso di tale tecnica offre un ulteriore vantaggio: la riduzione del numero di intestazioni del protocollo HTTP mandate sia in *upstream* (dal browser verso il server Web) sia in *downstream* (dal server Web verso il browser), che contribuiscono in misura rilevante all'aumento della quantità complessiva di dati mandati in aria.

Nella figura 2 è mostrato il vantaggio derivante dall'uso della tecnica GETALL al posto della GET dell'HTTP standard: invece di attendere l'arrivo di ogni oggetto prima di richiedere quello successivo, tutti gli oggetti contenuti nella pagina Web sono richiesti contemporaneamente.

A differenza dell'esempio mostrato in figura, in cui gli oggetti sono solo quattro, le pagine Web sono invece in genere composte da diverse decine di oggetti, per cui la tecnica GETALL risulta estremamente efficiente.

b) compressione dei dati in volo eseguita utilizzando gli standard di compressione previsti nel protocollo HTTP.

c) ricodifica delle immagini nel formato JPEG effettuata decodificando le immagini e ricodificandole con fattori di compressione più elevati.

(4) *Sono definiti lossless i formati di compressione che dopo un ciclo di compressione e di decompressione restituiscono il contenuto originale identico* (byte per byte).

(5) *In questo caso si trascurano i tempi di elaborazione sia lato client sia lato server, che sono sempre di qualche ordine di grandezza inferiori, e si suppone di usare connessioni* Keep-Alive. *Le connessioni* Keep-Alive, *o persistenti, permettono al Web browser di scaricare dal Web server diversi oggetti usando la stessa connessione TCP: esse sono state introdotte come estensione della versione 1.0 del protocollo HTTP, che prevedeva che per ogni oggetto scaricato andasse stabilita una nuova connessione TCP. Nel caso in cui non vengano invece usate le connessioni* Keep-Alive, *il tempo richiesto per ogni negoziazione* GET request/response *raddoppia, in quanto diventano due gli RTT: un RTT per aprire la connessione TCP tramite il* 3-way handshake *ed un RTT per la negoziazione* GET request/response *vera e propria.*

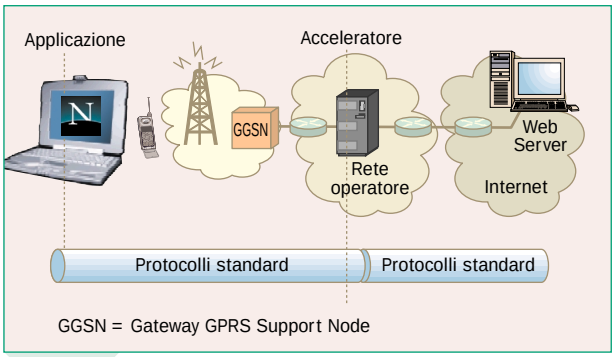


Figura 3 *Architettura di acceleratore server-only.*

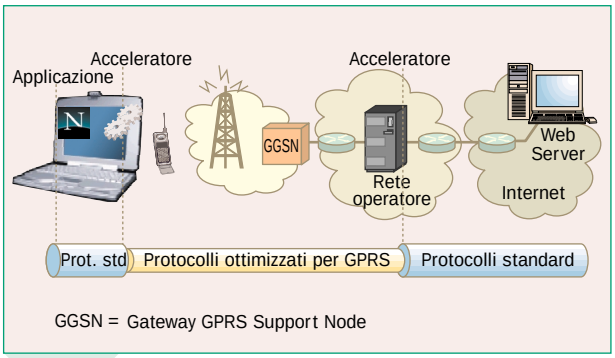


Figura 4 *Architettura di acceleratore client-server.*

d) riduzione del numero di colori contenuto nelle immagini GIF e rimozione delle eventuali animazioni.

4. Gli acceleratori/compressori per il GPRS

4.1 Aspetti generali

Tutti gli acceleratori GPRS sono oggetti che operano a livello di trasporto e applicativo e non richiedono, quindi, alcun intervento sulla parte radio trasmissiva e possono essere inseriti in qualunque punto della rete IP dell'operatore.

Sono oggi impiegati due tipi di schemi architetturali, che rappresentano due diverse soluzioni di compromesso tra efficienza e semplicità di installazione: il *server-only* e il *client-server*.

4.2 Architetture server-only

Le architetture *server-only*, schematizzate nella figura 3, richiedono solo l'installazione di un apposito server nella rete IP dell'operatore, che assume a tutti gli effetti la funzione di *proxy*.

La soluzione classica, chiamata anche *explicit proxy*, prevede che l'utente imposti esplicitamente sugli applicativi, per i quali desidera ottenere l'accelerazione, l'indirizzo del server di accelerazione.

In alternativa, è disponibile una soluzione denominata solitamente col termine di *transparent proxy* che prevede l'inserimento in rete di appositi apparati IP (chiamati *redirector*) che provvedono a reindirizzare le tipologie di traffico IP da accelerare verso il *proxy* acceleratore, senza richiedere alcun intervento sul lato utente.

Questo tipo di architettura non richiede l'installazione di alcun software sul dispositivo usato (computer portatili e palmari) e risulta perciò compatibile con tutti i dispositivi utilizzabili per l'accesso alla rete e con tutti i sistemi operativi più diffusi quali, ad esempio, Windows, MacOS, Linux.

La comunicazione tra il dispositivo di accesso e il server di accelerazione, che fisicamente transita sulla tratta radio, deve essere, però, realizzata necessariamente con protocolli standard, sui quali le tecniche di ottimizzazione applicabili hanno prestazioni più limitate.

	Server-only	Client-server
Livello trasporto (TCP)	Tuning dei parametri standard del TCP (per esempio numero dei segmenti iniziali slow start). Introduzione di un meccanismo di stima della banda compatibile con la semantica TCP	Sostituzione completa del TCP con un protocollo proprietario basato su UDP
Navigazione su Web (HTTP)	Compressione HTML. Riduzione della dimensione delle immagini GIF e JPEG	Richieste HTTP cumulative (GETALL). Uso di formati di compressione ad-hoc più efficienti

GIF = Graphics Interchange Format
HTML = Hyper Text Markup Language
HTTP = Hyper Text Transfer Protocol
JPEG = Joint Photographic Experts Group
TCP = Transmission Control Protocol
UDP = User Datagram Protocol

Tabella 1 *Tecniche di ottimizzazione applicabili ai protocolli di livello 4 e 7.*

4.3 Architetture client-server

Le architetture *client-server*, schematizzate nella figura 4, richiedono anche l'impiego di un software *client* da installare sul dispositivo usato per l'accesso.

Il *client* è configurato, di solito automaticamente, come *proxy* locale sugli applicativi per i quali si desidera l'accelerazione.

All'interno della configurazione del *client* è poi impostato l'indirizzo IP del server di accelerazione.

Il principale vantaggio presentato da questo tipo di architettura è che la comunicazione tra *client* e *server* di accelerazione, che fisicamente transita sulla tratta radio, può essere effettuata mediante appositi protocolli proprietari specificamente sviluppati per il funzionamento su GPRS.

Questa soluzione è quindi quella che garantisce il più sensibile miglioramento delle prestazioni e il massimo numero di applicazioni gestite.

La soluzione client-server ha però lo svantaggio di funzionare solo sulle piattaforme hardware e sui sistemi operativi che sono stati previsti dal singolo costruttore. I sistemi diventano così di tipo proprietario.

Tutte le soluzioni client-server supportano oggi, ad esempio, il sistema operativo Windows, mentre nessuna di esse gestisce i sistemi MacOS o quelli Linux.

4.4 Tecniche di ottimizzazione utilizzabili con le due architetture

Nella tabella 1, della pagina precedente, sono riassunte le tecniche di ottimizzazione applicabili ai protocolli di livello 4 e 7 prima esaminati, suddivise in base alle due architetture di acceleratori esistenti: server - only e client - server.

5. Conclusioni

La diffusione del servizio GPRS offerto dai maggiori operatori mobili di telecomunicazioni, unita alla maturità e alla stabilità di questa tecnologia trasmissiva, offre agli utenti la possibilità di continuare a usare le proprie applicazioni IP anche in mobilità.

Le prestazioni delle applicazioni IP, che partono già da un livello soddisfacente, garantito dalla buona disponibilità di banda offerta dal GPRS, possono essere ulteriormente migliorate tramite l'utilizzo degli acceleratori per GPRS.

Le tecniche di ottimizzazione dei protocolli e di compressione dei dati, utilizzate dagli acceleratori, migliorano le prestazioni delle applicazioni IP su reti radiomobili, rendendole confrontabili con quelle su linea fissa.

Bibliografia

[1] Stevens, R. : *TCP/IP Illustrated, Volume 1: The Protocols*. Addison-Wesley, 1994.

[2] *Transmission Control Protocol Darpa Internet Program Protocol Specification*. RFC 793, settembre 1981.

[3] Allman, M. et alii: *TCP Congestion Control*. RFC 2581, aprile 1999.

[4] Murhammer, M.W.; Atakan, O. : *TCP/IP Tutorial and Technical Overview*. IBM, ottobre 1998, disponibile all'URL: *http://www.redbooks.ibm.com*

[5] Bruno, G.; Mamino, D.: *Analisi preliminare del funzionamento del TCP su rete GPRS e cenni sull'implementazione dello stack TCP/IP in Solaris*. Telecom Italia Lab, Nota Tecnica DPC2001.02922, settembre 2001.

[6] Fielding, R. et alii: *Hypertext Transfer Protocol - HTTP/1.1*. RFC 2616, giugno 1999.

[7] *HTML 4.01 Specification*. W3C Recommendation, 24 dicembre 1999, disponibile all'URL *http://www.w3.org/TR/1999/REC-html401-19991224/*

Abbreviazioni

ASCII	American Standard Code for Information Interchange
CWND	Congestion WiNDow
GGSN	Gateway GPRS Support Node
GIF	Grafic Interchange Format
GPRS	General Packet Radio System
GSM	Global System for Mobile communications
HTML	HyperText Mark-up Language
HTTP	HyperText Transport Protocol
IP	Internet Protocol
JPEG	Joint Photograpic Experts Group
MIME	Multipurpose Internet Mail Extensions
SSTHRESH	Slow Start THRESHold
RTO	Retransmit TimeOut
RTT	Round Trip Time
RWND	Receiver WiNDow
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
URL	Uniform Resource Locator

Giorgio Bruno ha conseguito la laurea in Fisica ad indirizzo Cibernetico-Informatico presso l'Università degli studi di Torino nell'anno 1996, discutendo una tesi sulla compressione delle immagini. È stato assunto in Logicasiel (oggi Webegg), dove si è occupato di sistemi di assistenza online e della progettazione ed implementazione di un sistema di posta elettronica crittografata. Nel 2000 è stato assunto in CSELT (oggi TILAB), dove tuttora opera nell'ambito dell'area "Internet Technologies and Platforms" occupandosi della tematica relativa alle prestazioni delle applicazioni IP su GPRS, sia da un punto di vista teorico, attraverso lo studio dei vari protocolli utilizzati, sia da un punto di vista sperimentale con misure di laboratorio.

Fabio Mazzoli si è laureato nel 1998 in Ingegneria delle Telecomunicazioni presso l'Università degli studi di Roma "La Sapienza". Dallo stesso anno opera presso la linea "Sviluppo Sistemi di Commutazione" di TIM, dove ha inizialmente maturato esperienza nell'ingegnerizzazione di piattaforme e funzionalità di Core Network per le reti GSM e GPRS. Dal 2001 si occupa dello sviluppo di servizi, principalmente di quelli dati, contribuendo alla definizione dell'offerta dei servizi GPRS e acquisendo esperienza sulle tecnologie di ottimizzazione dei protocolli IP.

Aldo Vannelli si è laureato in Fisica con indirizzo Elettronico-Cibernetico presso l'Università degli studi di Roma "La Sapienza", discutendo una tesi sulla critto-compressione delle immagini. Dopo una breve esperienza nel campo della progettazione delle reti a pacchetto X.25, è entrato in Teleo, società dell'allora Gruppo Stet, dove si è occupato dal 1988 al 1992 di reti dati e tecnologie a supporto dei servizi di messaggistica elettronica X.400. Successivamente è stato assunto in SIP (oggi Telecom Italia) nella Direzione Clienti Business, dove inizialmente ha seguito le prime sperimentazioni di reti geografiche ad alta velocità in tecnologie Frame Realy e ATM (Asynchronous Transfer Mode), contribuendo allo sviluppo di vari progetti di ricerca a livello europeo (ATM Pilot, JAMES, TEN155). Nel 1996 ha assunto il coordinamento delle attività di definizione, progettazione e verifica sperimentale dei servizi dati a larga banda e in seguito dei servizi innovativi basati su tecnologia IP Gigabit Ethernet. In questo ambito ha svolto numerosi studi sulle metodologie di controllo della QoS (Quality of Service) dei servizi dati, in relazione alle problematiche di trasporto dei servizi multimediali. Da maggio 2001 opera nella linea "Sviluppo Sistemi di Commutazione" della Rete di Telecom Italia Mobile dove si occupa dello sviluppo dei servizi dati su tecnologie GPRS e UMTS.

Wireless LAN: tecnologie e applicazioni

MASSIMO COLONNA
GIOVANNA D'ARIA

La nascita nell'anno 1997 del primo standard IEEE 802.11 per le WLAN (Wireless Local Area Network) ha costituito il primo importante passo verso lo sviluppo di questa tecnologia che da allora ha trovato un sempre maggiore campo di impiego, evolvendosi continuamente per funzionalità, affidabilità e prestazioni. Inoltre, la gamma di apparati WLAN oggi sul mercato garantisce spesso che prodotti di manifatturiere diverse possano interoperare, svincolando l'utilizzatore di servizi su WLAN dalla scelta di un particolare produttore, così come già accade per i telefonini. Tutti questi fattori hanno favorito una sempre più rilevante riduzione dei prezzi e incoraggiato lo sviluppo di nuove applicazioni nel settore dell'accesso mobile a Internet, soprattutto in due aree: quella dei servizi offerti in reti private (in ambito residenziale, università, biblioteche, uffici, imprese, …) e quella dei servizi offerti in luoghi pubblici quali gli aeroporti, gli alberghi, i centri commerciali. L'articolo tratta entrambe le aree di applicazione citate, presentando a corredo alcune previsioni del mercato futuro. L'attività di standardizzazione in corso nel mondo, anche per quanto riguarda le frequenze, è quindi descritta in dettaglio, soffermandosi particolarmente sulle tecnologie IEEE 802.11 ed ETSI HIPERLAN/2 (HIgh PErformance Radio Local Area Network Type 2). Completa il quadro una breve descrizione di una terza tecnologia oggi disponibile, quella HomeRF.

1. Introduzione

Il settore delle *WLAN* sta emergendo come uno dei segmenti di mercato maggiormente in crescita nel panorama industriale delle comunicazioni.

I principali fattori che stanno determinando questo sviluppo sono molteplici: le migliorate prestazioni in termini di capacità; l'emergere di una tecnologia dominante, quella a standard IEEE 802.11; gli sforzi delle manifatturiere per garantire l'interoperabilità degli apparati; la riduzione dei prezzi; il supporto crescente alla tecnologia WLAN dei sistemi operativi e dei computer e l'adozione sempre più generalizzata anche nei settori *SME (Small Medium Enterprise)* e *SOHO (Small Office Home Office)*.

Anno dopo anno la produzione di apparati WLAN è notevolmente cresciuta, grazie alla disponibilità di una gamma sempre più estesa di prodotti standard e multi-vendor e alla costante discesa dei prezzi, che ha visto una riduzione superiore al 30 per cento rispetto a quello delle prime versioni di apparato. Secondo uno studio della società di analisi Cahners In-Stat/MDR [1], nel corso del 2002 le vendite di apparati WLAN aumenteranno del 75 per cento rispetto alle vendite dell'anno scorso, che avevano già visto un aumento del 23 per cento rispetto all'anno prima. Si prevede che le vendite di apparati WLAN passeranno dai 3,3 milioni di unità nel 2000 ai 23,6 milioni nel 2005[1] e che il mercato crescerà da

(1) *Fonte: Cahners In-Stat Group, 2001 [2].*

circa un miliardo di dollari nel 2000 fino a superare i 6,5 miliardi nel 2006[2] (figura 1).

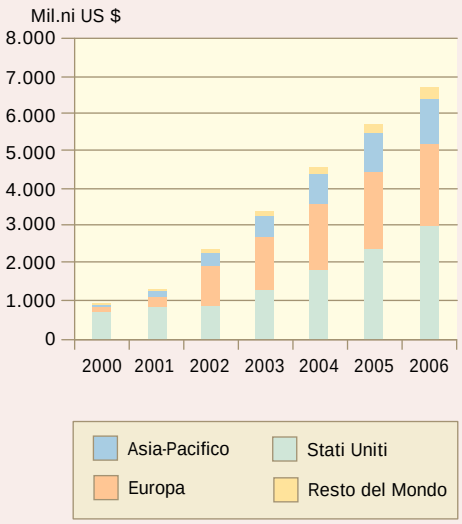


Figura 1 *Crescita delle vendite di apparati WLAN nel periodo 2000-2006.*

Le applicazioni nelle quali le WLAN hanno trovato finora impiego sono state soprattutto la realizzazione di reti locali da interno *(indoor)* per aziende o l'interconnessione di edifici *(reti di campus)*.

Più di recente l'adozione di soluzioni WLAN è stata anche estesa alla copertura di aree pubbliche *(hotspot)* quali aeroporti, alberghi, centri fieristici da parte di wireless ISP o agli ambienti domestici.

Relativamente alle aree pubbliche, negli Stati Uniti le WLAN sono infatti già installate in numerosi aeroporti, alberghi, caffè e ristoranti. Alla fine del 2001, il numero di installazioni era di circa 3mila, corrispondente all'80 per cento del totale delle installazioni mondiali, con una previsione di crescita di circa 35mila installazioni in tutto il mondo per il 2006[3] (figura 2).

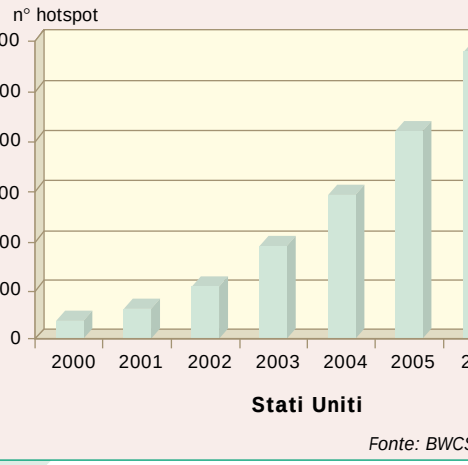


Figura 2 *Crescita del mercato americano degli hotspot nel periodo 2000-2006.*

In Europa, nonostante il più lento avvio rispetto agli Stati Uniti, sono ormai numerose le aree pubbliche (aeroporti, alberghi…) dotate di hotspot WLAN (figura 3) e sono realizzate prevalentemente nei Paesi scandinavi, mentre in altre Nazioni sono ancora in fase di sperimentazione.

In Gran Bretagna, per esempio, la British Airports Authority ha autorizzato una sperimentazione a Heathrow. L'operatore British Telecom ha inoltre di recente annunciato che, nel corso dell'anno, completerà 400 installazioni WLAN IEEE 802.11b in vari hotspot pubblici, compresi aeroporti e centri commerciali, che secondo le stime di British Telecom arriveranno a coprire circa il 30 per cento del mercato britannico potenziale. L'obiettivo è, infatti, quello di rendere operativi almeno 4mila siti entro i prossimi tre anni.

Gli hotspot pubblici permettono poi agli operatori mobili di utilizzare l'accesso WLAN sia come soluzione più rapidamente disponibile rispetto alle reti radiomobili di terza generazione *(3G)* per i servizi dati, sia come complemento alle stesse reti 3G e, in

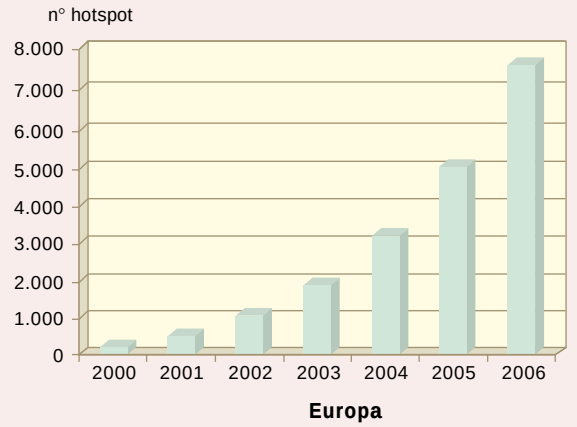


Figura 3 *Crescita del mercato europeo degli hotspot nel periodo 2000-2006.*

particolare, per fornire accessi dati ad alta velocità (oggi a 11 Mbit/s e nell'immediato futuro a 54 Mbit/s) in aree specifiche. Sono state avviate, a tal proposito, numerose iniziative tese a integrare le reti 3G con le WLAN, come per esempio il progetto H2U[4] di Ericsson e Telenor Mobil [4] basato su HIPERLAN/2 o l'attività di standardizzazione avviata da

(2) *Fonte: BWCS, 2001 [3].*

(3) *Fonte: BWCS, 2001[3].*

(4) *Il progetto H2U (sigla che mira a identificare i sistemi HIPERLAN/2 e UMTS) è stato promosso con l'obiettivo di valutare e provare gli aspetti tecnici, realizzativi e commerciali di soluzioni integrate HIPERLAN/2 e UMTS.*

UN MERCATO IN CAMMINO

È opinione comune degli analisti che il mercato delle WLAN sia in sensibile crescita: secondo Cahners In-Stat/MDR nel corso del 2002 le vendite di apparati WLAN aumenteranno del 75 per cento rispetto all'anno precedente, confermando così una tendenza già in atto in passato. Le vendite di apparati passeranno da 3,3 milioni di unità nel 2000 ai 23,6 milioni nel 2005, diffondendosi in molteplici aree di applicazione (residenziale, campus, imprese, hotspot pubblici, ...). Nelle aree pubbliche, solo negli Stati Uniti, il numero di hotspot crescerà da circa 3mila di fine 2001 fino a sfiorare i 30mila nel 2006; un analogo andamento nella crescita, anche se partendo da assai più bassi valori assoluti, si registrerà in Europa (BT da sola intende rendere operativi almeno 4mila siti hotspot entro i prossimi tre anni).

Inoltre, la diffusione delle WLAN in ambito privato, anche residenziale, sta facendo intravedere a molti operatori di telecomunicazione la possibilità di lanciare nuovi servizi di connettività locale, come dimostrato da recenti lanci di offerte commerciali.

ETSI, 3GPP e IEEE, finalizzate a definire le modalità di integrazione delle due reti.

Una crescita assai sostenuta è anche prevista dagli analisti per il settore domestico, dove per ora solo un numero minimo di abitazioni è dotato di LAN interna[5]. Nelle abitazioni possono essere oggi individuate diverse aree, ciascuna dedicata a un impiego specifico: telefonia; PC e relative periferiche; intrattenimento video e audio digitale *(TV e HiFi)*; automazione domestica (per il controllo e la sicurezza). Solitamente, però, le reti che collegano i terminali presenti all'interno di ciascuna area sono poco - o, più spesso, per nulla - interconnesse tra loro o alla rete di telecomunicazioni pubblica.

I sistemi cosiddetti *in-house* - e lo stesso *home networking* nel suo complesso - dovrebbero assumere un ruolo via via di maggior rilievo nelle comunicazioni, per l'evoluzione graduale di parte degli oggetti di uso domestico verso la possibilità di collegarsi a un qualche tipo di rete. L'obiettivo finale, per pervenire a un'efficiente rete domestica, sarà quello di predisporre una serie di reti, terminali e applicazioni, che interagiscano fra loro e che non siano dedicati a un unico tipo di impiego (figura 4).

Una delle possibili implementazioni di una rete LAN domestica, in grado di connettere tra loro i diversi sottosistemi, può essere realizzata, in tutto o in parte, con tecnologia WLAN, considerando naturalmente i benefici tecnico-economici e di *time-to-market* rispetto a tecnologie alternative (IEEE 1394 o Ethernet).

I nuovi sistemi in fase di sviluppo (IEEE 802.11a e HIPERLAN/2) consentiranno poi di superare le limitazioni di banda dei primi sistemi impiegati e permetteranno all'utilizzatore di poter usufruire di una vasta gamma di applicazioni, tra le quali anche quelle legate all'intrattenimento, che richiedono elevate velocità di trasmissione (tabella 1).

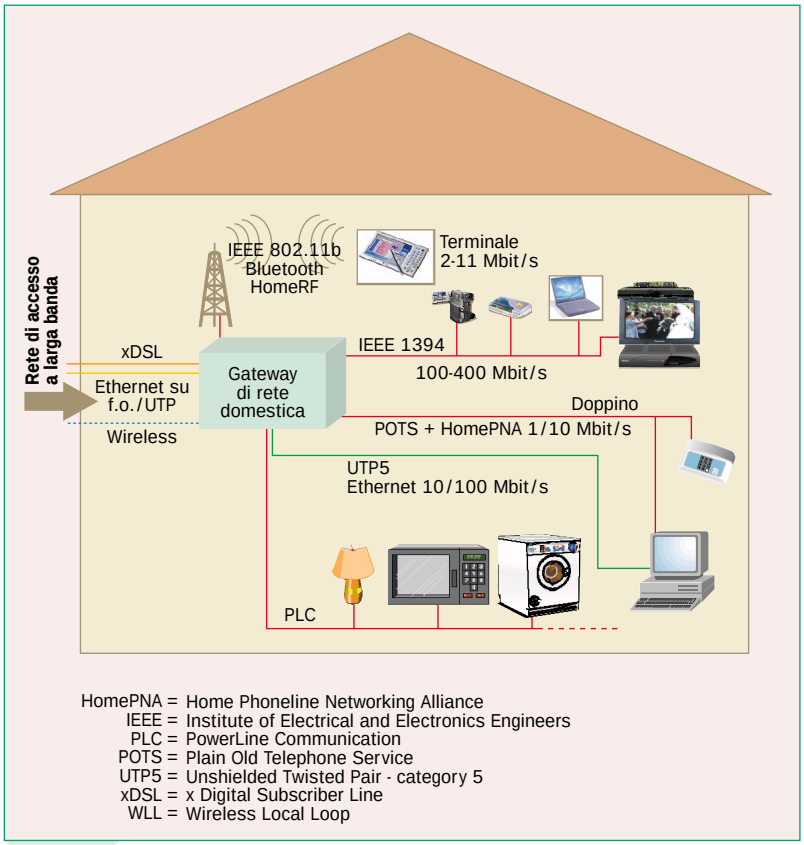


Figura 4 *Visione d'insieme delle tecnologie per Home Networking.*

Negli Stati Uniti, nel corso del 2001, le vendite di prodotti per l'*home networking* sono più che raddoppiate rispetto all'anno precedente, e i ricavi hanno sfiorato i 600 milioni di dollari. Già nel 2000 esse avevano avuto un incremento del 97 per cento rispetto all'anno 1999, raggiungendo i 290 milioni di dollari[6] e

(5) *Secondo, per esempio, un rapporto pubblicato da Yankee Group nel febbraio 2002, negli Stati Uniti, solo il 6 per cento delle abitazioni in cui è presente un PC è dotato di LAN domestica.*

(6) *Fonte: Cahners In-Stat Group, ottobre 2001.*

SERVIZI E APPLICAZIONI		
INFORMAZIONE E COMUNICAZIONE	INTRATTENIMENTO	Televisione e multimedia
		Foto/video
		Audio-HiFi
		Gioco
	SOCIO-CULTURALI	Accesso da postazione remota a una LAN aziendale
		Telelavoro, lavoro cooperativo
		Home news, Info-push e comunità virtuali
		Home banking, acquisti e pagamenti on-line
		Telemedicina, Teledidattica
	TELECOMUNICAZIONE	Integrazione con cellulari, PDA *(Personal Digital Assistant)*, per la mobilità
		Servizi telefonici
		Videocomunicazione
		E-mail
AUTOMAZIONE DOMESTICA	SICUREZZA	Telesoccorso
		Antintrusione, video controllo
		Segnalazione di incendio e di fughe gas
	GESTIONE AMBIENTE	Distribuzione energetica e consumi
	ROBOTICA DOMESTICA	Gestione automatizzata apparecchi di casa

Tabella 1 *Applicazioni per home networking.*

le previsioni per il futuro non fanno che avvalorare la tendenza attuale: il numero di unità abitative dotate in America di una struttura di rete domestica è destinato a crescere rapidamente nei prossimi anni (figura 5).

In questa situazione la penetrazione delle WLAN dovrebbe interessare il 20 per cento circa delle abitazioni.

2. Attività di standardizzazione

Negli ultimi anni sono stati compiuti considerevoli sforzi per standardizzare l'impiego delle tecnologie radio nelle reti locali. La prima organizzazione che ha avviato quest'attività è stata l'IEEE con il *Gruppo di Studio per le Wireless LAN* che nel 1997 ha pubblicato la prima versione di uno standard per applicazioni legate al trasporto di dati asincroni, chiamato 802.11.

Per il livello fisico *PHY (PHYsical layer)*, in particolare, erano previste inizialmente due versioni (802.11 DSSS e 802.11 FHSS) nel campo delle microonde, operanti entrambe nella banda *ISM (Industrial Scientific and Medical)* a 2,4 GHz e che impiegavano la tecnica dello *spread spectrum* per contrastare l'interferenza proveniente da altri apparati operanti nello stesso campo di frequenza (quali, per esempio, i forni a microonde)[7].

Per ottenere capacità più elevate, paragonabili a quelle delle LAN cablate, l'IEEE ha successivamente definito due ulteriori versioni del PHY, chiamate 802.11b e 802.11a, operanti, rispettivamente, nelle gamme a 2,4 GHz ed a 5 GHz.

Anche l'*ETSI (European Telecommunications Standards Institute)*, seppur con ritardo rispetto all'IEEE, ha avviato un'attività di standardizzazione sui sistemi WLAN. Con il progetto *ETSI BRAN (Broadband Radio Access Networks)* sono state, in particolare, completate di recente le specifiche di un sistema denominato *HIPERLAN/2* operante nella gamma a 5 GHz con l'obiettivo di fornire l'accesso ad alta velocità a differenti tipi di reti e la mobilità dei terminali.

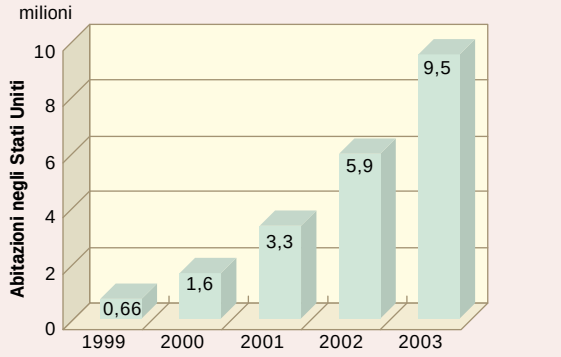


Figura 5 *Numero di abitazioni con home networking.*

In effetti, nell'ambito del Gruppo di Lavoro *RES 10 (Radio Equipment and Systems 10)*, l'ETSI aveva già definito lo standard *HIPERLAN Type 1* per reti ad-hoc nella gamma a 5 GHz. Sebbene lo standard HIPERLAN/1 sia stato introdotto per consentire il trasporto anche di servizi con vincoli di ritardo *(time-bounded)*, esso tuttavia non consente di controllare completamente e di garantire, quindi, una prefissata qualità del servizio. Queste limitazioni, unite alla massiccia presenza sul mercato di prodotti realizzati secondo lo standard IEEE 802.11b, hanno suggerito l'abbandono di questa tecnologia e hanno portato l'ETSI a definire la normativa HIPERLAN/2.

Accanto agli Enti di standardizzazione sono nate alcune Associazioni di costruttori con l'obiettivo di promuovere l'utilizzo delle WLAN. Il più importante di questi raggruppamenti è forse la *WECA (Wireless*

(7) *La versione base dello standard 802.11 comprende anche un livello PHY nella gamma dell'infrarosso (802.11 IR). La tecnologia a infrarosso standardizzata dall'IEEE usa una tecnica di trasmissione denominata* diffused infrared*, che, diversamente da altre, non richiede che i due apparati IR in comunicazione siano necessariamente puntati l'uno contro l'altro né che siano in vista. Tuttavia, proprio perché la ricezione è basata sulle riflessioni multiple del segnale, gli apparati realizzati con questa tecnologia richiedono per il funzionamento, ambienti con soffitti non troppo alti. Essi perciò possono essere utilizzati solo all'interno di ambienti singoli di dimensioni ridotte (non possono essere impiegati, per esempio, nelle hall degli alberghi, in stazioni ferroviarie o negli aeroporti) e questa limitazione ne spiega la ridotta presenza sul mercato.*

DIVERSE TECNICHE WLAN SI FRONTEGGIANO

Le principali normative internazionali della tecnologia WLAN sono l'IEEE 802.11 e l'ETSI HIPERLAN/2.

Lo standard IEEE 802.11 rappresenta una famiglia di sistemi con proprie peculiarità. In particolare, la versione IEEE 802.11b è quella oggi più diffusa: essa opera alle frequenze dei 2,4 GHz e consente una velocità trasmissiva fino a 11 Mbit/s. La versione IEEE 802.11a, invece, opera alle frequenze dei 5 GHz e offre una velocità trasmissiva fino a 54 Mbit/s. Altre versioni sono allo studio allo scopo di arricchire l'insieme dei sistemi di ulteriori funzionalità, quali, per esempio, la gestione della qualità del servizio, i meccanismi di sicurezza e di autenticazione, il controllo automatico della potenza in trasmissione, la selezione dinamica del canale radio.

Il sistema ETSI HIPERLAN/2 opera a 5 GHz e permette una velocità trasmissiva fino a 54 Mbit/s. Esso è già in grado di fornire, tra l'altro, le funzioni di controllo automatico della potenza in trasmissione e di selezione dinamica del canale, che rappresentano un requisito normativo necessario per l'utilizzo in Europa di apparati WLAN nella gamma dei 5 GHz. Inoltre, mentre, i sistemi IEEE 802.11 possono essere impiegati esclusivamente in reti IP (con il trasporto su Ethernet e PPP), lo standard ETSI HIPERLAN/2 definisce un sistema radio che può essere interconnesso, e quindi integrato in termini di servizi e applicazioni, anche a reti IEEE 1394, ATM e UMTS. Questa prestazione è resa possibile grazie a un'architettura flessibile e alla definizione di un insieme specifico di *convergence layer* di interconnessione.

Ethernet Compatibility Alliance) [5], della quale fanno parte, tra gli altri, 3Com, Lucent, Nortel, Nokia, Samsung, Philips, Cisco, e che persegue l'obiettivo primario di certificare l'interoperabilità tra apparati WLAN della famiglia IEEE 802.11 di costruttori diversi[8]. Gli apparati approvati da questa Associazione recano il marchio *Wi-Fi™ (Wireless Fidelity)*. In un futuro prossimo, compatibilmente alla disponibilità di apparati commerciali, anche gli apparati realizzati secondo lo standard IEEE 802.11a potranno fregiarsi di un marchio Wi-Fi che ne garantisca l'interoperabilità, denominato Wi-Fi5 (il numero cinque indica la gamma a 5 GHz in cui operano questi nuovi apparati).

Parallelamente all'attività di standardizzazione ufficiale, è stata intrapresa un'iniziativa, promossa da un vasto numero di aziende manifatturiere, tesa a definire le specifiche per l'interoperabilità di apparati radio rivolti essenzialmente a una clientela di tipo consumer che non presentano quindi requisiti stringenti in termini di capacità e di portata. I principali Consorzi sono *Bluetooth* e *HomeRF*.

Bluetooth [6] è una specifica, per un sistema ideato dalla Ericsson, che ha l'obiettivo di consentire la comunicazione radio tra una molteplicità di apparati (quali, ad esempio, PC, stampanti, telefoni, fax, auricolari) in ambiente domestico o *SOHO* e che non richiede alcuna infrastruttura di rete. La specifica è stata promossa da cinque aziende: la stessa Ericsson, e da Nokia, Intel, IBM e Toshiba, che nel 1998 hanno costituito il raggruppamento *SIG (Special Interest Group)*. Al gruppo di aziende promotrici hanno successivamente aderito, tra gli altri, anche 3Com, Lucent, Microsoft e Motorola, e oggi le società manifatturiere che complessivamente hanno adottato questa specifica sono oltre 4mila. Un approfondimento su questa tecnologia è disponibile in [7].

L'*HomeRF Working Group* [8] è un Gruppo di Studio avviato nel marzo 1998 cui hanno aderito diverse aziende che operano nel mondo dell'*ICT (Information & Communication Technology)*, tra cui Intel, Compaq, Proxim, Motorola e Siemens. Il Gruppo persegue l'obiettivo di definire uno standard di intercomunicabilità tra dispositivi di tipo consumer per il networking domestico (vedi riquadro a pagina 72). Lo standard, operante nella banda ISM a 2,4 GHz, è basato sul protocollo di accesso al canale chiamato *SWAP (Shared Wireless Access Protocol)*. Finora *HomeRF* non ha avuto un grande seguito applicativo se confrontato con le tecnologie IEEE 802.11b e Bluetooth.

3. Le frequenze

Le principali bande di frequenza assegnate dalla CEPT alle WLAN sono quelle a 2,4 GHz (banda ISM) [9] e a 5 GHz [10] (figura 6).

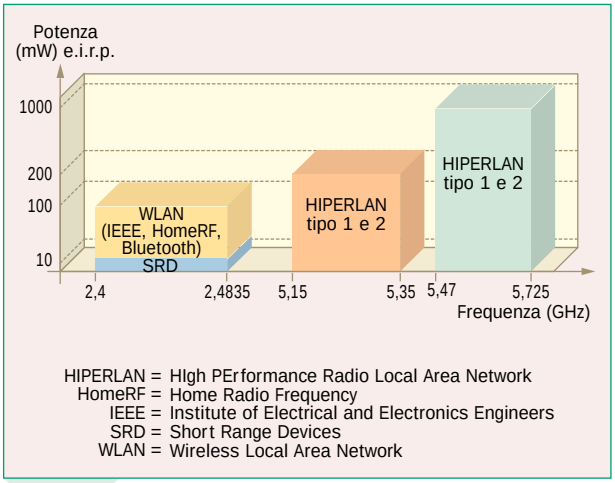
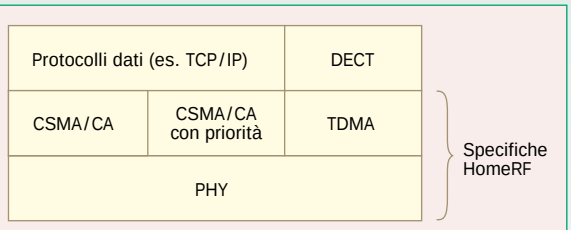


Figura 6 *Bande di frequenze CEPT per le WLAN.*

[8] *"WECA's mission is to certify interoperability of Wi-Fi (IEEE 802.11) products and to promote Wi-Fi as the global wireless LAN standard across all market segments".*

Il protocollo HomeRF

Il protocollo HomeRF, il cui modello di riferimento è mostrato nella figura A, opera a 2,4 GHz e utilizza la tecnica di *spreading Frequency Hopping Spread Spectrum*. La modulazione può essere di tipo 2-FSK o 4-FSK con una velocità trasmissiva rispettivamente di 1 o di 2 Mbit/s su canali di 1 MHz. La versione 2.0 dello standard, approvata dal Gruppo di Lavoro verso la metà dello scorso anno, consente di affasciare più canali con una velocità trasmissiva massima di 10 Mbit/s.



CSMA/CA = Carrier Sense Multiple Access with Collision Avoidance
DECT = Digital Enhanced Cordless Telecommunications
PHY = PHYsical layer
TCP/IP = Transmission Control Protocol/Internet Protocol
TDMA = Time Division Multiple Access

Figura A *Modello di riferimento del protocollo HomeRF.*

Il protocollo di accesso al canale è chiamato *SWAP (Shared Wireless Access Protocol)* e tratta in modo diverso i servizi fonici e quelli dati. Per i primi il protocollo è derivato dal DECT e quindi utilizza la tecnica di accesso multiplo a divisione di tempo *TDMA (Time Division Multiple Access)* con duplex a divisione di tempo, *TDD (Time Division Duplex)*; la voce è codificata con la tecnica ADPCM a 32kbit/s.

Per il servizio dati la tecnica adottata è, invece, la *CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance)* dello standard IEEE 802.11 con l'aggiunta di livelli diversi di priorità per la gestione separata dei flussi multimediali con vincoli stringenti sul jitter e sul ritardo.

La presenza congiunta dei due tipi di servizio è realizzata grazie alla definizione di una struttura di trama variabile (figura B, pagina 73).

In presenza di solo traffico dati, la trama, della durata di 20 ms, ha inizio con il salto di frequenza e la tecnica di accesso al canale è, come detto, la CSMA/CA[1] (figura B, riga 1, pag.73).

(1) *L'intervallo di tempo in cui l'accesso al canale avviene con tecnica CSMA/CA è chiamato* CP (Contention Period).

La banda ISM è compresa tra 2,4 e 2,4835 GHz. In essa possono operare apparati WLAN con un limite di potenza massimo pari a 100 mW *e.i.r.p. (equivalent isotropically radiated power)* e con modulazioni del tipo spread spectrum *(IEEE 802.11, Bluetooth e HomeRF)*. In questa banda le WLAN devono coesistere con tutti gli apparati radio a corto raggio e, in particolare, con gli *SRD (Short Range Devices)*, utilizzati nelle applicazioni industriali scientifiche e mediche ai quali la banda ISM era stata in origine assegnata.

La gamma a 5GHz è invece costituita da due bande separate: la prima compresa tra 5,150 e 5,350 GHz, la seconda tra 5,47 e 5,725 GHz.

In particolare, nella prima banda possono operare esclusivamente apparati per applicazioni in interni *(indoor)* con un limite di potenza massimo pari a 200 mW e.i.r.p. Nella seconda possono, invece, operare anche apparati per applicazioni in ambienti esterni *(outdoor)* con un limite di potenza pari a 1 W e.i.r.p.

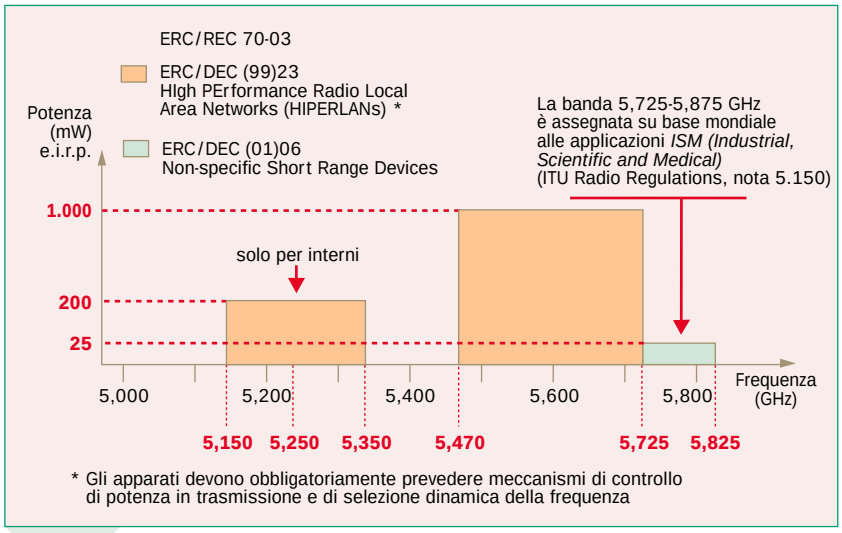


Figura 7 *Regolamentazione CEPT per l'Europa delle frequenze a 5 GHz.*

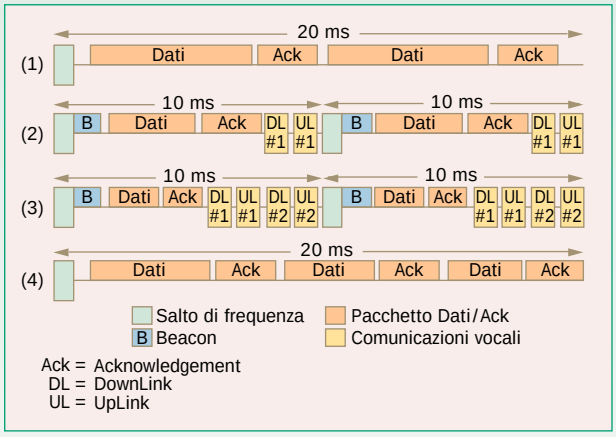


Figura B *Struttura di trama del protocollo SWAP.*

In presenza di traffico fonico la trama si riduce a 10 ms (in tal modo il ritardo dei servizi vocali è contenuto entro questo tempo) e include, oltre al CP, un *CFP* (*Contention-Free Period*)[2] la cui presenza viene notificata a tutti i terminali da un pacchetto di Beacon inviato in broadcast subito dopo il salto di frequenza da un Connection Point, il terminale che gestisce l'accesso multiplo al canale per le comunicazioni vocali. Il CFP è costituito da un massimo di 8 intervalli di tempo *(time slot)* per direzione trasmissiva (TDMA/TDD) per le comunicazioni vocali. In particolare il primo slot per ciascuna direzione trasmissiva è dedicato alla comunicazione in *DownLink*, ovvero nella direzione dal *Connection Point* ai terminali, mentre il secondo è dedicato a quella in *UpLink*, ovvero nella direzione dai terminali al *Connection Point* (figura B, riga 2).

Il numero di time slot varia in base al numero di comunicazioni attive (figura B, riga 3).

Quando tutte le comunicazioni vocali terminano, la trama ritorna a una durata di 20 ms (figura B, riga 4).

I sistemi *HomeRF* permettono di realizzare sia reti *ad-hoc* sia reti *infrastructure*. Nelle prime, poiché la gestione dell'accesso al canale è distribuita tra tutti i terminali, possono essere forniti solo i servizi di dati. Nelle seconde, invece, poiché il controllo del mezzo è gestito da un Connection Point, è possibile fornire entrambi i tipi di servizio.

Una rete può essere composta al massimo da centoventisette nodi (voce e dati) e può gestire fino a otto comunicazioni vocali full-duplex contemporanee.

[2] *L'intervallo di tempo in cui l'accesso al canale avviene con tecnica TDMA/TDD è chiamato CFP.*

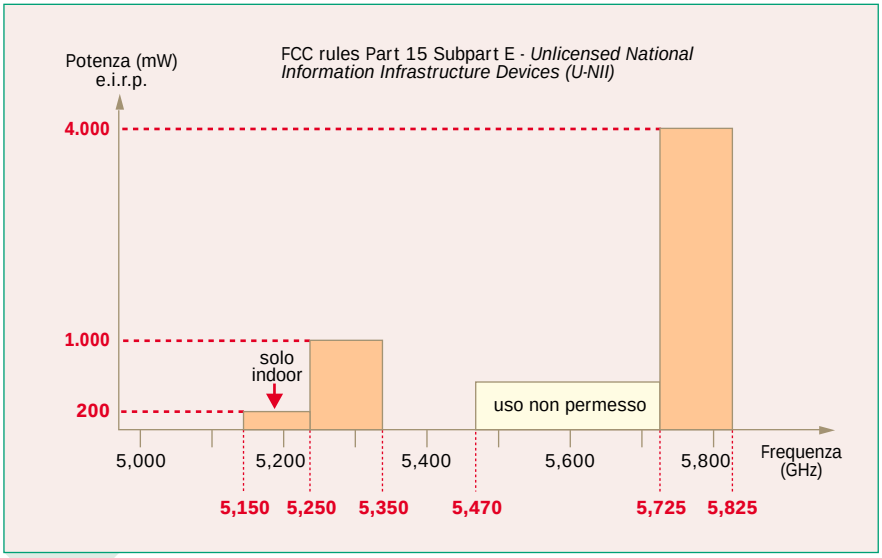


Figura 8 *Regolamentazione FCC negli Stati Uniti delle frequenze a 5 GHz.*

Per ridurre il più possibile le interferenze e per ottimizzare l'uso dello spettro disponibile, devono essere impiegati meccanismi di *TPC (Transmit Power Control)* e di *DFS (Dynamic Frequency Selection)*.

La gamma a 5 GHz è assegnata esclusivamente agli apparati conformi alle specifiche ETSI per HIPERLAN e quindi in Europa non è permesso oggi l'impiego di sistemi realizzati secondo lo standard IEEE 802.11a.

Nelle figure 7 e 8 è riportata una visione della regolamentazione a 5 GHz, rispettivamente, in Europa e negli Stati Uniti.

4. Tecnologia IEEE 802.11

Con riferimento al modello ISO/OSI, lo standard IEEE 802.11 definisce il livello PHY e il livello *MAC (Medium Access Control)*. L'interfaccia del livello MAC verso il livello Rete è costituita dal

LA SFIDA IN ATTO TRA GLI STANDARD DELLE WLAN

Mentre oggi e nell'immediato futuro, lo standard dominante è l'IEEE 802.11b, appare invece aperta, soprattutto in Europa, la sfida tra gli standard IEEE 802.11a ed ETSI HIPERLAN/2. Al momento lo standard americano sembra favorito, non fosse altro che per la disponibilità sul mercato di apparati commerciali, seppure nella versione non ancora utilizzabile in Europa. HIPERLAN/2 riuscirà, tuttavia, molto verosimilmente a ritagliarsi settori di applicazione specifici in cui siano richiesti servizi con requisiti di qualità sui ritardi e sulla banda, quali i servizi video e voce, integrati ai servizi dati. Un ulteriore fattore determinante sarà il prezzo a regime degli apparati, di quelli, in particolare, destinati agli utilizzatori privati, oggi non ancora prevedibile.

livello *LLC (Logical Link Control)* specificato dallo standard IEEE 802.2 [11] (figura 9).

Lo standard 802.11 prevede finora quattro diversi

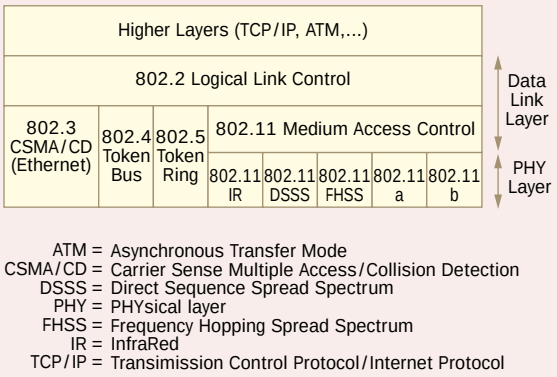


Figura 9 *Protocolli IEEE 802.11 e relazioni con alcuni degli altri protocolli della famiglia IEEE 802.*

livelli PHY nel campo delle microonde e un unico livello MAC. Le principali caratteristiche della tecnologia 802.11 sono:

- *supporto per i servizi asincroni:* il protocollo MAC fornisce due tipi di servizi: asincroni e *contention free*. La fornitura dei primi servizi è obbligatoria mentre quella dei secondi è opzionale e non è presente in nessun apparato oggi disponibile sul mercato;
- *mobilità dei terminali*: per comunicare tra loro o con la rete esterna, le stazioni utilizzano il miglior punto di accesso, *AP (Access Point)*, cioè quello che consente a ciascuna di esse di avere il miglior rapporto segnale-rumore. Per favorire gli spostamenti all'interno di un'area, lo standard prevede la possibilità che le stazioni effettuino periodicamente misure di potenza ricevuta su tutte le frequenze disponibili in modo da consentire di trovare il migliore AP al quale connettersi;
- *meccanismi di power saving per i terminali portatili:* questi meccanismi consentono a una stazione di passare al funzionamento a basso consumo di potenza, quando il trasmettitore e il ricevitore sono spenti, e di uscirne periodicamente per ricevere i dati a essa destinati;
- *meccanismi di sicurezza:* lo standard definisce strumenti e procedure per garantire la privacy delle comunicazioni, *WEP (Wired Equivalent Privacy)*, e per prevenire l'accesso non autorizzato alla rete.

4.1Il livello MAC

Il livello MAC [12] fornisce i servizi asincroni tramite la *DCF (Distributed Coordination Function)* che utilizza la tecnica di accesso multiplo al canale denominata *CSMA/CA (Carrier Sense Multiple Access/Collision Avoidance)*.

Secondo la DCF (figura 10), una stazione può avviare la trasmissione di un pacchetto di dati su un determinato canale solo dopo aver rilevato che il canale è libero per un tempo superiore a un valore detto *DIFS (Distributed InterFrame Space)*.

In caso contrario la trasmissione è rimandata e la stazione avvia un processo di back-off, durante il quale il relativo temporizzatore *(timer)*[9] è decrementato solo quando la stazione rileva che il canale è libero e dopo aver aspettato per un tempo pari a un DIFS mentre è bloccato quando la stazione rileva che il canale è occupato.

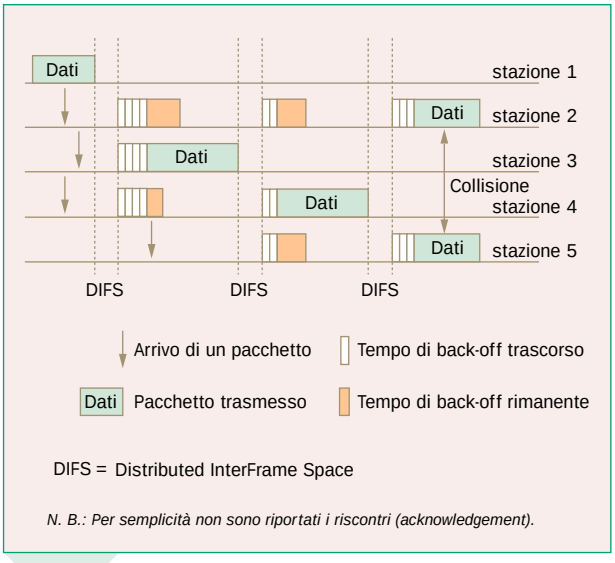


Figura 10 *Meccanismo di accesso al canale CSMA/CA.*

(9) *Il valore del temporizzatore è scelto casualmente tra zero e un valore massimo chiamato* CW (Contention Window).

La tecnica DSSS nei sistemi IEEE 802.11

Le tecniche Spread Spectrum sono state introdotte nelle WLAN con l'obiettivo primario di combattere le interferenze prodotte da altri apparati operanti nella stessa gamma di frequenza.

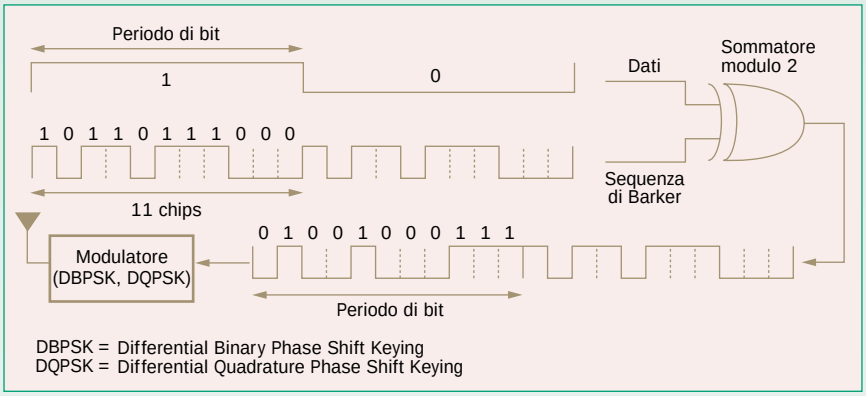


Figura A *Schema a blocchi del trasmettitore DSSS.*

Con la tecnica *DSSS (Direct Sequence Spread Spectrum)*, ogni bit di informazione, prima di essere modulato e trasmesso sul canale, viene sommato modulo 2 (funzione XOR) a una sequenza di *spreading*, lunga 11 bit (chip)[1], generata con cadenza pari a 11 Mbit/s (detta *chipping clock*). All'uscita del sommatore, la sequenza binaria risultante ha una velocità che è undici volte maggiore della velocità della sequenza di informazione e ciò corrisponde, nel dominio della frequenza, a un segnale il cui spettro è "spalmato" *(spread)* su un campo di frequenze allargato e con un livello di potenza ridotto (figura A). In ricezione l'operazione di correlazione del segnale ricevuto con la stessa sequenza (operazione di despreading) consente di recuperare il segnale utile e, se sul canale è presente un segnale interferente, di ridurre la potenza di quest'ultimo nella banda del segnale utile mitigandone notevolmente gli effetti (figura B).

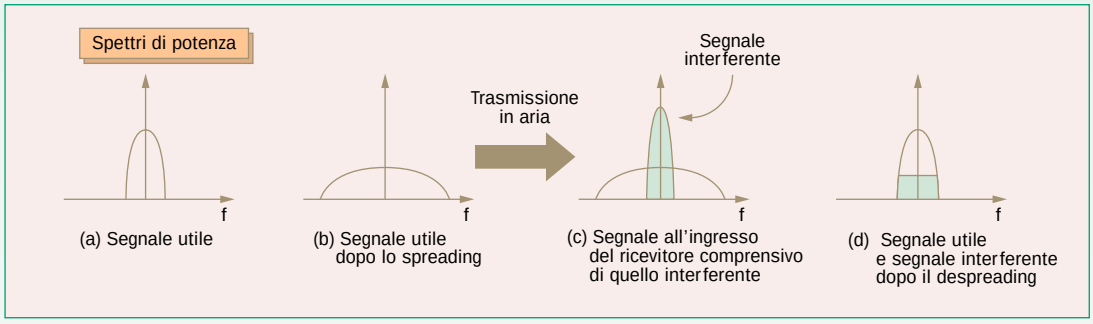


Figura B *Funzionamento della tecnica DSSS in presenza di un segnale interferente.*

DBPSK	
bit	**Transizione di fase**
0	0
1	π
DQPSK	
bit	**Transizione di fase**
00	0
01	π/2
11	π
10	3π/2

DBPSK = Differential Binary Phase Shift Keying
DQPSK = Differential Quadrature Phase Shift Keying

Tabella A *Associazione tra i bit all'ingresso del modulatore e le transizioni di fase del segnale trasmesso.*

La trasmissione sul canale del segnale dopo lo spreading avviene utilizzando una modulazione di tipo *DPSK (Differential Phase Shift Keying)*, in particolare di quella *DBPSK (Differential Binary Phase Shift Keying)* oppure *DQPSK (Differential Quadrature Phase Shift Keying)*. L'associazione dei bit della sequenza all'ingresso del modulatore con le transizioni di fase del segnale trasmesso è riportata nella tabella A.

(1) *La sequenza impiegata è una sequenza di Barker di 11 bit costituita dai simboli (chip) {+1, –1, +1, +1, –1, +1, +1, +1, –1, –1, –1}.*

Quando il *timer* si azzera la stazione è autorizzata a trasmettere. Se due o più stazioni trasmettono contemporaneamente, si producono collisioni tra i messaggi.

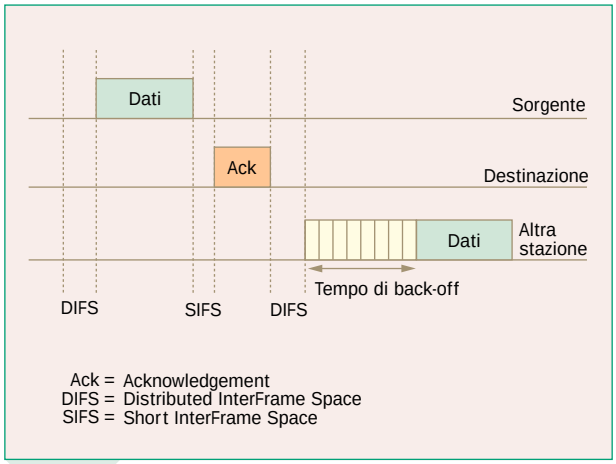


Figura 11 *Trasmissione di pacchetto con riscontro.*

Contrariamente a quanto accade nelle reti cablate - per esempio quelle a standard IEEE 802.3 ovvero quelle Ethernet, che fanno uso della tecnica *CSMA/CD (Carrier Sense Multiple Access/Collision Detection)* - le collisioni non possono essere rilevate dai terminali. Per notificare alla stazione trasmittente la corretta ricezione di un pacchetto, la stazione ricevente invia, perciò, un riscontro *(acknowledgement)*. Per evitare la contesa anche nella trasmissione del riscontro, l'invio di quest'ultimo avviene dopo un tempo pari a un valore detto *SIFS (Short InterFrame Space)*, più breve di un DIFS, dalla fine della ricezione (figura 11).

Un problema frequente negli ambienti interni, causato dall'attenuazione dovuta agli ostacoli - quali pareti o soffitti - è quello dei nodi nascosti, ovvero di quelle stazioni che non sono capaci di rilevare l'una la trasmissione dell'altra.

Come si vede anche dalla figura 12, le stazioni A e B, pur appartenendo alla stessa cella, sono al di fuori dei rispettivi raggi di copertura e quindi non sono in grado di rilevare l'una le trasmissioni dell'altra. Entrambe, quindi, sentendo il canale sempre libero, trasmettono contemporaneamente, generando collisioni che, in situazioni di elevato traffico, determinano un forte degrado nelle prestazioni.

Per far fronte a questo problema, il protocollo MAC prevede un meccanismo basato sullo scambio di due brevi pacchetti di controllo: *RTS (Ready To Send)* e *CTS (Clear To Send)*. Il primo è inviato dalla stazione trasmittente, mentre il secondo è inviato da quella ricevente in risposta al primo. Entrambi i messaggi contengono un campo che specifica il tempo necessario a completare la trasmissione dei dati.

Tutte le stazioni della cella, comprese quelle nascoste rispetto a una delle due coinvolte nella comunicazione, ricevono almeno uno dei due pacchetti di controllo ed evitano l'accesso al mezzo per il tempo necessario al completamento della trasmissione (figura 12). L'*overhead* aggiuntivo dovuto a questi due pacchetti è più che compensato dal miglioramento delle prestazioni ottenibile con l'eliminazione delle collisioni.

I servizi *contention free* - la cui introduzione è opzionale - sono invece realizzati dalla *PCF (Point Coordination Function)* che si alterna con la DCF. In una rete che contiene questa prestazione è sempre presente un *PC (Point Coordinator)* che periodicamente, dopo aver atteso per un tempo pari ad un *PIFS (Priority InterFrame Space)*, più corto di un DIFS, prende il controllo del mezzo trasmettendo un particolare pacchetto, chiamato *Beacon*, che informa i terminali dell'inizio del *CFP (Contention-Free Period)* e della sua durata.

Durante il CFP, il PC utilizza un meccanismo di polling, ovvero interroga tutte le stazioni che hanno richiesto servizi *contention free* e le abilita alla trasmissione dei propri dati (figura 13).

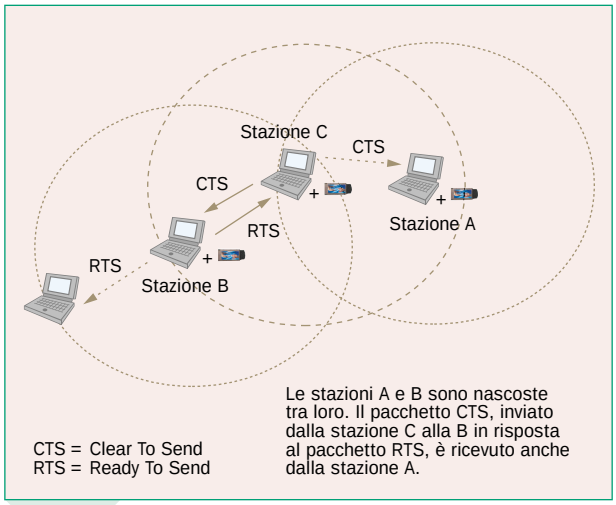


Figura 12 *Nodi nascosti.*

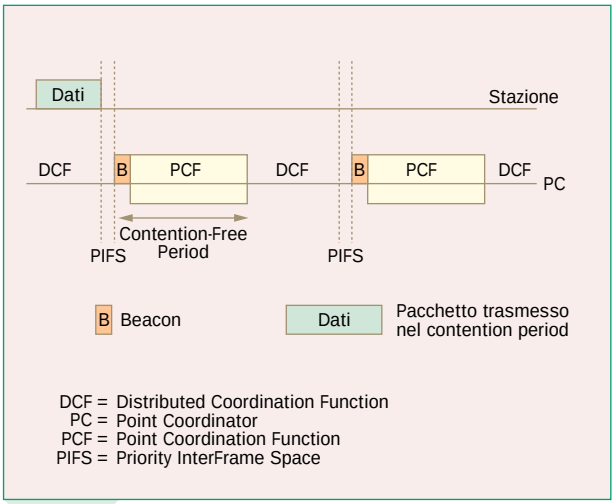


Figura 13 *Alternanza tra Point Coordination Function e Distributed Coordination Function.*

La tecnica FHSS nei sistemi IEEE 802.11

Con la tecnica FHSS, la portante a radiofrequenza sulla quale sono trasmessi i dati non è fissa ma varia in un insieme comprendente un dato numero di frequenze, che in Europa è pari a 79, secondo una sequenza, chiamata *sequenza di hopping* e determinata in modo pseudocasuale (figura A). In questo modo un eventuale interferente non danneggia tutta la trasmissione, ma solo parte di essa ovvero solo quella porzione il cui spettro va a sovrapporsi a quello dell'interferente. La velocità con cui il segnale passa da un canale all'altro deve essere di almeno 2,5 salti al secondo, in accordo con quanto stabilito per l'Europa in [26].

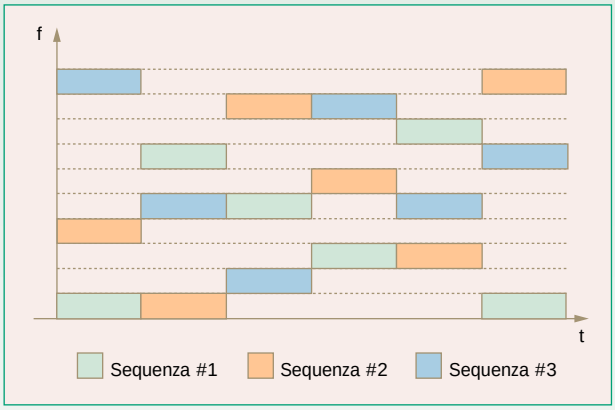


Figura A *Trasmissione secondo la tecnica FHSS.*

2-GFSK	
bit	**± f_d (kHz)**
0	+ 160
1	- 160
4-GFSK	
bit	**± f_d (kHz)**
10	+ 216
11	+72
01	- 72
00	- 216

GFSK = Gaussian Frequency Shift Keying

Tabella A *Associazione tra i bit all'ingresso del modulatore e le variazioni di frequenza della portante.*

Lo schema di modulazione utilizzato è del tipo *GFSK (Gaussian Frequency Shift Keying)*, che associa i bit di informazione a una variazione (f_d) positiva o negativa della frequenza della portante (F_c). Con il 2-GFSK un 1 binario determina una variazione positiva della frequenza della portante ($F_c + f_d$), mentre uno 0 binario determina una variazione negativa ($F_c - f_d$). Con il 4-GFSK, invece, sono possibili quattro variazioni associate ciascuna a una coppia di bit (00, 01, 11, 10) (tabella A). Allo scopo di ridurre l'occupazione spettrale in aria, i bit, prima di essere inviati al modulatore, attraversano un filtro passabasso con risposta all'impulso di tipo Gaussiano, caratterizzato da un B*T pari a 0,5.

4.2 Livelli PHY

Le specifiche 802.11 DSSS e FHSS [12], per sistemi operanti nella banda ISM a 2,4 GHz, prevedono l'impiego rispettivamente delle tecniche di trasmissione *DSSS (Direct Sequence Spread Spectrum)* e *FHSS (Frequency Hopping Spread Spectrum)*.

Più in particolare, i sistemi in accordo con la norma 802.11 DSSS utilizzano le modulazioni DBPSK e DQPSK per una velocità trasmissiva *(data rate)* rispettivamente di 1 e 2 Mbit/s (vedi riquadro a pagina 75), mentre i sistemi che seguono lo standard 802.11 FHSS utilizzano le modulazioni 2-GFSK e 4-GFSK[10] con una velocità trasmissiva ancora di 1 e 2 Mbit/s (vedi riquadro in questa stessa pagina).

Per i sistemi DSSS l'intera banda ISM è suddivisa in tredici canali, le cui frequenze centrali sono separate a passi di 5 MHz. Poiché l'occupazione spettrale di questi sistemi è di 22 MHz[11], i canali sono parzialmente sovrapposti tra di loro (figura 14). Per i sistemi FHSS sono invece disponibili settantanove canali[12], ciascuno di 1 MHz, suddivisi in tre insiemi di ventisei canali, costituiti in modo da evitare periodi di collisione prolungati tra due sequenze di hopping appartenenti a ciascun insieme.

Lo standard 802.11b [13] costituisce l'evoluzione dell'802.11 DSSS ed è quello realizzato su tutti gli apparati oggi offerti dal mercato. Anch'esso utilizza la tecnica DSSS che, combinata con lo schema di modulazione *CCK (Complementary Code Keying)*, consente una velocità trasmissiva di 11 Mbit/s mantenendo la

(10) *Questo schema di modulazione è opzionale.*

(11) *In ambito IEEE l'ampiezza di banda definita è quella compresa tra i due zeri del lobo principale.*

(12) *La banda disponibile è compresa tra 2402 e 2480 MHz.*

La tecnica CCK (Complementary Code Keying)

La tecnica *CCK (Complementary Code Keying)*, proposta in normativa da Harris Semiconductor e Lucent Technologies, è stata adottata nello standard IEEE 802.11b per consentire velocità di trasmissione di 11 e di 5,5 Mbit/s, grazie alla possibilità che essa offre di interoperare con la tecnologia DSSS a 1 e 2 Mbit/s mantenendo in particolare la stessa banda.

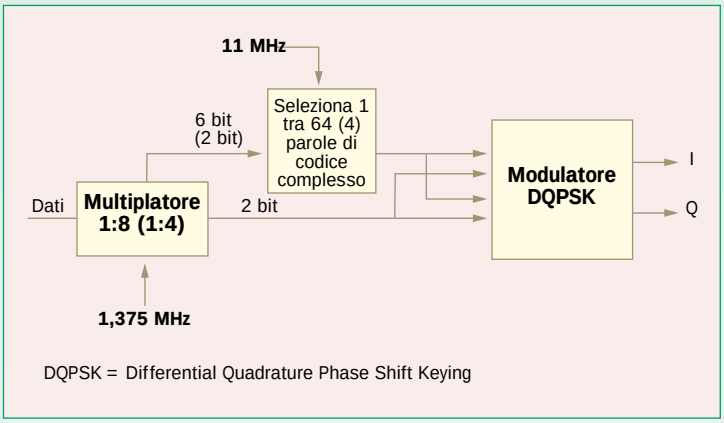


Figura A *Schema a blocchi del modulatore CCK per sistemi a 11 Mbit/s (i valori tra parentesi si riferiscono ai sistemi a 5,5 Mbit/s).*

La tecnica CCK utilizza sequenze di spreading formate da 8 chip complessi e ottenute mediante la seguente formula:

$$c=\left\{e^{j(\varphi_1+\varphi_2+\varphi_3+\varphi_4)},e^{j(\varphi_1+\varphi_3+\varphi_4)},e^{j(\varphi_1+\varphi_2+\varphi_4)},-e^{j(\varphi_1+\varphi_4)},e^{j(\varphi_1+\varphi_2+\varphi_3)},e^{j(\varphi_1+\varphi_3)},-e^{j(\varphi_1+\varphi_2)},e^{j\varphi_1}\right\}$$

I parametri delle fasi φ_1, φ_2, φ_3 e φ_4 sono determinati in modo opportuno a partire dai bit di informazione (tabella A).

CCK 11 Mbit/s	
Coppie di bit	**Parametri di fase**
(d_0, d_1)	φ_1
(d_2, d_3)	φ_2
(d_4, d_5)	φ_3
(d_6, d_7)	φ_4

Tabella A *Relazione tra i parametri di fase e le coppie di bit di informazione per i sistemi a 11 Mbit/s.*

Più precisamente i bit di informazione, generati a una velocità di 11 o di 5,5 Mbit/s, sono raggruppati in blocchi rispettivamente di 8 bit o di 4 bit, secondo la velocità di trasmissione, generando così i simboli di informazione a una velocità di 1,375 Mbaud. I due bit più significativi di ciascun blocco (d_0 e d_1) sono inviati direttamente al modulatore DQSK (figura A) e determinano il valore del parametro φ_1. I restanti sei bit (due nei sistemi a 5,5 Mbit/s) sono impiegati per selezionare la sequenza di spreading (figura A). Le coppie di bit (d_2, d_3), (d_4, d_5) e (d_6, d_7) determinano i valori assunti dai parametri di fase φ_2,φ_3 e φ_4 secondo le relazioni mostrate in tabella B e in tabella C (solo la coppia - d_2, d_3 - nei sistemi a 5,5 Mbit/s). Tra tutte le sessantaquattro sequenze disponibili (quattro nei sistemi a 5,5 Mbit/s) ne viene selezionata una sola costituita da 8 chip complessi e da una velocità di 11 Mbit/s *(operazione di spreading)*.

CCK 11 Mbit/s	
Valori delle coppie di bit	**Valori dei parametri di fase**
00	0
01	$\pi/2$
10	π
11	$3\pi/2$

Nota: Il mapping coincide con quello di un sistema QPSK con codifica binaria

Tabella B *Relazione tra i parametri di fase e i valori delle coppie di bit di informazione per i sistemi a 11 Mbit/s.*

Questa sequenza è inviata al modulatore DQPSK che, come detto, identifica la fase φ_1 in base ai valori assunti dalla coppia (d_0, d_1) (tabella D). Ai simboli dispari è aggiunta un'ulteriore rotazione di fase di 180° (π).

In ricezione l'operazione di correlazione tra il segnale ricevuto e le sessantaquattro possibili sequenze (quattro nei sistemi a 5,5 Mbit/s) consente di determinare le coppie di bit (d_2, d_3), (d_4, d_5) e (d_6, d_7) e nei sistemi a 5,5 Mbit/s la coppia (d_2, d_3) (operazione di *despreading*), mentre la fase del segnale consente di determinare la coppia di bit (d_0, d_1).

CCK 5,5 Mbit/s		
Coppia di bit		Parametro di fase
(d_0, d_1)		φ_1
bit d_i	Valore del bit d_i	Parametri di fase
d_2	0	$\varphi_2 = \pi/2$
	1	$\varphi_2 = 3\pi/2$
$\forall$ d_2, d_3		$\varphi_3 = 0$
d_3	0	$\varphi_4 = 0$
	1	$\varphi_4 = \pi$

Tabella C *Relazione tra i parametri di fase e i bit di informazione per i sistemi a 5,5 Mbit/s.*

Valori della coppia di bit (d_0, d_1)	Transizione di fase (simboli pari)	Transizione di fase (simboli dispari)
00	0	π
01	$\pi/2$	$3\pi/2$
11	π	0
10	$3\pi/2$	$\pi/2$

Tabella D *Codifica della coppia di bit (d_0, d_1) secondo lo schema DQPSK.*

stessa occupazione spettrale dei sistemi a standard 802.11 DSSS (vedi riquadro a pagina 78). La modulazione impiegata consente poi sia di variare in maniera dinamica la velocità da 5,5 a 2 o a 1 Mbit/s sulla base della qualità del collegamento radio sia di garantire la compatibilità con i sistemi 802.11 DSSS.

Lo standard 802.11a [14], definito per operare nella gamma a 5 GHz, impiega la tecnica *OFDM (Orthogonal Frequency Division Multiplexing)* per contrastare gli effetti delle riflessioni multiple. Gli schemi di modulazione e codifica (chiamati modi fisici), variano sulla base della qualità del collegamento radio e consentono di utilizzare una capacità compresa tra un minimo di 6 Mbit/s e un massimo di 54 Mbit/s (tabella 2) con passo di canalizzazione di 20 MHz[13] (vedi riquadro a pagina 80).

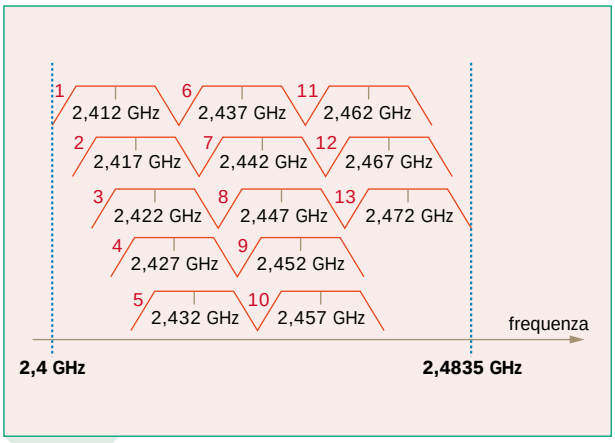


Figura 14 *Canali nella gamma a 2,4 GHz per i sistemi 802.11 DSSS e 802.11b.*

Modo	Schema di modulazione	Rate del codice per la protezione da errori	Velocità trasmissiva [Mbit/s]
1	BPSK	1/2	6
2	BPSK	3/4	9
3	QPSK	1/2	12
4	QPSK	3/4	18
5	16-QAM	1/2	24
6	16-QAM	3/4	36
7	64-QAM	2/3	48
8	64-QAM	3/4	54

N.B.: i modi 1, 3 e 5 sono obbligatori; gli altri sono opzionali

BPSK = Binary Phase Shift Keing
QAM = Quadrature Amplitude Modulation
QPSK = Quadrature Phase Shift Keing

Tabella 2 *Modi fisici della norma IEEE 802.11a.*

4.3 Attività di standardizzazione in corso

Il Gruppo di Lavoro IEEE per le Wireless LAN sta mettendo a punto nuove norme tecniche per

[13] *Più precisamente, la distanza tra le portanti centrali di due canali adiacenti è di 20 MHz, mentre la banda a 3 dB definita nella maschera di spettro in trasmissione è di 18 MHz.*

La tecnica OFDM (Orthogonal Frequency Division Multiplexing)

L'idea posta alla base della tecnica *OFDM (Orthogonal Frequency Division Multiplexing)* è quella di suddividere un flusso di dati ad alta velocità in un certo numero di flussi paralleli a velocità più bassa, ciascuno dei quali modula una sottoportante separata e distinta. Scegliendo le frequenze delle sottoportanti in modo che la loro distanza sia pari al reciproco del periodo di simbolo, le sottoportanti risultano ortogonali. Lo spettro del segnale modulato su ciascuna sottoportante presenta, così, un attraversamento dello zero in corrispondenza della frequenza delle altre sottoportanti (figura A).

Nei sistemi IEEE 802.11a e HIPERLAN/2, i bit all'uscita del codificatore sono raggruppati in blocchi di 1, 2, 4 o 6 bit cui corrispondono i simboli delle costellazioni BPSK, QPSK, 16-QAM o 64-QAM, a seconda del modo trasmissivo impiegato[1]. Ogni simbolo modula una delle quarantotto sottoportanti *(data subcarriers)* disponibili, alle quali ne sono aggiunte altre quattro *(pilot subcarriers)*, utilizzate per irrobustire il processo di demodulazione coerente nei confronti degli offset di frequenza e del rumore di fase. Un blocco di quarantotto simboli e di quattro simboli pilota costituisce un simbolo OFDM.

Per prevenire fenomeni di *ISI (InterSymbol Interference)*, ogni simbolo OFDM è preceduto da un intervallo di guardia la cui durata (0,8 μs) consente di ottenere buone prestazioni su canali con un *delay spread* di 250 ns. La separazione tra le portanti è pari a 0,3125 MHz, per una banda complessivamente occupata da 16,6 MHz.

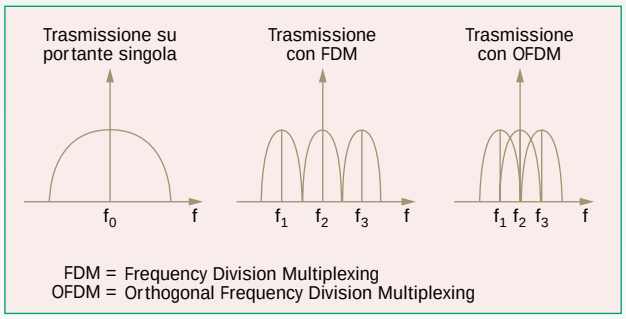


Figura A *Rappresentazione semplificata degli spettri delle tecniche FDM e OFDM.*

Negli ultimi anni la tecnica OFDM ha suscitato un interesse sempre crescente nelle comunicazioni radio per due motivi. Il primo è costituito dalle eccellenti prestazioni in presenza di affievolimenti selettivi in frequenza rispetto alle tecniche convenzionali che impiegano una portante singola in trasmissione. Questo tipo di affievolimento corrompe, infatti, solo una piccola percentuale delle portanti per cui gli errori generati dal canale possono essere facilmente recuperati con la codifica di canale. Il secondo motivo riguarda la generazione delle portanti OFDM che può essere realizzata completamente per via numerica tramite una operazione di *FFT (Fast Fourier Transform)*. I recenti progressi nella tecnologia *VLSI (Very Large Scale Integration)* hanno reso facilmente disponibili sul mercato a prezzi contenuti i chip che realizzano tale operazione.

(1) *La corrispondenza tra i bit in ciascun blocco e i simboli è realizzata secondo la codifica di Gray.*

migliorare quelle già esistenti o per aggiungere a queste ultime nuove caratteristiche.

Qui di seguito sono elencate le più importanti norme oggi allo studio:

- *IEEE 802.11e.* Aggiunge al MAC meccanismi per la gestione della *QoS (Quality of Service)* e la definizione di classi di servizio;
- *IEEE 802.11g.* Costituisce un'evoluzione dello standard 802.11b, in quanto opera sempre nella banda ISM a 2,4 GHz, ma permetterà di raggiungere velocità trasmissive fino a 54 Mbit/s;
- *IEEE 802.11h.* Aggiunge allo standard 802.11a le funzioni di *TPC* e *DFS* richieste dalla CEPT per l'impiego in Europa dei sistemi WLAN a 5 GHz;
- *IEEE 802.11i.* Migliora i meccanismi di sicurezza e di autenticazione previsti oggi dal MAC[14].

5. Tecnologia ETSI HIPERLAN/2

Lo standard ETSI HIPERLAN/2 definisce un sistema radio che può essere utilizzato per l'accesso a diverse reti dorsali *(core networks)*. Quest'impiego è reso possibile grazie a un'architettura flessibile che definisce i livelli PHY e *DLC (Data Link Control)* in modo indipendente dalla *core network* e grazie a un insieme di *Convergence Layer* che consentono l'accesso a tali reti.

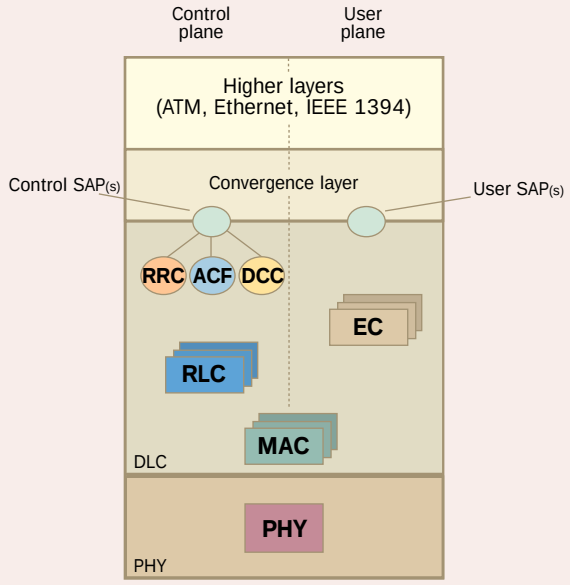
Le caratteristiche di diversi *Convergence Layer* sono

(14) *Diversi studi, il più famoso dei quali condotto presso l'Università di Berkeley [15], hanno messo in luce, negli ultimi mesi, la debolezza del meccanismo di sicurezza com'è ora definito dallo standard (WEP).*

state già normalizzate o sono in fase di definizione per l'*interworking* tra una rete d'accesso, che utilizzi la norma HIPERLAN/2, e una rete IP - Ethernet o *PPP (Point to Point Protocol)* - ATM e IEEE 1394 (figura 15).

Le principali caratteristiche dello standard HIPERLAN/2, in aggiunta a quelle già presenti nella norma IEEE 802.11, sono:

- *Protocollo orientato alla connessione.* In una rete che impieghi HIPERLAN/2 i dati sono inviati su connessioni punto-punto o punto-multipunto, instaurate prima della trasmissione, utilizzando appositi messaggi di segnalazione.
- *Meccanismi di QoS e di gestione di servizi con vincoli temporali o di errore.* A ogni connessione può essere assegnato un particolare valore di QoS, in termini, ad esempio, di banda, jitter, tasso di errore o, più semplicemente, indicando un livello di priorità. Questi meccanismi facilitano lo svolgimento di servizi, quali quello fonico o video in combinazione con quello di trasmissione di dati.
- *Handover tra celle.* Anche in HIPERLAN/2 le stazioni, durante gli spostamenti, possono effettuare misure su tutte le frequenze disponibili in modo da rilevare il miglior punto di accesso alla rete (AP). In aggiunta però, prima di poter effettuare il cambio di AP, le stazioni devono avviare una vera e propria procedura di handover attraverso la quale si ha il trasferimento dal vecchio al nuovo AP di tutte le connessioni con i relativi parametri di QoS e di sicurezza.
- *Supporto a meccanismi di DFS (Dynamic Frequency Selection).* Come richiesto dalla CEPT, il sistema assegna automaticamente a ogni cella le frequenze migliori tra quelle disponibili scelte in base al rapporto segnale-interferente. È così possi-



ATM = Asynchronous Transfer Mode
ACF = Association Control Function
DCC = DLC Connection Control
DLC = Data Link Control
EC = Error Control
IEEE = Institute of Electrical and Electronics Engineers
MAC = Medium Access Control
PHY = PHYsical layer
RRC = Radio Resource Control
RLC = Radio Link Control

Figura 15 *Modello di riferimento del protocollo HIPERLAN/2.*

La Dynamic Frequency Selection in HIPERLAN/2

Uno degli obiettivi delle specifiche tecniche di HIPERLAN/2 è quello di far sì che il sistema operi in modalità *Plug-and-Play* e senza necessità di pianificazione frequenziale ed è a questo fine che le specifiche prevedono un meccanismo di *DFS (Dynamic Frequency Selection)*. Lo scopo è quello di evitare le interferenze da parte di altri apparati, sia dello stesso tipo, sia di tipo diverso, che utilizzino lo stesso spettro di frequenze favorendone un uso il più possibile uniforme.

La funzione di DFS deve essere basata su misure di potenza di segnale sia all'AP sia al terminale associato, e in entrambi i casi sia sul proprio canale operativo, sia su altri canali. Infatti, se è vero che la selezione automatica del canale di frequenza da parte dell'AP rappresenta il primo passo all'accensione dell'apparato, è anche vero che nel tempo, per sopraggiunti motivi di interferenza, l'AP sia costretto a spostarsi dal canale inizialmente selezionato. Analogamente, possono esserci terminali che, trovandosi in particolari situazioni di interferenza, non siano più in grado di comunicare con l'AP in modo efficiente.

Nel meccanismo di DFS specificato in HIPERLAN/2, sia l'AP sia il terminale devono quindi essere in grado di effettuare misure di potenza di segnale ricevuto a una data frequenza; inoltre, per quanto detto sopra, l'attivazione della misura può avvenire sia su richiesta dell'AP, sia su iniziativa del terminale. L'algoritmo con cui l'AP effettua la scelta di cambiare non è invece specificato dallo standard, per cui ogni manifatturiera ha facoltà di realizzarne uno proprio.

bile che più reti condividano lo spettro disponibile evitando all'operatore di pianificare le frequenze. (Vedi riquadro a pagina 81).

5.1 Convergence Layer

Per consentire di normalizzare un *DLC (Data Link Control)*, indipendente dalle caratteristiche della rete alla quale il sistema HIPERLAN/2 è connesso, il *CL (Convergence Layer)* deve svolgere due funzioni: deve anzitutto trasferire e classificare le richieste di servizio dei livelli superiori sulle connessioni DLC e deve, poi, convertire i pacchetti di lunghezza fissa o variabile, provenienti dai livelli superiori, nelle *PDU (Protocol Data Unit)* utilizzate dal DLC, tramite meccanismi di segmentazione e di riassemblaggio, chiamati *SAR (Segmentation And Reassembly)*.

Per il *BRAN* sono stati definiti due tipi di *CL*: *cell-based* [16] e *packet-based* [17]. Il primo è impiegato per l'interconnessione a reti ATM [18], mentre il secondo è utilizzato per l'interconnessione a tutte le reti a pacchetto (in particolare oggi, a quelle Ethernet [19] e IEEE 1394 [20]).

La flessibilità e l'apertura verso altre tecnologie sono assicurate dalla suddivisione di ogni CL in due parti: una *common part*, comune alle diverse tecnologie, che realizza essenzialmente le funzionalità di SAR e un certo numero di *SSP (Service Specific Part)* che sono specifiche di ogni tecnologia.

5.2 Livello Data Link Control

Secondo lo standard HIPERLAN/2, l'accesso al mezzo è gestito in maniera centralizzata ed è basato sulla tecnica TDMA/TDD con una trama MAC di durata pari a 2 ms. All'interno della trama, i singoli intervalli di tempo *(time slot)* sono allocati dinamicamente sulla base delle richieste dei terminali.

La trama è suddividibile in cinque fasi (figura 16) [21]: *broadcast*, *downlink*, *direct link*[15], *uplink* e *random access*.

La *fase di broadcast* trasporta il *BCH (Broadcast CHannel)*, l'*FCH (Frame CHannel)* e l'*ACH (Access feedback CHannel)*.

Il BCH è utilizzato dall'Access Point per inviare a tutti i terminali le informazioni di base della cella (identificativo, livello di potenza,…).

L'FCH interviene nelle fasi di *downlink*, di *direct link* e di *uplink*, cioè indica quali sono i terminali autorizzati a ricevere e/o a trasmettere in quelle fasi, il tipo di contenuto dei dati - dati di utente, riscontri *(ack)*, richieste di risorse,… - e il numero di intervalli di tempo *(slot)* assegnati.

L'ACH contiene invece le risposte ai tentativi di accesso al canale effettuati dai terminali nella fase di accesso casuale *(random access)* nella trama precedente.

Le *fasi di downlink*, *uplink* e *direct link*, con durata variabile di trama in trama, hanno la stessa struttura di base e differiscono solo per la sorgente e per la destinazione delle informazioni. Nella prima, le informazioni (dati e controllo) sono inviate dall'*access point* ai terminali; nella seconda *(uplink)* sono trasmesse dai terminali all'AP. Nel *direct link*, infine, i messaggi sono scambiati tra gli stessi terminali senza il passaggio dall'AP.

Le informazioni scambiate sono organizzate in blocchi di lunghezza variabile chiamati *cell trains*. Ogni *cell train* può contenere due diversi tipi di *PDU (Protocol Data Unit)*. Le *LCH (Long transport CHannel)* PDU hanno una lunghezza di 54 byte (di questi, 48 per il payload) e trasportano essenzialmente i dati di utente veri e propri, mentre le *SCH (Short transport CHannel)* PDU hanno una lunghezza di 9 byte e veicolano solo informazioni di controllo (riscontri, richiesta di risorse, …).

La *fase di random access* è costituita da PDU, in numero variabile di trama in trama a seconda del numero di terminali presenti nella cella, chiamate *RCH (Random access CHannel)* PDU, utilizzate dai terminali per accedere alla rete la prima volta, per richiedere risorse trasmissive, ovvero durante la procedura di *handover*.

Gli RCH sono canali a contesa e utilizzano un processo di risoluzione delle collisioni simile a quello definito dallo standard IEEE 802.11. Ogni tentativo di accesso deve, perciò, ricevere una conferma positiva o negativa nella trama successiva tramite l'ACH.

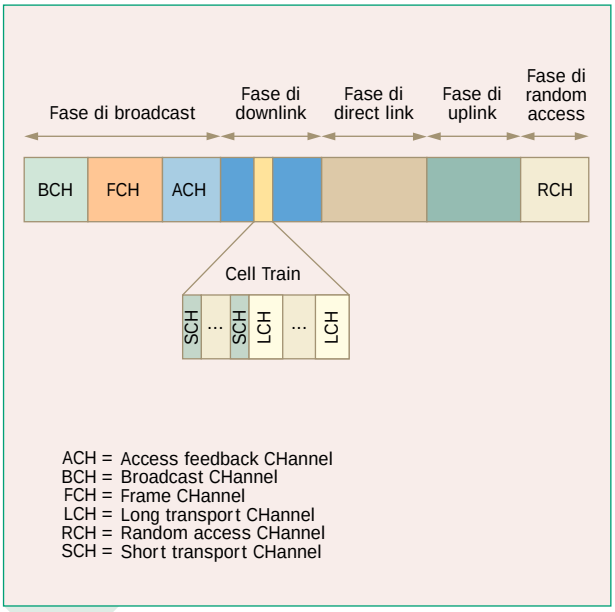


Figura 16 *Struttura della trama MAC HIPERLAN/2.*

La funzione *EC (Error Control)* effettua il rilevamento e la correzione degli errori di ricezione dovuti alla tratta radio, in aggiunta al *FEC (Forward Error Correction)*. Essa deve anche assicurare che i pacchetti siano trasmessi al CL nella sequenza corretta.

Nella modalità di funzionamento *acknowledged*, il recupero degli errori è realizzato tramite un meccanismo *ARQ (Automatic Repeat reQuest)* con il quale la sta-

(15) *La fase di direct link è presente nel tipo di funzionamento direct mode e deve essere sempre fornita nelle reti domestiche (home profile [22]). La modalità di funzionamento alternativa, chiamata centralized mode, prevede solo le fasi di downlink e di uplink e deve essere sempre utilizzata nelle reti per ambienti office (business profile [23]).*

zione ricevente notifica a quella trasmittente gli identificativi delle LCH PDU ricevute errate. La tecnica ARQ, specificata per il sistema HIPERLAN/2, è

Modo	Schema di Modulazione	Rate del codice per la protezione da errori	Velocità trasmissiva [Mbit/s]
1	BPSK	1/2	6
2	BPSK	3/4	9
3	QPSK	1/2	12
4	QPSK	3/4	18
5	16-QAM	9/16	27
6	16-QAM	3/4	36
8	64-QAM	3/4	54

N.B.:l'unico modo opzionale è quello 8.

BPSK = Binary Phase Shift Keing
QAM = Quadrature Amplitude Modulation
QPSK = Quadrature Phase Shift Keing

Tabella 3 *Modi fisici di HIPERLAN/2.*

denominata *SRPB (Selective Repeat with Partial Bitmap);* essa utilizza "mappe" i cui bit sono associati ciascuno a una determinata PDU e il cui valore ne indica la corretta o l'errata ricezione mediante, rispettivamente, il valore 1 o quello 0. Questa tecnica consente alla stazione ricevente di riscontrare più PDU con un solo messaggio ARQ.

La trasmissione dei riscontri può essere disabilitata per tutte quelle connessioni di utente che hanno stringenti vincoli temporali (modalità di funzionamento *unacknowledged*).

Il protocollo *RLC (Radio Link Control)* [24] permette di fornire un servizio di trasporto per le funzioni di *DCC (DLC Connection Control)*, *RRC (Radio Resource Control)* e *ACF (Association Control Function)*, per le funzioni, cioè, di gestione delle connessioni (apertura, modifica, chiusura), gestione delle risorse radio (selezione dinamica delle frequenze, economizzatori di consumi, *handover*, …) e per l'associazione dei terminali alla rete radio.

5.3 Livello PHY

Il *PHYsical layer* di HIPERLAN/2 [25] è simile a quello IEEE 802.11a, grazie all'attività di armonizzazione delle normative svolto dai due Enti di standardizzazione. Anche in questo caso è utilizzata la tecnica OFDM con schemi di modulazione e di codifica delle sottoportanti variabili in base alla qualità del collegamento radio; la velocità trasmissiva è compresa tra 6 e 54 Mbit/s con canali di ampiezza pari a 20 MHz. Le differenze rispetto a quanto prescritto dalla norma 802.11a, consistono unicamente nel rate del codice impiegato nel modo 5 e nell'assenza del modo 7 (tabella 3). Nello standard HIPERLAN/2, inoltre, è opzionale solo la modulazione 64-QAM.

6. Reti e pianificazione

Con le WLAN possono essere realizzate due diversi tipi di rete: *infrastructure* e *ad-hoc*.

Nelle reti del tipo *infrastructure* è sempre presente un AP, eventualmente connesso alla rete cablata (figura 17); attraverso esso transitano tutti i pacchetti, compresi quelli appartenenti a comunicazioni che riguardano due stazioni della stessa cella. In questo secondo caso l'AP svolge esclusivamente la funzione di ripetitore.

Una rete del tipo *ad-hoc* è invece costituita solo da stazioni terminali; in essa lo scambio di pacchetti avviene direttamente tra le stazioni coinvolte (figura 18). Per queste reti, se realizzate con apparati a stan-

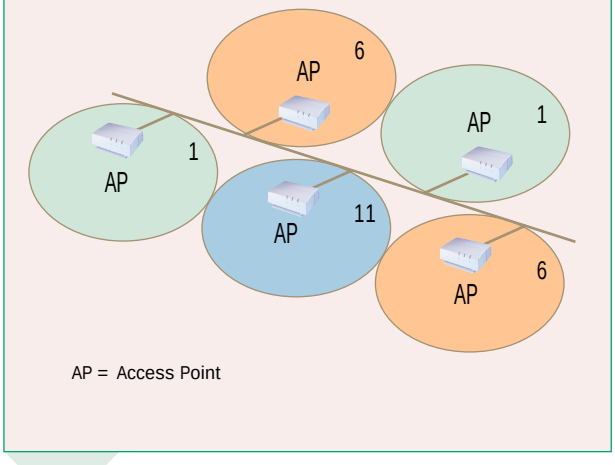


Figura 17 *Rete del tipo infrastructure.*

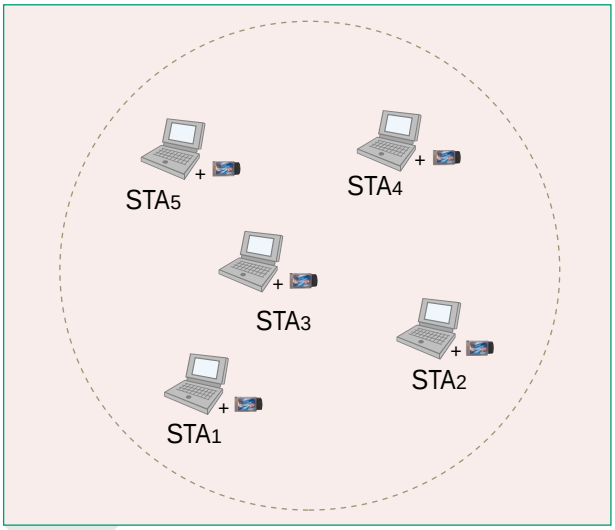


Figura 18 *Rete del tipo ad-hoc.*

dard HIPERLAN/2, è necessario che una delle stazioni, chiamata *CC (Central Controller)*, gestisca l'allocazione della banda utilizzata dalle altre.

Nell'assegnare le frequenze a una rete con apparati 802.11b occorre tener presente che i tredici canali disponibili nella banda, anche se in parte sovrapposti, possono essere raggruppati in blocchi di due o tre canali non sovrapposti (figura 14 a pagina 79). In una copertura cellulare occorre quindi impiegare i tre canali di ciascun blocco (ad esempio 1, 6 e 11) in modo da evitare interferenze tra le celle e quindi riduzioni delle prestazioni in termini di traffico smaltito e di copertura. Gli stessi criteri devono essere applicati nel caso di più operatori presenti nella stessa area.

Nella gamma all'interno dei 5 GHz, la pianificazione delle frequenze da parte dell'operatore non è necessaria in quanto essa è realizzata in modo automatico dagli apparati grazie ai meccanismi di *Dynamic Frequency Selection*. Le frequenze scelte sono quelle che consentono comunque, anche in presenza di più reti cellulari nella stessa area, di rendere massimo il rapporto segnale-interferente e quindi di migliorare le prestazioni.

7. Conclusioni

Nell'articolo sono state presentate le principali tecnologie e applicazioni relative ai sistemi WLAN. Molto alte sono oggi le aspettative di sviluppo commerciale di questi sistemi, in numerosi settori e in ambito sia pubblico *(hotspot)* sia privato *(home networking)*. Le tecnologie e gli standard disponibili sono numerosi e diversi in termini sia di prestazioni sia di gamme di frequenza impiegate.

Oggi e nell'immediato futuro, lo standard dominante è senz'altro l'IEEE 802.11b operante nella gamma a 2,4 GHz e con una velocità trasmissiva fino a 11 Mbit/s, accompagnato, nel futuro più prossimo, dalle diverse evoluzioni per esso previste.

Appare invece aperta, soprattutto in Europa, la sfida tra gli standard HIPERLAN/2 e IEEE 802.11a, operanti entrambi nella gamma a 5 GHz e con una velocità trasmissiva fino a 54 Mbit/s, anche se al momento l'IEEE 802.11a sembra essere favorito, non fosse altro che per la disponibilità sul mercato di apparati commerciali ormai numerosi, seppure in versione rispondente alle normative americane.

Tuttavia, grazie alla presenza nello standard HIPERLAN/2 di tecniche per la gestione della qualità del servizio, questi sistemi riusciranno molto verosimilmente a ritagliarsi settori di applicazione specifici in cui sia richiesto di fornire servizi con requisiti di qualità sui ritardi e sulla banda, quali i servizi video e voce, integrati ai servizi dati.

Un ulteriore fattore determinante sarà, infine, rappresentato dal prezzo a regime degli apparati e, in particolare, di quelli destinati agli utilizzatori privati, oggi non ancora prevedibile.

Bibliografia

[1] *WLAN Chipset Market - The Incredible Journey Is Just Beginning*. In-Stat/MDR, marzo 2002.

[2] *Life, Liberty and WLANs: Wireless Networking Brings Freedom to the Enterprise*. Cahners In-Stat/MDR, novembre 2001.

[3] *Wireless LANs and the threat to Mobile Revenues*. BWCS, 2001.

[4] Gaarder, K.; Skolt, E.: *H2U - a joint prestudy by Telenor Mobil, Telenor R&D, and Ericsson*. IST Mobile Summit 2001, Barcellona (Spagna), 9-12 settembre 2001.

[5] http://www.wi-fi.com/

[6] http://www.bluetooth.org/

[7] Gerla, M.; Zanella, A.: *Bluetooth: una nuova tecnologia per reti radio personali*. «Notiziario Tecnico Telecom Italia», Anno 10, n. 2, settembre 2001, pp. 82-96.

[8] http://www.homerf.org/

[9] *ERC Decision of 12 March 2001 on harmonised frequencies, technical characteristics and exemption from individual licensing of Short Range Devices used for Radio Local Area Networks (RLANs) operating in the frequency band 2400 - 2483.5 MHz*. ERC/DEC/(01)07, marzo 2001.

[10] *ERC Decision of 29 November 1999 on the harmonised frequency bands to be designated for the introduction of High Performance Radio Local Area Networks (HIPERLANs)*. ERC/DEC/(99)23, novembre 1999.

[11] *IEEE Standard for Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks - Specific requirements - Part 2: Logical Link Control*. IEEE 802 LAN/MAN Standards Committee, IEEE Std 802.2-1998, 1998.

[12] *IEEE Standards for Information technology -- Telecommunications and information exchange between systems -- Local and metropolitan area network -- Specific requirements -- Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*. IEEE 802 LAN/MAN Standards Committee, IEEE Std 802.11-1999, 1999.

[13] *Supplement to IEEE Standard for Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks - Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Higher - Speed Physical Layer Extension in the 2.4 GHz Band*. IEEE 802 LAN/MAN Standards Committee, IEEE Std 802.11b-1999, 1999.

[14] *Supplement to IEEE Standard for Information*

technology - Telecommunications and information exchange between systems - Local and metropolitan area networks - Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications - High-speed Physical Layer in the 5 GHz Band. IEEE 802 LAN/MAN Standards Committee, IEEE Std 802.11a-1999, 1999.

[15] N. Borisov, I. Goldberg, D. Wagner: *Intercepting Mobile Communications: The Insecurity of 802.11*, 7th Annual International Conference on Mobile Computing and Networking, Roma, giugno 2001.

[16] *Broadband Radio Access Network (BRAN); HIPERLAN Type 2; Cell based Convergence Layer; Part 1: Common Part*. ETSI BRAN, TS 101 763-1 v1.1.1, aprile 2000.

[17] *Broadband Radio Access Network (BRAN); HIPERLAN Type 2; Packet based Convergence Layer; Part 1: Common Part*. ETSI BRAN, TS 101 493-1 v1.1.1, aprile 2000.

[18] *Broadband Radio Access Network (BRAN); HIPERLAN Type 2; Cell based Convergence Layer; Part 2: UNI Service Specific Convergence Sublayer (SSCS)*. ETSI BRAN, TS 101 763-2 v1.1.1, aprile 2000.

[19] *Broadband Radio Access Network (BRAN); HIPERLAN Type 2; Packet based Convergence Layer; Part 2: Ethernet Service Specific Convergence Sublayer (SSCS)*. ETSI BRAN, TS 101 493-2 v1.1.1, aprile 2000.

[20] *Broadband Radio Access Network (BRAN); HIPERLAN Type 2; Packet based convergence layer; Part 3: IEEE 1394 Service Specific Convergence Sublayer (SSCS)*. ETSI BRAN, TS 101 493-3 v1.1.1, settembre 2000.

[21] *Broadband Radio Access Network (BRAN); HIPERLAN Type 2; Data Link Control (DLC) Layer; Part 1: Basic Data Transport Functions*. ETSI BRAN, TS 101 761-1 v1.2.1, novembre 2000.

[22] *Broadband Radio Access Network (BRAN); HIPERLAN Type 2; Data Link Control (DLC) Layer; Part 4: Profile for Home Environment*. ETSI BRAN, TS 101 761-5, luglio 2001.

[23] *Broadband Radio Access Network (BRAN); HIPERLAN Type 2; Data Link Control (DLC) Layer; Part 3: Profile for Business Environment*. ETSI BRAN, TS 101 761-3, settembre 2000.

[24] *Broadband Radio Access Network (BRAN); HIPERLAN Type 2; Data Link Control (DLC) Layer; Part 2: Radio Link Control (RLC) sublayer*. ETSI BRAN, TS 101 761-2 v1.2.1, aprile 2001.

[25] *Broadband Radio Access Network (BRAN); HIPERLAN Type 2; Physical (PHY) layer*. ETSI BRAN, TS 101 475 v1.2.2, febbraio 2001.

[26] *Electromagnetic compatibility and Radio spectrum Matters (ERM);Wideband Transmission systems; Data transmission equipment operating in the 2,4 GHz ISM band and using spread spectrum modulation techniques; Part 1: Technical characteristics and test conditions*. ETSI, EN 300 328-1 v1.3.1, dicembre 2001.

Abbreviazioni

3GPP	3rd Generation Partnership Project
ACF	Association Control Function
ACH	Access feedback CHannel
ADPCM	Adaptive Differential Pulse Code Modualation
AP	Access Point
ARQ	Automatic Repeat reQuest
ATM	Asynchronous Transfer Mode
BCH	Broadcast CHannel
BPSK	Binary Phase Shift Keying
BRAN	Broadband Radio Access Network
CC	Central Controller
CCK	Complementary Code Keying
CEPT	Conférence Européenne des Administrations des Postes et des Télécommunications
CFP	Contention-Free Period
CL	Convergence layer
CPCS	Common Part Convergence Sublayer
CSMA/CA	Carrier Sense Multiple Access/Collision Avoidance
CSMA/CD	Carrier Sense Multiple Access/Collision Detection
DBPSK	Differential Binary Phase Shift Keying
DCC	DLC Connection Control
DFS	Dynamic Frequency Selection
DIFS	Distributed InterFrame Space
DLC	Data Link Control
DPSK	Differential Phase Shift Keying
DQPSK	Differential Quadrature Phase Shift Keying
DSSS	Direct Sequence Spread Spectrum
EC	Error Control
e.i.r.p.	equivalent isotropically radiated power
ERC	European Radiocommunications Committee
ETSI	European Telecommunications Standards Institute
FCC	Federal Communications Commission
FCH	Frame CHannel
FEC	Forward Error Correction

FFT	Fast Fourier Transform
FHSS	Frequency Hopping Spread Spectrum
FSK	Frequency Shift Keying
GFSK	Gaussian Frequency Shift Keying
HIPERLAN/2	HIgh PErformance Radio Local Area Network Type 2
HomePNA	Home Phoneline Networking Alliance
HomeRF	Home Radio Frequency
IEEE	Institute of Electrical and Electronics Engineers
IP	Internet Protocol
IR	InfraRed
ISI	InterSymbol Interference
ISM	Industrial Scientific and Medical
ISO/OSI	International Standard Organization/Open System Interconnection
ISP	Internet Service Provider
ITU	International Telecommunication Union
LAN	Local Area Network
LCH	Long transport CHannel
LLC	Logical Link Control
MAC	Medium Access Control
OFDM	Orthogonal Frequency Division Multiplexing
PC	Point Coordinator
PCF	Point Coordination Function
PDA	Personal Digital Assistant
PDU	Protocol Data Unit
PHY	PHYsical layer
PIFS	Priority InterFrame Space
PLC	PowerLine Communications
POTS	Plain Old Telephone Service
PPP	Point to Point Protocol
RES	Radio Equipment and Systems
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RCH	Random access CHannel
RF	Radio Frequency
RLC	Radio Link Control
RRC	Radio Resource Control
SAP	Service Access Point
SAR	Segmentation And Reassembly
SCH	Short transport CHannel
SIFS	Short InterFrame Space
SIG	Special Interest Group
SME	Small Medium Enterprise
SOHO	Small Office Home Office
SRD	Short Range Devices
SRPB	Selective Repeat with Partial Bitmap
SSCS	Service Specific Convergence Sublayer
SSP	Service Specific Part
SWAP	Shared Wireless Access Protocol
TCP	Transmission Control Protocol
TDD	Time Division Duplexing
TDMA	Time Division Multiple Access
TPC	Transmit Power Control
UMTS	Universal Mobile Telecommunications System
UNI	User Network Interface
U-NII	Unlicensed-National Information Infrastructure
UTP	Unshielded/Unscreened Twisted Pair
VLSI	Very Large Scale Integration
WECA	Wireless Ethernet Compatibility Alliance
WEP	Wired Equivalent Privacy
Wi-Fi	Wireless-Fidelity
WLAN	Wireless Local Area Network
WLL	Wireless Local Loop
xDSL	x Digital Subscriber Line

Giovanna D'Aria ha conseguito la laurea con lode in Ingegneria Elettronica presso il Politecnico di Torino nel luglio 1984. Dall'ottobre dello stesso anno lavora in Telecom Italia Lab (già CSELT), dove svolge attività di ricerca nel settore dei sistemi radio, sia fissi sia mobili, partecipando a numerosi gruppi di normativa e di ricerca nazionali e internazionali. Attualmente è coinvolta nel settore Home and Office Networks, dove si interessa di sistemi radio nella rete d'accesso e di WLAN. È autore di numerose pubblicazioni e inventore di un brevetto nel campo delle trasmissioni numeriche. Ha partecipato alla organizzazione di conferenze ed eventi scientifici, tra cui l'*ECRR (European Conference on Fixed Wireless Systems and Networks)* di cui, da numerose edizioni, è membro del Comitato Scientifico.

Massimo Colonna ha conseguito la laurea in Ingegneria Elettronica nel luglio 1997 presso il Politecnico di Torino. Dal settembre dello stesso anno è in Telecom Italia Lab (già CSELT) dove, nell'ambito dell'area Home and Office Networks, svolge attività di ricerca sui sistemi d'accesso radio per servizi a larga banda e sulle tecnologie e i sistemi per wireless LAN, relativamente agli aspetti legati alle prestazioni, al test su apparati, al progetto delle coperture, al dimensionamento di rete e alle valutazioni economiche.

Il Sistema GPS

DUILIO CORATELLA

Il sistema GPS consente di determinare con elevata precisione la posizione in tre dimensioni di un qualsiasi punto della superficie terrestre. Oltre agli scopi militari per cui è nato, esso permette lo sviluppo di innumerevoli applicazioni nei più diversi settori. Il presente articolo si pone l'obiettivo di descrivere in quale modo sia possibile effettuare misure con precisione che, in certe condizioni e con particolari tecniche, può essere addirittura centimetrica a partire da satelliti distanti oltre 20mila chilometri dalla superficie terrestre.

Dopo una breve analisi dei diversi segmenti (spaziale, di controllo, di utente), nell'articolo è descritto il segnale GPS emesso dai satelliti e il funzionamento di un ricevitore. Si analizzano quindi le diverse fonti di errore intrinseco che inficiano l'accuratezza raggiungibile e si descrive il principio su cui si basa la tecnica di correzione differenziale, che consente di eliminare o di attenuare l'effetto di alcuni errori incrementando la precisione del posizionamento.

Tra gli operatori radiomobili, TIM ha di recente avvertito l'esigenza di pianificare la copertura cellulare sul territorio con un'accuratezza sempre maggiore, ricorrendo al GPS e alla tecnica di correzione differenziale per determinare le coordinate delle proprie antenne.

In un prossimo articolo saranno illustrate la rete di stazioni GPS di riferimento TIM e l'architettura di sistema in grado di erogare i dati di correzione differenziale GPS (utili per topografi, geologi, ricercatori, aziende municipalizzate, aziende per la realizzazione di grandi infrastrutture, ...).

In ulteriori articoli sarà trattato il tema della localizzazione per reti radiomobili, descrivendo le principali tecniche oggi standardizzate in ambito ETSI e 3GPP.

1. Introduzione

Il GPS, o più precisamente il sistema *GPS - NAVSTAR (Global Positioning System - NAVigation Satellite Timing And Ranging)* è nato negli Stati Uniti negli anni Settanta come sistema militare su progetto della *US Navy*, poi sviluppato dal *DoD (Department of Defence)* americano attraverso il *GPS Joint Program Office* della *USAF Space Division*.

Il sistema consente di determinare le coordinate di un qualsiasi punto sulla superficie terrestre, in qualsiasi istante, e con qualsiasi condizione atmosferica, purché un numero sufficiente di satelliti sia in visibilità radio (in realtà esso funziona fino a quote di qualche decina di km). In più per tale punto, esso fornisce anche un riferimento temporale assoluto (tempo GPS) con un livello di precisione molto elevata. Il sistema è oggi gestito dal DoD, che oltre a effettuare il controllo e la manutenzione della rete satellitare, stabilisce con quali precisioni il sistema è fruibile (gratuitamente) per applicazioni civili.

Come ogni sistema satellitare, il sistema GPS è suddiviso in tre blocchi principali o segmenti: *spaziale*, di *controllo* e di *utente* [1].

Il *segmento spaziale* (figura 1) è basato su una costellazione di ventiquattro satelliti, disposti su sei orbite quasi circolari inclinate di 55° sul piano

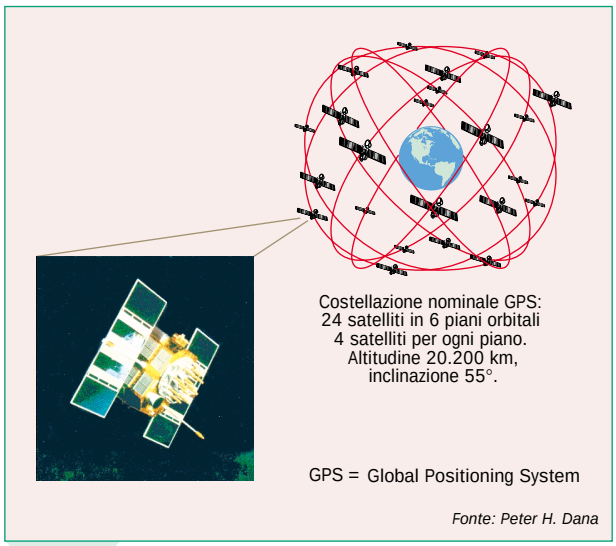


Figura 1 *Schema della costellazione GPS e immagine di un satellite.*

equatoriale, a intervalli di longitudine di 60°, e ad un'altezza (media) di circa 20.200 km dalla superficie terrestre.

Il periodo di rivoluzione di ciascun satellite è pari a 11 h 58 min 2,0455 s, ovvero ogni satellite compie due orbite complete in poco meno di un giorno solare, in modo che, per un punto qualsiasi della Terra, tutta la costellazione si ripresenta quotidianamente con un anticipo di un po' meno di quattro minuti rispetto al giorno precedente. I parametri orbitali adottati fanno sì che, in ogni istante e in ogni luogo, nell'ipotesi di assenza di ostacoli, siano visibili almeno quattro satelliti. I livelli ricevuti al suolo sono però ottimizzati per angoli di elevazione intorno ai 40° sull'orizzonte (come sarà mostrato più avanti, nella figura 6 del paragrafo 2).

I satelliti oggi in orbita, realizzati dalla *Lockheed Martin*, sono stati lanciati in date diverse. Essi hanno dimensioni di circa 130x180x200 cm e un peso di circa 800 kg. Utilizzano pannelli solari con una superficie complessiva di circa sette metri quadrati, in grado di fornire una potenza di circa 700 W. Dal momento che le misure di distanza effettuate dai ricevitori a terra sono ricavate da misure di intervalli temporali, gli orologi a bordo dei satelliti sono estremamente precisi (costituiscono cioè un livello primario di riferimento). Ciascun satellite ha a bordo quattro orologi atomici (due al cesio e due al rubidio) aventi stabilità media pari a una parte su 10^{12}; ciò comporta che essi perdono un secondo ogni 10^{12} secondi, ovvero ogni 317mila anni circa. Il GPS costituisce quindi un sistema di sincronizzazione di livello superiore a quello stabilito dalla specifica ITU-T G.811 per i *PRC (Primary Reference Clocks)* [7], che è pari a una parte su 10^{11}. Per la trasmissione si utilizzano due trasmettitori in banda L.

La stabilizzazione dei satelliti è assicurata da sistemi giroscopici. A bordo sono anche presenti motori in grado di imprimere alle orbite piccole correzioni lungo tre assi ortogonali. Le scorte di carburante sui satelliti limitano la vita media a non più di 7÷8 anni.

Il *segmento di controllo* verifica lo stato di funzionamento dei satelliti e ne aggiorna le relative orbite. È costituito da una serie di stazioni di terra distribuite lungo la fascia equatoriale[1], come riportato in figura 2:

- cinque *Monitor Station* per il controllo dei satelliti, situate a Colorado Springs (Colorado, Stati Uniti), Isole Hawaii e Isola Kwajalein (Oceano Pacifico), Isola di Ascensiòn (Oceano Atlantico) ed Isola Diego Garcia (Oceano Indiano);
- tre *Upload Station* per la trasmissione in banda S verso i satelliti dei comandi di controllo e delle informazioni da inserire nel *Navigation Message* (descritto nel paragrafo 2) destinato ai ricevitori; esse sono co-locate con le Monitor Station delle isole di Ascensiòn, Diego Garcia e Kwajalein;
- una *Backup Station* situata a Sunnyvale (Stati Uniti);
- una *Master Control Station* a Colorado Springs (Colorado, Stati Uniti), per il controllo dell'intero sistema.

I satelliti sono soggetti a una forza di gravità che è circa il sei per cento di quella al suolo, e ruotano con

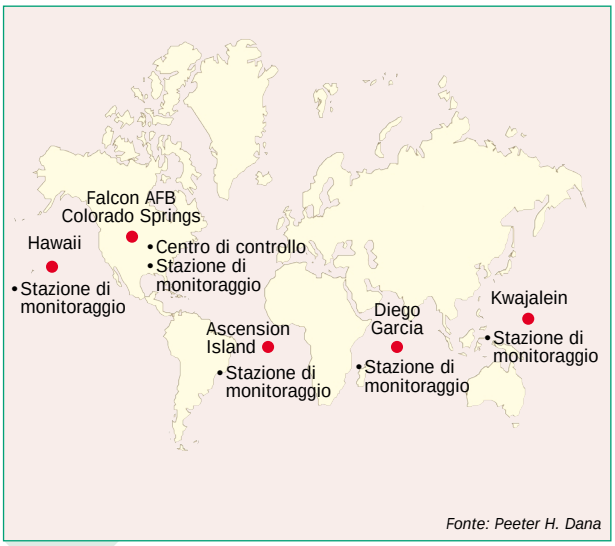


Figura 2 *Le stazioni di terra del sistema GPS.*

una velocità tangenziale pari a circa 3,9 km/s, ovvero di circa 14.040 km/h (valore circa dodici volte superiore alla velocità tangenziale di rotazione della terra). Essendo il periodo di rivoluzione di ciascun satellite

(1) *Alla latitudine 0° il satellite passa per il punto con un'elevazione maggiore sull'orizzonte ed è quindi visibile (e di conseguenza monitorabile) per un intervallo di tempo maggiore che ad altre latitudini.*

MISURARE PER CONOSCERE: LA STRATEGICITA' DEL SISTEMA GPS

Da sempre l'uomo ha avvertito la necessità di misurare l'ambiente che lo circonda. Nelle varie epoche storiche la cartografia ha infatti assunto una fondamentale importanza per lo sviluppo delle attività umane. Più in generale, la scienza si fonda sul principio in base al quale per comprendere la realtà (un qualsiasi fenomeno o una grandezza fisica) è necessario caratterizzarla attraverso la definizione di unità di misura, costruendo una metrica e definendo una metodologia di misura.

Il sistema GPS - nato circa trenta anni fa negli Stati Uniti sulla base di specifiche esigenze militari - consente di determinare con straordinaria precisione (errore anche inferiore al centimetro) la localizzazione di un qualsiasi punto della superficie terrestre, e permette per la prima volta di definire un sistema di riferimento, sia spaziale che temporale, valido in tutto il mondo. Per queste caratteristiche il sistema ha velocemente catalizzato un forte interesse specie dal punto di vista commerciale.

Recentemente, sulla base delle pressioni esercitate dalle imprese che utilizzano questo sistema per i più svariati usi civili e, soprattutto, a seguito della conclusione della guerra fredda, il governo degli Stati Uniti ha consentito di migliorare, in misura significativa, la precisione del posizionamento ottenibile di un ricevitore mediante sistemi GPS per uso civile, anche di basso costo.

È, perciò, prevedibile nel breve termine un'ulteriore rapida crescita del mercato dei servizi e dei prodotti relativi alle misure di posizionamento.

pari a circa dodici ore, le *Monitor Station* equatoriali determinano due volte al giorno posizione, altezza, velocità e orbita dei satelliti, e le trasmettono alla *Master Station* situata a Colorado Springs (Colorado). Qui, il sistema di controllo elabora i dati di correzione delle *effemeridi* (descritte nel paragrafo 2), che sono poi inviati ai satelliti dalle tre *Upload Station*; gli stessi dati sono inseriti nel messaggio di navigazione (descritto nel paragrafo 2) e sono rispediti dai satelliti verso tutti i ricevitori, per le opportune correzioni.

Il *segmento di utente* è costituito dall'insieme dei ricevitori GPS terrestri.

In figura 3 è mostrato uno schema tipico del ricevitore che è costituito dai seguenti elementi fondamentali:

- antenna omnidirezionale;
- orologio (oscillatore al quarzo);
- unità di generazione dei codici (copia delle sequenze emesse da ciascun satellite, da usare per la correlazione con i segnali ricevuti);
- unità di elaborazione, costituita da un microprocessore che elabora in tempo reale i segnali ricevuti per determinare la posizione del punto e che, eventualmente, apporta le correzioni differenziali;
- unità di memorizzazione dei dati per le successive post-elaborazioni;
- unità di alimentazione (costituita da batterie o da un alimentatore esterno).

Nel paragrafo 2 saranno descritti i segnali GPS e il modo con cui il ricevitore sincronizza il proprio orologio interno al tempo GPS, mentre nei paragrafi 3 e 4 sarà trattato in dettaglio il funzionamento del ricevitore.

2. Il segnale GPS

Ciascun satellite invia due segnali L1 ed L2, modulati sulle frequenze portanti multiple di quella fondamentale degli oscillatori atomici di bordo (f_0 = 10,23 MHz):

$f_{L1} = 154 \times f_0 = 1.575{,}42$ MHz ($\lambda_1 = 19{,}05$ cm)
$f_{L2} = 120 \times f_0 = 1.227{,}60$ MHz ($\lambda_2 = 24{,}45$ cm)

Secondo la specifica [5] i segnali sono compresi in due bande di ampiezza pari a 20,46 MHz, centrate su L1 ed L2, e sono polarizzati circolarmente *RHCP (Right-Hand Circularly Polarized)*.

La tecnica di modulazione è del tipo *spread spectrum* a modulazione di fase, che consente la

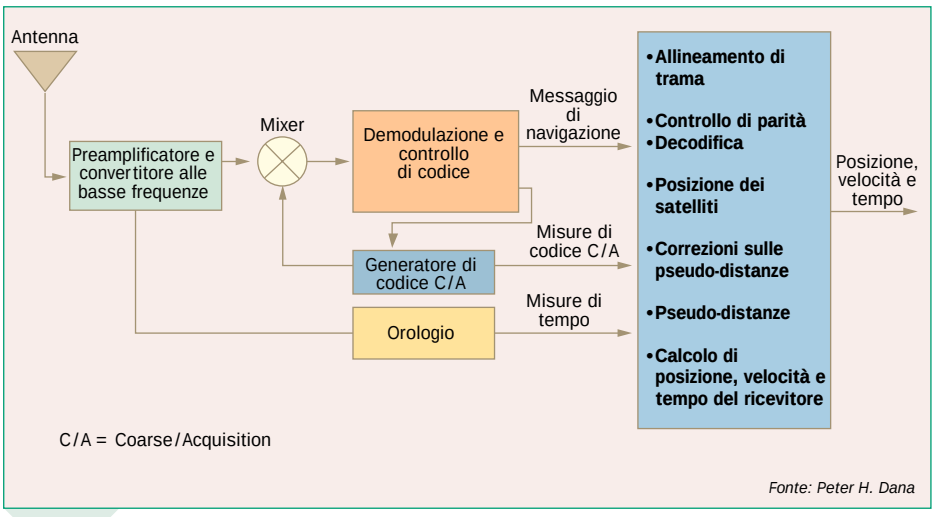


Figura 3 *Schema di un ricevitore GPS.*

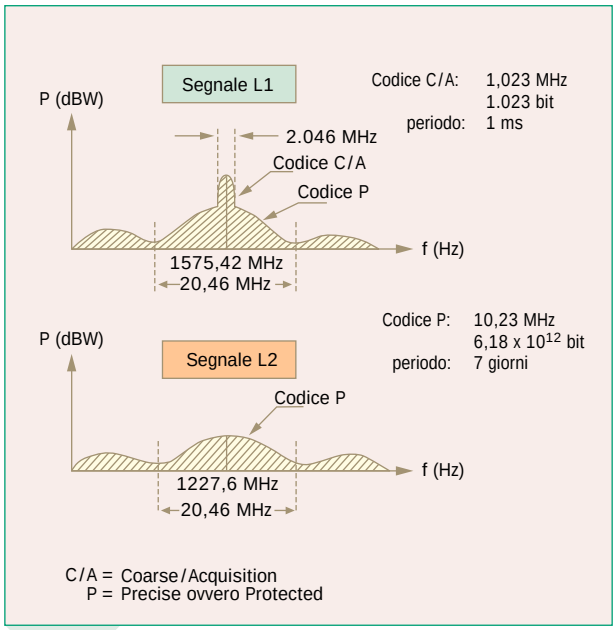


Figura 4 *Spettro dei segnali GPS.*

decodifica del segnale utile anche quando esso è totalmente immerso nel rumore. In figura 4 è mostrato lo spettro dei due segnali L1 ed L2.

Ogni satellite emette due sequenze pseudocasuali *PRN (Pseudo Random Noise)* diverse *(codice Gold)*[2], che consentono anche di identificare ciascun satellite, più un messaggio informativo.

Queste sequenze sono:

- *C/A (Coarse/Acquisition)*, di pubblico dominio: sequenza di 1.023 bit che si ripete con una cadenza di 1 ms a una velocità di cifra pari a 1,023 Mbit/s ($f_{C/A} = f_0/10 = 1{,}023$ MHz), utilizzato nell'*SPS (Standard Positioning System);*
- *P (Precise o Protected)*, cifrabile (nel qual caso è chiamata Y) e usata solo a scopi militari: sequenza di $6{,}187104 \times 10^{12}$ bit che si ripete ogni sette giorni con una velocità di cifra pari a 10,23 Mbit/s ($f_P = f_0 = 10{,}23$ MHz), utilizzato nel *PPS (Precise Positioning System).*
- *NAVigation Message*: sequenza di 1.500 bit, trasmessa con una frequenza di 50 bit/s, contenente informazioni utili al ricevitore (descritto successivamente).

La tecnica di modulazione utilizzata è la *BPSK (Binary Phase Shift Keying)*[3]. Lo schema di principio della modulazione è riportato in figura 5 [1].

Il segnale portante L1 è modulato da due sequenze *(bit train)*: la prima costituita dalla somma modulo due del codice P (o Y) e del *navigation message* e l'altra costituita dalla somma modulo due del codice C/A e del *navigation message*. Il segnale portante L2 è invece modulato dalla sola sequenza ottenuta come somma modulo due del codice P (o Y) e del *navigation message.*

I livelli minimi dei segnali che da specifica [5] devono essere garantiti in ricezione sono indicati nella tabella 1.

Essi variano in funzione dell'angolo di elevazione del satellite sull'orizzonte, in modo da avere un massimo intorno ai 40°, come riportato in figura 6. I livelli massimi in ricezione non superano generalmente i –153,0 dBW per il codice civile C/A e –155,5 dBW per il codice protetto P (anche nella sua versione cifrata Y) sulla portante L1, e –158,0 dBW per entrambi i codici sulla portante L2 [5].

Il *NAVigation message* è suddiviso in cinque sottosequenze di 300 bit ciascuna, che contengono le seguenti informazioni [5]:

- dati per la correzione dell'errore dovuto all'offset degli orologi atomici di bordo (sottosequenza 1);
- dati delle *effemeridi* - ovvero delle funzioni matematiche che descrivono le orbite dei satelliti con elevata precisione, consentendo al ricevitore di conoscere in anticipo quanti e quali satelliti sono visibili a una certa ora in un determinato luogo - valide per diverse ore (sottosequenze 2 e 3);

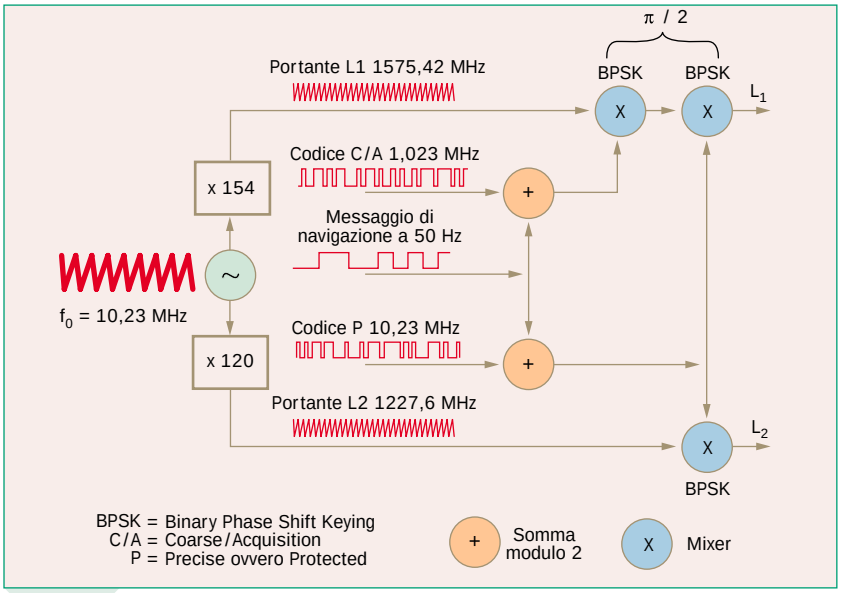


Figura 5 *Schema di modulazione dei segnali GPS.*

(2) *Il codice Gold è un codice generato mediante la somma modulo-due di due sequenze spread spectrum.*

(3) *La trasmissione BPSK avviene modulando la fase della portante con un segnale binario: la fase resta invariata alla trasmissione di un 1, mentre è invertita per la trasmissione di uno 0 (o viceversa).*

canale	livello minimo di potenza in ricezione [dBW]	
	C/A	P (Y)
L1	– 160,0	– 163,0
L2	– 166,0	– 166,0

C/A = Coarse/Acquisition
L = portante
P = codice Preciso ovvero Protetto
Y = versione cifrata del codice Protetto (P)

Tabella 1 *Livelli minimi in ricezione per i segnali GPS.*

- altri dati, tra cui i parametri per la correzione dell'errore dovuto ai ritardi ionosferici e le informazioni temporali *UTC (Universal Time Coordinated)* (sottosequenza 4);
- l'*almanacco*, che rappresenta una versione semplificata delle effemeridi, valida solo per poche ore, ma che consente tra l'altro di individuare i codici PRN di ciascun satellite senza dover procedere per tentativi (sottosequenza 5).

Le sottosequenze 4 e 5 sono poi strutturate ciascuna in venticinque pagine diverse di dati (analogamente alle sottopagine di una pagina del televideo). Per decodificare l'intero *navigation message* è necessario un tracciamento continuo per almeno 30 s (1.500 bit a 50 bit/s).

A titolo esemplificativo, vediamo il processo di trasformazione eseguito dal ricevitore per ricavare il tempo UTC mediante l'uso dei parametri contenuti nel *navigation message* (figura 7) [1, 5]. Il tempo UTC viene utilizzato per marcare le posizioni (coordinate) individuate dal ricevitore, per cui ogni rilievo di posizione è determinato dalla

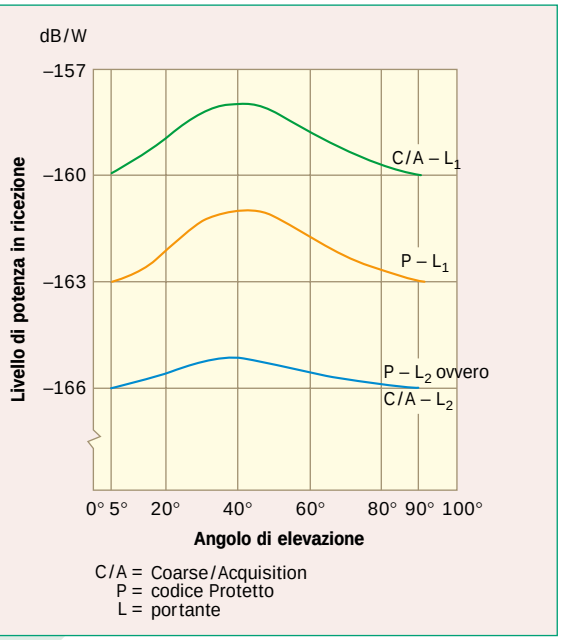


Figura 6 *Livelli minimi dei segnali GPS in ricezione.*

terna di coordinate latitudine, longitudine e quota, più il tempo UTC della misura[4]. Inoltre, in tal modo il ricevitore sincronizza il suo orologio interno.

Il monitoraggio del riferimento temporale nel sistema GPS è demandato al segmento di controllo. Il tempo di sistema è riferito al tempo zero UTC, definito dall'*USNO (United States Naval Observatory)* coincidente con la mezzanotte del 5 gennaio 1980. L'informazione di tempo elementare è costituita da un contatore (a 19 bit) definito *TOW (Time Of Week)*, che conta da 0 a 403.199 unità elementari in cui è suddivisa una settimana[5].

Nella sottosequenza 1 del *navigation message*, il satellite invia al ricevitore questo contatore e

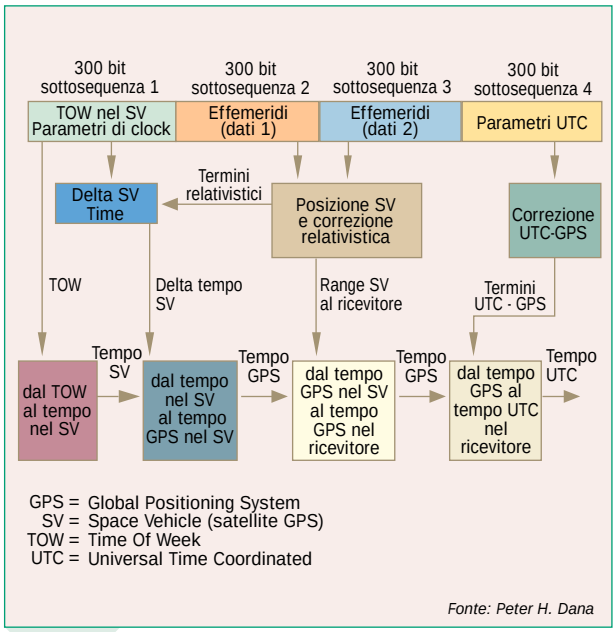


Figura 7 *Determinazione del tempo GPS e del tempo UTC.*

alcuni fattori di correzione (trasmessi dalle *upload station* a ciascun satellite, come descritto nel paragrafo 1), che gli consentono di determinare il tempo del satellite. Per valutare il tempo GPS il ricevitore utilizza ulteriori termini di correzione (legati alla posizione e all'orbita dei satelliti ed agli effetti relativistici dovuti al moto relativo satellite-ricevitore solidale con la terra) contenuti nelle sottosequenze 2 e 3 del *navigation message*. Infine, per passare dal tempo GPS (che è un

(4) *L'informazione temporale è utilizzata anche nel caso in cui sulla misura rilevata debba essere applicata la correzione differenziale, descritta successivamente nel paragrafo 6.*

(5) *L'unità elementare citata è la sottosequenza X1, di lunghezza pari a 15.345.000 bit, che viene sommata modulo 2 all'ulteriore sottosequenza X2$_i$ per generare il codice P (che si ripete ogni settimana). 403.200 di tali sottosequenze durano 6,187104 x 10^{12} bit, che è la lunghezza della sequenza P.*

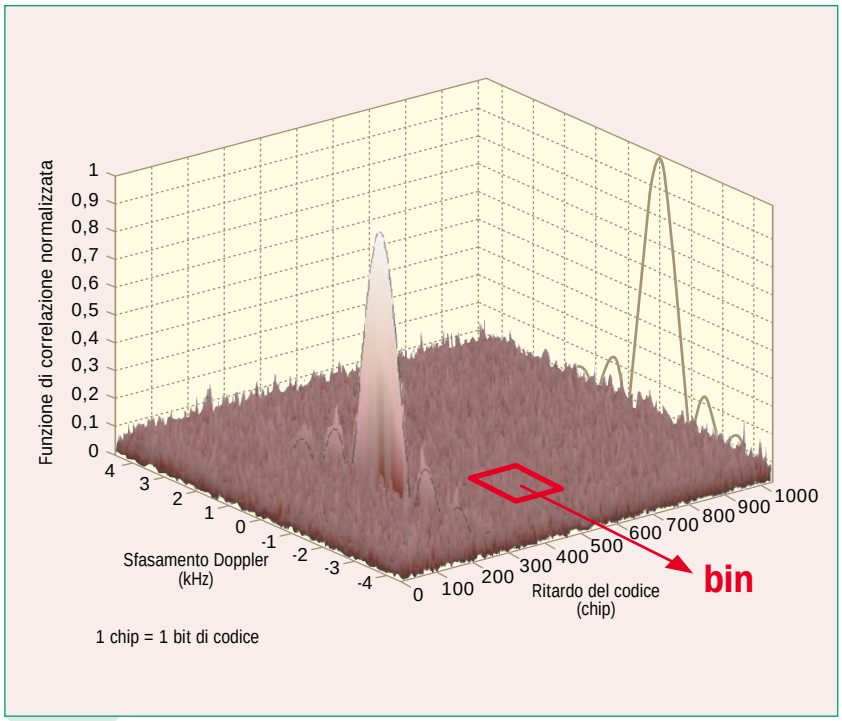


Figura 8 *Spazio di ricerca del segnale GPS per l'acquisizione iniziale.*

tempo continuo) al tempo UTC (che come già detto viene periodicamente corretto), il ricevitore utilizza le informazioni di correzione, contenute nella sottosequenza 4 nella pagina 18 del messaggio, che consentono di raggiungere un errore massimo di 90 ns. I due riferimenti temporali sono attualmente sfasati tra di loro di circa 15 s.

3. Fasi di acquisizione e di tracciamento

All'accensione, prima di effettuare il *fixing* (ovvero prima di determinare la propria posizione) un ricevitore GPS deve acquisire il maggior numero di satelliti della costellazione. Nel caso di accensione dopo molto tempo o quando operi in un'area geografica assai diversa da quella in cui operava prima dello spegnimento, il ricevitore non dispone di informazioni utili per ottimizzare la ricerca.

Esso deve perciò:

- cercare tutti i satelliti in visibilità effettuando una correlazione tra i codici pseudocasuali ricevuti e le repliche dei codici stessi generate localmente;
- demodulare il *navigation message* di un satellite agganciato, per ottenere alcuni parametri quali: il riferimento temporale, le posizioni orbitali (almanacco, effemeridi), le correzioni ionosferiche da applicare. Deve anche mantenere nel tempo i satelliti agganciati.

La catena di ricezione contiene un certo numero di canali (tipicamente non inferiore a dodici) e un numero elevato di correlatori che operano in parallelo e che effettuano la correlazione tra il codice pseudocasuale ricevuto da ciascun satellite e una copia generata localmente. Un rivelatore di picco consente di determinare il valore massimo della funzione di correlazione e di agganciare il satellite.

La ricerca è effettuata contemporaneamente sulla finestra dei *code delays* (ovvero su un periodo temporale pari a 1023 tempi di bit) e su quella degli scostamenti di frequenza (e cioè su una banda pari a 8,4 kHz, generata dall'effetto Doppler causato dal moto dei satelliti e dalla rotazione terrestre), impiegando blocchi elementari detti *bins* (figura 8) [6].

Se si suddivide lo spazio delle frequenze di 8,4 kHz in 40 bins e si analizza ciascun bin, ad esempio, in un tempo medio di 1 ms, nel peggiore dei casi (esplorazione completa di tutta l'area di ricerca) l'acquisizione dei satelliti richiede circa 40 s.

Dopo l'acquisizione, il ricevitore demodula il *navigation message* per determinare sia i parametri necessari per le successive acquisizioni e per il tracciamento *(tracking)*, sia i riferimenti temporali per effettuare le misure di pseudodistanza *(pseudo-range)*, necessarie per determinare la posizione *(fixing)*.

Il tempo complessivo necessario per effettuare il *fixing* dipende quindi dal livello di aggiornamento dei dati del *navigation message*. Esso si definisce come *TTFF (Time To First Fix)* e dipende dal *TSLF (Time Since Last Fix)*.

Nella tabella 2 sono indicate le durate tipiche riportate in letteratura, ottenute come valori medi per un ricevitore GPS di classe intermedia.

4. Determinazione della posizione di un punto (fixing): misure di codice e misure di fase

Il *fixing* è ottenuto a partire dalla misura delle distanze del ricevitore dal maggior numero di satelliti in visibilità. Per valutare queste distanze è necessario misurare il cosiddetto *TOA (Time Of Arrival)*, ovvero il tempo che il segnale radio, emesso dal satellite, impiega a raggiungere il ricevitore in un preciso istante temporale.

tipo di avvio	TSLF	TTFF
snap start	< 30 min	2÷3 s
hot start	30 min < TLSF < 2÷3 h	8÷10 s
cold start	> 2÷3 h	> 20 s

TSLF = Time Since Last Fix
TTFF = Time To First Fix

Tabella 2 *Valori tipici del Time to First Fix, al variare del Time Since Last Fix.*

La distanza può essere determinata in due modi: mediante una misura di pseudodistanza oppure con una misura di fase. Nel primo caso si utilizza lo stesso codice pseudocasuale utilizzato per l'acquisizione, e si considerano l'istante di ricezione del picco di correlazione più le informazioni decodificate dal *navigation message* che consentono di risalire all'istante di emissione del segnale da ciascun satellite. La differenza temporale è moltiplicata per la velocità della luce nel mezzo, che, da specifica [5], risulta essere

$$v = 299.792.458 \text{ m/s}$$

Si ottengono perciò equazioni del tipo:

$$\sqrt{\left[X_S(t_0)-X_P\right]^2+\left[Y_S(t_0)-Y_P\right]^2+\left[Z_S(t_0)-Z_P\right]^2}=\int_{t_0}^{t} v\cdot d\tau \quad (1)$$

dove

- X_s, Y_s, Z_s sono le coordinate del satellite i-esimo;
- X_p, Y_p, Z_p sono le coordinate incognite del punto;
- $t - t_0$ (intervallo di integrazione) è il tempo che il segnale impiega per arrivare dal satellite al ricevitore.

Nell'equazione (1) è stato utilizzato l'integrale della velocità nel tempo invece del semplice prodotto della velocità per il tempo, perché la velocità del segnale non è costante (cioè pari alla velocità della luce nel vuoto), ma varia con le caratteristiche fisiche degli strati atmosferici da esso attraversati.

L'istante t che compare nella relazione precedente rappresenta l'istante di arrivo del segnale a terra misurato dall'orologio del satellite. In realtà il ricevitore effettua la misura all'istante $t' = t + \Delta t$ dove Δt è lo sfasamento tra l'orologio integrato al suo interno e quello del satellite. Infatti, nonostante la sincronizzazione (descritta nel paragrafo 2), l'orologio del ricevitore essendo di basso costo è soggetto a deriva *(drift)*.

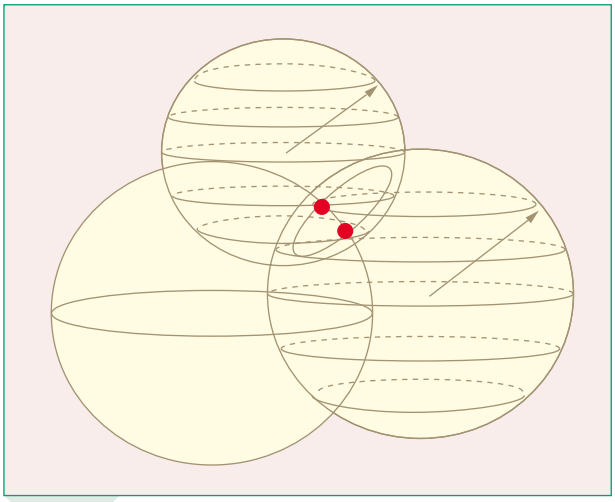


Figura 9 *Determinazione del punto come intersezione di tre sfere.*

Per determinare la posizione di un punto nello spazio le incognite sono perciò quattro: le tre coordinate del punto X_p, Y_p e Z_p oltre allo sfasamento dt. Sono quindi necessarie quattro equazioni del tipo (1), ovvero occorre effettuare misure da almeno quattro satelliti. Nel caso in cui sia sufficiente determinare la posizione sul piano, sono sufficienti tre equazioni.

Il punto è determinato, in maniera intuitiva, come l'intersezione di sfere aventi come centro la

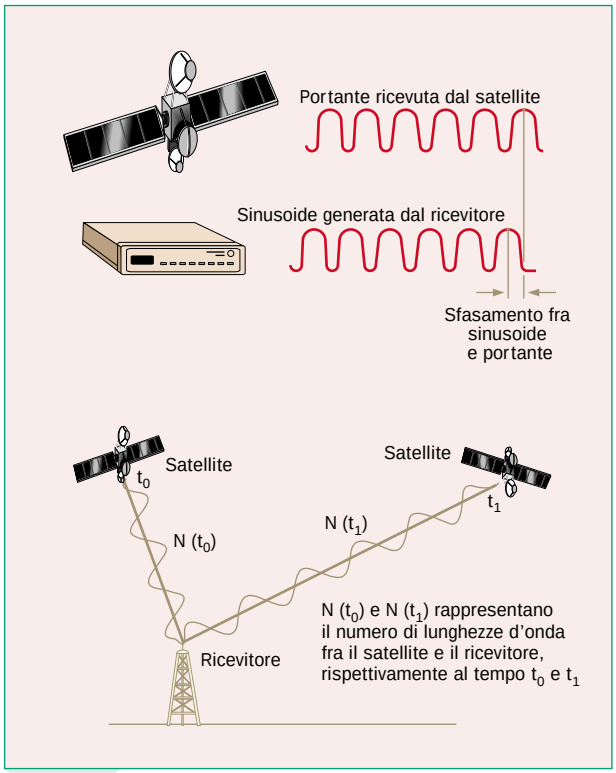


Figura 10 *Misure di fase.*

posizione dei satelliti utilizzati; le misure ottenute dal quarto satellite servono anche per discriminare tra i due punti che si otterebbero con tre satelliti (figura 9).

Nei casi pratici, anche per effetto degli errori, la soluzione è determinata solo in maniera approssimativa. I ricevitori sono in grado di ricevere i segnali provenienti da più di quattro satelliti (ricevitori con otto o con dodici canali), per cui si ottiene un sistema sovradimensionato di n equazioni (n > 4) con quattro incognite, la cui soluzione è ottenuta utilizzando il *metodo dei minimi quadrati*.

La precisione teorica ottenibile con questo metodo (prescindendo dagli errori sistematici, che sono esaminati nel paragrafo successivo) è limitata dalla velocità di cifra del segnale con cui si opera.

Nel caso si utilizzi il codice *C/A (Coarse / Acquisition)*, la cui velocità di cifra è pari a 1,023 Mbit/s, la durata di un bit è pari a circa 97,75 µs, tempo in cui alla velocità della luce il segnale percorre circa 293 m. È possibile d'altra parte misurare lo sfasamento temporale dei due segnali con precisione pari all'1

PRINCIPIO DI FUNZIONAMENTO DEL GPS

All'accensione, un ricevitore GPS effettua una ricerca dei satelliti nel proprio campo visibile, cercando di sintonizzarsi sui segnali da essi trasmessi. La ricerca è molto veloce nel caso in cui esso disponga di alcune informazioni ricevute dagli stessi satelliti (nel corso di una precedente sessione di misura), opportunamente memorizzate e periodicamente aggiornate.

Per determinare un punto dello spazio il ricevitore deve disporre dei segnali di almeno quattro satelliti (le incognite sono le tre dimensioni spaziali più il tempo). Qualora sia sufficiente una misura nel piano, si possono utilizzare i segnali provenienti da tre satelliti. A livello applicativo sono, inoltre, disponibili programmi software e/o strumenti inerziali (ad esempio l'odometro) che consentono di tracciare posizione, velocità e accelerazione, sopperendo a momentanee interruzioni nella ricezione dei segnali dai satelliti.

Tutti i ricevitori GPS determinano la propria posizione misurando le pseudodistanze dai satelliti; questi valori sono ricavati misurando le differenze temporali tra gli istanti di generazione dei segnali da parte di ciascun satellite (noti a valle della sincronizzazione iniziale degli orologi) e quelli di ricezione. I ricevitori più sofisticati sono in grado di misurare anche lo sfasamento tra i segnali portanti emessi dai satelliti e le repliche degli stessi segnali generati localmente nel ricevitore.

Le accuratezze raggiungibili variano a seconda della complessità e del costo del ricevitore. Dopo la soppressione di un particolare errore (descritto al paragrafo 5.1) inserito dal Governo degli Stati Uniti (proprietario del sistema GPS), oggi anche i ricevitori più semplici (del costo di qualche centinaio di €) consentono di raggiungere un'accuratezza pari a circa 15 – 20 m nel 67 per cento dei casi.

I ricevitori più costosi (ad esempio per applicazioni geodetiche) consentono accuratezze dell'ordine del centimetro (ottenute mediante correzione differenziale).

Il principio della correzione differenziale, utilizzato anche in elettronica, consente di ridurre gli effetti di elementi di errore che compaiono con lo stesso segno e con ampiezza uguale o simile su due misure (*termini di modo comune*). Sottraendo tra loro le misure, questi termini scompaiono ovvero sono drasticamente attenuati.

per cento della durata di un bit. L'errore teorico di posizione - prescindendo, come si è detto, dagli effetti degli errori sistematici - è quindi di circa 2,93 m.

In alternativa all'uso dei codici, per la valutazione della distanza tra il satellite e il punto da rilevare, possono essere utilizzate le misure effettuate sulla fase dell'onda portante del segnale.

Il sistema si basa sul confronto tra la fase della portante generata dal satellite e ricevuta a terra, e quella di un analogo segnale generato dal ricevitore. Lo scostamento rilevato è la frazione di lunghezza d'onda tra i due segnali, mentre il numero intero di cicli d'onda costituisce un'ulteriore incognita, definita *ambiguità iniziale*, che è determinata con altri metodi, ad esempio mediante le misure di pseudodistanza (figura 10).

Miscelando le due portanti si ottiene una serie di battimenti, la cui fase corrisponde alla differenza di fase delle due portanti.

In ciascun istante t può essere impiegata un'equazione del tipo:

$$\frac{\sqrt{\left[X_S(t_0)-X_P\right]^2+\left[Y_S(t_0)-Y_P\right]^2+\left[Z_S(t_0)-Z_P\right]^2}}{\lambda}=F(t)+N+f\cdot d_t \qquad (2)$$

dove, nel secondo membro dell'uguaglianza (che esprime la differenza misurata tra le fasi dei segnali del satellite e del ricevitore), *F(t)* è la fase dei battimenti, *N* è l'ambiguità iniziale, *f* è la frequenza, *dt* è lo sfasamento tra i due orologi (sul satellite e nel ricevitore).

La precisione teorica (sempre al netto degli effetti degli errori intrinseci) è legata in questo caso alla capacità di discriminare l'inizio di un ciclo (ovvero, ad esempio, il passaggio per lo zero della portante). Supponendo di poter distinguere un centesimo della lunghezza d'onda (che come già detto al paragrafo 2 è pari a circa 19 cm) si ottiene un'accuratezza pari a circa 2 mm.

5. Analisi degli errori presenti nel sistema GPS

Le accuratezze finora indicate sono valori solo teorici, che non tengono in considerazione l'effetto di un certo numero di errori. Alcuni di essi sono legati a fenomeni fisici, altri sono invece dovuti a limitazioni tecnologiche.

Nei punti che seguono sono descritte brevemente le maggiori fonti di errore ed è indicata la tecnica eventualmente utilizzabile per eliminarli o, quanto meno, per ridurne gli effetti [2]. Nel punto 5.8 è indicato, a titolo di esempio, il valore medio dell'errore complessivo che inficia l'accuratezza di un tipico ricevitore GPS per impieghi civili o per quelli militari. Nel paragrafo 6 è spiegato il principio della tecnica di correzione differenziale.

5.1 Errori nella determinazione della posizione dei satelliti e "Selective Availability"

Fino al 30 aprile 2000 il *DoD (Department of Defence)* statunitense ha limitato la precisione ottenibile in tempo reale dai ricevitori di uso civile, inserendo nel messaggio di navigazione alcuni errori *ad hoc* nel valore delle effemeridi.

Questi errori, che influiscono solo sulla determinazione delle misure di pseudodistanza mediante il codice C/A, sono noti con l'acronimo *S/A (Selective Availability)*. A partire dal 1° maggio 2000, su mandato dell'allora Presidente degli Stati Uniti, Bill Clinton, il DoD ha soppresso questo errore, riservandosi comunque la facoltà di reinserirlo, anche su scala locale, qualora lo ritenesse opportuno (in occasione ad esempio di crisi internazionale o di conflitto). In ogni caso, questi errori sono di tipo sistematico e rientrano nella classe degli errori cosiddetti di *modo comune*, eliminabili cioè utilizzando tecniche differenziali.

5.2 Offset degli orologi dei satelliti e del ricevitore

L'intera costellazione satellitare, per poter effettuare il *fixing,* deve essere sincronizzata sul cosiddetto *GPS time*, correggendo il *ritardo temporale relativistico* dovuto all'elevata velocità di rotazione dei satelliti rispetto al suolo (14.040 km/h).

Gli orologi atomici di bordo (oscillatori al cesio e al rubidio) hanno una stabilità pari a circa una parte su 10^{12} su un arco temporale di 24 ore. L'errore temporale accumulato in un giorno è quindi di circa $8{,}64 \times 10^{-8}$s, intervallo di tempo in cui il segnale radio percorre circa 26 m. Nell'algoritmo che consente di determinare le suddette correzioni si ipotizza poi che le orbite dei satelliti siano perfettamente circolari; in realtà il raggio presenta invece un'escursione massima di circa il 2 per cento (ovvero circa 400 km). L'errore non compensato sugli orologi atomici di bordo può arrivare perciò a circa 46 ns.

Anche le misure temporali possono essere alterate dal DoD mediante l'introduzione di un errore (S/A), che comunque si elimina con le tecniche differenziali.

Gli orologi inseriti nei ricevitori non sono molto costosi e quindi precisi, per cui l'errore temporale da essi introdotto è maggiore rispetto a quello sopra descritto. Anche questo errore è, però, di modo comune e quindi è eliminabile con tecniche differenziali.

5.3 Ritardi dovuti a riflessioni ionosferiche

Nell'attraversare la ionosfera il segnale è curvato a causa di riflessioni microscopiche. L'effetto è variabile sia nel breve termine (fluttuazioni istantanee) sia nel medio e nel lungo termine (variazioni stagionali).

La *divergenza ionosferica* provoca i seguenti effetti:

- la velocità del segnale radio diminuisce in proporzione al numero di elettroni liberi presenti nella ionosfera; il ritardo risulta direttamente proporzionale all'inverso del quadrato della frequenza della portante;
- la fase della portante è invece anticipata della stessa entità.

Il fenomeno, poi, non è uniforme nel tempo e differisce a seconda delle latitudini: normalmente la ionosfera si comporta in modo stabile nel tempo nelle zone temperate, mentre ha caratteristiche radioelettriche molto variabili in prossimità dell'equatore e dei poli.

Una maniera per eliminare, o quantomeno per ridurre, l'errore consiste nel definire il rallentamento del segnale dovuto a condizioni ionosferiche medie per ciascuna zona della terra e per ciascun periodo di tempo, e nell'inviare i parametri medi nella sottotrama quattro del *navigation message*.

Questi parametri sono utilizzati dai ricevitori di minor costo per correggere i ritardi quando effettuano misure della posizione in modo assoluto (non differenziale). Se si utilizza, invece, la tecnica differenziale è possibile ridurre in modo più soddisfacente questi errori, purché la distanza tra il ricevitore, di cui si vuol conoscere la posizione, e il ricevitore di riferimento non superi i 20÷30 km (supponendo costanti in tale area le caratteristiche ionosferiche).

Ricevitori più sofisticati detti *a doppia frequenza* sono in grado di correggere da soli in modo soddisfacente i ritardi in oggetto in quanto possono effettuare misure su entrambe le frequenze f_{L1} ed f_{L2}. Il rallentamento di un'onda elettromagnetica che attraversa la ionosfera è infatti inversamente proporzionale al quadrato della frequenza, per cui, confrontando i tempi di arrivo dei due segnali, si risale al rallentamento che ciascuno di essi ha subito.

In questo caso [5], detto quindi PR il valore di *pseudodistanza corretto* da determinare, e PR_{L1} e PR_{L2} le misure di pseudodistanza effettuate alle due frequenze f_{L1} ed f_{L2}, vale la relazione:

$$PR = \frac{PR_{L2} - \left(\frac{f_{L1}}{f_{L2}}\right)^2 \cdot PR_{L1}}{1 - \left(\frac{f_{L1}}{f_{L2}}\right)^2} \qquad (3)$$

5.4 Ritardi dovuti a rifrazioni troposferiche

Nell'attraversare la troposfera, il segnale è rallentato a causa di numerose microscopiche rifrazioni con le particelle che la compongono. Anche questi effetti sono variabili sia nel breve termine sia nel medio e nel lungo termine (variazioni stagionali) e dipendono dalle escursioni di temperatura, umidità e pressione.

L'errore, analogamente a quello ionosferico, si attenua considerando il rallentamento del segnale dovuto a condizioni troposferiche medie per ciascuna zona della terra e per ciascun periodo di tempo, e inviando i parametri medi nella sottotrama quattro del *navigation message*. In alternativa l'errore è eliminabile con la tecnica differenziale.

5.5 Percorsi multipli in ricezione

L'errore dovuto ai percorsi multipli *(multipath)* è quello più difficilmente eliminabile.

Il segnale trasmesso dai satelliti subisce, infatti, una serie di riflessioni e rifrazioni indotte dagli

ACCURATEZZA DEL GPS DAL PUNTO DI VISTA STATISTICO

Gli errori che influenzano la determinazione della posizione non sono costanti ma variabili nel tempo. Nel valutare perciò la precisione di uno strumento GPS occorre fornirne l'accuratezza, che deve essere però sempre associata a un certo *intervallo di confidenza* [3].

In un generico esperimento ripetibile per la misurazione di una grandezza fisica, il risultato ottenuto è sempre affetto da errori intrinseci (oggettivi e soggettivi), che consentono solo di stimare un valore atteso o presunto della grandezza stessa. Si associa in pratica al valore stimato un intervallo di confidenza, oppure si esprime la percentuale di volte in cui, nel caso di successive ripetizioni dell'esperimento di misura, si ottiene statisticamente il valore indicato.

A questo scopo sono necessarie le seguenti ipotesi:

a) La densità di probabilità (ddp) dell'errore in ciascuna delle tre dimensioni è gaussiana.

Quest'ipotesi non è vera per l'errore di *selective availability* introdotto artificialmente che è deterministico (paragrafo 5.1). Gli errori residui per misure differenziali (paragrafo 6) sono gaussiani se le misure sono mediate su un intervallo temporale sufficientemente lungo; per brevi intervalli di misura (qualche minuto) è invece dominante il contributo di errore dovuto ai percorsi multipli in ricezione, che è quasi costante.

b) La densità di probabilità dell'errore sul piano orizzontale è circolare.

Se l'errore sui due assi del piano è gaussiano, la *ddp* bidimensionale è una superficie a forma di campana le cui sezioni piane sono in genere delle ellissi. Solo nel caso in cui le gaussiane lungo i due assi presentano le stesse deviazioni standard, la sezione piana della superficie è una circonferenza.

In queste ipotesi, per la grandezza misurata x si fornisce abitualmente il valore:

$$x = \mu \pm \sigma$$

dove μ è il valore medio della curva gaussiana e σ rappresenta la deviazione standard. Le misure si addensano cioè intorno all'asse di simmetria della densità di probabilità gaussiana che rappresenta il valore medio μ e, in circa il 67 per cento dei casi, esse sono comprese tra (μ - σ) e (μ + σ).

La precisione ottenibile deve essere perciò riferita alla percentuale di volte in cui mediamente la precisione è raggiunta, e può essere espressa in termini di:

- **RMS** — **deviazione standard, ovvero radice quadrata della media degli errori quadratici, nota anche come 1s;**
- **2DRMS** — **due volte *(twice Distance)* il valore della lunghezza espressa in RMS, usato anche per indicare misure bidimensionali;**

ostacoli che si ritrovano sul cammino diretto verso il ricevitore. Esso quindi si scompone in più segnali che seguono percorsi diversi e che giungono al ricevitore in istanti anch'essi differenti.

Nel caso di misure statiche e non in tempo reale, l'errore può essere ridotto facendo crescere la durata della rilevazione, in modo da eliminare statisticamente i contributi delle componenti di segnale più deboli (che quindi hanno subito un maggior numero di riflessioni).

Per le stazioni fisse si adoperano antenne di buona qualità che abbiano una scarsa sensibilità per bassi angoli di incidenza dei segnali, e che utilizzano un piano di massa o un *choke ring*[6]. Questa

[6] *Il segnale ricevuto è composto da una componente diretta e una riflessa, relativa ai percorsi multipli. Un piano di massa* choke ring *è realizzato mediante diversi anelli conduttori concentrici che circondano il centro di fase dell'antenna, consentendo di cancellare (o attenuare) la componente riflessa. La cancellazione è solitamente più efficace per sorgenti di segnale prossime allo zenith (verticali), mentre è minima se prossime all'orizzonte.*

•CEP raggio di un cerchio centrato nell'antenna del ricevitore, contenente il 50 per cento dei punti su cui mediare per ottenere la misura sul piano;

•R95 raggio di un cerchio centrato nell'antenna del ricevitore, contenente il 95 per cento dei punti su cui mediare per ottenere la misura sul piano;

•SEP raggio di una sfera centrato nell'antenna del ricevitore, contenente il 50 per cento dei punti su cui mediare per ottenere la misura nello spazio.

Nella maggior parte dei casi la precisione di un ricevitore GPS è indicata in termini di *RMS (Root Mean Square)* sul piano orizzontale, o come *CEP (Circular Error Probable)*. Alcune volte si trovano precisioni espresse in *2DRMS*. Le misure possono essere inoltre mono-, bi- o tridimensionali, a seconda che si voglia determinare solo una quota, una posizione orizzontale su di un piano, o una posizione nello spazio.

Nella tabella A si riportano i principali tipi di precisione con cui si caratterizza uno strumento GPS.

Nella tabella B si riportano i fattori moltiplicativi, riportati in letteratura, per la conversione tra questi tipi di misura, determinati statisticamente sulla base di un campione di circa due milioni di punti rilevati mediante GPS con correzione differenziale e validi per circa il 62 per cento dei casi. Per la loro determinazione sono state considerate le seguenti ipotesi: errori lineari gaussiani; ddp circolare nel piano; PDOP/HDOP = 2,1 e VDOP/HDOP = 1,9.

dimensioni	tipo di precisione	probabilità %	utilizzo tipico
1	RMS	68	misure di quote
2	CEP	50	misure sul piano
2	RMS	63÷68	misure sul piano
2	R95	95	misure sul piano
2	2DRMS	95÷98	misure sul piano
3	RMS	61÷68	misure tridimensionali
3	SEP	50	misure tridimensionali

2DRMS = twice Distance RMS (oppure 2 Dimensional RMS)
CEP = Circular Error Probable
RMS = Root Mean Square
SEP = Spherical Error Probable

Tabella A *Principali tipi di precisione con cui si caratterizzano le misure con uno strumento GPS.*

	a RMS di quota	a CEP sul piano	a RMS sul piano	a R95 sul piano	a 2DRMS sul piano	a RMS in 3-D	a SEP in 3-D
da RMS di quota	1	0,44	0,53	0,91	1,1	1,1	0,88
da CEP sul piano	2,27	1	1,2	2,1	2,4	2,5	2,0
da RMS sul piano	1,89	0,83	1	1,7	2	2,1	1,7
da R95 sul piano	1,1	0,48	0,59	1	1,2	1,2	0,96
da 2DRMS sul piano	0,91	0,42	0,5	0,83	1	1,1	0,85
da RMS in 3-D	0,91	0,4	0,48	0,83	0,91	1	0,79
da SEP in 3-D	1,14	0,5	0,59	1,04	1,18	1,27	1

2DRMS = twice Distance RMS oppure 2 Dimensional RMS
CEP = Circular Error Probable
RMS = Root Mean Square
SEP = Spherical Error Probable

Tabella B *Fattori moltiplicativi per la conversione tra diversi tipi di misura.*

soluzione consente di schermare le riflessioni dal suolo; inoltre, per eliminare i segnali riflessi che giungono al ricevitore con angoli elevati occorre posizionare il ricevitore lontano dagli eventuali ostacoli.

Gli ostacoli vicini, qualora siano costituiti da buoni conduttori, possono infatti far spostare il centro di fase dell'apparato di ricezione. In casi estremi essi possono causare errori superiori a 15 m.

Per i ricevitori mobili possono essere utilizzate correzioni standard, quando siano noti i comportamenti del ricevitore in presenza di *segnali multipath minimi* (condizioni ottimali) o *elevati* (condizioni standard).

5.6 Rumore al ricevitore

Anche gli strati più bassi dell'atmosfera, densi di vapore acqueo, determinano ulteriori perturbazioni, tanto maggiori quanto più spesso è lo strato da attraversare, cioè quanto minore risulta l'angolo di elevazione del satellite (i ricevitori GPS

scartano in modo automatico i segnali provenienti da satelliti con angoli di elevazione minori di 10° sull'orizzonte). Il contributo dovuto a questa perturbazione è definito rumore in ingresso al ricevitore e si presenta sia nelle misure di pseudodistanza sia in quelle di fase ed è difficilmente eliminabile.

In questa fonte di errore ricade anche quello legato alle caratteristiche tecnologiche dello stesso ricevitore. L'uso in passato di soli uno o due canali per l'aggancio di più satelliti rendeva, ad esempio, più lunga e complessa l'acquisizione, con perdita di sensibilità del ricevitore. Oggi normalmente un ricevitore commerciale dispone di otto o dodici canali di *tracking* e di una batteria di correlatori che consentono di rilevare più velocemente picchi di massima correlazione tra il segnale ricevuto e quello generato localmente.

Anche il software limitato dei vecchi ricevitori, basati su microprocessori a otto bit, ha contribuito in passato a limitare le prestazioni: con misure di distanze di oltre 20mila km è richiesta, infatti, una precisione di almeno dieci cifre. L'uso di chip molto più evoluti consente oggi di contenere questo tipo di errore al di sotto del metro.

5.7 Errori dovuti alla disposizione geometrica dei satelliti nel campo visibile

Gli errori finora descritti sono quelli di natura fisica o tecnologica, i cui contributi si sommano tra loro e pregiudicano l'accuratezza di ogni singola misura che il ricevitore effettua sul segnale di ogni satellite.

È presente, però, anche un errore legato alla modalità di esecuzione del rilievo (posizionamento), che si basa sulla triangolazione di più misure.

Come avviene per tutte le misure ottenute per triangolazione, la precisione del posizionamento mediante GPS è limitata dalla disposizione nel campo visibile dei satelliti monitorati dal ricevitore.

Il parametro *GDOP (Geometric Dilution Of Precision)* indica in ogni istante la bontà della configurazione satellitare e costituisce un fattore moltiplicativo del valore quadratico medio dei singoli errori.

Il caso di un GDOP pari a uno corrisponderebbe all'evento (solo ideale) in cui i satelliti fossero distribuiti in tutte le possibili direzioni dello spazio. In realtà per un ricevitore posto su un qualsiasi punto della superficie terrestre, anche su mare aperto, non è possibile osservare i satelliti nel semispazio al di sotto di esso, per cui i migliori valori di GDOP sono difficilmente inferiori a due (questo è anche uno dei motivi per cui la precisione planimetrica ottenibile è sempre maggiore della precisione altimetrica).

Satelliti che siano visti sotto un basso angolo di elevazione non consentono poi misure precise, a causa del maggiore rallentamento del segnale dovuto all'attraversamento di strati atmosferici più spessi. Se i satelliti, nel corso dei moti orbitali da essi descritti e a causa degli ostacoli che limitano la visibilità ottica del ricevitore, tendono a raggrupparsi in un certo settore sferico, il GDOP cresce; un buon valore di GDOP generalmente non è superiore a 4.

Il GDOP può essere scomposto secondo la classica relazione vettoriale nella componente verticale *VDOP (Vertical Dilution Of Precision)* e planimetrica *HDOP (Horizontal Dilution Of Precision)*; questa seconda componente può essere ulteriormente scomposta nelle sue componenti ortogonali orientate *NDOP (North-axis Dilution Of Precision)* e *EDOP (East-axis Dilution Of Precision)*.

5.8 Determinazione dell'errore medio complessivo

In tabella 3 sono riportati alcuni valori tipici presenti in letteratura per gli errori descritti nei precedenti sottoparagrafi, relativi a ricevitori di medie prestazioni (uno civile, l'altro militare). L'errore di disponibilità selettiva *(selective availability)*, pur non essendo più presente, secondo quanto affermato nel paragrafo 5.1, è citato a titolo di confronto.

Poiché le fonti di errore sono tra loro scorrelate, l'errore equivalente, definito *UERE (User Equivalent Range Error)*, è ottenuto come valore quadratico medio dei singoli errori (si sommano i quadrati degli errori, e si calcola la radice quadrata del totale). Ad esso occorre applicare il fattore moltiplicativo del GDOP (descritto nel precedente paragrafo 5.7), assunto in tabella pari a quattro, per ottenere il valore dell'accuratezza orizzontale 2DRMS, ovvero con un intervallo di confidenza pari al 95÷98 per cento (si veda il riquadro a pagina 96).

6. Tecniche di correzione differenziale degli errori GPS

6.1 Generalità

Mediante l'utilizzo di due o più ricevitori è

Sorgente di errore	Errore tipico (metri) codice civile C/A	codice militare P
disponibilità selettiva	24,0	0,0
ionosfera	7,0	0,01
troposfera	0,7	0,7
clock dei satelliti ed effemeridi	3,6	3,6
percorsi multipli in ricezione	1,8	1,2
rumore in ingresso al ricevitore	1,5	0,6
UERE	25,4	3,9
Accuratezza nel piano (intervallo di confidenza 2DRMS), per GDOP=4	101,5	15,6

2DRMS = twice Distance RMS oppure 2 Dimensional RMS
GDOP = Geometric Dilution of Precision
UERE = User Equivalent Range Error

Fonte: Internet

Tabella 3 *Valori tipici degli errori in metri per i ricevitori civili e militari.*

CORREZIONI DIFFERENZIALI PER MISURE GPS

La correzione differenziale può essere effettuata sulle misure determinate dai ricevitori GPS in due modi: in *post processing* o in *real time*. In entrambi i casi, il ricevitore è dotato di un opportuno software di elaborazione, che sincronizza temporalmente le misure in modo che sia le misure dirette sia i parametri di correzione siano relativi allo stesso istante temporale, misurato avendo come riferimento il tempo GPS o il tempo *UTC (Universal Time Coordinated)*. I due riferimenti hanno una precisione molto elevata e sono sfasati di circa 15 s.

Nella modalità in *post processing* i parametri di correzione sono organizzati in archivi *(file)*, ciascuno dei quali contiene in genere i dati relativi a un'ora (ad esempio dalle 10:00 alle 10:59 e 59 secondi) secondo lo standard *RINEX (Receiver INdependent EXchange format)*.

I calcoli sono di norma effettuati utilizzando un PC sul quale sono installati gli opportuni software, che accettano in ingresso le misure dirette effettuate dal ricevitore GPS e i suddetti file di correzione.

Nella modalità in *real time* la correzione differenziale è effettuata direttamente nel ricevitore GPS, nel cui *firmware* è installato il programma di calcolo. I parametri di correzione sono inviati al ricevitore o mediante trasmissione radio (canali UHF) oppure più di recente, mediante l'utilizzo della trasmissione dati su rete radiomobile. Il protocollo standard utilizzato è l'*RTCM SC-104* (che ha assunto il nome dal *Radio Technical Commission for Marittime services*, Special Committee 104 che lo ha definito).

P	distanza misurata tra satellite e ricevitore con la tecnica della pseudodistanza, affetta da errore
C	distanza misurata tra satellite e ricevitore con la tecnica della fase, affetta da errore
R	distanza effettiva (range) tra il satellite ed il ricevitore
E_S	errore nel posizionamento del satellite, indotto dal DoD come selective availability
E_{CS}	errore dovuto agli offset degli orologi a bordo dei satelliti
E_I	errore dovuto al ritardo del segnale causato dalle riflessioni ionosferiche
E_T	errore dovuto al ritardo del segnale causato dalle rifrazioni troposferiche
E_{CR}	errore dovuto all'offset dell'orologio del ricevitore
E_{MP}	errore dovuto al percorso multiplo in ricezione per la misura della pseudodistanza
E_{MC}	errore dovuto al percorso multiplo in ricezione per la misura di fase
E_{NP}	errore dovuto al rumore del ricevitore sulle misure della pseudodistanza
E_{NC}	errore dovuto al rumore del ricevitore sulle misure di fase
λ	lunghezza d'onda della portante L1 (19,05 cm) o L2 (24,45 cm)
N	numero intero di cicli che costituiscono l'ambiguità iniziale

DoD = Department of Defence

Tabella A *Grandezze caratteristiche degli errori presenti nel sistema GPS.*

Per comprendere a livello di esempio come sia possibile eliminare gli errori descritti al paragrafo 5, è opportuno introdurre le seguenti grandezze [4] riportate in tabella A.

Per un generico ricevitore GPS valgono le due relazioni:

$$P = R + E_S + E_{CS} + E_I + E_T + E_{CR} + E_{MP} + E_{NP} \qquad (1)$$
$$C = R + E_S + E_{CS} - E_I + E_T + E_{CR} + E_{MC} + E_{NC} + \lambda \cdot N$$

Se due ricevitori ricevono il segnale dallo stesso satellite, può essere applicata un'*equazione alle differenze singole* che rappresenta la differenza di cammino ottico percorso dal segnale per raggiungere i due ricevitori. La posizione di uno dei ricevitori è assunta come nota e si determina la posizione dell'altro rispetto al primo.

Se le osservazioni avvengono nello stesso istante è possibile eliminare o ridurre gli errori nel calcolo della posizione dei satelliti e quelli dovuti agli sfasamenti degli orologi di bordo. Differenziando le (1) e indicando con P_{ri} e C_{ri} le misure riferite al ricevitore i-esimo, possiamo infatti scrivere:

$$dP_r = P_{r2} - P_{r1} = dR + dE_I + dE_T + dE_{CR} + dE_{MP} + dE_{NP}$$
$$dC_r = C_{r2} - C_{r1} = dR - dE_I + dE_T + dE_{CR} + dE_{MC} + dE_{NC} + \lambda \cdot dN \qquad (2)$$

Può essere osservato che nella seconda relazione compare il termine relativo all'ambiguità iniziale (numero intero di cicli d'onda). In realtà le due osservazioni non avvengono esattamente nello stesso istante. È necessario tuttavia che lo sfasamento temporale sia contenuto entro pochi millisecondi nel caso di misure di pseudodistanza, e in pochi microsecondi nelle misure di fase.

Nelle relazioni sopra riportate sono ancora presenti i contributi di errore relativi alla deriva degli orologi dei ricevitori, ai percorsi multipli e al rumore di ricezione, nonché l'errore legato all'ambiguità iniziale.

Se due ricevitori ricevono contemporaneamente i segnali da una stessa coppia di satelliti, è poi possibile scrivere le *equazioni alle differenze doppie*. Esse possono essere ottenute sottraendo tra loro le differenze singole ottenute con ciascun satellite. Indicando con:

dP_r^{sr}, dC_r^{sr} **le differenze per il satellite con maggiore elevazione, assunto come riferimento,**
dP_r^{sa}, dC_r^{sa} **le differenze valutate secondo le relazioni (2) per un altro satellite tracciato,**

si ottengono le seguenti equazioni:

$$ddP_r^s = dP_r^{sa} - dP_r^{sr} = ddR + ddE_I + ddE_T + ddE_{MP} + ddE_{NP}$$
$$ddC_r^s = dC_r^{sa} - dC_r^{sr} = ddR - ddE_I + ddE_T + ddE_{MC} + ddE_{NC} + \lambda \cdot ddN \qquad (3)$$

Queste differenze doppie consentono di eliminare l'errore dovuto alla deriva degli orologi dei ricevitori. I due contributi ddE_I e ddE_T, relativi ai ritardi dovuti rispettivamente agli effetti ionosferici e troposferici, possono ritenersi trascurabili per distanze tra i ricevitori inferiori a 20÷30 km. Restano solo i contributi di errore causati dai percorsi multipli e dal rumore in ingresso ai ricevitori, oltre all'ambiguità iniziale.

Gli errori residui dovuti ai ritardi iono- e troposferici possono essere eliminati (o ulteriormente ridotti) in base a quanto già chiarito nei paragrafi 5.3 e 5.4 (tabella B).

errori ionosferici	Per **ricevitori a doppia frequenza** si possono effettuare contemporaneamente misure di fase su entrambe le portanti L1 ed L2, e stimare il ritardo ionosferico tenendo conto del fatto che esso è inversamente proporzionale alla frequenza. Per **ricevitori a singola frequenza** il ritardo si stima ricorrendo ai parametri medi inviati dai satelliti nel subframe 4 del *Navigation Message*.
errori troposferici	Si possono correggere applicando modelli empirici che si basano sulla stima delle caratteristiche troposferiche ottenuta mediante misure di pressione, temperatura e umidità effettuate a terra (utili in caso di *post-processing*).

Tabella B *Sistemi impiegati per eliminare gli errori residui dovuti ai ritardi iono- e troposferici.*

Per eliminare l'ambiguità iniziale, presente esclusivamente nelle misure di fase, è sufficiente determinare le differenze doppie in due successivi istanti di tempo t_1 e t_2 (*epoche*) e sottrarle tra loro, ottenendo le cosiddette equazioni alle differenze triple:

$$dddC_r^s(t) = ddC_r^s(t_2) - ddC_r^s(t_1) = dddR + dddE_{MC} + dddE_{NC} \qquad (4)$$

nelle quali, come si è già detto, sono stati trascurati i contributi di errore dovuti agli effetti iono- e troposferici.

Le uniche incognite rimaste, implicite nella misura del dddR (differenza tripla applicata alle distanze effettive tra i satelliti ed i ricevitori), sono le differenze lungo i tre assi tra le coordinate del ricevitore mobile e quelle del ricevitore fisso. Queste equazioni possono essere utilizzate ogniqualvolta, nel corso di una sessione di misura, ricompare l'ambiguità iniziale a causa del momentaneo sganciamento del segnale ricevuto da un satellite (fenomeno noto come *cycle slip*).

Il calcolo delle differenze triple richiede un tempo sensibilmente maggiore rispetto alle differenze singole e doppie e presuppone l'utilizzo di ricevitori GPS portatili di elevata qualità, in grado cioè di effettuare misure di fase.

possibile elaborare i dati raccolti contemporaneamente da ciascuno di essi e correggere le misure effettuate, in modo da eliminare o da ridurre drasticamente tutti gli errori di *modo comune*. La posizione di uno dei ricevitori *(riferimento)* è assunta come nota ed esatta, in quanto determinata mediante strumenti di altissima precisione su tempi molto lunghi, in modo da eliminare il più possibile tutte le fonti di errore. Mediante altri ricevitori portatili *(rover)* si acquisiscono le misure relativamente a quella del primo ricevitore, determinando quindi le distanze dalla stazione di riferimento (definite in ambito topografico come le *basi*).

Il metodo differenziale è applicabile sia alle misure di pseudodistanza (precisione di qualche metro) sia a quelle di fase (precisione centimetrica). La correzione è tanto più efficace quanto minore risulti sia lo sfasamento temporale tra i dati GPS grezzi ed i dati di correzione (comunque non superiore a 30 s), sia la lunghezza della *base* (distanza tra il rover e il riferimento).

Per le misure di fase sussiste inoltre un limite di applicabilità del metodo differenziale: la distanza tra GPS di riferimento e GPS mobile non deve superare i 30 km, in modo da contenere la differenza tra i ritardi con cui il segnale arriva ai due ricevitori sotto una lunghezza d'onda.

Altro parametro che condiziona l'accuratezza raggiungibile con la correzione differenziale è il tempo di misura, che può variare da qualche secondo *(rilievi cinematici)* a qualche ora *(rilievi statici)*.

Le differenze sono effettuate più volte ottenendo le cosiddette *differenze multiple*, che consentono di eliminare, o di ridurre drasticamente, alcune delle sorgenti di errore. Gli unici errori che non sono eliminabili con la tecnica differenziale sono il rumore al ricevitore (dovuto all'attraversamento del segnale nella bassa atmosfera) e quello relativo al percorso multiplo in ricezione. Questi due tipi di errore possono però essere stimati dal sistema prima di effettuare la misura, per essere eliminati successivamente mediante una postelaborazione.

Nel riquadro a pagina 99 si riporta qualche dettaglio matematico sulla tecnica differenziale.

6.2 Equazioni alle differenze singole

Se due ricevitori ricevono contemporaneamente il segnale da uno stesso satellite, può essere impiegata un'*equazione alle differenze singole*, che rappresenta la differenza di cammino ottico percorso dal segnale per raggiungere ciascuno di essi. La posizione di uno dei ricevitori è assunta come nota e con questa si determina la posizione dell'altro rispetto al primo (figura 11). Se le osservazioni avvengono nello stesso istante è possibile eliminare o ridurre l'errore sull'orologio del satellite.

In modo analogo, considerando un solo ricevitore e le misure dei segnali che provengono da due satelliti, si può ridurre l'entità dell'errore dell'orologio del ricevitore (figura 12).

6.3 Equazioni alle differenze doppie

Se due ricevitori tracciano contemporaneamente gli stessi due satelliti, è possibile utilizzare le *equazioni alle differenze doppie* che sono ottenibili sottraendo tra loro le differenze singole ottenute dai due ricevitori con ciascun satellite (figura 13). Le differenze doppie consentono di eliminare o di attenuare gli errori atmosferici, nell'ipotesi che le condizioni ionosferiche e troposferiche siano pressoché costanti nella zona in cui sono eseguite le misure (le distanze tra i ricevitori non devono essere superiori a qualche decina di km).

Le differenze doppie sono definite come *osservabili fondamentali*, poiché sono quelle utilizzate dalla maggior parte dei software per l'elaborazione dei dati GPS.

6.4 Equazioni alle differenze triple

Per eliminare l'ambiguità iniziale insita nelle misure di fase (descritta nel paragrafo 4) è sufficiente determinare le differenze doppie in due

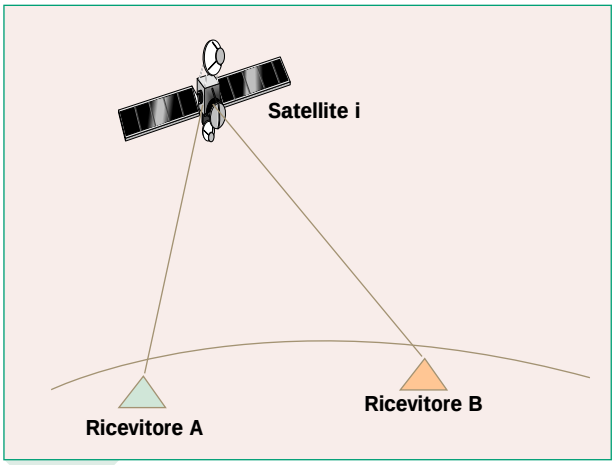


Figura 11 *Differenze singole con due ricevitori e un satellite.*

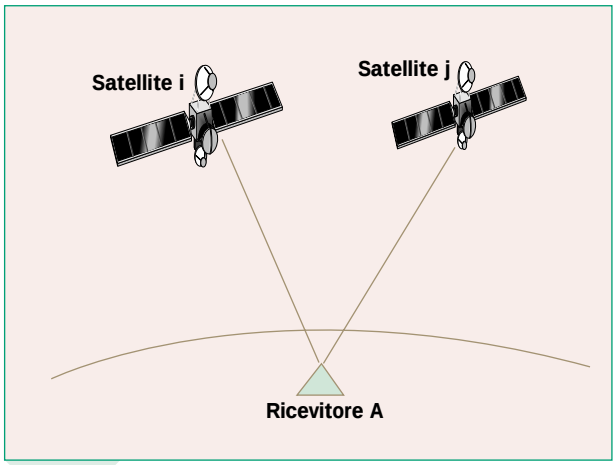


Figura 12 *Differenze singole con un ricevitore e due satelliti.*

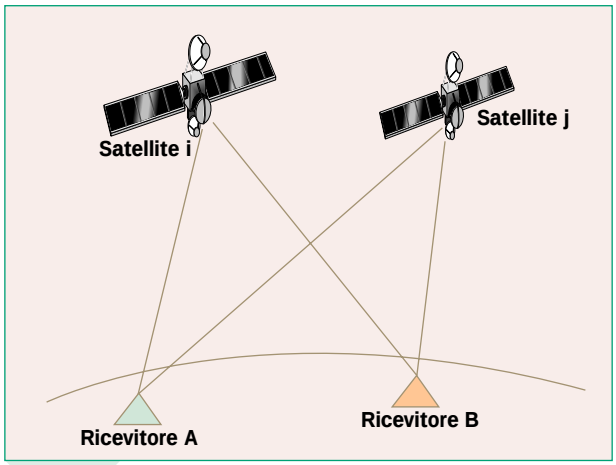


Figura 13 *Differenze doppie.*

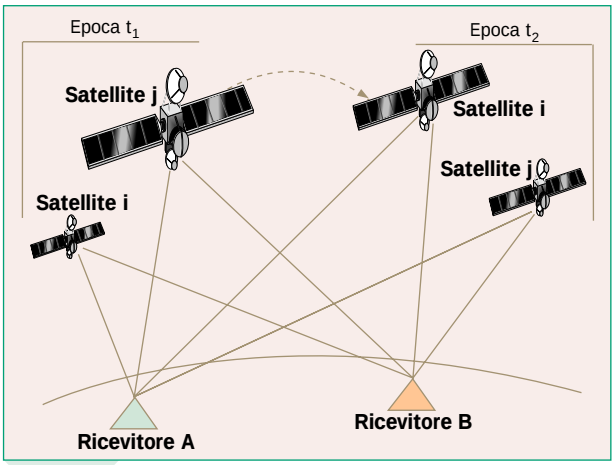


Figura 14 *Differenze triple.*

successivi istanti di tempo t_1 e t_2 (usualmente definiti *epoche*) e sottrarle tra loro, ottenendo così le cosiddette *equazioni alle differenze triple* (figura 14). Queste equazioni possono essere utilizzate quando, nel corso di una sessione di misura, ricompare l'ambiguità iniziale a causa del momentaneo sganciamento del segnale ricevuto da un satellite (fenomeno noto come *cycle slip*).

Il calcolo delle differenze triple richiede un tempo più elevato rispetto a quello necessario per determinare le differenze singole o doppie e presuppone che anche il ricevitore *GPS (rover)*, dovendo effettuare misure di fase, sia di qualità elevata.

7. Conclusioni

Nel corso dell'ultimo decennio si è assistito ad una sempre più spiccata integrazione tra i settori delle telecomunicazioni e dell'informatica, che ha consentito di definire nuovi servizi a valore aggiunto.

La telefonia mobile consente oggi di rendere disponibili e fruibili ovunque le informazioni, con il vantaggio di ottimizzare i processi gestionali e decisionali. Una delle informazioni da sempre considerata come fondamentale per le attività umane è quella relativa alla determinazione di una posizione (di un oggetto ovvero di sé stessi) e alla sua relazione con il mondo circostante (*georeferenziazione*).

Fonte: Nikon Instruments

Figura 15 *Rilevamento GPS per la realizzazione di infrastrutture.*

Fonte: Compaq

Figura 16 *Un esempio di navigatore GPS integrato in un computer palmare.*

Il presente articolo ha descritto il sistema GPS, finora il più avanzato per la determinazione delle suddette informazioni di posizione e di tempo.

Dopo l'analisi dei segmenti di cui esso si compone (costellazione satellitare, controllo, di utente), è stato descritto il segnale inviato dai satelliti e le funzionalità principali di un ricevitore GPS per acquisire il suddetto segnale e per determinarne la posizione, indicando il livello di precisione teorica della misura. Sono state, quindi, analizzate le principali fonti di errore che intervengono sulla precisione e le modalità di correzione (eliminazione o attenuazione) di tali effetti, con particolare riguardo alla tecnica di correzione differenziale.

In un successivo articolo, che sarà pubblicato in un prossimo numero del Notiziario Tecnico, si descriveranno i motivi che hanno indotto TIM a dotarsi di una rete di stazioni GPS di riferimento per la correzione differenziale. Sarà anche trattata la soluzione architetturale adottata per consentire l'erogazione di servizi commerciali, rivolti a una clientela costituita da professionisti della misura, quali, ad esempio, geologi, topografi, ricercatori, aziende per la realizzazione di infrastrutture e opere civili (figura 15).

Non si esclude la possibilità che in un prossimo futuro nuovi prodotti e applicazioni, come quello riprodotto in figura 16, possano avvicinare all'uso del GPS anche il mercato di massa.

Abbreviazioni

BPSK	Binary Phase Shift Keying
2DRMS	twice Distance Root Mean Square (oppure 2 Dimensional Root Mean Square)
C/A	Coarse/Acquisition
CEP	Circular Error Probable
DoD	Department of Defence
EDOP	East-axis Dilution Of Precision
GDOP	Geometric Dilution Of Precision
GPS	Global Positioning System
HDOP	Horizontal Dilution Of Precision
NAVSTAR	NAVigation Satellite Timing And Ranging
NDOP	North-axis Dilution Of Precision
PPS	Precise Positioning System
PRC	Primary Reference Clock
PRN	Pseudo Random Noise
RHCP	Right Hand Circularly Polarized
RINEX	Receiver INdependent EXchange format
RMS	Root Mean Square
RTCM	Radio Technical Commission for Maritime service
SC-104	Radio Technical Communication for Maritime services Special Commitee-104
S/A	Selective Availability
SEP	Spherical Error Probable
SPS	Standard Positioning System
SV	Space Vehicle (satellite)
TOA	Time Of Arrival
TOW	Time Of Week
TSLF	Time Since Last Fix
TTFF	Time To First Fix
UERE	User Equivalent Range Error
UTC	Universal Time Coordinated
VDOP	Vertical Dilution Of Precision

Bibliografia

[1] Dana, P.H.: *Global Positioning System Overview*. www.colorado.edu/geography/gcraft/notes/gps/gps_f.html, settembre 1994.

[2] Parkinson, B.W.; Spilker, J.J.: *GPS: Theory and Applications* - cap. 11: *GPS Error Analysis*. American Institute of Aeronautics and Astronautics, gennaio 1996.

[3] Diggelen, F.v.: *GPS Accuracy*. Lies, Damn Lies and Statistics. www.gpsworld.com/columns/9805innov.html, 1998.

[4] samsvl@nlr.nl: *Some Theory on GPS Range Measurements*. http://home-2.worldonline.nl/~samsvl/theory.htm, marzo 2000.

[5] Arinc Research Corporation El Segundo, CA (USA): *ICD-GPS-200C - NAVSTAR GPS Space Segment / Navigation User Interfaces*. Ottobre 1993.

[6] Diggelen, F.v.; Abraham, C.: *IndoorGPSTM Technology White Paper*. Presentato al CTIA Wireless-Agenda, Dallas, maggio 2001.

[7] *Timing characteristics of Primary Reference Clocks*. ITU-T Recommandation G.811, settembre 1997.

Duilio Coratella ha conseguito la laurea con lode in Ingegneria Elettronica indirizzo Telecomunicazioni presso il Politecnico di Bari nell'anno 1993, discutendo una tesi su dispositivi ottici non lineari su guida d'onda per reti di telecomunicazioni. Nel 1995 è stato assunto nell'Area Rete della Direzione Generale di TIM. Dal 1995 al 1997 si è occupato di analisi e valorizzazione dei parametri di cella e della definizione ed analisi dei dati statistici di traffico e di qualità della rete GSM. Dal 1997 al 1999 ha coordinato un gruppo di lavoro per l'ottimizzazione della rete TACS in termini di efficienza, riduzione della congestione e distribuzione delle risorse radio sul territorio, ed è stato membro del SATIG nell'ambito del GSM MoU. Dal 1999 al 2000 si è occupato di tecniche e di sistemi per il *positioning* su rete radiomobile ed ha seguito lo sviluppo della rete GPS di TIM, collaborando con TILAB al progetto di una piattaforma per l'erogazione dei dati GPS e, con le funzioni di marketing di TIM, per la definizione dei relativi servizi commerciali. Opera ora nel Client Management di Rete, dove, nell'ambito del processo di innovazione prodotti e servizi, cura i rapporti tra le linee di sviluppo e di esercizio e le aree commerciali e di marketing, coordinando gli studi di fattibilità e collaborando alla definizione delle specifiche funzionali. È membro del Comitato Tecnico TIM-ASI (Agenzia Spaziale Italiana) nel cui ambito segue le attività di integrazione della Rete GPS di TIM nel Sistema EUREF (EUropean REference Frame).

Costi sociali e ambientali delle tecniche di scavo

LUCA GIACOMELLO
PAOLO TROMBETTI

L'obiettivo di questo articolo è di mettere in evidenza i vantaggi legati all'utilizzo delle tecnologie impiantistiche innovative rispetto alla tecnica tradizionale di scavo a cielo aperto, prendendo in considerazione sia il costo legato all'incremento del traffico viario sia il costo di impatto ambientale relativo ai materiali e all'energia utilizzati nei lavori, alle emissioni gassose, ai rilasci liquidi e ai rifiuti solidi prodotti.

1. Introduzione

Nel settore della costruzione e della manutenzione dei sottoservizi, lo sviluppo e la diffusione delle tecnologie alternative rappresenta una novità non solo sul piano strettamente tecnologico, ma anche, e soprattutto, sul piano del differente impatto che queste tecniche impiegate per l'esecuzione dei lavori comportano per la collettività.

Da circa dieci anni Telecom Italia ha avviato una politica di investimenti, tuttora in corso, per la ricerca e lo sviluppo di tecnologie, a basso impatto ambientale, alternative agli scavi tradizionali nella posa di infrastrutture di telecomunicazione.

Sebbene Telecom Italia risulti essere il primo gestore europeo e il terzo al mondo, dopo Stati Uniti e Giappone, per chilometri di infrastrutture realizzate con sistemi innovativi, *directional drilling*, va rilevato come, a fronte dell'impegno profuso, gli sforzi non siano stati tenuti nell'opportuna considerazione dagli Organismi governativi e dagli Enti locali.

Una maggiore incisività da parte delle Istituzioni, infatti, si rivela fondamentale per una diffusione più ampia di queste soluzioni che favoriscono lo "sviluppo sostenibile".

Per un gestore di telecomunicazioni, quale Telecom Italia, che ha adottato verso l'ambiente un atteggiamento proattivo, indirizzato a rendere minimo l'impatto negativo e ad accrescere l'impiego delle opportunità rese disponibili dallo sviluppo sostenibile, è assai importante conoscere in modo dettagliato l'impatto sociale e ambientale associato alle proprie attività e operare di concerto con gli Enti locali per valutare opportunamente tutti gli interventi, considerando sia la validità economica, sia la riduzione dell'impatto ambientale. Questo obiettivo si traduce, in pratica, nella esigenza di sviluppare strumenti mirati a misurare le esternalità associate ai processi e ai prodotti gestiti.

Le esternalità, rappresentano, infatti, gli effetti, generati da un bene nel corso del ciclo di vita, i cui costi non sono sostenuti solo da quanti partecipano alla fruizione del bene stesso ma si ripercuotono sull'intera collettività.

La valutazione delle esternalità consente, quindi, di tradurre economicamente l'impatto sociale e ambientale associato ai propri processi industriali.

In questo articolo è mostrato un modello di valutazione tecnico ed economico che, partendo dagli effetti legati alla predisposizione dei cantieri consente

di stimare le esternalità associate a tecniche impiantistiche quali scavo a cielo aperto, perforazione orizzontale guidata - *no-dig* o *trenchless technologies*[1] - microtrincea e minitrincea. Sono in particolare esaminati alcuni parametri quali: il costo di installazione, il costo legato all'incremento del traffico viario e il costo dell' impatto ambientale relativamente a un *case study* rappresentativo di una situazione tipica di quelle affrontate in pratica da Telecom Italia.

Per valutare il costo legato all'incremento del traffico viario è stato utilizzato un algoritmo riprese dalla letteratura [1] che permette di stimare l'influenza dei lavori di scavo -quindi delle parzializzazioni della carreggiata- sulla portata effettiva delle strade.

Per valutare il costo legato all'impatto ambientale sono stati presi in considerazione due fattori: quello provocato dall'utilizzo delle diverse tecnologie di scavo e l'altro attribuibile all'incremento del traffico, che porta a un aumento delle emissioni dei gas di scarico per i veicoli coinvolti nell'area occupata dai cantieri.

La valutazione di questi due fattori è stata basata sui risultati ottenuti dagli inventari, stilati a valle di un'analisi di tipo *LCA (Life Cycle Assessment)*, e di una successiva elevazione a valore del costo relativo all'impatto sull'ambiente mediante il metodo *EPS (Environmental Priority Strategies)* [4].

In questo articolo si vuole, infatti, mostrare uno strumento per effettuare considerazioni di carattere socio-ambientale con indicazioni di costo, senza effettuare confronti specifici tra le diverse tecnologie. Lo studio può fornire al progettista indicazioni utili circa la scelta della tecnica di scavo che meglio si adatta a ogni singolo caso di posa delle infrastrutture.

Nella presente analisi sono state, infine, tralasciate alcune componenti di costo di difficile valutazione attraverso modelli previsionali, quali il costo di tipo sociale sopportato dalla collettività (costo legato a diseconomie esterne, cioè a interferenze tra il cantiere e le attività economiche della zona considerata) e il costo di rischio legato all'impiego della tecnologia di scavo considerata (costo derivante dai danni procurati a persone e a cose).

2. Tecnologie impiegate nella realizzazione degli impianti

Le tecnologie esecutive analizzate nel seguito del documento sono quelle di:

- scavo a cielo aperto;
- perforazione orizzontale guidata;
- microtrincea;
- minitrincea.

Nei diagrammi di flusso relativi alle diverse tecniche utilizzate per realizzare gli impianti, i riquadri tratteggiati indicano l'insieme delle fasi considerate nelle analisi LCA.

È anche chiarito che l'analisi eseguita non tiene conto del tipo di infrastruttura posata; l'esame è stato, infatti, indirizzato verso la valutazione delle differenti tipologie di scavo e delle operazioni a esse correlate.

2.1 La tecnica di scavo a cielo aperto

L'intervento con scavo a cielo aperto *(open-cut)* costituisce il sistema tradizionalmente impiegato nella realizzazione degli impianti. Esso è relativo alla realizzazione di una trincea nel terreno sul fondo della quale sono posate le infrastrutture che contengono i cavi per telecomunicazioni. Questo tipo di intervento comporta l'utilizzo di una serie di mezzi e di attrezzature per la movimentazione di grandi quantità di materiale da e verso l'area del cantiere.

La tecnica in questione si articola generalmente nelle seguenti fasi principali, come illustrato nel diagramma a blocchi di figura 1:

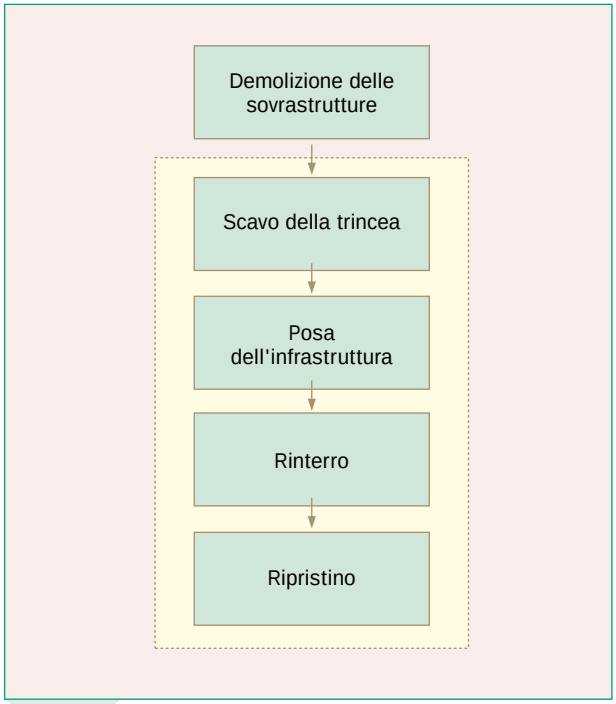


Figura 1 *Diagramma a blocchi dello scavo a cielo aperto.*

- rimozione delle sovrastrutture esistenti (ad esempio della pavimentazione stradale o pedonale);
- cavo della trincea sino alla profondità operativa;
- esecuzione delle operazioni di posa;
- rinterro;
- ripristino.

La tecnica di scavo tradizionale presenta una serie di inconvenienti quali: l'inquinamento atmosferico e acustico legato ai macchinari utilizzati nel cantiere di scavo; la rimozione di grandi volumi di terra destinata a discarica nel corso dei lavori; il consumo di risorse naturali legato all'impiego di mate-

(1) *Con il termine "no-dig" o "trenchless technologies" si intende l'insieme delle tecniche di ispezioni, messa in opera, sostituzione e rinnovamento di infrastrutture sotterranee per reti di telecomunicazione, gas, acquedotti, energia e fognarie, mediante l'impiego di macchine e robot senza scavo a cielo aperto o, in alcuni casi, limitando lo scavo alla realizzazione delle due buche in partenza e in arrivo.*

TECNOLOGIE IMPIANTISTICHE PER LA POSA DI INFRASTRUTTURE SOTTERRANEE

Numerosi sono gli aspetti economici e ambientali di installazione di sistemi impiegati nella posa di infrastrutture sotterranee da Telecom Italia.
Le tecniche oggi impiegabili sono quelle relative allo:

- *scavo a cielo aperto* che rappresenta il sistema tradizionalmente impiegato e che consiste nella realizzazione di una trincea nel terreno di opportune dimensioni, sul fondo del quale sono posate le infrastrutture per telecomunicazione;
- *directional drilling o perforazione orizzontale guidata* che è una tecnica che comporta la posa di infrastrutture nel sottosuolo senza la realizzazione di scavi a cielo aperto, o limitandone l'impiego alle sole buche di partenza e di arrivo dell'infrastruttura;
- *microtrincea* che consiste nella posa di uno o più cavi a fibra ottica, opportunamente realizzati, in un solco poco profondo (da 7 a 10 cm) e assai stretto (da 10 a 12 mm);
- *minitrincea* che è una trincea di dimensioni più contenute (da 7 a 10 cm di larghezza e da 30 a 35 cm di profondità), realizzata con un'attrezzatura che consente di eseguire in maniera automatica diverse fasi operative.

riale inerte necessario per procedere al rinterro, con conseguenti trasporti aggiuntivi che determinano un ulteriore fattore di impatto ambientale.

Le operazioni di scavo a cielo aperto non sono poi praticabili agevolmente in tutte le realtà, ad esempio nei centri storici delle città; nei piccoli centri abitati, specie nei casi in cui la larghezza delle strade è assai ridotta, nelle zone con traffico intenso, in tutti quei casi, quindi, dove la riduzione della velocità di avanzamento dei veicoli può causare disagi non trascurabili per le attività commerciali e turistiche. Quest'insieme di inconvenienti ha portato a individuare sistemi di scavo differenti impiegati per realizzare le infrastrutture.

Nei due paragrafi che seguono sono presentate tre tecnologie innovative, che formano l'oggetto di una valutazione sia dal punto di vista economico sia da quello ambientale.

2.2 I sistemi di perforazione orizzontale guidata

Il termine *tecniche no-dig* identifica una serie di sistemi (quali ad esempio quelli: *directional drilling*, *microtunnelling*), utilizzati nella realizzazione degli impianti che permettono di mettere in opera tubi e infrastrutture sotterranee mediante macchine e robot, riducendo il ricorso, evitando gli scavi a cielo aperto.

Nel seguito si farà riferimento in particolare alla tecnica denominata perforazione orizzontale guidata *(directional drilling)* [2], [3].

A seconda che si impieghino o no fluidi per la perforazione, la tecnologia può essere classificata a secco o ad umido. L'analisi svolta riguarda la perforazione orizzontale guidata a secco, che costituisce una tecnica oggi largamente diffusa e impiegata nella maggior parte delle operazioni di posa delle infrastrutture.

Essa prevede una perforazione eseguita mediante un martello pneumatico montato su una trivella rotante. L'avanzamento avviene per la spinta esercitata dal martello: in questo caso, per effetto della spinta, il terreno è compresso lungo le pareti del foro. Una miscela lubrificante, a base di acqua è utilizzata per raffreddare l'utensile. Il diametro della trivellazione è eventualmente aumentato, in funzione del diametro dell'infrastruttura da posare, tramite il passaggio successivo di alesatori di diametro opportuno.

La trivellazione guidata può essere eseguita - in relazione al tipo di progetto - dalla superficie, da una buca o da un pozzetto e consente di superare ostacoli architettonici e naturali, limitando i punti di intervento per la realizzazione degli impianti ai soli cantieri di lancio e di arrivo.

Prima di effettuare la perforazione occorre eseguire una serie di indagini, quali ad esempio l'introspezione, mediante radar, della natura del sottosuolo e della presenza di altri impianti *(indagine litologica)* che consentano di ricostruire la situazione del sottosuolo nel tratto interessato dalla posa dei tubi. È così possibile evitare danni ad altre infrastrutture preesistenti e ottimizzare il percorso dell'infrastruttura.

Le fasi principali del processo di scavo, riassunte anche graficamente nel diagramma di flusso di figura 2, sono elencati qui di seguito:

- indagini preliminari con mappatura del sottosuolo (radar, litologica);
- delimitazione delle aree di cantiere;
- realizzazione del foro pilota;
- eventuale alesatura del foro pilota e contemporanea posa dell'infrastruttura.

Questo sistema, impiegato nella realizzazione degli impianti, non comporta alcuno scavo preliminare ma richiede solo di effettuare buche di partenza e di arrivo, diversamente dal caso prima presentato relativo ad impianti realizzati con tecniche tradizionali. Esso non comporta, quindi, di demolire prima e poi di ripristinare le sovrastrutture esistenti.

2.3 La tecnica della microtrincea

La tecnica della microtrincea *(microtrench)* assai innovativa, comincia a diffondersi solo in questi ultimi tempi e si presenta come particolarmente adatta per la posa di tratti di rete di accesso.

L'impiego di questa tecnica consente di avere cantieri che interessano superfici molto limitate, con

evidenti benefici dal punto di vista dell'intralcio alla viabilità.

La tecnologia qui considerata consiste essenzialmente nel praticare un solco nell'asfalto di 7-10 cm di profondità e 10-12 mm di larghezza. All'interno di questo scavo è dapprima posata l'infrastruttura, costituita da cavi ottici di dimensioni ridotte e da due cordoli sovrastanti di protezione e, successivamente, è effettuata una sigillatura mediante materiale bituminoso.

Le attrezzature impiegate (in particolare il "tagliasfalti") sono di dimensioni ridotte e consentono di allestire un cantiere in uno spazio estremamente contenuto.

Il ciclo necessario per la realizzazione delle infrastrutture può quindi essere sintetizzato nelle fasi seguenti (figura 3):

- realizzazione del solco di microtrincea;
- lavaggio e pulizia del solco, con acqua e con aria compressa;
- asciugatura del solco;
- posa del cavo e delle protezioni ad esso relative, costituite da due cordoli polimerici;
- sigillatura del solco con materiale bituminoso.

A differenza della perforazione orizzontale, con questa tecnica è asportata, una quantità trascurabile del sottosuolo direttamente con le operazioni di pulitura del solco.

2.4 La tecnica della minitrincea (minitrench)

Questa tecnica di scavo, introdotta molto di recente, riduce il volume dello scavo adottando una tecnica di fresatura.

In questo caso le operazioni da svolgere in successione sono le seguenti (figura 4):

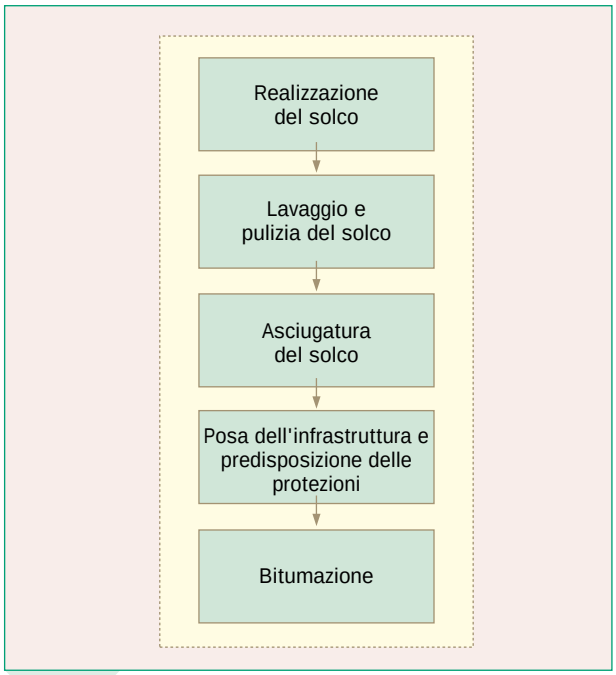


Figura 3 *Diagramma a blocchi della tecnica della microtrincea.*

- realizzazione del solco mediante fresatura;
- posa dell'infrastruttura;
- riempimento del solco mediante la speciale malta cementizia.

Sono state messe a punto due tecniche di realizzazione: la prima, automatica in linea, utilizza un'unica attrezzatura per consentire di eseguire tutte le operazioni sopra descritte e la seconda per la cui attuazione sono svolte in successione le diverse fasi operative.

In questo articolo si è preferito descrivere la

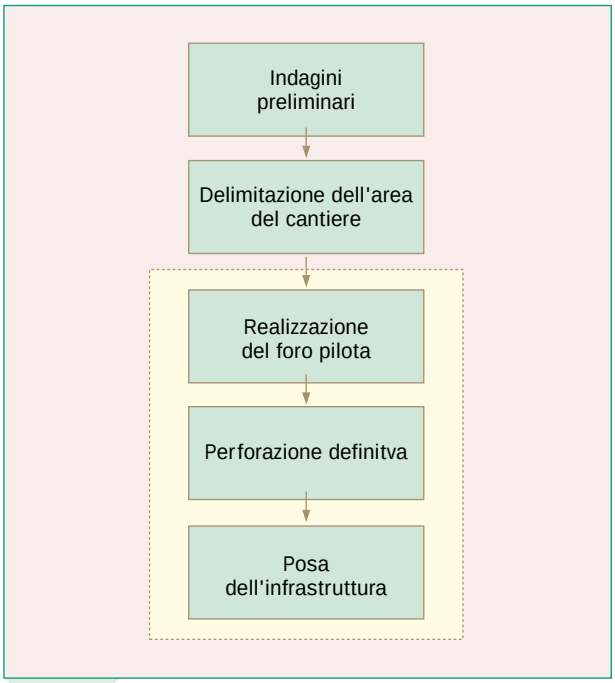


Figura 2 *Diagramma a blocchi della perforazione orizzontale guidata.*

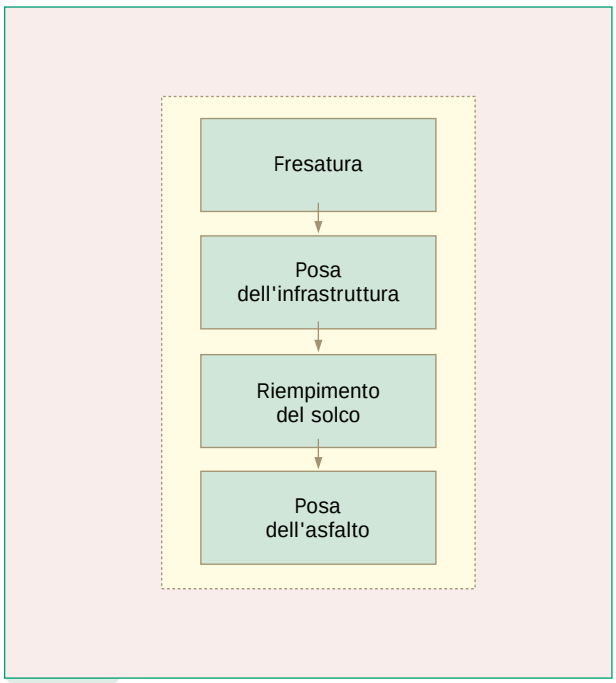


Figura 4 *Diagramma a blocchi di una minitrincea.*

MODELLO ECONOMICO AMBIENTALE

Il modello economico e ambientale, applicato a una situazione tipica, quali quelle che affronta quotidianamente un gestore di reti di servizi, consente di stimare le esternalità (come i costi a carico della collettività generati dall'apertura dei cantieri) associate alle diverse tecniche di installazione.

Rispetto allo scavo tradizionale, la perforazione orizzontale guidata, la microtrincea e la minitrincea risultano, in particolare, essere assai più vantaggiose in termini di riduzione dell'impatto ambientale e del costo legato alla congestione del traffico viario.

Per quanto attiene ai costi di installazione, le tecniche innovative, quando tecnicamente applicabili, risultano essere in alcuni casi economicamente vantaggiose.

I gestori di reti auspicano quindi per il futuro una maggiore presenza propositiva dei soggetti pubblici preposti alla formalizzazione delle normative che contemplino l'impiego di tecniche con un impatto ambientale limitato e, nel contempo, l'intervento delle Istituzioni e degli Enti capaci di stimolare queste soluzioni con azioni concrete, quali gli ecoincentivi e le defiscalizzazioni, in modo da consentire il raggiungimento di benefici per l'intera collettività.

prima tecnica, quella automatica, che prevede di interrare un'infrastruttura di posa (due tubi di diametro pari a 40 mm) posati in una trincea di larghezza di 7-10 cm e con profondità di 30-35 cm.

Con questo sistema la posa dell'infrastruttura e il riempimento sono eseguiti in un'unica operazione, senza soluzione di continuità; il ripristino dell'asfalto avviene invece in un momento successivo. L'utilizzo di questa tecnica, rispetto agli interventi successivi con più attrezzature, consente di ridurre i consumi e, conseguentemente, anche di contenere l'impatto ambientale associato a queste operazioni.

Con questo tipo di scavo si produce materiale da destinare a discarica anche se, tenendo conto delle dimensioni ridotte della trincea, si tratta di quantità trascurabili.

3. Modello economico e ambientale delle diverse tecniche di posa

È possibile ora passare all'esame di alcune componenti del costo legato alla presenza dei cantieri per l'esecuzione dei lavori di scavo, confrontando dal punto di vista economico, laddove possibile, le tre tecniche di posa prima descritte.

Le principali componenti di costo, considerate qui di seguito, sono:

- *costo di installazione;*
- *costo legato all'incremento del traffico viario;*
- *costo di impatto ambientale.*

Non è stato possibile tenere presente - per la mancanza di informazioni e per l'impossibilità di valutare alcuni parametri - il costo sociale, cioè quello sostenuto dalla collettività (legato a diseconomie esterne provocate dalle interferenze tra il cantiere e le attività economiche della zona considerata), nonché il costo di rischio, cioè quello derivante da danni procurati a persone o a cose legati all'impiego di una determinata tecnica esecutiva di scavo. Questi due tipi di costo sono comunque influenzati direttamente dalle dimensioni del cantiere.

Allo scopo di esemplificare la metodologia utilizzata per la stima dei costi con dati numerici significativi, nell'articolo è stato esaminato un caso pratico con l'obiettivo di mettere in luce i vantaggi legati all'utilizzo di tecniche di scavo innovative.

Le ipotesi relative al tipo di cantiere a cui si riferiscono le diverse stime di costo sono riportate in modo particolareggiato nei paragrafi specifici che seguono.

Tutti i dati sono riferiti a un'area di cantiere presente in una strada con elevata densità di traffico in una zona urbana semicentrale. È stata anche considerata una larghezza della carreggiata da ciglio a ciglio che misuri circa 11 m, con due corsie per senso di marcia ciascuna larga 2,70 m.

3.1 Costo d'installazione

Il costo di installazione è rappresentato dall'ammontare complessivo delle risorse occorrenti per effettuare l'intervento costruttivo o di manutenzione e per il successivo ripristino della sede viaria in termini di: materiali, mezzi d'opera, lavorazioni, canoni e concessioni.

Questi costi sono valutati attraverso le classiche operazioni di computo e di stima basate sull'impiego di listini dei prezzi praticati dagli Enti appaltanti.

3.2 Costo legato all'incremento del traffico viario

Il costo dovuto alla riduzione della sede stradale è quello che occorre sostenere quando i cantieri aperti per la posa di un sottoservizio occupano, in parte o completamente, la piattaforma viaria.

In funzione della configurazione geometrica dei lavori da eseguire, le interferenze possono essere di tre tipi:

- *parallele*: quando il cantiere presenta uno sviluppo prevalentemente monodimensionale e parallelo all'infrastruttura viaria, senza che esso causi mai la completa interruzione del traffico che si svolge nella strada interessata dai lavori;
- *trasversali:* quando il cantiere si sviluppa trasversalmente alla strada, potendo anche causare un'interruzione totale del traffico che su di essa si svolge;
- *miste*: quando il cantiere presenta sia tratti paral-

leli sia trasversali all'arteria, potendo così causare anche un'interruzione totale del traffico che su di essa si svolge.

Il costo legato alla congestione del traffico viario può essere considerato, in modo approssimato, come somma di due elementi che incidono sul singolo utente che utilizza quel tratto stradale: anzitutto il costo del maggiore tempo di percorrenza; in secondo luogo il costo del maggior consumo di carburante.

Queste grandezze possono essere valutate mediante un modello di calcolo previsionale che si basa sull'influenza che parzializzazioni della carreggiata possono avere sulla portata effettiva di traffico, distinguendo tra arterie extra-urbane di grande comunicazione e strade urbane.

La metodologia di calcolo analizzata è basata su considerazioni di carattere geometrico e cinematico.

Per l'applicazione del modello è stato considerata un'interferenza di tipo parallela al percorso stradale e un cantiere situato in una zona urbana semicentrale.

Possono essere impiegati, tuttavia, in queste valutazioni, anche altri modelli (modelli distributivi del traffico), basati su funzioni di impedenza che tengano conto delle velocità medie possibili che hanno i veicoli in funzione delle condizioni di blocco del traffico o di restringimento temporaneo della carreggiata (stato di occupazione, condizione del manto stradale, presenza di ostacoli). Questi modelli, pur essendo più rigorosi, sono molto complessi in quanto tengono conto di un numero elevato di variabili e sono quindi di difficile applicazione, anche perché in letteratura non esistono dati sufficienti, tali da consentirne l'impiego nei casi che si presentano in pratica.

Dal modello [1] si ricava la seguente espressione per il calcolo del costo legato agli effetti della presenza di un cantiere:

$$C_{rc} = C_{mpt} + C_{mc}$$

con

- C_{rc} - costo legato agli effetti della realizzazione di un cantiere;
- C_{mpt} - costo del maggior tempo di percorrenza;
- C_{mc} - costo del maggior consumo di carburante.

La valutazione del costo, legato al maggior tempo di percorrenza, è eseguita seguendo il processo di seguito descritto. Il modello di riferimento scelto è quello indicato in bibliografia.

- *passo 1*: si valuta la "portata" della strada considerata, funzione delle caratteristiche intrinseche e dell'ubicazione nell'ambiente cittadino (centro, semicentro e periferia) ad essa relative. La portata è trasformata tenendo conto che il valore varia nell'arco delle ventiquattro ore e consente di ottenere un valore di portata media giornaliera, denominato Q_{me} ed espresso in veicoli per ora, preso come riferimento per le elaborazioni successive;
- *passo 2*: la portata Q_{me} è confrontata con la riduzione che si ha in presenza di un ostacolo sulla carreggiata, costituito dal cantiere. Questa variazione di portata è indicata con $n_r Q_r$, dove n_r è il numero di corsie che rimangono percorribili anche in presenza dello scavo. Quando la riduzione di portata è minore di quella Q_{me} si ha la congestione del traffico con formazione di code. Questo è il caso di interesse per il calcolo del costo del maggior tempo di percorrenza;
- *passo 3*: si calcola il valore del tempo di attesa in coda T_a per un singolo veicolo, sulla base della stima del numero di veicoli presenti in coda e del valore della portata ridotta che varia in funzione della tecnologia di scavo considerata;
- *passo 4*: si misura il tempo di attraversamento T_{atr} del cantiere, sulla base della lunghezza che esso ha e su una stima della velocità di attraversamento dello stesso;
- *passo 5*: si valuta il tempo totale t_{ic} accumulato per veicolo a causa del cantiere sommando t_a (attesa in coda) e t_{atr} (attraversamento);
- *passo 6*: si calcola il numero totale N_v di veicoli che subiscono questo ritardo per l'intero periodo dei lavori, moltiplicando la portata media Q_{me} dell'arteria per la durata dell'apertura del cantiere;
- *passo 7*: si ricava la perdita di tempo T_t complessiva per tutti i veicoli che percorrono l'arteria moltiplicando il numero totale di veicoli N_v per il valore di t_{ic} (ritardo per il singolo veicolo accumulato) ricavato in precedenza;
- *passo 8*: si passa dal tempo al costo considerando un valore monetario associabile all'unità di tempo V_{amt}, ottenuto con valutazioni statistiche basate su dati ISTAT e relativi al guadagno orario medio dell'utente. Il valore di V_{amt} è fissato in 0,044€/minuto. Il costo complessivo C_{mpt} da attribuire al cantiere per la perdita di tempo è quindi il prodotto tra V_{amt} e la perdita di tempo totale T_t:

$$C_{mpt} = 0{,}044 \text{ €} \; T_t$$

3.3 Costo dell'impatto ambientale

Il costo dell'impatto ambientale quantifica gli effetti negativi dovuti ai fenomeni di inquinamento delle risorse quali acqua, aria, suolo e consumo di risorse non rinnovabili mediante l'utilizzo di un'analisi LCA.

Con queste valutazioni non si considerano gli impatti di tipo visivo o acustico ma solo quelli di consumo di risorse e di produzione di emissioni legati ai processi e ai prodotti utilizzati.

La stima dell'impatto ambientale è stata effettuata considerando gli effetti negativi legati all'utilizzo delle diverse tecnologie di scavo e all'incremento del traffico viario.

Dai risultati degli inventari sono stati elaborati gli indici di impatto ambientale economico con la metodologia EPS [4]. Essa consiste nell'effettuare un'associazione tra i dati di emissioni e dei consumi di risorse e i vari effetti ambientali (effetto serra, acidificazione, buco dell'ozono), per poi passare a elaborare un indice unico di impatto che è in pratica una somma pesata dei diversi contributi all'impatto ambientale forniti da tutti gli effetti considerati. Questo indice globale, contrariamente ad altri metodi *(EcoPoint, EcoIndicator)* che utilizzano unità

di misura adimensionali, è monetizzabile ed è esprimibile in euro.

Il costo di impatto ambientale può essere perciò considerato come somma di due elementi:

- costo d'impatto ambientale legato all'utilizzo delle tecnologie di scavo;
- costo riconducibile all'incremento del traffico viario.

4. Caso allo studio: applicazione del modello alle tecniche per la realizzazione degli impianti per le telecomunicazioni

Il modello economico e ambientale mostrato in precedenza è applicato qui di seguito per confrontare lo scavo tradizionale con le tre tecniche realizzative innovative prima descritte (perforazione orizzontale guidata, microtrincea e minitrincea).

4.1 Ipotesi di calcolo

Per questo studio sono state formulate le seguenti ipotesi:

- strada in zona semicentrale composta da due corsie per senso di marcia, ciascuna delle quali larga 2,70 m. Ogni senso di marcia ha quindi una carreggiata larga 5,40 m;
- larghezza del cantiere dello scavo tradizionale (parallelo) pari a 4,80 m dal ciglio della strada (distanza dalla mezzeria 60 cm). In presenza di un cantiere le corsie si riducono perciò a due;
- larghezza del cantiere per la perforazione orizzontale guidata, microtrincea e minitrincea (parallela) di 1,80 m dal ciglio della strada (distanza dalla mezzeria 3,60 m). Sono quindi tre le corsie in presenza del cantiere. Inoltre, l'area occupata dal cantiere nel caso di utilizzo di perforazione orizzontale guidata è limitata alla zona di partenza della perforazione - coincidente con quella di posizionamento della macchina - e a quella di arrivo della perforazione, coincidente con la buca di uscita della perforazione. Nella valutazione si è ritenuto comunque che, anche in questo caso, il numero di corsie residue si riduca a tre in quanto il disagio ai veicoli causato dal cantiere è paragonabile a quello attribuibile alle tecniche di microtrincea o di minitrincea;
- lunghezza dello scavo di 150 m;
- durata dei lavori di sette giorni lavorativi per uno scavo tradizionale, di tre giorni per perforazione orizzontale guidata e minitrincea, di un giorno per l'esecuzione di lavori con la tecnica microtrincea;
- impiego dei valori riportati in bibliografia [1] per valutare la portata tipica della strada.

La riduzione di corsie, legata alla presenza del cantiere con i differenti tipi di scavo considerati, è mostrata nella figura 5.

Esempio di scavo tradizionale.

4.2 Stima dei costi

4.2.1 Costo d'installazione

Nella tabella 1 sono riportati i costi d'installazione, ricavati a partire dai dati disponibili in Telecom Italia. I valori sono quelli medi, ottenuti dai costi relativi agli scavi per la rete di accesso e per quella di trasporto, che sono leggermente differenti tra loro.

Nel costo di installazione della tecnologia microtrincea è anche compreso quello relativo all'infrastruttura posata (cavo in fibra ottica).

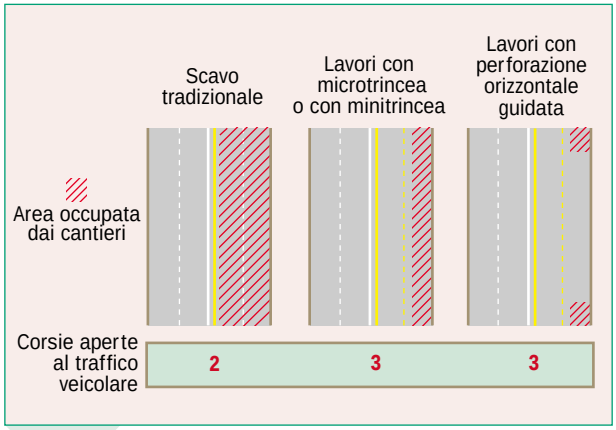


Figura 5 *Occupazione dei cantieri al variare della tecnica di posa.*

Tipo di intervento		
Perforazione orizzontale guidata	Microtrincea	Minitrincea
– 29	– 78	– 64

Tabella 1 *Risparmi percentuali rispetto al costo con lo scavo tradizionale.*

4.2.2 Costo del maggior tempo di percorrenza

Per questo tipo di costo, sono stati ottenuti i valori riportati nella tabella 2. Per i dati relativi alla portata media giornaliera ed a quella ridotta, causata dai lavori, sono forniti i campi di escursione, poiché i vari coefficienti di riduzione sono disponibili in letteratura come intervalli di valori e non come valore unico.
Nella tabella 3, è riportato il confronto dei dati di costo per il maggior tempo di percorrenza C_{mpt}, riferiti alle tecniche di scavo analizzate nell'esempio esaminato.

Parametro	Descrizione	VALORI CALCOLATI NEL MODELLO		
		Scavo tradizionale	Perforazione orizzontale guidata e minitrincea	Microtrincea
Q_{me}	Portata media sulle 24 ore (veicoli/h)	612-1374	612-1374	612-1374
Q_r	Riduzione della portata in presenza di ostacolo per singola corsia (veicoli/h)	445-633	543-759	543-759
n_r	Numero di corsie residue	2	3	3
t_a	Tempo di attesa in coda (s)	284,3	158,1	158,1
t_{atr}	Tempo di attraversamento (s)	36	36	36
t_{ic}	Tempo totale per veicolo (s)	320	194	194
N_v	Totale dei veicoli coinvolti dal cantiere	102816	44064	14688
T_t	Ritardo accumulato per la durata del cantiere (h)	9149,5	2375,8	791,9
C_{mpt}	Costo totale associato al tempo T_t (€)	24.000	6.400	2.100

Tabella 2 *Variazione dei parametri per diversi tipi di scavo legati alla diversa percorrenza per veicolo.*

4.2.3 Costo del maggior consumo di carburante

I valori associati al numero di veicoli e al ritardo complessivo sono quelli già calcolati nel computo del maggior tempo di percorrenza (tabella 3). Per ciò che concerne i dati di costo per durata unitaria di funzionamento, è stata considerata una potenza media del motore pari a 50kW, un consumo medio specifico di 50 g/kWh e un costo indicativo del carburante pari a 1,084 €/litro. La tabella 4 mostra i risultati complessivi del calcolo, sempre riferiti alla lunghezza totale del cantiere.

Sommando i due elementi di costo (legati al maggior tempo di percorrenza e al maggior consumo di carburante), si ottiene il costo globale generato dall'incremento del traffico viario a causa della presenza del cantiere riportato nella tabella 5.

Tipo di intervento		
Perforazione orizzontale guidata	Microtrincea	Minitrincea
– 74	– 91	– 74

Tabella 3 *Risparmi percentuali dei costi per il maggior tempo di percorrenza, riferiti a uno scavo tradizionale.*

Tipo di intervento		
Perforazione orizzontale guidata	Microtrincea	Minitrincea
– 74	– 91	– 74

Tabella 5 *Risparmi percentuali dei costi globali dovuti al traffico viario, riferiti a uno scavo tradizionale.*

Tipo di intervento		
Perforazione orizzontale guidata	Microtrincea	Minitrincea
– 74	– 91	– 74

Tabella 4 *Risparmi percentuali dei costi per il consumo di carburante, riferiti a uno scavo tradizionale.*

4.2.4 Costo di impatto ambientale legato all'utilizzo delle diverse tecniche di scavo

Per quanto riguarda il costo di impatto ambientale associato alle tecniche di scavo, i dati raccolti per effettuare l'analisi sono essenzialmente il consumo di combustibile delle varie attrezzature usate per l'esecuzione dei lavori e la composizione dei materiali utilizzati, così come descritto nei processi riportati in precedenza. Si è deciso, in particolare, di tralasciare la fase di "posa dell'infrastruttura" in quanto considerata troppo variabile in funzione delle caratteristiche peculiari del cantiere. A queste informazioni sono poi associate, tramite l'utilizzo di banche dati, i consumi e le emissioni legate alla produzione dei combustibili e dei vari prodotti (bitumi, collanti).

I risultati ottenuti sono riportati in tabella 6.

Il costo relativo all'impatto ambientale, legato allo scavo di tipo tradizionale, è superiore a quello

Tipo di intervento		
Perforazione orizzontale guidata	Microtrincea	Minitrincea
– 84	– 74	– 82

Tabella 6 *Risparmi percentuali dei costi per il consumo dei materiali, riferiti ad uno scavo tradizionale.*

attribuibile alle tecniche innovative. L'origine dell'impatto per le diverse tecniche di scavo è connessa soprattutto al consumo complessivo di combustibili fossili, necessari per alimentare le attrezzature utilizzate. Nel caso dello scavo tradizionale il consumo è più di quattro volte superiore rispetto agli scavi con minitrincea, di oltre sette volte se confrontato con il consumo relativo a uno scavo con perforazione orizzontale guidata e di sedici volte rispetto allo scavo microtrincea.

Per quanto riguarda poi i consumi energetici, le emissioni in aria e i rilasci in acqua, la tecnica che causa l'impatto più contenuto è il microtrincea, seguita dalla perforazione orizzontale guidata e dalla minitrincea. Nel caso di scavi eseguiti con il microtrincea si ha infatti un maggior consumo di risorse naturali, dovuto al cordolo in gomma utilizzato per il riempimento e considerato in questa analisi; in particolare si hanno i consumi di bauxite, ferro e zinco.

Queste differenze si ritrovano anche nella fase di valorizzazione del singolo impatto, e causano gli scostamenti dell'indice globale EPS *(Environmental Priority Strategies)*, mostrati nella tabella 6.

Occorre sottolineare, da ultimo, che nel metodo EPS utilizzato per la stima dei costi ambientali non sono considerati i rifiuti solidi prodotti.

Questo parametro tuttavia è di particolare rilievo nel caso degli scavi, per quanto riguarda la terra destinata a discarica. Dai dati di inventario emerge infatti che, per ogni metro lineare di scavo, sono avviati a discarica oltre 480 kg di terra nel caso dello scavo tradizionale, contro i circa 40 kg dello scavo con minitrincea. Questa differenza comporta anche un consumo aggiuntivo di risorse, legato al prelievo di nuovo materiale da utilizzare per il ripristino (ghiaia, sabbia, oltre che bitume). L'ulteriore utilizzo di materie prime è computato nell'analisi LCA.

Nel caso di scavi con perforazione orizzontale guidata o con microtrincea non si hanno invece rifiuti solidi di questo tipo, in quanto con la prima tecnica non si effettua asportazione di materiale del sottosuolo durante la perforazione, mentre con la seconda il quantitativo asportato è di entità trascurabile.

4.2.5 Costo d'impatto ambientale legato all'incremento del traffico viario

Il calcolo dell'impatto ambientale legato all'incremento del traffico è stato effettuato partendo dai dati relativi al maggior consumo di carburante dei veicoli in coda che variano a seconda della tecnologia di scavo impiegata. L'applicazione a tali quantità di una valutazione LCA e la successiva monetizzazione mediante la metodologia EPS dei risultati ottenuti ha permesso di quantificare il costo dell'impatto ambientale attribuibile alle emissioni in aria dovute alla combustione del carburante. I risultati di costo ottenuti sono presentati nella tabella 7.

Come si vede nella tabella, anche in questo caso il costo ambientale legato all'incremento del traffico viario, legato alla scavo tradizionale, è circa il triplo di quello relativo a scavi con perforazione orizzontale guidata o con la minitrincea e di nove volte rispetto a scavi con microtrincea.

Tipo di intervento		
Perforazione orizzontale guidata	Microtrincea	Minitrincea
– 74	– 90	– 74

Tabella 7 *Risparmio percentuale del costo di impatto ambientale legato all'incremento del traffico riferito a uno scavo tradizionale.*

L'origine dei singoli impatti e la conseguente stima dei costi deve essere attribuita al maggior tempo di percorrenza in coda dei veicoli e al conseguente incremento di emissioni legate ai gas di scarico.

Due sono i fattori che influenzano il costo: la riduzione delle corsie e la durata del cantiere.

Nello scavo tradizionale la strada risulta infatti percorribile solo su una corsia per senso di marcia e il cantiere è aperto per circa una settimana. Per scavi con altre tecniche le corsie si riducono a tre, ma nel caso della perforazione orizzontale guidata e della minitrincea i lavori del cantiere durano tre giorni, contro una sola giornata impiegata nel microtrincea.

Sommando i due elementi di costo calcolati, si ottiene il costo d'impatto ambientale riportato in tabella 8.

4.3 Analisi dei risultati

Nella tabella 9 sono riportati, in percentuale, i

Tipo di intervento		
Perforazione orizzontale guidata	Microtrincea	Minitrincea
– 75	– 90	– 74

Tabella 8 *Risparmi percentuali del costo totale di impatto ambientale riferito a uno scavo tradizionale.*

risparmi legati all'utilizzo di tecniche innovative rispetto allo scavo a cielo aperto, sia per componente di costo, sia per il costo totale. Questa indicazione è la più importante, in quanto i valori assoluti sono rappresentativi solo se associati alle ipotesi poste alla base del caso in esame.

Un confronto percentuale permette, invece, di avere un'indicazione sulle differenze di costo associate all'utilizzo delle varie tecniche a parità di condizioni iniziali.

In termini assoluti, la tabella 9 mostra che il costo dello scavo a cielo aperto risulta il triplo rispetto a quello legato alla perforazione orizzon-

Tipologia di costo	Tipo di intervento		
	Perforazione orizzontale guidata	Microtrincea	Minitrincea
installazione	– 29	– 78	– 64
Costo legato all'incremento del traffico viario	– 74	– 91	– 74
Impatto ambientale	– 75	– 90	– 74
Risparmio percentuale	**– 69**	**– 90**	**– 73**

Tabella 9 *Risparmi percentuali dei costi tra le diverse tecniche di posa riferito allo scavo tradizionale.*

tale guidata, quasi il quadruplo rispetto alla minitrincea e dieci volte di quello relativo allo scavo microtrincea.

Come è mostrato nel grafico di figura 6, il costo legato alla congestione del traffico viario ha un impatto notevole sul costo totale per tutte le tecniche di scavo considerate. Esso incide, in particolare, notevolmente per la tecnologia di scavo a cielo aperto in quanto l'impatto sul traffico è maggiore: la strada risulta percorribile solo su una corsia per senso di marcia e di conseguenza si può avere la formazione di code in ampie fasce orarie, con maggiori tempi di percorrenza impiegati dagli automobilisti e un maggior consumo di carburante.

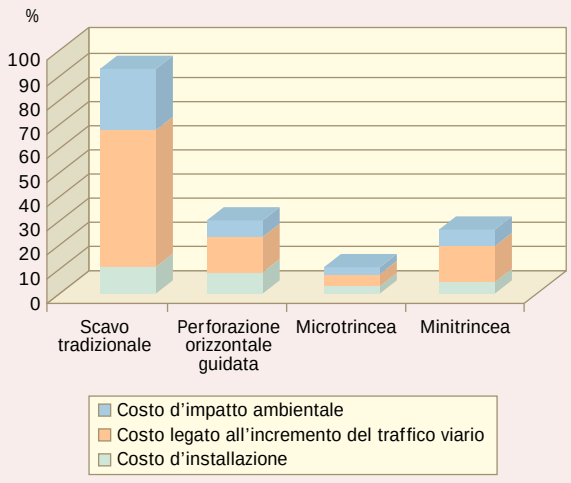


Figura 6 *Confronto percentuale tra i costi per diverse tecniche di scavo.*

Le altre tecniche di scavo rendono invece disponibili complessivamente tre corsie e, quindi, la formazione di coda riguarda essenzialmente le ore di punta. L'incidenza di questa voce si riflette in un impatto minore, anche se pur sempre considerevole, sul costo totale.

Anche per quanto riguarda il costo legato all'impatto ambientale, la componente predominante è rappresentata dalle emissioni associate all'incremento del traffico viario.

Le condizioni prima indicate influenzano sensibilmente lo scavo tradizionale mentre la discriminante principale tra le tecniche innovative è legata alla durata dei cantieri. Nell'ambito di questa componente temporale non sono stati considerati i costi legati all'impatto visivo e all'inquinamento acustico.

I costi sociali e ambientali, essendo valori "virtuali", non sarebbero d'altra parte propriamente sommabili ai costi oggi sostenuti da una generica azienda, quali per esempio quelli d'installazione.

Uno scavo con microtrincea.

Si ritiene, tuttavia, conveniente effettuare questa somma per avere un indicatore di costo unico che permetta di valutare le esternalità associate alle singole attività.

Per ciò che concerne, infine, il costo di installazione, l'utilizzo di tecniche innovative tipo no-dig (a parità di parametri progettuali e facendo una valutazione complessiva dell'intervento), consentono un risparmio rispetto alla realizzazione di attraversamenti con scavi a cielo aperto, sui quali incide in misura considerevole l'esecuzione dei ripristini e dell'ambiente per le aree interessate dai lavori (demolizione e ripristino della pavimentazione stradale, maggiore incidenza dei costi legati alla manodopera, maggiore peso del costo di scavo in corrispondenza di roccia o di determinate tipologie di terreno).

Foto: Sirti

Perforazione orizzontale guidata in ambito urbano.

Questi aggravi di costo si riflettono, in particolare, maggiormente nella tecnica della minitrincea e in quella microtrincea per le quali il costo di installazione ha un'incidenza relativamente bassa sul costo totale; l'applicazione di tali tecniche sono però ancora limitati. Un'ultima considerazione riguarda la profondità dello scavo a cielo aperto il cui costo risulta crescere naturalmente in maniera drastica con l'aumentare della profondità.

5. Conclusioni

In questo articolo è stato presentato un modello di valutazione tecnica ed economica che, tenendo conto degli impatti legati all'apertura dei cantieri sulla vivibilità normale, consente di stimare le esternalità associate alle diverse tecniche impiegate nella realizzazione degli impianti. Il modello è stato applicato a un esame rappresentativo di una situazione tipica di quella che in genere si trova ad affrontare e risolvere un gestore di una rete di telecomunicazioni.

Sono stati in particolare esaminati il costo di installazione, quello legato all'incremento del traffico viario e quello di impatto ambientale per quattro differenti tipologie di possibili tecniche utilizzate nelle realizzazioni degli impianti: scavo a cielo aperto, perforazione orizzontale guidata, microtrincea e minitrincea.

Lo scopo del modello è di permettere di conoscere i costi esterni associati alle attività di scavo; queste valutazioni rappresentano uno dei passi fondamentali per un'azienda che voglia perseguire la compatibilità tra crescita competitiva e tutela dell'ambiente. In generale è possibile osservare come le esternalità (costo legato all'incremento del traffico viario e costo d'impatto ambientale) rappresentino un valore tutt'altro che trascurabile se rapportato ai costi di installazione, gli unici effettivamente sostenuti. In particolare i costi sociali e ambientali incidono notevolmente nel caso si impieghi la tecnica di scavo a cielo aperto, in quanto la strada risulta percorribile solo su una corsia per senso di marcia e la durata dei lavori è maggiore. Si può prevedere la formazione di code durante buona parte della giornata, con maggiore perdita di tempo subita dagli utenti della strada, maggior consumo di carburante e maggiori emissioni nocive in aria.

Le altre tecniche rendono, invece, disponibili complessivamente tre corsie e, quindi, la formazione di code si verifica solo nelle ore di punta.

Si ricorda che le esternalità sono state generate da modelli di previsioni, quindi, pur essendo indicative dei risparmi economici ottenibili con l'utilizzo

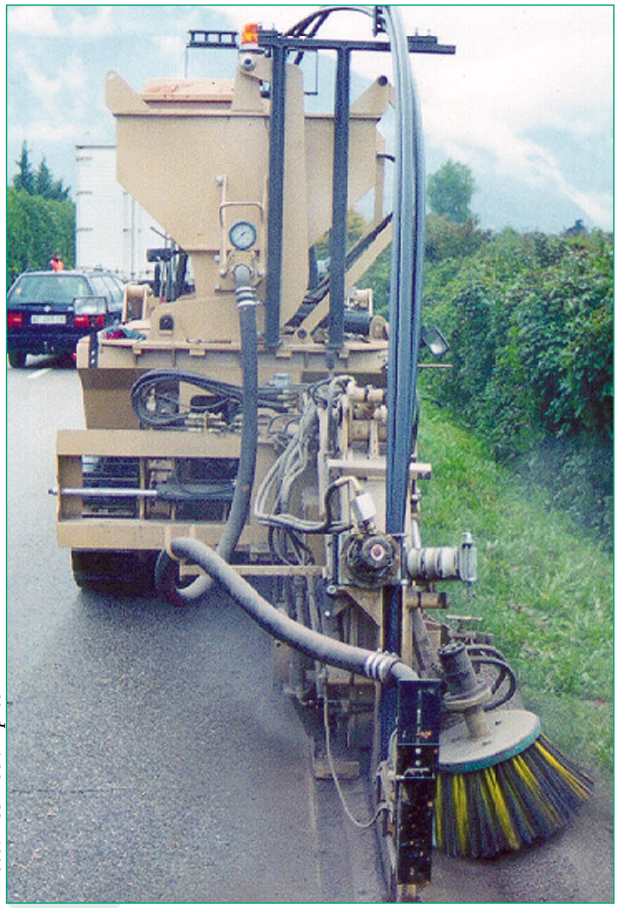


Foto: Tesmec /Alpitel

Esempio di scavo con minitrincea.

di tecniche innovative, non sono da considerare stime "esatte" in valore assoluto, ma utili per confronti diretti tra i diversi sistemi di scavo impiegabili.

Considerando, inoltre, che da diversi settori dell'opinione pubblica e delle Istituzioni sia nazionali sia internazionali si stanno avanzando richieste circa l'inserimento delle esternalità all'interno dei valori di mercato dei "beni", e i risultati riportati nell'articolo forniscono un'indicazione preziosa sull'entità dei costi che un'azienda, in un futuro prossimo, dovrebbe sostenere di fronte alla necessità di effettuare nuovi investimenti per realizzare nuovi impianti.

Telecom Italia considera il rispetto dell'ambiente come uno dei valori di riferimento e ritiene di poter contribuire concretamente a migliorarne la qualità, seguendo la strada dello sviluppo sostenibile.

Telecom Italia, è disponibile a condividere le proprie esperienze con altri gestori di servizi a rete sia con le Istituzioni sia con gli Enti preposti alla tutela dell'ambiente.

L'impiego diffuso di queste tecnologie può rappresentare un aspetto fondamentale per il miglioramento della vivibilità soprattutto nei centri urbani. In tale ottica si rende opportuno un ruolo da protagonista degli Enti responsabili per adoperarsi con azioni mirate (sgravi fiscali, incentivi, direttive di legge) che incoraggino l'utilizzo di queste nuove tecniche che favoriscono il vivere civile.

Abbreviazioni

EPS	Enviromental Priority Strategies
LCA	Life Cycle Assessment

Luca Giacomello si è laureato in Ingegneria Elettronica nel 1992 presso il Politecnico di Torino. Ha poi ottenuto uno speciale diploma di Quality Management e nel 1995 un master in telecomunicazioni al COREP - Politecnico di Torino. Ha iniziato l'attività operativa nel settore qualità; ha sviluppato alcuni studi sui sistemi di gestione ambientale nel campo delle valutazioni delle prestazioni di prodotti e servizi usando il metodo *LCA (Life Cycle Assessment)*. Dal 2001 lavora sull'applicazione dei servizi di home networking. È autore di alcuni articoli sull'impiego del sitema LCA applicato alle telecomunicazioni. È inoltre membro del Gruppo Ambiente in ambito CEI e del gruppo di lavoro Ecolabel -LCA, nella Commissione Ambiente dell'*UNI (Unione Normativa Italiana)*. È socio fondatore dell'associazione italiana per lo sviluppo dell'LCA.

Bibliografia

[1] Chirulli, R.; Caruso, A.: *Un modello di analisi tecnico-economica nel confronto tra directional drilling e scavo a cielo aperto*. Atti del Convegno "Stato dell'Arte e nuove possibilità applicative del Directional Drilling", Bari, 11-12 maggio 1998.

[2] Colonna, P.; Tragni, O.: *Il directional drilling: la tecnica, i campi di impiego, le nuove possibilità applicative*. Atti del Convegno "Stato dell'Arte e nuove possibilità applicative del Directional Drilling", Bari, 11-12 maggio 1998.

[3] *Norma Tecnica sulle perforazioni orizzontali guidate*. Telecom Italia, Ediz. aprile 2001.

[4] *EPS-Calculations, Environmental Priority Strategies*. Federation of Swedish Industries, N. G. Westerlund, settembre 1995.

Paolo Trombetti si è diplomato in Meccanica di precisione; assunto in SIP nel 1987, fino al 1994 si è occupato dello studio e progettazione dei materiali accessori dei cavi in rame, effettuando anche valutazioni di tipo economico. Dal 1994 segue la stesura delle specifiche e della valutazione economica dei materiali per la realizzazione delle infrastrutture della rete di TLC in fibra ottica e rame, compresi gli aspetti installativi di tali materiali. Ha partecipato alla stesura delle norme del progetto Socrate. Cura la definizione delle norme di realizzazione della rete di accesso in fibra ottica, con particolare riguardo alle nuove tecnologie di posa dei cavi e delle infrastrutture (trenchless tecnology, microtrincea, minitrincea, tecnologie radar, ecc.) per la realizzazione della rete in ambito urbano ed extraurbano; di tale attività segue gli aspetti di sperimentazione, sviluppo e valutazione economica. Nel 1998 ha ricoperto la carica di segretario della *I.A.T.T. (Italian Association for Trenchless Tecnology)* l'Associazione Italiana delle tecnologie trenchless, e dal 2000 ne é il Presidente. Ha partecipato come docente a corsi sulle tecnologie impiegate da Telecom Italia nella realizzazione delle infrastrutture per reti di TLC. Ha collaborato all' organizzazione, ed è stato relatore, di diversi convegni nazionali ed internazionali sulle trenchless technology. Ha pubblicato, su riviste del settore, diversi articoli sull'impiego delle tecnologie trenchless nel campo delle telecomunicazioni. È Rapporteur presso l'*ITU (International Telecommunication Union)* del SG6 per le tecnologie alternative di posa delle infrastrutture di telecomunicazione. Nel Comitato Tecnico 306 del *CEI (Comitato Elettrotecnico Italiano)* è responsabile di un gruppo di lavoro per la redazione delle norme d'impiego delle perforazioni orizzontali guidate (posa di cavi e infrastrutture di telecomunicazione ed energia), e per quelle relative all'impiego del radar per l'introspezione del sottosuolo. Frequenta il corso di laurea di Economia Aziendale, presso la 3ª Università di Roma.

The harmony of innovation and profit in access

XIV^th International Symposium on Services and Local accesS

Seul, Corea, 14 - 18 aprile 2002

Francesco Costantino, Paolo Impiglia, Francesco Silletta

"The last year has seen the telecommunications world undergo tremendous change. The era of deregulation and increased competion has profoundly changed the structure and nature of the industry....While the fundamental direction of the industry continues towards the broadband services future, current economic realities are providing interesting challenges."

Così ha aperto i lavori Frank Mellor, Presidente dell'International Technical Committee.

Tra danze folcloristiche e rulli di tamburi si è aperta domenica 14 aprile di quest'anno a Seoul (Corea del Sud), la quattordicesima edizione dell'*International Symposium on Services and Local access* denominata *ISSLS 2002*.

La Conferenza, pur rimanendo tra gli eventi biennali più interessanti e più densi di contenuti sul tema della rete di accesso e dei servizi, ha posto in luce come l'apertura del mercato delle telecomunicazioni abbia anche contribuito a rendere più rara la volontà di condividere informazioni da parte degli autori delle memorie. Ciò non ostante, anche quest'anno, sono stati presentati alcuni interventi particolarmente interessanti, sia nelle sessioni dedicate agli aspetti di mercato e dei servizi, sia nelle sessioni più propriamente tecniche.

Un workshop particolare è stato dedicato alla presentazione dello stato della larga banda in Corea, e gli elementi di maggior rilievo emersi nell'esposizione sono stati riportati nel riquadro di pagina 118. Nella presentazione del workshop il Presidente e Amministratore delegato di Korea Telekom, Sang-Chul Lee, ha enfatizzato il ruolo strategico della larga banda, in particolare dell'ADSL, nello sviluppo socioeconomico del Paese.

Al symposium hanno partecipato circa 450 persone e di questi una metà era costituita da coreani, segno evidente dell'elevato interesse a livello locale sul tema trattato. Buona è stata la presenza a livello internazionale, anche se più ridotta rispetto alle precedenti edizioni. Per il Gruppo Telecom Italia, oltre agli estensori di questa relazione, ha partecipato Carlo Mazzetti membro dell'International Technical Committee e Chairman della sessione Business Aspects. Nel corso della Conferenza è stata presentata una memoria sui servizi *wholesale* di accesso a larga banda da Francesco Costantino e Francesco Silletta su: "Telecom Wholesale Service scenarios for new Broad Band access to IP and Data Network Services".

L'importanza dell'accesso nella rete (sotto il poster del quattordicesimo simposio ISSL 2002, Seoul, Corea).

1. Aspetti di business

Nelle sessioni maggiormente orientate all'analisi degli aspetti di mercato legati ai servizi di accesso, gli interventi presentati sono riconducibili a tre aree geografiche: Stati Uniti d'America, Europa - e in particolare a Nord Europa - e Corea. Le tematiche affrontate nel Convegno riguardano l'analisi tecnico-economica di servizi basati su tecnologie con elevato impatto in termini di investimenti iniziali, quali *VDSL (Very high bit-rate Digital Subscriber Line)* e *GbE (Gigabit Ethernet)*, e l'analisi di servizi ibridi in quanto utilizzano sia la rete fissa sia quella mobile.

1.1 Scenario USA

Particolare interesse ha suscitato l'intervento di *M. Schwartz* di *Telcordia (The changing landscape for incumbent service providers)*, che ha presentato un quadro sugli impatti dell'alta competizione in cui si opera oggi nel mercato nordamericano, e le strategie che gli operatori, in particolare quelli tradizionali, stanno adottando per conservare le proprie quote di mercato.

L'analisi finanziaria delle prestazioni degli operatori degli Stati Uniti ha mostrato profondi cambiamenti negli scenari di business, essenzialmente causati dal passaggio degli utenti della telefonia fissa verso quella

mobile, e alla progressiva migrazione dei ricavi dai servizi in fonia verso quelli dei dati.
La contrazione dei ricavi sulla telefonia fissa è anche causata dalla riduzione del traffico Internet di tipo *dial-up*, a fronte di un incremento di accessi *xDSL (x Digital Subscriber Line)* e *FTT-x (Fiber To The-x)*, e ad un trasferimento graduale verso la trasmissione digitale a pacchetto.
La riduzione dei ricavi è stata, poi, fortemente condizionata dalla diminuzione dei prezzi per i servizi di trasporto di dati, dovuta sia a una diminuzione dei costi per megabit trasportato, sia a un aumento della competizione.
Le contromisure che si stanno adottando sono basate fondamentalmente sui seguenti fattori:

- *riduzione dei costi operativi:* soprattutto sui *Contact Center* e sugli interventi presso la sede del cliente, introducendo portali Web mediante i quali sono gestiti gli ordini, il tracciamento delle attività di manutenzione e la fatturazione. Altre aree di intervento riguarderanno l'adozione di soluzioni basate sul trasporto della voce su rete dati che utilizzano xDSL, con un utilizzo congiunto quindi dello stesso doppino nella rete d'accesso;
- *sviluppo di reti IP (Internet Protocol):* è previsto che esse siano gestite con l'integrazione di più tecnologie di trasporto su un unico strato *MPLS (MultiProtocol Label Switching)*. Le tecnologie, basate su GbE e *PON (Passive Optical Network)*, potranno in futuro consentire una riduzione nei costi dell'accesso per la crescita tumultuosa della disponibilità di nuovi apparati di multiplazione. La rete dati dell'operatore dominante può così essere utilizzata per integrare le reti fissa e mobile, con la rete Internet per la fornitura di servizi avanzati, attraverso l'adozione di un unico *back-bone* e l'introduzione di nodi di controllo del servizio e di piattaforme di gestione innovative;
- *fornitura di servizi dati avanzati:* lo sviluppo di contenuti e di applicazioni a valore aggiunto per il cliente finale costituiranno una grande opportunità per i *network provider*, nella fornitura di servizi con un'elevata redditività basati su un'offerta di alcune caratteristiche primarie quali, ad esempio: qualità, sicurezza, servizi di *directory*, gestione di piattaforme avanzate per la fornitura del servizio.

Un momento del Symposium.

1.2 Sviluppo di servizi xDSL

Numerosi interventi (soprattutto da parte di operatori del Nord Europa, in particolare Telenor, che ha presentato diversi articoli) sono stati indirizzati verso le analisi tecnico-economiche per lo sviluppo di servizi a larga banda in aree poco competitive, mostrando da un lato un certa attenzione, anche politica, per lo sviluppo della larga banda come servizio universale, e, allo stesso tempo, un'attenta e puntuale valutazione dei ritorni economici sugli investimenti.
I servizi interattivi, la larga banda, la trasmissione televisiva digitale diffusiva e il telelavoro sono, infatti, considerati fattori rilevanti per lo sviluppo di questi servizi nelle regioni periferiche e rurali. In diversi Paesi europei (e non solo), sono stati manifestati interessi del mondo politico per lo sviluppo di servizi a larga banda, in quanto anche in quei casi in cui essi non fossero economicamente convenienti nel breve periodo, questo tipo di investimenti sono, infatti, considerati un fattore trainante per lo sviluppo generale del territorio interessato dall'innovazione.
Per valutare queste problematiche è stata condotta una dettagliata analisi tecnico-economica per lo sviluppo di servizi a larga banda in aree non competitive. Quest'indagine è stata effettuata nell'ambito del progetto *IST (Information Society Technology)* chiamato *TONIC (TechnO ecoNomICs of IP optimised network and services)* ed è stata presentata da *Kalhagen* di *Telenor (Provision of broadband services in non-competitive areas in Western European countries)*.
Per quest'anno sono stati studiati tre modelli relativi rispettivamente al nord, al centro e al sud dell'Europa, definiti sulla base della densità della popolazione e delle caratteristiche socio-economiche e demografiche.
In particolare, per quanto riguarda l'analisi del modello definito per i Paesi del sud

IL SUCCESSO DEI SERVIZI A LARGA BANDA IN COREA DEL SUD

La Corea del Sud è il Paese che ha, su base mondiale, il più alto grado di penetrazione dei servizi a larga banda. Circa 23 milioni di abitanti, metà della popolazione (che oggi è, infatti, di circa 49 milioni), utilizzano Internet. Gli utenti che impiegano i servizi a larga banda sono circa 8 milioni (figura A), e di questi il 65 per cento utilizza i servizi xDSL, mentre il resto impiega i servizi *CATV (CAble TeleVision)* e ISDN.

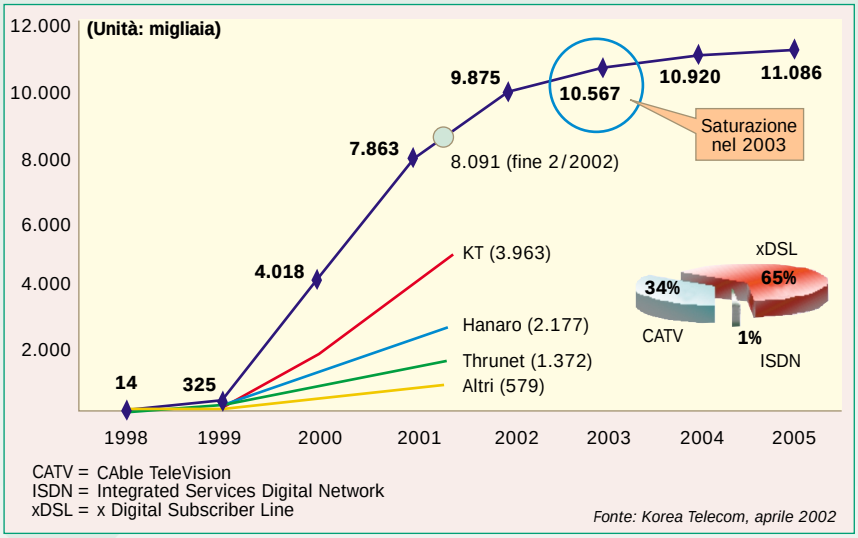


Figura A *Previsione di crescita del mercato a larga banda in Corea.*

L'operatore dominante, *Korea Telekom*, ha superato i quattro milioni di clienti connessi a larga banda, mentre i concorrenti, e in particolare *Hanaro* e *Thrunet*, raggiungono l'altra metà degli utenti a larga banda.

I fattori che hanno determinato il successo dei servizi Internet nella Corea del Sud sono essenzialmente tre: l'elevata crescita della domanda; la rapidità della fornitura del servizio e l'intervento da parte dello Stato nel promuovere e nell'incentivare la diffusione dei servizi Internet.

Una così grande spinta verso l'utilizzo esteso dei servizi Internet è dovuta alla popolarità dei servizi legati ai giochi on-line, all'introduzione di Internet nelle scuole e all'impiego crescente di Internet per le transazioni finanziarie (che attraverso Internet raggiunge il 70 per cento circa dell'intero volume di transazioni del mercato finanziario coreano).

	1	2	3	4	5
Collegamento a Internet (ore/mese)	Corea 16,2	Canada 10,5	USA 9,6	Germania 8,2	Giappone 7,6
Pagine visitate (visite/mese)	Corea 2.164	Hong Kong 1.123	Germania 818	Giappone 788	Canada 755

Tabella A *Utilizzo dei servizi Internet nei Paesi più evoluti.*

Gli utenti Internet coreani sono in testa alla classifica mondiale per l'utilizzo di Internet, con circa 16,2 ore di utilizzo medio mensile dei servizi Internet (tabella A).

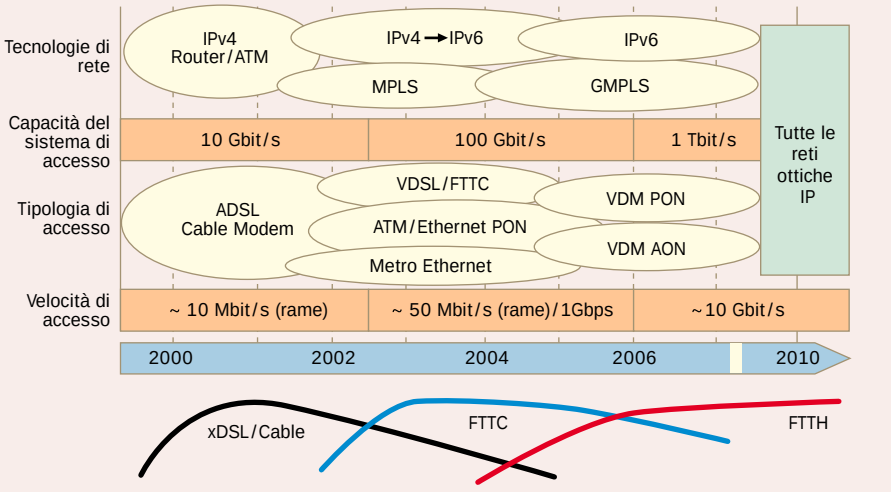
L'Autorità nazionale coreana per le telecomunicazioni ha poi completato il processo di liberalizzazione del mercato delle telecomunicazioni creando un mercato competitivo. Il Governo coreano ha, anche, promosso il progetto *e-Korea*, che ha consentito la realizzazione di un'infrastruttura Internet nazionale (gestita da Korea Telecom) che raggiunge circa 11.500 scuole e 20.700 agenzie.

Sebbene facciano impressione i dati di diffusione di Internet nella Corea del Sud, resta il problema del *Digital Divide*, che, a detta del presidente di *Korea Telecom*, *Sang Chul Lee*, non è causato dal prezzo ma dalla copertura del servizio sul territorio, in quanto essa non è stata ancora completata in alcune aree geografiche del Paese. Una possibile soluzione è rappresentata dal servizio Wireless LAN, che consentirà di estendere l'area di copertura dei servizi a larga banda ad aree impervie, difficilmente raggiungibili e che presenterebbero con i sistemi tradizionali costi di gestione elevati.

Il Presidente di Korea Telecom ha anche annunciato che nel corso dei prossimi anni saranno effettuati investimenti per integrare i servizi *wireline* e *wireless*. Questo nuovo modo di accesso integrato è stato identificato da Sang-Chul Lee con il concetto di *Logical Local Loop* ovvero l'accesso all'ultimo miglio "virtuale", che consente al Cliente l'utilizzo del servizio a larga banda in modo trasparente alla rete di accesso fissa (servizi xDSL, CATV o fibra) o attraverso la rete mobile (servizi Wireless LAN e 3G).

Il Ministro per le telecomunicazioni della Corea, intervenuto all'inaugurazione del convegno ISSLS, ha ribadito il successo dei servizi Internet in Corea e l'azione svolta dalla regolamentazione per la liberalizzazione del mercato. Nel corso del Convegno ha poi annunciato che l'attività di regolamentazione proseguirà, in modo da garantire la libera concorrenza del mercato delle telecomunicazioni e da estendere il più possibile l'utilizzo dei servizi a larga banda fra la popolazione coreana per ridurre, così, il *digital divide*. Il Ministero sta considerando, inoltre, la possibilità che per il 2005 venga esteso il servizio universale anche ai servizi a larga banda *(broadband universal service)*.

È stato, infine, presentato il piano di evoluzione della rete di accesso, dei sistemi di trasmissione impiegati nella rete di trasporto e dei servizi di rete che saranno utilizzati nel prossimo decennio. Le caratteristiche di maggior rilievo sono sintetizzate nella figura B.



ADSL = Asymmetric Digital Subscriber Line
AON = All Optical Network
ATM = Asynchronous Transfer Mode
FTTC = Fiber To The Curb
FTTH = Fiber To The Home
GMPLS = Generalized Multi Protocol Label Switching
IP = Internet Protocol
IPv4 = Internet Protocol versione 4
IPv6 = Internet Protocol versione 6
MPLS = Multi Protocol Label Switching
PON = Passive Optical Network
VDM = Video Digital Multimedia
VDSL = Very high bit-rate Digital Subscriber Line
xDSL = x Digital Subscriber Line

Figura B *Piano di sviluppo della rete di accesso coreana a larga banda.*

dell'Europa, è stata posta in evidenza un'elevata sensibilità dei risultati a parametri critici, quali i costi delle opere civili, le distanze tra gruppi di clienti, la penetrazione dei servizi a larga banda ed i prezzi. I risultati conseguiti mettono in luce la necessità di interventi governativi a supporto degli investimenti necessari in alcune aree del Paese dove, come si è detto, un'analisi del mercato potenziale da parte dei Service Provider difficilmente porterebbe allo sviluppo di servizi a larga banda.

Francesco Costantino di Telecom Italia interviene al dibattito seguito a una presentazione.

1.3 Sviluppo di servizi WLAN

Per quanto riguarda la tecnologia *WLAN (Wireless Local Area Network)* sono emersi essenzialmente due orientamenti per l'introduzione in rete. Il primo riguarda la possibilità di individuare modelli di business per gli operatori di sistemi mobili di terza generazione *(3G)*, che prevedono lo sviluppo di servizi basati inizialmente sull'*UMTS (Universal Mobile Telecommunications System)*, (2002÷2003), e successivamente (2004÷2005) su WLAN in aree caratterizzate da un'alta densità abitativa.
I risultati ottenuti nell'ambito del progetto *TONIC* hanno permesso di rilevare vantaggi significativi per un operatore 3G, che potrebbe così permettere anche servizi a larga banda tramite WLAN, migliorando l'offerta al cliente nel suo complesso. L'incremento degli investimenti, richiesto per la componente WLAN, è, infatti, trascurabile rispetto al valore totale, a fronte dei ricavi ottenibili.
Un secondo modello di business riguarda invece gli operatori di rete fissa che, dopo aver sviluppato già servizi a larga banda su xDSL, intendono ampliare le quote di mercato attraverso la fornitura di soluzioni *wireless*.

Gyeongbokgung Palace costruito sull'acqua nel 1394 sotto la dinastia Joseon (1392 - 1910).

È stato segnalato, in particolare, che due operatori in Corea stanno per avviare un massiccio sviluppo di servizi WLAN, dopo un periodo sperimentale iniziato nell'ottobre 2001 con circa quarantadue interventi *spot* su tutto il territorio nazionale, in modo da integrare l'offerta a larga banda su *ADSL (Asymmetric Digital Subscriber Line)* con una soluzione wireless in ambienti quali alberghi, aeroporti, centri congressi, centri commerciali.
Pur non essendo ancora definita la posizione dell'autorità locale in merito alla soluzione WLAN che opera nella gamma dei 5 GHz, è stato effettuato un primo esame per la definizione di una strategia di transizione da una soluzione basata sulla tecnologia che opera a 2,4 GHz a una allocata nella banda dei 5 GHz.

1.4 Sviluppo di servizi basati su VDSL

Diversi interventi hanno riguardato alcuni scenari di business nello sviluppo di servizi basati sulla tecnologia VDSL. L'introduzione di servizi basati su VDSL richiede, infatti, investimenti di rilievo, soprattutto in determinate aree geografiche, per cui è fondamentale determinare le strategie di sviluppo dei servizi nel breve, medio e lungo termine.
Da un punto di vista tecnico è stata messa in luce la necessità di integrare il più possibile le tecnologie xDSL in un'unica architettura.
Per quanto riguarda l'analisi economica delle soluzioni di rete oggi in discussione, *Nils Kristian Elnegaard* e *Kjell Stordahl* di *Telenor* hanno presentato un contributo *(Deciding on the right timing of VDSL roll-outs: a real options approach)* su alcuni approcci innovativi *(real option)*, adottati spesso nel mondo finanziario, per pervenire a modelli che permettano di valutare l'impatto,

che hanno sui flussi di cassa investimenti ingenti, la cui allocazione temporale influenza notevolmente tutta l'analisi.
Un altro contributo di *Haga*, *Johannessen*, *Meinich*, *Thrane*, *Ling*, *Andersson*, *Myrvold*, di *Telenor*, *(Towards deployment of VDSL for interactive broadband services: experiences from a large-scale consumer market trial)* ha illustrato come sia stato condotto un market trial di servizi a larga banda su VDSL che ha interessato circa 750 clienti e che ha permesso di rilevare un certo interesse per servizi di accesso a Internet, con connessione sempre attiva *(always-on)* ad alta velocità, forniti assieme a più canali video in una stessa abitazione. In questo caso, non sono state ritenute soddisfacenti soluzioni basate su ADSL.
Nella presentazione è stato anche sottolineato che i principali temi da approfondire riguardano il cablaggio domestico e quanto legato alla realizzazione di impianti nelle installazioni di sistemi VDSL da cabinet.

2. Aspetti tecnici

Le sessioni più propriamente tecniche sono state centrate sulle tecnologie xDSL e sull'accesso in fibra ottica. In particolare sul tema ADSL molto interesse ha suscitato lo studio di *Kyung-Ah Han*, *Young-Chul Cha* e *Jae-Jin Lee*, ricercatori del *Korea Telecom Access Network Laboratory (xDSL Network Evolution Strategies to Minimize Crosstalk of FTTC-ADSL)*.
Esso è dedicato all'analisi degli effetti di interferenza di linee ADSL.
L'indagine conclude evidenziando che l'utilizzo in una stessa rete in rame di terminazioni ADSL sia da centrale sia da cabinet - situazione tipica in regime di *sub-loop unbundling* - ha i principali svantaggi e le maggiori limitazioni nelle interferenze causate dalla diafonia del portante, generate dalla seconda soluzione *(da cabinet)* nei confronti della prima *(da centrale)*. A un'analoga conclusione, che prendeva in esame anche gli effetti di altre tecnologie, quali HDSL e ISDN giunge un contributo congiunto presentato da *Nedev*, *Daley*, *McLaughlin* e *Laurenson de*lla *Fujitsu Telecomunications* e dell'*Università di Edimburgo (The Impact of Crosstalk in an Unbundled Environment on ATM and IP)*.
In una ulteriore memoria di *Sven Symalla* e *Thomas Kessler* di *T-Systems GmbH (Spectral Management in an Unbundled Environment)* lo stesso aspetto è stato preso in considerazione dal punto di vista della gestione dello spettro: partendo dall'ipotesi che la gestione corretta di queste situazioni è particolarmente importante in scenari in cui è fornito in misura estesa l'*unbundling*; nel testo è proposta una possibile soluzione attenuando la potenza emessa dagli apparati ADSL da cabinet.
La situazione è, comunque, di gestione non agevole in quanto, rispondendo il presentatore della memoria a una domanda al riguardo, è stato confermato che, di fatto, in Germania non è stato ancora deciso completamente come risolvere questo problema (in particolare modo chi dovrebbe gestire l'attenuatore e, poi, come definirne i valori).
Una soluzione alternativa proposta, riguarda l'impiego di terminazioni VDSL nelle soluzioni da cabinet e il non utilizzo di ADSL nel *sub-loop*.
Un'importante sessione è stata interamente dedicata alla tecnologia VDSL. *Don Clarke*, di *BT Exact Technologies (The FS-VDSL Committee, specifications Status and Next Steps)* ha illustrato i progressi compiuti dal Gruppo internazionale, denominato FS-VDSL, di cui è Presidente. Questo Gruppo è stato costituito con l'obiettivo di produrre specifiche relative alla definizione degli aspetti di architettura, di sistema, di caratteristiche degli apparati ubicati nella residenza del cliente, di O&M e fornitura, e d'interoperabilità tra diversi costruttori di componenti *(chipset)* VDSL, in modo da poter offrire servizi su una coppia in rame quali, ad esempio, programmi multipli video ovvero voce e dati ad alta velocità.
Le prime specifiche saranno pubblicate prima dell'estate 2002 e saranno successivamente presentate in *ITU-T (International Telecommunication*

Interno del Palazzo reale Deokougung Palace, costruito sotto la dinastia Joseon.

Union-Telecommunication standardization sector).
Durante l'intervento lo stesso Clarke ha dichiarato che BT stima che in Gran Bretagna i clienti raggiunti dalla tecnologia xDSL nel 2006 saranno approssimativamente 5 milioni.
Nella stessa sessione dedicata all'xDSL, *Piotr Korolkievicz* di *Ericsson Svezia (DSL Interoperability)* ha affermato che il "segreto" legato al successo delle tecnologie xDSL potrà risiedere nel raggiungimento dell'interoperabilità.
Questa svolta dovrebbe tradursi sia in un impegno da parte dei costruttori di *chipset* normalizzati, sia nella definizione delle architetture e delle soluzioni per nuovi servizi che possano garantire un'interoperabilità di tipo *any-to-any*.
Riguardo al primo punto è stato anche confermato che è stata risolta la maggior parte dei problemi d'interoperabilità riguardanti la tecnologia ADSL ma che bisogna essere molto cauti nella definizione dei nuovi standard (per esempio ADSL+). Per l'*SHDSL (Symmetric High density Digital Subsriber Line)*, invece, si è molto vicini alla soluzione finale ma bisognerà dedicare molta attenzione a nuove applicazioni e funzioni da richiedere ai sistemi.
Un ruolo molto importante continueranno a svolgerlo poi gli Organismi e gli Enti di standardizzazione come ad esempio l'*ETSI (European Telecommunications Standarda Institute)* e il DSL Forum.
La sessione dedicata esclusivamente all'accesso in fibra ottica ha visto la presentazione di tre interventi, tutti inerenti al tema *BPON (Broadand Passive Optical Network)*, nel corso dei quali, però, non è stata ancora indicata alcuna soluzione orientata ad applicazioni commerciali, ma sono stati mostrati solo studi di fattibilità tecnica.
La soluzione legata all'impiego della fibra nella rete di accesso è stata presentata come una possibilità di raggiungere anche la clientela di servizi a larga banda con architetture ibride.
In questo caso la soluzione che è sembrata essere più idonea prevede un accesso finale realizzato in VDSL. Su questa soluzione sono stati presentati alcuni risultati relativi a prove in campo in Norvegia, ad esempio, di *Telenor da Haga, Johannessen et alii (Towards deployment of VDSL for interactive broadband services: experiences from a large-scale consumer market trial).* Questi risultati confermano che il futuro dell'accesso e dei servizi a larga banda è legato al connubio tra la fibra e l'xDSL.

Il logo del prossimo Congresso ISSL 2004 che si svolgerà a Edimburgo, in Scozia.

3. Conclusioni

La conferenza è sembrata riscuotere l'interesse dei partecipanti, che hanno avuto modo, oltre che di assistere alla presentazione dei trentanove articoli suddivisi in undici sessioni diverse, di poter prendere visione su quanto esposto da diciannove aziende fornitrici di apparati tra cui *ALCATEL, CISCO, NORTEL, LG ELECTRONICS* e *SAMSUNG.*
Nella cerimonia di chiusura i rappresentanti scozzesi, vestiti con un kilt tradizionale, hanno ricevuto dal Comitato organizzatore il passaggio delle consegne per la preparazione della prossima edizione del Convegno, che si terrà a Edimburgo, ospitato da *BT Exact Technologies* nel 2004 (www.issls-council.org/issls04.htm).

Abbreviazioni

ADSL	Asymmetric Digital Subscriber Line
CATV	CAble TeleVision
FTTx	Fiber To The x
GbE	Gigabit Ethernet
IP	Internet Protocol
MPLS	MultiProtocol Label Switching
PON	Passive Optical Network
SHDSL	Symmetric High density Digital Subscriber Line
UMTS	Universal Mobile Telecommunications System
VDSL	Very high bit-rate Digital Subscriber Line
WLAN	Wireless Local Area Network
xDSL	x Digital Subsciber Line

Francesco Costantino, Paolo Impiglia:
Telecom Italia Domestic Wireline
Francesco Silletta: Telecom Italia Lab

La gestione, leva competitiva degli operatori ICT

NOMS 2002

Management Solution for the New Communications World

Firenze, 15 - 19 aprile 2002

Guido Bruno

La gestione del prossimo futuro o di futura generazione cambierà il modo con cui le aziende erogheranno i propri servizi? Potrà entrare nel novero delle soluzioni determinanti per incrementare i margini di profitto delle aziende? Più in particolare, la gestione delle reti e dei servizi così come finora pensata e progettata sarà ancora utile per la prossima generazione di servizi oppure dovrà cambiare radicalmente paradigma? E ancora, il contesto industriale è pronto per le nuove sfide oppure siamo in presenza ancora di evidenti limiti tecnici che solo con più continue prove e sperimentazioni in campo sapranno dare una svolta significativa al

Graphic Design - www.giannisossi.it

Firenze, tra presente e futuro: esperti si sono dati appuntamento per comprendere le direttrici evolutive della gestione (NOMS 2002).

Spesso nel quotidiano ci interroghiamo sul futuro e sull'evoluzione di ciò che riteniamo consolidato e conseguenza di scelte avvenute nel recente passato. Diviene quindi spontaneo confrontarsi con tecnici e manager di altri paesi, per porre a fattor comune le reciproche esperienze e i cambiamenti che condizioneranno gli investimenti dei gestori di telecomunicazione nei prossimi anni. Uno dei fattori distintivi per un operatore continuerà ad essere sempre più la "gestione delle reti, dei servizi e dell'operatività", e questa Conferenza ha cercato di formulare risposte pratiche a temi che si intravedono all'orizzonte.

Enrico Bagnasco *(General Co-Chair)* saluta i congressisti nel corso della cerimonia di apertura del NOMS 2002.

"valore imprenditoriale" legato al concetto di "servizio di gestione"?

Questi e numerosi altri quesiti, altrettanto stimolanti e attuali, sono stati trattati nel corso dell'ottava edizione del Network Operations and Management Symposium *(NOMS, www.noms.org/2002)*, un appuntamento biennale di rilevanza mondiale nel quale gli esperti fanno il punto su come evolve il settore e discutono le linee di indirizzo del prossimo biennio su temi riguardanti, in particolare, la gestione della rete e dei servizi di telecomunicazioni. Il Convegno è sponsorizzato dall'IEEE e dall'IFIP.

A questa edizione hanno partecipato più di trecento esperti provenienti sia in rappresentanza dei principali operatori e delle industrie manifatturiere e di software ICT sia appartenenti all'ambito accademico, universitario e a centri di studio.

A far da cornice questa volta è stata Firenze; *Telecom Italia Lab* ha svolto i ruoli di curatore e coordinatore dell'evento, in particolare con *Enrico Bagnasco* in qualità di General Co-Chair e con *Guido Bruno* come Publicity Chair.

Qui di seguito sono riportati i punti di maggior rilievo emersi durante il Symposium.
L'evoluzione della rete e dei servizi *(Next Generation Network)* verterà, nei prossimi anni, su alcuni dei seguenti paradigmi e soluzioni: soft-switch, IPV6, MPLS, GMPLS, Cellular IP, WLAN-WiFi, Ultra Wideband, Network Based Security, Content Delivery Network and Architecture e ad-hoc Networking.
I fattori abilitanti per l'evoluzione degli *OSS (Operation Support System)* saranno: il paradigma *NGOSS (Next Generation Operation System and Software)*, lo sviluppo del mercato del *componentware/COTS (Commercial Off The Shelf Components)*, le tecnologie software mature per applicazioni rivolte a grandi gestori e ad imprese *(carrier-class)* quali JAVA, CORBA, ed XML.

Stefano Pileri, key-note speaker, presenta: "Management solutions for the new communication world".

Le tecnologie software di riferimento saranno Java-based, CORBA; mentre le tecnologie emergenti sono: *XML (Extensible Markup Language)*, *SOAP (Simple Object Access Protocol)* e *J2EE (JAVA 2EE)*. Per la crescita dell'impiego *(easy-to-integrate e easy-to-use)* di queste ultime tecnologie è necessario uno standard e una verifica delle prestazioni di modularità e di scalabilità.
Le sfide che persistono riguardano diversi obiettivi quali:

- standardizzare le tecnologie emergenti, le architetture di gestione e il workflow;
- normalizzare i modelli informativi di rete e interfacce di gestione;
- fornire il *service activation and provisioning:* automatico e veloce in ambiente eterogeneo (multivendor, multitecnologie);
- gestire il *traffic engineering*, la *QoS (Quality of Service)*, il *BoD (Bandwidth on Demand)*. Questi temi in particolare saranno sempre più le chiavi di successo per la gestione integrata dell'inventario di rete *(inventory management)* la fatturazione *(billing)* e la gestione della sicurezza in rete *(security)*;
- introdurre, in modo sempre più pervasivo, il concetto di agenti-mobili *(mobile agent)*, in modo da semplificare le attività gestionali della rete e dei servizi, automatizzando, dove possibile, i lavori più ripetitivi.

Nella Comunità scientifica e accademica ha suscitato particolare fermento il tema dell'*information modeling:* in particolare constatare che "purtroppo" esistono diversi Enti coinvolti nella standardizzazione (IETF, ITU, DTMF, TMF, OMG…) e quindi diversi approcci e modelli (CIM, DEN-ng, UML, ...).
Se, in particolare, l'obiettivo comune è quello di rappresentare le informazioni che devono essere scambiate tra i diversi attori coinvolti (ISP, System Integrators, Network Providers, Standardization Bodies, …) in maniera indipendente dalla tecnologia, può essere osservato che sono presenti ancora carenze sull'efficacia di modellizzazione della semantica delle interfacce e delle regole di business.
Un punto di vista fortemente rimarcato è stato quello di evitare di forzare un modello unico (la generalità può indurre inadeguatezza), ma piuttosto un modello "federato" (ossia un modello in grado di condividere in modo federale i dati distribuiti fisicamente e geograficamente su un gran numero di banche dati aziendali) che possa coprire in modo più flessibile le caratteristiche specifiche dei diversi domini.
L'integrazione di *multivendor component-based OSS* diventerà un obbligo: il focus è sull'interoperabilità e la riduzione dei costi, piuttosto che sul *plug&play*. La *MDA (Model Driven Architecture)*, promossa dall'OMG, è stata indicata come una risposta valida all'esigenza di integrare tra loro le diverse tecnologie *middleware* oggi disponibili (CORBA, EJB, XML, SOAP, NET) per permettere l'interoperabilità delle applicazioni ottenendo, così, un'integrazione indipendente dalla tecnologia.
Inoltre, *OSS/J (OSS through JAVA iniziative)*, promossa da un consorzio privato, mira a definire un approccio allo sviluppo software di componenti basato sul modello *(framework)* J2EE, tramite la fornitura di specifiche, API Java standard e strumenti per il testing. L'obiettivo è definire una API aperta - basata sulla piattaforma J2EE - per lo sviluppo di applicazioni di gestione. Con l'API si vuole, infatti, offrire un più rapido e semplice sviluppo di applicazioni con i vantaggi derivanti dall'uso della tecnologia Java.
I "driver" chiave per la realizzazione di nuovi paradigmi gestionali riguardano l'adozione di soluzioni *NGOSS-compliant* (certificati del tipo *Next Generation Operation System and Software)* caratterizzate, in particolare, da un *open bus*, in grado di integrare componenti software distribuite, uno *shared data model*, che permette di condividere

dati comuni a più componenti e un *soft workflow*, che consente di coordinare l'interlavoro tra componenti distinte tramite l'effettiva possibilità di portare all'esterno le politiche di *workflow* dalle componenti medesime. È stato inoltre sottolineato che la riduzione dei costi operativi si ottiene attraverso un'attenta progettazione e attuazione di politiche rivolte all'automazione della creazione di rete, all'assicurazione del *testing* e della qualità dei servizi, al *work force management*, all'automazione spinta del processo di acquisizione, all'integrazione dell'archivio di rete e all'integrazione delle soluzioni di ERP.

Infine, il tema, peraltro non innovativo, del *Policy-based Management* - ossia alle tecniche rivolte alla definizione di soluzioni di supporto che limitino l'intervento straordinario del personale operativo nelle attività gestionali - ha posto in evidenza, dopo circa un decennio di analisi e di esperimenti, l'attuale livello di non-maturità delle soluzioni in essere, confermando che l'argomento, seppur affascinante, è ancora di difficile e rischiosa applicabilità da parte di operatori per la lunga distanza *(carrier-class)*. Molto interessanti sono anche stati gli interventi dei keynote speakers. *Stefano Pileri* (Responsabile della rete di *Telecom Italia Domestic Wireline*) ha presentato una panoramica sull'innovazione sistemistica e operativa di Domestic Wireline che mira a centralizzare il ruolo della gestione, che diventa il fattore abilitante di efficienza e di eccellenza verso il cliente.

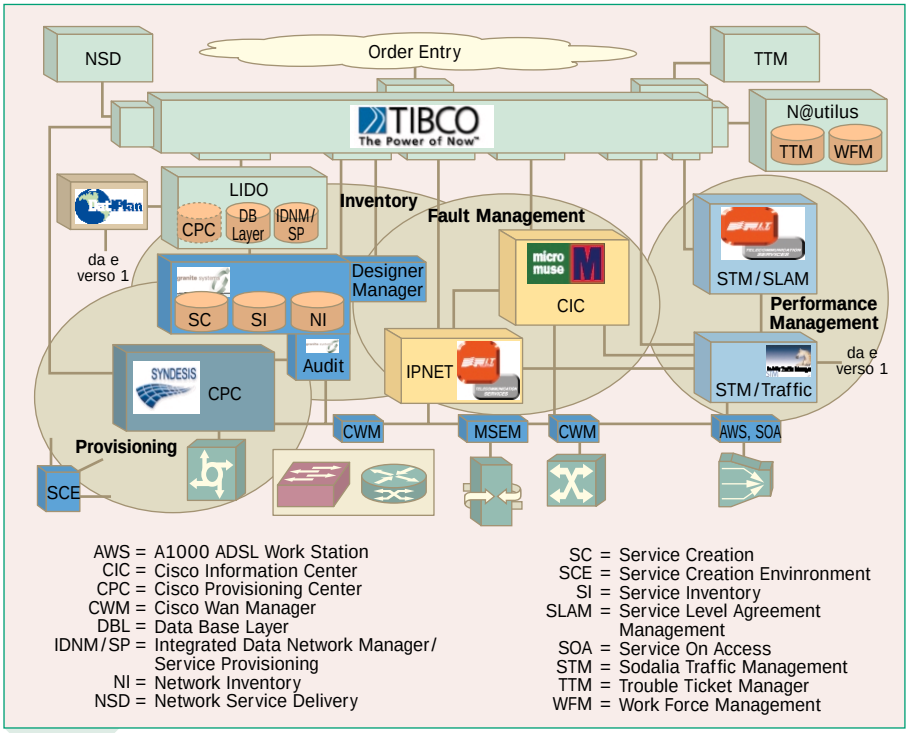


Figura 1 *L'architettura di prossima generazione per la gestione della rete di Domestic Wireline di Telecom Italia.*

Secondo Pileri le architetture di gestione di prossima generazione, *NGOSS (Next Generation Operation System and Software)*, permetteranno nel breve periodo di accelerare i processi operativi di *Provisioning* e di *Assurance* a un ritmo prima ritenuto sfidante. Con le infrastrutture gestionali in fase di allestimento (figura 1), il target d'impresa di Telecom Italia (Domestic Wireline) è il seguente:

- tempo di *delivery* ridotto di circa 4 minuti per ordinativo;
- tempo di correlazione tra differenti allarmi o segnalazioni di anomalia ridotto di circa 30 minuti per fonte *(root cause)*;
- tempo di creazione di un *trouble ticket* ridotto di 15 minuti per disservizio;
- incremento globale del 10 per cento dell'efficienza del processo di *assurance*;
- incremento globale del 20 per cento dell'efficienza del processo di archiviazione della rete.

La capacità di smaltimento dei lavori da parte di Domestic Wireline con l'introduzione della piattaforma gestionale di nuova generazione è ripor-

Massima capacità di ordini per giorno	12.1999	12.2001	08.2002	06.2003
Voce	27.000	41.000	50.000	50.000
Trasporto	2.500	3.500	3.500	5.000
Larga banda	0	3.000	5.000	10.000
Automazione di processo	**12.1999**	**12.2001**	**08.2002**	**06.2003**
Voce (%)	68	72	85	85
Trasporto (%)	20	45	45	70
Larga banda (%)	0	70	95	95

Tabella 1 *Capacità di smaltimento dei lavori della Domestic Wireline di Telecom Italia con la nuova piattaforma gestionale.*

tata in tabella 1. Questi temi saranno trattati in un prossimo numero del Notiziario Tecnico.
Eric Fremont (Senior V. President di *Verizon Communications*) ha presentato la soluzione OSS (in particolare la parte di front-end Business-to-Business) che Verizon, nata dalla fusione di *Bell Atlantic*, *Nynex* e *GTE*, ha messo in esercizio in un tempo assai breve, a partire da un'architettura di gestione estremamente variegata e ramificata nelle sue componenti sistemistiche e di software iniziali (figura 2). Questa soluzione comprende un nutrito portafoglio di servizi wholesale che garantisce l'interazione automatica verso altri operatori e verso gli Internet Service Provider, pur partendo da una situazione iniziale complessa di sistemi ridondanti e duplicati.
Young Hyun Cho (Vice President di *Korea Telekom*) ha approfondito le soluzioni gestionali e operative (figura 3) che hanno consentito a Korea Telecom di attivare circa 4,5 milioni di clienti xDSL in due anni e mezzo (con un ritmo attuale di 10mila nuovi clienti attivati al giorno). La Corea del Sud è il Paese oggi col più alto grado di penetrazione dei servizi a larga banda su base mondiale: circa 23 milioni di abitanti, il 50 per cento della popolazione, utilizza oggi Internet. I clienti che usano i servizi broadband sono circa 8 milioni, di cui il 65 per cento di tipo xDSL, mentre il resto utilizza i servizi CATV e ISDN. Tra i successi che hanno determinato

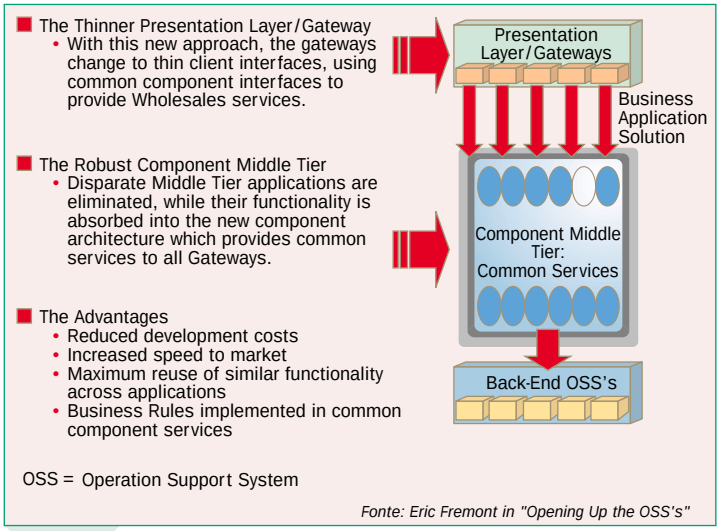


Figura 2 *Modello Verizon per la fornitura di servizi wholesale.*

l'esplosione dei servizi Internet in Corea del Sud, è senz'altro da annoverare la rapidità della fornitura. Gli utenti Internet coreani sono al primo posto nella classifica mondiale per l'utilizzo del servizio, con una media di circa 16,2 ore di impegno al mese.
Nel dibattito di fine symposium *(Distinguished Expert Panel Speaker)* coordinato da *Maurizio Dècina* (*Politecnico di Milano / CEFRIEL*) e che ha visto come protagonisti *Nim Cheung*, *Telcordia Technologies*, Stati Uniti; *Keith J. Willets*, *TeleManagement Forum*, Gran Bretagna; *Makoto Yoshida*, *University of Tokyo & NTT-AT* Exec. Advisor, Giappone; *Mark Basham*, *RHK*, Stati Uniti, sono state evidenziate le maggiori dirompenti tendenze tecniche e gestionali che nel prossimo futuro influenzeranno non poco le scelte dei principali operatori del settore. Le situazioni tecnologiche che costituiscono un fattore chiave di successo e allo stesso tempo una leva prioritaria su cui investire nel prossimo futuro sono in sintesi: la crescita del traffico Internet (dagli attuali 10 Gbit/s ai previsti 100 Tbit/s entro il 2010), l'evoluzione del networking per le imprese e per l'utenza residenziale di prossima generazione, l'evoluzione dei terminali mobili (*Bluetooth*, *smart-phone*, *web-phone*, *wearable computer*), l'evoluzione della banda disponibile per singola

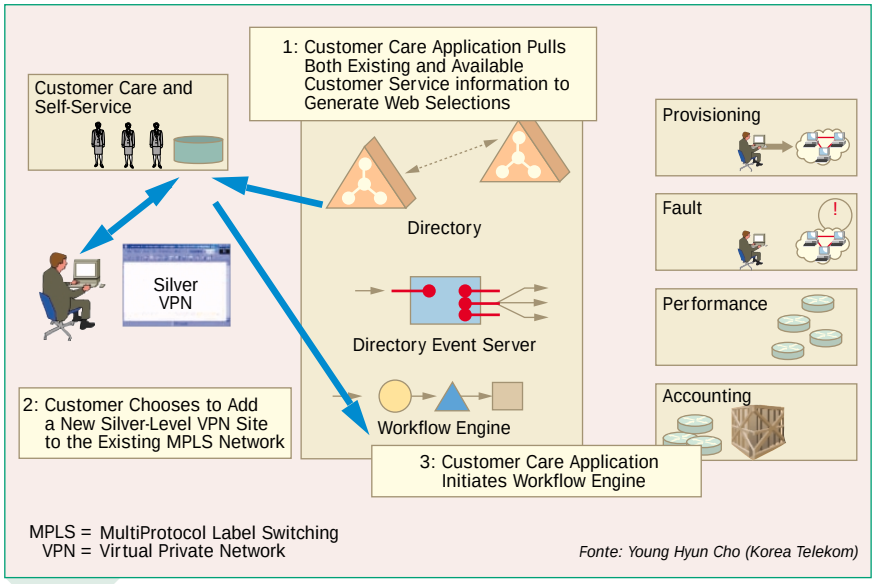


Figura 3 *Processo di Service Delivery "interattivo" sviluppato dalla Korea Telecom.*

Maurizio Dècina coordina il dibattito tra i "Distinguished Expert Panel" al termine della Conferenza.

applicazione e servizio e le soluzioni di rete dette "all IP".
Tra gli interventi nel *Distinguished Expert Panel*, la gestione leva competitiva degli operatori ICT, si ricorda quello di Makodo Yoshida (NTT) che ha illustrato le leve abilitanti tecnologiche e sistemistiche per il lancio della prossima generazione di servizi multimediali (chiamata *BROBA*, *http://www/broba/cc*) che avranno effetti dirompenti sui criteri di fatturazione, sulla gestione della sicurezza, sulla gestione dei contenuti nonché sulla componentizzazione veloce di "servizi di gestione" dedicati alla fornitura, all'assurance e alla qualità del servizio offerto.
TILAB ha contribuito molto attivamente alla parte tecnica presentando tre memorie: *Sergio Sabato* di *TILAB* e *Carlo Filangeri* - di *Telecom Italia Domestic Wireline* - con ASSO (a Web-portal application for Operational Process Assessment and Organizational Improvement); *Giuseppe Ricucci* con un intervento su Generic architecture for a management system of automatic switched transport network; *Fabrizio*

Lo stand di TILAB dove erano mostrati NetDoctor, CnoSS e ASSO.

Verroca, infine, con Real time monitoring and operational assistance for mobile network.
TILAB ha poi coordinato la sessione sul *Service Provisioning (Luigi Artusio)*, e ha assicurato la partecipazione alla tavola rotonda su *Managing 3G and M-Commerce* con *Giorgio Castelli*.
Enrico Ronco ha, infine, presentato il tutorial *eTOM: the business process framework for information and communication service provider.*
TILAB era anche presente all'esposizione con uno stand nel quale è sembrato che abbiano riscosso un grande interesse alcune delle soluzioni gestionali innovative da essa presentate: *NetDoctor (Monitoraggio real-time e Assistenza all'operatività di reti mobili)*, *CnOSS (Configuration Management di reti dati)*, *ASSO* (ambiente di analisi e supporto all'*ASSesment Operativo*) e *Improver* (servizio di consulenza per il miglioramento dei processi e dell'operatività nelle telecomunicazioni).
Il prossimo appuntamento per il NOMS è previsto nel 2004, in Corea.

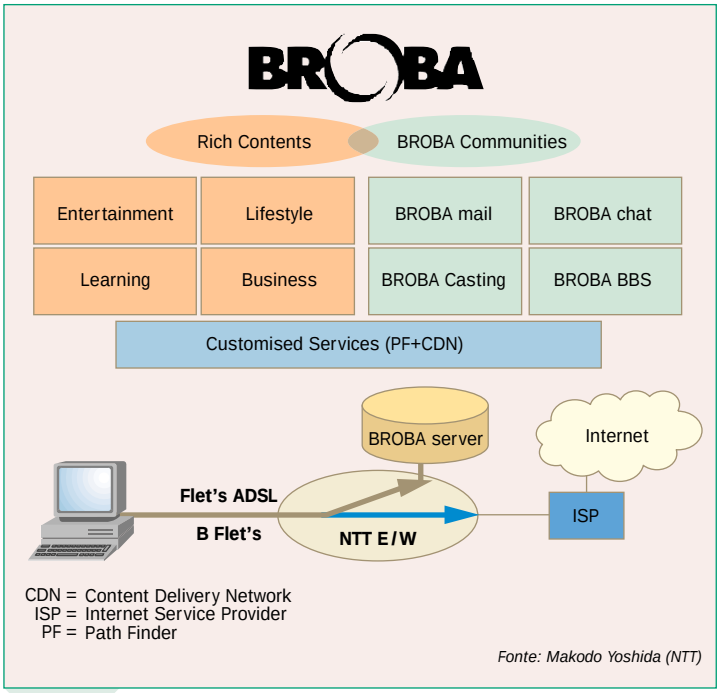


Figura 4 *Modello di riferimento della futura architettura BROBA di NTT.*

Guido Bruno,
Telecom Italia Lab

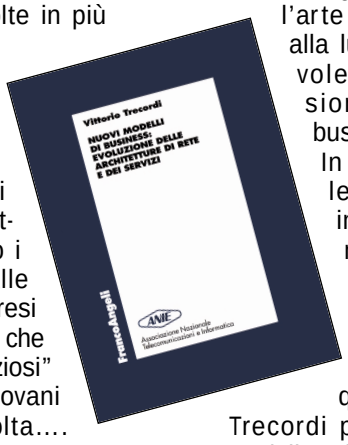
Libri

Vittorio Trecordi

NUOVI MODELLI DI BUSINESS: EVOLUZIONE DELLE ARCHITETTURE DI RETE E DEI SERVIZI

Editore: FrancoAngeli
Milano, 2002
pp. 520, € 38,00
Lingua: italiana

I tecnici che operano da parecchi (o più) anni nelle telecomunicazioni ricorderanno di sicuro che, periodicamente, ricevevano in omaggio da alcune importanti industrie multinazionali, un libro, a volte in più volumi. Queste pubblicazioni, molto chiare ed esaurienti, raccoglievano, in maniera organica una serie di testi, preparati da esperti assai qualificati del settore, che approfondivano i diversi aspetti legati alle telecomunicazioni, compresi quelli economici. Ricordo che questi volumi "assai preziosi" erano contesi da noi giovani ingegneri e qualche volta.... sparivano misteriosamente.



Purtroppo, forse per le difficoltà finanziarie delle industrie, si è persa poco alla volta la consuetudine di aggiornare e distribuire questi testi.

Due anni fa l'*Associazione Nazionale Telecomunicazioni e Informatica* dell'ANIE, ha colmato la lacuna prendendo l'iniziativa di pubblicare un libro in due volumi, *Telecomunicazioni e liberalizzazione in Italia,* documentato e completo, la cui redazione fu affidata ad Aldo Roveri (Professore di Reti per telecomunicazioni presso l'Università La Sapienza di Roma). Il testo approfondiva segnatamente i temi legati allo sviluppo delle reti e dei servizi.

Più di recente l'ANIE ha affidato lo stesso compito "a un noto esperto di tecnologia, di reti e dei servizi da essi derivati" e, al tempo stesso, un divulgatore di prestigio: Vittorio Trecordi, Professore presso il Politecnico di Milano e persona ben nota ai lettori di questa rivista.

Trecordi ha presentato di recente, nel corso di una Conferenza organizzata dall'ANIE, sullo scenario *dell'Information and Communication Technology,* un testo ponderoso, di oltre cinquecento pagine, nel quale dà una risposta esauriente e aggiornata a ogni nostro dubbio sull'ICT che riguardi lo stato dell'arte "delle tecnologie, alla luce della consapevolezza delle dimensioni dei modelli di business".

In questo volume, i lettori troveranno, infatti, una panoramica completa della situazione attuale e delle tendenze in atto nel settore nel quale operiamo.

Trecordi prende, infatti, le mosse dall'evoluzione delle tecnologie di base e dalla situazione socioeconomica mondiale e, più in particolare, da quella relativa al nostro Paese, per soffermarsi, poi, sui fattori che maggiormente incidono sugli scenari attuali del settore e sui dati significativi, necessari per delineare un quadro aggiornato del mercato ICT.

Approfondisce, successivamente, i temi relativi ai nuovi modelli di offerta al mercato, legati specificatamente al settore ICT, e, in questo quadro, esamina anzitutto l'impatto di Internet sulla Società, dedicando un ampio spazio ad applicazioni, quali, ad esempio l'e-business e il commercio elettronico.

Nel libro viene, anche, considerato il ruolo svolto dagli attori che oggi ruotano all'interno del mercato ICT dando rilievo all'analisi delle attese dei clienti che, continueranno, naturalmente, a svolgere un ruolo centrale nello sviluppo futuro.

Trecordi aggiorna e approfondisce anche i temi trattati nel libro prima citato di Aldo Roveri. Dal confronto dei due testi si può osservare quanto abbiano inciso le novità che, in solo due anni, si sono affacciate nel nostro settore, nonostante la crisi degli investimenti e come sia cambiata o si vada modificando la stessa struttura della rete pubblica delle telecomunicazioni.

Un testo di rilievo, quindi, che non richiede ai lettori conoscenze specialistiche e che si legge senza difficoltà, perché i singoli argomenti trattati sono esposti con chiarezza, ordinati in maniera logica e privi di formule, di sigle ricorrenti o, più in generale, di tecnicismi. Con questo libro è, perciò, possibile chiarire compiutamente e con rapidità i dubbi tecnici che ci si presentano ogni giorno.

Un'ultima piacevole sorpresa: ogni capitolo è arricchito da una nutrita bibliografia, con citazione di articoli, spesso disponibili nei siti Internet, che permettono agli interessati di approfondire i singoli argomenti trattati nel libro.

Un volume, dunque, da leggere e da tenere in buon'evidenza sulla scrivania.

Rocco Casale

Enzo Pontarollo

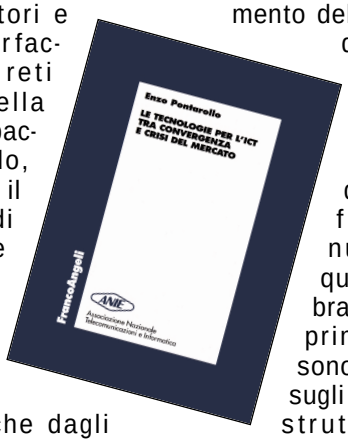
LE TECNOLOGIE PER L'ICT TRA CONVERGENZA E CRISI DEL MERCATO

Editore: FrancoAngeli
Milano, 2002
pp. 263, € 28,00
Lingua: italiana

Un appuntamento biennale, quello organizzato l'11 giugno di quest'anno *dall'Osservatorio sul Mercato e sugli investimenti sulla filiera ICT dell'ANIE (Associazione Nazionale Telecomunicazioni e Informatica),* per fare il punto con i tecnici del settore sugli scenari dell'ICT a livello mondiale e per discutere le strategie e i progetti per la ripresa nel nostro Paese.
A questi incontri un ospite atteso è Enzo Pontarollo, economista, professore all'Università Cattolica del Sacro Cuore di Milano e noto esperto di problemi industriali che, anche in quest'occasione, ha presentato un libro che ci aggiorna su quanto è successo nell'ICT nell'ultimo biennio e ci fornisce alcune chiavi di lettura su quanto possiamo aspettarci nel prossimo futuro. L'analisi che l'autore compie scava subito in profondità, per individuare e per mettere in evidenza i punti critici di maggior rilievo sui quali soffermarsi a riflettere.
Il libro, come i due precedenti, è, infatti, diretto a quanti vogliono conoscere i nuovi scenari nei quali operiamo e sentono, allo stesso tempo, la necessità di fare ordine tra le proprie idee, specie quando sono chiamati a scegliere gli elementi di maggior valore, tra le più importanti novità che si annunciano nel settore.
Nel volume è messa in rilievo, innanzi tutto, la caduta dell'illusione, nutrita nel recente passato, "che la nostra economia (quella dell'ICT) possa svilupparsi in modo rettilineo o esponenziale" e la constatazione che "essa, come gli altri comparti dell'economia, è di tipo ciclico".
"Siamo a valle dell'esplosione della bolla speculativa, afferma Pontarollo, che aveva fatto lievitare, oltre ogni ragionevole attesa, i titoli della *new economy,* fino a quando le attese di crescita hanno mostrato tutta la loro fragilità generando un profondo pessimismo, ulteriormente accentuato dopo gli attentati dell'11 settembre 2001".
E' un punto di partenza obbligato che chiarisce la differenza dell'analisi di oggi, da quella presentata dal professore solo due anni fa nel volume precedente.
In questo scenario assai buio s'intravede uno spiraglio di luce. Esiste, infatti, qualche elemento positivo che l'autore ci indica: "la domanda dei servizi di telecomunicazione è meno influenzata dal ciclo economico rispetto ad altri settori e, anche nel difficile 2001, essa è cresciuta un po' ovunque".
Pontarollo passa, poi, ad analizzare il mercato mondiale delle telecomunicazioni, giustificando l'interpretazione delle modifiche in esso rilevabili, anche con l'ausilio di numerose tabelle di sintesi riprese da alcune fonti autorevoli.
Questa ricognizione mostra, in particolare, l'andamento decrescente nel tempo del mercato mondiale dei sistemi tradizionali, impiegati nella rete pubblica delle telecomunicazioni, assai più marcato, naturalmente, per i sistemi di commutazione.
Di interesse è l'analisi sull'andamento della trasmissione e dell'accesso nelle reti pubbliche fisse: Pontarollo fa notare che nello scorso anno in Europa "si è verificata una prima accelerazione della domanda di accesso a banda larga, ancora allo stato embrionale, per cui occorreranno alcuni anni di intensi investimenti prima che quest'area si trasformi in un mercato di massa".
Da questo primo esame emerge una buona notizia per l'Italia: tra il 2000 e il 2001 il nostro Paese ha avuto in Europa la crescita percentuale maggiore in termini di linee di banda larga installate occupando, anche, una posizione di rilievo nella classifica delle terminazioni in servizio, sempre a banda larga.
L'autore si sofferma, successivamente, sulla *Next Generation Network* e osserva che sembra diventare sempre più imminente la convergenza tra le diverse reti, specie tenendo conto dell'aumento del traffico di dati. Conclude, poi, rilevando che convenga investire al più presto nelle reti di prossima generazione specie se si osserva l'aumento rilevato negli ultimi anni per il traffico dati (dall'agosto del 2000 esso ha superato negli Stati Uniti quello fonico, mentre per

Libri

l'Italia il sorpasso è previsto per la fine del 2002).
In un successivo capitolo del libro, l'autore valuta lo "stato di salute" della telefonia mobile e, in particolare, indaga sulle cause dell'insuccesso del WAP e sulle difficoltà che i gestori incontrano nell'introduzione del GPRS. Per favorire il successo dei nuovi servizi, Pontarollo suggerisce di concordare, tra costruttori e operatori, un'interfaccia comune per reti fisse e mobili nella commutazione di pacchetto, garantendo, allo stesso tempo, il *roaming* tra reti di diversi operatori e specificando, infine, in quale elemento della rete esso debba essere eseguito.



Ritiene, infatti, che dagli operatori vada ripetuto lo sforzo di normalizzazione spinta, compiuto per il GSM.
Un altro problema cruciale riguarda il livello dei prezzi dei servizi usufruibili con il GPRS. Nel libro si ricorda che i criteri oggi impiegati dai gestori sono basati sulla durata dell'impegno della singola connessione. I maggiori operatori hanno, però, già cominciato a introdurre una tariffazione basata sul numero di bit trasferiti. Pontarollo propone di far pagare ai clienti, in più, una quota fissa, indipendente dal tempo di utilizzo del collegamento. Non si può non concordare con questa proposta. A parere dello scrivente, sarebbe opportuno introdurre al più presto anche una tariffa differente, legata ai contenuti, e, quindi, alla qualità del servizio offerto o richiesto. (Quest'ipotesi sembrerebbe già allo studio da parte di alcuni operatori).
Per quel che riguarda la liberalizzazione della telefonia, nel libro si ricordano le tre fasi attraversate dal settore: guerra sui prezzi; differenziazione del portafoglio di offerte; aggregazione sia orizzontale (estensione geografica del mercato servito) sia verticale (ampliamento del pacchetto di offerte di prodotti, allargato mediante partnership e accordi).
Nel nostro Paese, probabilmente per la difficoltà nel reperire finanziamenti per nuovi investimenti, quest'evoluzione sembra essersi fermata alla prima fase. (Nel testo sono anche riportati dati sugli investimenti in infrastrutture effettuati dai gestori italiani).
Come procedere allora? In quale direzione conviene muoversi? Nel libro viene messo in evidenza che lo sgonfiamento della bolla speculativa, legata ai titoli della new economy, ha posto tutte le imprese del settore di fronte a una sensibile caduta di ordini. La crisi è molto grave e preoccupante, ma può essere scorto un segnale debole positivo: quasi tutte le industrie, manifatturiere e di servizio, nonostante le difficoltà, hanno continuato a investire nella ricerca e nello sviluppo.
Risulteranno quindi premiate le imprese produttrici di apparati e di sistemi che sapranno, nel momento della ripresa, presentarsi con un "portafoglio di tecnologie, prodotti e servizi" in grado di assecondare il processo di convergenza in atto. In proposito, l'autore suggerisce un modello nel quale le imprese siano in grado di garantire un'offerta completa di soluzioni, attraverso partnership e alleanze strategiche. O, meglio, le multinazionali dovrebbero sempre più evolvere verso la figura dei *System Integrator.*
Un quadro, dunque, che mette in luce che il settore, nonostante il pessimismo corrente, segue un cammino di crescita e che esso è già in grado di rispondere con proposte innovative alle attese che via via si manifesteranno nel mercato.
Come nella precedente edizione il volume è, infine, completato da una serie di schede che descrivono le caratteristiche peculiari delle singole aziende che operano nelle telecomunicazioni.
Si suggerisce la lettura del libro a quanti vogliano riflettere sulla situazione attuale e, al tempo stesso, abbiano la necessità di individuare soluzioni innovative, per affrontare la domanda, manifesta o latente, che nel prossimo biennio si manifesterà nell'ICT.
E' auspicabile, poi, che Enzo Pontarollo, assieme all'ANIE, ci "regali" fra due anni un nuovo testo aggiornato, per stimolarci ancora a riflettere su come affrontare i nuovi scenari che allora si presenteranno.

Rocco Casale

Franco Battaglia

ELETTROSMOG: *UN'EMERGENZA CREATA AD ARTE*

Editore: Leonardo Facco
e-mail: leofacco@tin.it
Treviglio (BG), marzo 2002
pp. 128, € 10,32
Lingua: italiana

Il libro, uscito di recente, riporta i molti interventi dell'autore, pubblicati sui quotidiani nazionali, che trattano i problemi legati all'elettrosmog. Franco Battaglia è oggi professore di Chimica Fisica presso l'Università di Roma Tre e ha svolto un'intensa attività di ricerca, in particolare al Max Planck Institut di Göttingen come "Research Associate", e negli Stati Uniti alla "State University of New York" di Buffalo e all'"University of Rochester", dove ha anche conseguito il Ph.D. in Chimica Fisica. L'autore, nella dedica del libro a sua figlia Cleis, precisa che la guida di tutta la sua attività di studioso è stata sempre indirizzata verso l'obiettivo di "consegnare ai nostri figli un mondo migliore di quello che c'è stato consegnato dai nostri genitori". Il Sapere di Franco Battaglia continua ad essere posto al servizio della "vera" conoscenza scientifica, di quella "riconosciuta ufficialmente e scientificamente accreditata ". E questa conoscenza, sostiene con forza l'autore, porta a concludere, dopo molti anni di ricerche in tutto il mondo, che, per quanto attiene alle reti e ai telefoni cellulari, il "problema dell'elettrosmog non esiste".

Una conferma autorevole a questa tesi la dà Umberto Veronesi (Direttore dell'Istituto Europeo di Oncologia e Ministro della Sanità nella precedente legislatura) che, presentando il libro, mette in evidenza lo sforzo di Franco Battaglia in questi ultimi anni "per difendere la verità sulle posizioni della scienza in tema di effetti sanitari dei campi elettromagnetici".

Nel libro, come si è detto, sono riprodotti numerosi brevi editoriali, articoli o interventi, scritti sempre con un taglio divulgativo e con un linguaggio semplice ed efficace, con i quali l'autore riporta le conoscenze più recenti delle ricerche. Risponde, così, a quanti, per fini non sempre trasparenti, cercano di alimentare la paura dei campi elettromagnetici e si ostinano a sostituire il dubbio (sempre presente in ogni affermazione della scienza), con finte certezze che, precisa Battaglia, "distolgono l'attenzione dai problemi reali che causano veri e propri disastri a livello ambientale".

Nel volume sono anche riprodotti numerosi documenti preparati dalla comunità scientifica accreditata, come l'appello al Presidente Ciampi, firmato da oltre duecento uomini di Scienza, esperti nei campi della radioprotezione, dell'oncologia, della pediatria, della medicina, della fisica, della biologia, e quasi tutti docenti in Università italiane. È anche riportata una nota di protesta contro "l'oscurantismo del Governo", di oltre millecinquecento studiosi, tra i quali i Premi Nobel Renato Dulbecco, Rita Levi Montalcini, Silvio Garattini, Tullio Regge, Angelo Spena, Edoardo Boncinelli, ed è ripreso un intervento di Angelo Renato Ricci (Presidente onorario della Società Italiana di Fisica e già Presidente della Società Europea di Fisica), che "mette al servizio del Paese le proprie competenze e la propria esperienza, partendo da un pensiero totalmente libero da posizioni ideologiche".

In tutto il libro si avverte lo sforzo e la determinazione di Franco Battaglia nella difesa del patrimonio delle conoscenze scientifiche, specie quando esse sono sacrificate a favore d'interessi particolari. Il testo rappresenta, quindi, una sintesi del rapporto, in Italia, tra informazione e problemi sanitari da campi elettromagnetici e, per i non addetti ai lavori, costituisce una fonte preziosa per distinguere il mito dalla realtà dei fatti, per come sono noti alla Comunità scientifica.

Rocco Casale

Bruno Giussani

SENZA FILI: *L'EQUIVOCO DELL'INTERNET MOBILE, E COME USCIRNE*

Editore: Fazi
Roma, dicembre 2001
pp. 374, € 23,24
Lingua: italiana

Ci muoviamo sempre più in scenari sugli sviluppi dell'ICT dai contorni incerti. Negli ultimi tempi sono stati, infatti, compiuti numerosi errori di valutazione sullo sviluppo dei servizi legati a Internet sia sulla rete fissa sia su quella mobile. Dopo l'euforia manifestata da molti operatori di reti mobili per acquisire le frequenze necessarie per introdurre l'UMTS, si è passati successivamente a un pessimismo spropositato e irrazionale. Oggi molti degli addetti al settore sono, quindi, alla ricerca di analisi che consentano di orientare tempi di offerta, sviluppi tecnologici e proposta di nuovi servizi al mercato.

Utile a questo scopo può essere la lettura del libro, *L'equivoco dell'Internet mobile, e come uscirne,* pubblicato di recente da Bruno Giussani, uno svizzero che è oggi direttore dell'Innovazione e della Comunicazione Strategica di Telefonica 3G Mobile e che ha un passato da giornalista, avendo scritto di tecnologia per quotidiani e settimanali in Europa, come il "Sole 24 Ore".
L'autore desidera fornire ai lettori alcune chiavi di lettura dei dati oggi disponibili, per cercare di ridurre in futuro nuovi errori di valutazione. A questo scopo, Giussani ha utilizzato molti "rapporti di ricerca, sondaggi, relazioni di analisti e altri documenti provenienti da numerose, e assai diverse, fonti".
Nel libro è anche riportato un numero elevato di opinioni diverse, acquisite prevalentemente intervistando alcuni tra i maggiori esperti del settore (basti pensare che le prime cinque pagine del libro sono dedicate ai ringraziamenti alle persone incontrate).
Mano mano che sono riportati questi pareri, Giussani approfondisce e mette assieme queste diverse opinioni e questi dati per proporre ai lettori possibili scenari sul prossimo futuro. Il libro sembra soprattutto voler rispondere a un quesito che circola tra quanti si occupano oggi dell'evoluzione delle reti per il trasporto dell'informazione e, in particolare, delle reti mobili: l'UMTS sarà un'evoluzione del telefono mobile o sarà qualcosa di diverso, più simile alla crescita di Internet?
L'autore, anche se non dà giudizi definitivi, sembra orientarsi verso questa seconda possibilità. Ritiene, infatti, che gli operatori delle reti sembrano giocare un ruolo di secondo piano in questi nuovi scenari. Il successo futuro dell'UMTS sarà, a suo avviso, maggiormente legato alla capacità degli ISP di fornire nuovi servizi e alle innovazioni dei produttori di portatili, che dovranno sforzarsi di offrire terminali per queste nuove applicazioni.
Una conferma viene dall'analisi del fallimento del WAP, dal successo dell'I-mode giapponese e dalla crescita tumultuosa dei messaggi SMS. In questi casi, infatti, sono risultate vincenti la facilità di trasmettere l'informazione e le caratteristiche dei terminali, ma soprattutto la risposta a un bisogno degli utenti.
Resta aperto il problema di quanta informazione possa essere inviata sui terminali mobili (che qualcuno oggi comincia a chiamare nomadici) e quanta, d'altra parte, debba essere trasmessa su quelli fissi. Un segnale a larga banda, come per esempio un *video-clip,* impegnerà, infatti, la rete mobile per un paio di minuti.
L'autore fornisce anche alcune indicazioni sulle stime dei proventi del servizio, valutate da diverse fonti, e mette in rilievo che esse variano di un ordine di grandezza e più!
In quel continuo scavare per cercare di fornire ai lettori ulteriori spunti di riflessione, Giussani si addentra in temi specifici legati ai singoli servizi per i consumatori e per le aziende, quali quelle del *mobile entertainment,* della pubblicità, delle questioni legate all'esigenza della riservatezza.
Un altro tema esaminato nel libro riguarda i terminali: l'autore osserva che i portatili - sempre più piccoli - non sono idonei per un impiego con le nuove potenzialità che saranno offerte con le nuove applicazioni e si chiede se in futuro dovrà essere proposto un solo tipo di cellulare da utilizzare per tutte le applicazioni o se si useranno terminali diversi. Giussani sembra preferire la seconda soluzione. Un'immagine, infatti, non si potrebbe ricevere, senza perdita in definizione, con schermi di cellulari che tendono ad avere dimensioni sempre più ridotte.
Nel libro, poi, un intero capitolo è dedicato ai problemi di sicurezza dell'informazione che nei sistemi *wireless* diventano un fattore assai più critico che nei sistemi per rete fissa.
In conclusione, Giussani riesce a scrivere - per ingegneri e non - mantenendo sempre alta l'attenzione del lettore, con uno stile che ricorda i suoi trascorsi giornalistici. Purtroppo come gli alimenti (anche se la scadenza non è riportata in copertina) il libro ha una vita breve. Va letto presto perché la materia trattata tende a invecchiare assai rapidamente.

Rocco Casale