Sicurezza delle reti VoIP: tecniche di anomaly detection

SICUREZZA

Paolo De Lutiis, Dario Lombardo

Le infrastrutture NGN ed IMS generano un'enorme quantità di dati sotto forma di log ed eventi di vario genere, ma generalmente tali informazioni non sono completamente utilizzate, almeno dal punto di vista della sicurezza. L'utilizzo di tali informazioni rende possibile l'adozione di strumenti innovativi di monitoraggio e controllo, con minimo impatto sulle infrastrutture quindi sui costi di deployment. Sulla base di queste premesse, nel 2006 il gruppo di Security Innovation ha avviato un progetto per la sicurezza dei servizi VoIP, denominato SAD (Sip Anomaly Detection), che ha permesso la prototipazione di un sistema atto ad ampliare la surveillance della piattaforma VoIP di Telecom Italia sulla base delle informazioni generate dalle piattaforme di assurance e billing. L'articolo fornisce una panoramica delle principali problematiche di sicurezza dei servizi IMS (VoIP in particolare) e di come queste possono essere affrontate con strumenti di analisi specifici come SAD, adatti sia per gli operatori di rete fissa che mobile. Sono descritti gli scenari di rischio più significativi, le relative problematiche di detection e i principali riscontri ottenuti durante la sperimentazione effettuata in campo sulla piattaforma Alice Voice.

1 Introduzione

L'architettura delle NGN (Next Generation Network) si basa sul principio di convergenza dei servizi telefonici, offerti tradizionalmente dalle reti a commutazione di circuito, e dei servizi dati, offerti dalle reti a commutazione di pacchetto come Internet, su un'unica infrastruttura di rete basata su protocollo Internet Protocol (IP). Voice over Internet Protocol (VoIP) è la tecnologia che

permette agli utenti di effettuare chiamate telefoniche utilizzando il protocollo IP permettendo quindi il superamento della rete telefonica tradizionale Public Switched Telephone Network (PSTN). Oltre alla possibilità di abbassare i costi di gestione e delle infrastrutture, il nuovo paradigma VoIP, e SIP (Session Initiation Protocol) in particolare [canal], permette la creazione di nuovi servizi di telecomunicazione quali ad esempio informazioni di Presence, chat voce-video-testo, messaggistica multimediale, o multiparty conference, che rappresentano il reale valore aggiunto delle piattaforme NGN.

Purtroppo, come ogni altra tecnologia di comunicazione, anche il VoIP è suscettibile di problematiche di sicurezza e in particolare di DoS (Denial of service) e SPAM che possono minare l'usabilità dell'intera NGN. A tale proposito un termine apposito per lo SPAM è stato coniato in ambiente VoIP: Spam Over Internet Telephony (SPIT). In particolare lo SPAM consiste in messaggi di disturbo ovviamente non desiderati che consumano risorse di spazio e di computazione. Si prevede che tale fenomeno presto potrebbe divenire un grosso problema per gli operatori VoIP e i loro clienti.

Lo sviluppo dei servizi di telecomunicazioni per le reti di prossima generazione richiede quindi la definizione e l'implementazione di apposite soluzioni di sicurezza, che vanno dalle soluzioni per l'autenticazione, alla protezione della segnalazione e della riservatezza delle comunicazioni.

Queste soluzioni devono rispondere a stringenti requisiti, in modo da garantire elevati standard di sicurezza, almeno comparabili con quelli che contraddistinguono le tradizionali (es. PSTN) soluzioni di telecomunicazione, sia di rete fissa che mobile. In tale ambito un aspetto fondamentale per la sicurezza delle reti NGN e VoIP è legato alla necessità per l'operatore di controllare lo stato di salute della rete e dei servizi con appositi strumenti di monitoring, in modo da poter identificare in modo tempestivo i possibili problemi e quindi poter intervenire con le contromisure adatte a ripristinare un adeguato livello del servizio.

1.1 *Principali minacce di sicurezza*

In questo articolo prenderemo in considerazione due tra le principali tipologie di minacce per la sicurezza delle reti VoIP, gli attacchi DoS e i fenomeni di SPIT.

Tipicamente un attacco di tipo Denial-of-Service rende una rete, un terminale, o altri componenti dell'infrastruttura di rete, indisponibili agli utenti. Un attacco di questo tipo opera creando situazioni di sovraccarico di lavoro per un'infrastruttura tali che questa non riesce più ad erogare i servizi per i quali è stata progettata. Nel caso specifico della telefonia IP, attacchi DoS si tramutano in difficoltà, abbassamento della qualità o addirittura impossibilità di effettuare o ricevere chiamate. In particolare il DoS può interessare i protocolli di segnalazione causando ritardi nelle chiamate o impossibilità di stabilire le sessioni. Un attaccante potrebbe anche mettere a segno un "media DoS", ossia un Denial-of-Service che coinvolge tutta la rete, rendendo impossibile oltre che lo stabilire la chiamata anche il solo uso del telefono e degli altri dispositivi connessi alla rete: nella rete telefonica tradizionale questo si traduce nel "telefono isolato" [1].

Un DoS può anche coinvolgere il gateway di collegamento tra la rete VoIP e la rete telefonica tradizionale, che può essere saturato e rendere perciò impossibile la comunicazione con l'esterno malgrado la connettività interna non ne risenta. Va inoltre sempre considerato il caso più semplice di DoS, ovvero quando la rete non implementa una configurazione sufficiente di qualità del servizio per il traffico voce in una rete voce-dati integrata. In questo caso anche solo il contemporaneo accesso ad un sito web con il trasferimento di grandi documenti FTP può portare ad un effettivo DoS sul traffico VoIP con la conseguente quasi impossibilità di effettuare qualsiasi telefonata.

Essendo il VoIP una tecnologia di comunicazione, essa è influenzata e nello stesso tempo influenza gli aspetti sociali tra gli individui di una comunità. In questo senso l'integrazione di servizi voce e dati sulla stessa rete potrà dare luogo

alla nascita di fenomeni di spamming simili a quelli che ora affliggono i sistemi di posta elettronica. Lo SPAM, definito come trasmissione di una grande quantità di messaggi non desiderati, è una delle minacce più conosciute del mondo Internet [16]. Tale scenario si potrebbe ripresentare oggi con il VoIP: decine o centinaia di messaggi vocali per utente (ma non solo, essendo il VoIP il reale enabler di applicazioni multimediali), inviati sulle reti IP che vanno ad intasare caselle vocali e server VoIP. Per avere un'idea dello scenario indesiderato a cui potrebbero essere destinate le nostre reti di telecomunicazione si pensi a spot pubblicitari che interrompono le conversazioni, bombardamenti di chiamate indesiderate, tentativi di phishing vocale (vishing) e segreterie intasate da messaggi molesti che reclamizzano prodotti di ogni genere [13]. Le stesse persone che oggi generano SPAM non possono non vedere con interesse i sistemi e le tecnologie VoIP (e altre applicazioni/ servizi multimediali): telemarketing e truffe on-line, ma anche hacker interessati a sfidare la sicurezza delle emergenti NGN. In ambito normativo e di standardizzazione [7, 8, 9] la problematica viene considerata molto seriamente, come ad esempio in [6], dove sono analizzate le cause e i mezzi per generare SPIT. Già oggi sono disponibili strumenti che possono essere utilizzati per generare SPIT in modo massivo, basti pensare ai botnet utilizzati oggi per lo SPAM della posta elettronica (si veda il proof-of-concept "accademico" [10] che realizza un sistema di generazione di SPIT basato su IRC). Vedasi box spit per ulteriori dettagli sulla tematica.

1.2 *Strumenti di controllo della sicurezza delle reti*

Il livello di sofisticatezza degli attacchi a sistemi o risorse di rete è aumentato drasticamente negli ultimi anni, complice la sempre maggiore connettività tra i sistemi e la dipendenza da servizi di rete distribuiti. Inoltre la crescente disponibilità di tool free automatici di intrusione e la possibilità di sfruttare appositi script già pronti all'uso stanno drasticamente incrementando i metodi di attacco conosciuti, ed è probabile che l'attenzione degli autori di tali tool si rivolga presto alle caratteristiche delle reti NGN e dei servizi VoIP.

Tra gli strumenti di sicurezza a disposizione degli SP e operatori, l'Intrusion Detection System (IDS) è una tecnologia che permette di monitorare, e a volte prevenire, tentativi di intrusione o comunque di manomissione di un sistema o delle risorse di una rete. I sistemi IDS sono, a tutti gli effetti, un importante strumento di controllo dello stato di una rete e permettono agli amministratori di avere sempre un quadro completo degli eventi che possono in qualche modo influire sull'usabilità dei servizi erogati [3].

Gli IDS permettono l'automazione delle attività di monitoraggio e sono in grado di produrre specifici allarmi od anche una risposta automatica per ogni attività sospetta: l'obiettivo dell'intrusion detection è di identificare tutti gli attacchi che effettivamente avvengono e ignorare quegli eventi che, anche se possono avere le caratteristiche di un attacco, in realtà non lo sono (i falsi positivi). Infatti, i requisiti fondamentali per un IDS sono la tempestività e la correttezza delle risposte rispetto agli stimoli provenienti dalle reti sotto osservazione.

1.3 *Tipologie di IDS*

L'intrusion detection system utilizza diverse tecniche di valutazione per verificare se è avvenuta o meno un'intrusione o un tentativo di attacco. In particolare ci sono due categorie di approcci [2]: "signature-based" che identifica dei pattern corrispondenti ad attacchi noti (conoscenza a priori) ed "anomaly-based" che identifica deviazioni significative da un comportamento atteso e considerato corretto (conoscenza a posteriori).

L'approccio signature-based consiste nell'analizzare l'attività di un sistema cercando eventi che corrispondono ad un pattern di eventi predefiniti

che descrivono un attacco già noto. I pattern corrispondenti ad attacchi conosciuti sono detti signature, da cui il nome del metodo.

Il principio su cui si basa un anomaly detection system è di rilevare deviazioni significative rispetto ad un comportamento normale atteso (baseline). I sistemi che si basano sull'anomaly detection partono dall'assunzione che gli attacchi di sicurezza modificano sensibilmente l'attività "normale" di un sistema (ad esempio di una rete VoIP) e che dunque possono essere rilevati mettendo in luce queste differenze. In particolare, gli IDS anomaly-based costruiscono, su base statistica, profili che rappresentano il normale comportamento di una rete collezionando i dati storici, raccolti osservando per un certo periodo la normale attività del sistema. In fase di monitoraggio per la rilevazione degli attacchi, i sensori collezionano i dati relativi agli eventi che si verificano e, sulla base di opportune misure, determinano quando l'attività monitorata devia dalla baseline. L'approccio statistico di anomaly-detection comporta la scelta ed analisi di metriche base per pattern di traffico conosciuti e il settaggio di una soglia di allarme nel caso avvengano consistenti variazioni nei pattern di traffico sotto controllo. Una possibile baseline potrebbe essere calcolata come la media dei valori caratteristici del sistema (es. numero di chiamate/ora) con una tolleranza pari alla deviazione standard.

Un sistema signature-based è efficiente ed è in grado di rilevare attacchi generando un ridotto numero di falsi allarmi. I sistemi di questo tipo traggono le loro conclusioni sulla base del pattern-matching: possono fornire un messaggio di allarme ogni qualvolta viene rilevata l'occorrenza di una determinata signature (o impronta dell'attacco) all'interno dei flussi di comunicazione posti sotto osservazione.

Gli IDS signature-based sono però in grado di rilevare solo quegli attacchi che si conoscono e quindi, perché siano effettivamente efficaci, è necessario aggiornarli costantemente con le signature dei nuovi attacchi quando questi divengono noti. Sono quindi sistemi che non hanno la facoltà di riconoscere quegli attacchi denominati 0day (zero-day), cioè quegli attacchi che non sono già stati osservati in passato e per cui non esiste una signature. Inoltre questi sistemi richiedono comunque una fase di configurazione iniziale e un costante costo di gestione richiesto per il loro aggiornamento.

I sistemi anomaly-based sono invece più difficili da configurare perché richiedono, per ogni sistema da monitorare, una descrizione del comportamento osservato e di quello che ci si attende da esso. L'output dei sistemi di IDS anomaly-based produce generalmente conclusioni basate su correlazioni statistiche tra il comportamento attuale e quello atteso. Sono più soggetti ai falsi positivi generati da comportamenti anormali ma non provocati intenzionalmente da un attaccante. Il loro vantaggio è che sono in grado di rilevare attacchi non basati su una conoscenza pregressa ma in base alle logiche interne del sistema. Per questo motivo sono adatti a monitorare i contesti tecnologici nuovi per cui non esiste una *knowledge-base* vasta.

I risultati migliori si ottengono comunque combinando i due metodi.

2 Proposta per un sistema innovativo di monitoraggio: SAD

Nel 2006 il gruppo di Security Innovation ha avviato un progetto sulla security VoIP denominato SAD (Sip Anomaly Detection). L'obiettivo di questo progetto era quello di prototipare un sistema di sicurezza per ampliare la surveillance della piattaforma VoIP. La scelta è stata quella di sviluppare un sistema di anomaly detection, perché questo tipo di sistema è quello che meglio si presta ad un contesto nuovo come quello delle reti NGN e dei servizi IMS. Per questo motivo, data la mancanza di una base di dati su cui costruire delle signature di attacchi su VoIP, l'approccio anomaly detection si è rivelato quello più proficuo. Anche la natura sociale delle comunicazioni che transitano in questo tipo di reti ha portato ad optare per la soluzione anomaly detection. Nonostante ciò sono stati, comunque, in-

tegrati alcuni controlli di tipo pattern-matching al fine di ottenere risultati migliori.

Il sistema di rilevamento anomalie SAD ha come fine quello di rilevare situazioni anomale che interessano la rete SIP nella sua interezza. Lo scopo non è quello di rilevare fenomeni puntuali, ma di identificare quei fenomeni il cui verificarsi si può ripercuotere negativamente su tutta la rete. È questo il caso di attacchi DoS (alla rete SIP o alla rete IP con ripercussioni sul VoIP), errate configurazioni (per esempio dei server SIP), bug sui software, ecc. Per questo motivo l'elemento caratterizzante il sistema SAD è il modello che è stato costruito per rappresentare l'intera rete telefonica VoIP in modo da poterne osservare solo i macro-comportamenti. Tale modello è descritto più dettagliatamente nel paragrafo relativo all'architettura del sistema.

2.1 *SAD: principali use case*

Descriviamo ora alcuni casi di possibili problemi nella rete VoIP che possono essere affrontati tramite l'uso di un sistema di anomaly detection come SAD. Questi casi sono di interesse per l'operatore in quanto possono creare gravi problemi all'intera rete e devono essere individuati ed isolati nel più breve tempo possibile. Per questo tipo di fenomeni non è sufficiente guardare una porzione limitata della rete. È necessario invece avere una visione globale della rete, ampia sia per numero di utenti coinvolti che per tempo di osservazione. Alcune anomalie possono essere infatti notate solo se si osserva l'evoluzione del comportamento degli utenti nel tempo oppure solo correlando l'azione di molti utenti contemporaneamente.

2.1.1 DOS

Con DOS intendiamo identificare sia gli attacchi di Denial of Service tradizionali (quindi non mirati alla rete VoIP ma alla rete IP), sia quelli che hanno la piattaforma VoIP come target. Il loro obiettivo è quello di creare un traffico così elevato da rendere problematiche le operazioni regolari dei sistemi. Quando l'attacco non è mirato alla rete VoIP, è possibile comunque che essa ne subisca gli effetti in quanto il canale di trasporto è condiviso tra dati e voce. Se l'attacco è invece mirato alla piattaforma VoIP, esso tende ad esaurire le risorse degli apparati di servizio (elementi IMS, SBC, ecc).

L'effetto generalizzato di questo tipo di attacchi è quello di impedire le comunicazioni. Per un sistema che osserva il loro andamento questa anomalia appare abbastanza evidente: la frequenza di chiamata degli utenti si riduce in maniera anormale. Anche qualora si sia in presenza di un DOS alla rete di trasporto, l'effetto è comunque lo stesso, in quanto si impediscono le comunicazioni VoIP.

Nel caso di un attacco DOS portato a livello della rete ed infrastruttura VoIP (ad esempio utilizzando il canale di segnalazione SIP) il sistema SAD osserva un'anomalia che è tanto più grave quanto più volume di traffico viene generato dal potenziale attaccante. Nel caso di un attacco DOS verso gli elementi dell'infrastruttura come ad esempio il server SIP, l'effetto è trasversale su molti utenti e quindi più essere rilevato in maniera abbastanza semplice dal sistema di monitoring in maniera proporzionale dal volume di traffico osservato. Un altro scenario rilevabile con SAD è rappresentato in *Figura 1* . Questo scenario è simile ad un attacco DDoS generato da una botnet: molti utenti malevoli generano chiamate verso un unico utente, al fine ad esempio di saturare la sua banda. In questo caso non è necessario generare flussi di banda elevati, ma è possibile rilevare l'anomalia per il fatto che la vittima si trova al centro di più flussi contemporaneamente, cosa che in condizioni normali non avviene (a meno di analizzare il caso di un call center). Il caso di un attacco condotto contro la rete di trasporto è rilevato in termini di minore, anomalo, utilizzo delle risorse VoIP. Il sistema di monitoraggio si aspetta che il comportamento della rete segua il modello appreso durante la fase di learning, e quindi, ad esempio, che gli utenti utilizzino maggiormente il loro telefono durante le ore centrali dei giorni festivi.

Nel momento in cui si dovesse misurare un

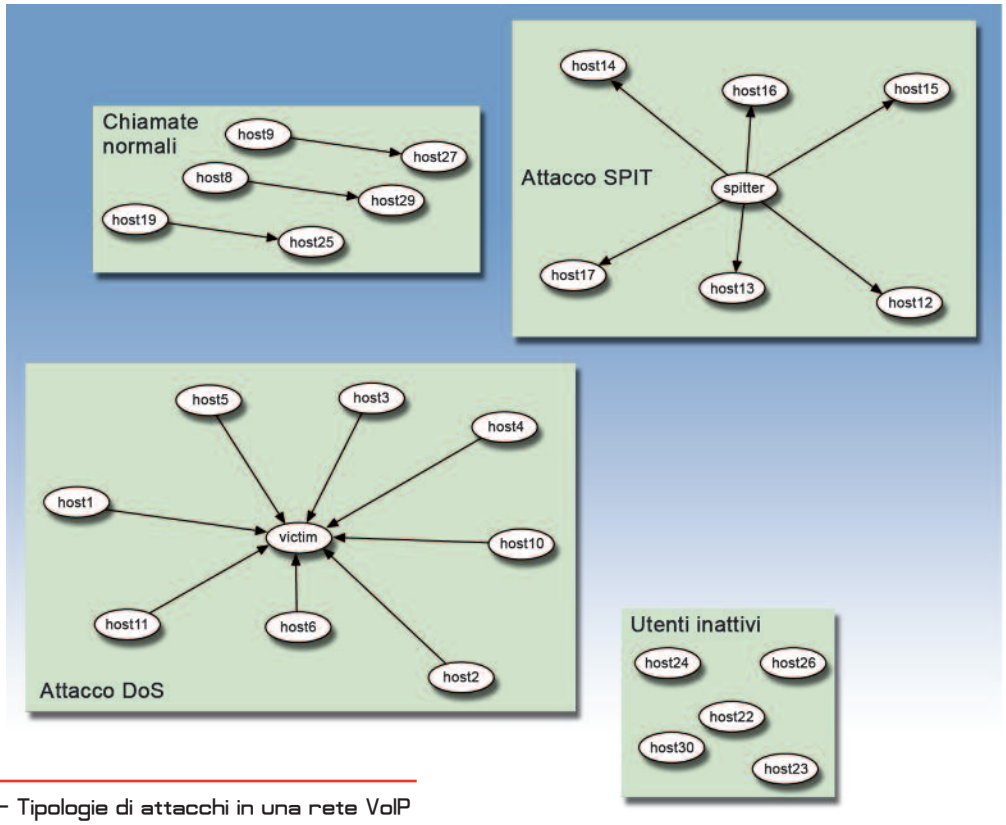


Figura 1 – Tipologie di attacchi in una rete VoIP

sensibile calo del numero di chiamate ecco che verrebbe sollevato l'allarme, potenzialmente dovuto ad un attacco DOS.

2.1.2 SPAM/SPIT

Lo SPAM viene generalmente usato per veicolare ogni genere di informazioni pubblicitarie e/o promozionali, virus e malware in genere e ultimamente, integrato con tecniche di phishing e di social engineering, ai fini di carpire informazioni riservate (ad esempio codici di accesso bancari) ad utenti sprovveduti. Esistono già da tempo fenomeni di spam telefonico (ad esempio pubblicità di offerte commerciali da parte di call center), ma si prevede che con la migrazione verso servizi VoIP da parte dei service provider, il fenomeno possa avere un impatto molto più significativo. Si parla quindi di Spam over Internet Telephony o SPIT.

I servizi VoIP di nuova generazione non veicoleranno solo la voce, ma anche altri servizi, ad esempio Instant Messaging, Presence Service e Voice Mailbox. In questa ottica lo SPIT diventa ancora più rilevante, a causa di alcuni fattori che lo rendono più economico e semplice per gli attaccanti:

- Costi bassi di chiamata;
- Possibilità di usare protocolli di segnalazione semplici e facilmente interfacciabili con hardware a basso costo (SIP Scripting);
- Possibilità di generare richieste di INVITE massive in un breve lasso di tempo;
- Possibilità di sfruttare botnet di macchine compromesse per lanciare attacchi;
- Presenza in rete di elenchi di possibili vittime grazie a servizi di directory pubbliche come ENUM;
- Maggior impatto sulla vittima: a differenza delle email che possono essere cancellate senza leggere il corpo del messaggio (vanificando così l'intento dello spammer), le chiamate indesiderate causano un indubbio disagio all'utente.

Esistono ovviamente diversi scenari di attacco SPIT. Con SAD è possibile rilevare anche tale tipologia di attacco. Si prenda ad esempio lo scenario SPIT riportato in *Figura 1*. Lo schema

Breve storia dello SPIT

Ad oggi non sono stati resi noti molti attacchi SPIT reali, in parte perché le tecnologie VOIP si stanno affermando solo di recente, ed in parte perché i servizi e le reti VoIP sono ancora isolate tra loro e poco interconnesse (ma è anche possibile che gli operatori non siano disponibili a pubblicare informazioni di questo genere). Inoltre, mentre qualche notizia arriva dal Giappone e dagli USA, poco si conosce del contesto Europeo e ancora meno dall'Italia. Infatti una ricerca condotta da NetIQ [15] su 66 responsabili IT italiani di medie o grandi imprese che utilizzano o hanno in programma di installare sistemi VoIP, rivela che più della metà degli intervistati (59%) ha definito come "bassa" o "molto bassa" la possibilità che virus o worm attacchino il proprio sistema VoIP e i rischi di SPIT sono considerati minimi dagli intervistati: solo il 12% e il 18% rispettivamente definiva queste minacce alla sicurezza come "gravi" o "molto gravi": segno che lo SPIT non si è ancora diffuso in Italia.
La nascita della problematica dello SPIT risale al febbraio del 2004, quando in Giappone diversi utenti ricevettero chiamate preregistrate durante le quali era reclamizzato un sito per adulti. Da allora lo SPIT ha iniziato a diffondersi e ha raggiunto sia gli Stati Uniti che l'Europa, arrivando nell'aprile del 2006 a coinvolgere anche Skype, il software VoIP forse più noto e diffuso.
I primi fenomeni SPIT documentati sono quindi avvenuti in Giappone[1], dove le tecnologie VoIP sono già piuttosto utilizzate (si calcolano 10 milioni di utenti VoIP, escluso Skype, nel 2008).
Secondo un articolo della Nikkei Communications del 2005, il VoIP provider SoftBankBB (4,6 milioni di clienti) ha denunciato tre eventi SPIT:

- Febbraio 2004: Messaggi commerciali che pubblicizzavano un sito per adulti
- Agosto 2004: Squilli di disturbo (comunicazione interrotta appena il ricevente sganciava la cornetta) con frequenze di 6000 chiamate al giorno (rilevate) complessive sulla rete VoIP
- Novembre 2004: Richieste telefoniche di informazioni personali

Malgrado tutti gli eventi siano stati contraddistinti da caller-id differenti, SoftBankBB è riuscita a risalire al colpevole degli attacchi SPIT ed isolarlo dalla propria rete.
Durante il 2006 sono stati registrati negli USA due fenomeni interessanti, uno di Vishing (Phishing via VoIP) e uno di SPIT tradizionale. Infatti, oltre al telemarketing, lo Spit può essere utilizzato in maniera anche più efficace per il phishing: è possibile simulare la chiamata di una banca o di una qualsiasi altra organizzazione, invitando la potenziale vittima a rivelare a voce i propri dati sensibili. Il caso di Vishing si è verificato nel mese di aprile. La società Cloudmark [14] ha rilevato due diverse segnalazioni: in entrambi i casi, i messaggi fraudolenti avvertivano di un problema con il conto bancario e fornivano un numero di telefono per risolverlo. Il numero telefonico connetteva la potenziale vittima ad un sistema di risposta vocale identico a quello della banca presa di mira, riutilizzando quindi le stesse tecniche del phishing tradizionale, che utilizza siti web fasulli ma identici nella veste grafica a quelli delle istituzioni o società dalle quali le mail sembrano provenire. La voce guida invitava a fornire informazioni personali come il numero di conto corrente e la password.
Il caso di SPIT puro è invece accaduto nella rete (overlay) di SKYPE: diversi clienti SKYPE hanno riportato (in un blog [12]) di essere stati interrotti durante conversazioni private da un'altra chiamata, rivelatasi poi un messaggio registrato pubblicitario. Immediatamente dopo anche la chiamata iniziale veniva abbattuta.
Sempre negli USA, nel 2007, un caso di SPIT si è verificato con successo alla Columbia University [18] di New York, dove ignoti sono riusciti ad accedere al proxy del sistema centrale dell'ateneo e a contattare in sequenza tutte le linee, facendo ascoltare un messaggio a chiunque sollevasse la cornetta. >

1 I casi reali di SPIT riferiti al Giappone sono stati illustrati da NEC Europe durante incontri diretti. Sono comunque disponibili su Internet alcuni riferimenti al caso SoftBankBB.

In Europa uno dei primi casi documentati è stato segnalato in Olanda ed è riportato in [17], dove si descrive un attacco SPIT per veicolare una serie di messaggi di tipo religioso originato dagli USA e destinato a specifiche Corporate Olandesi.
Più di recente, nel periodo 4 Settembre 2008 – 10 Settembre 2008 è stato registrato un attacco SPIT di dimensioni rilevanti verso numeri telefonici di diversi paesi europei e non (principalmente Germania e Israele, ma anche Francia, Finlandia e forse Italia) e originato nella rete del SP tedesco Freenet.
L'attacco prevedeva l'invio massivo di richieste di INVITE. I terminali degli utenti attaccati hanno squillato per lungo tempo (anche 40 minuti) a intervalli di qualche decina di minuti, molto spesso a tarda notte. Molti utenti hanno richiamato il numero 5199362832664 dal quale provenivano queste chiamate. Sembra si tratti di un numero peruviano (almeno dal prefisso 51) che una volta chiamato risponde per circa tre minuti e poi chiude la chiamata. Alcuni utenti, principalmente israeliani hanno ricevuto chiamate analoghe dal numero 99362832664 che sembrerebbe essere un numero afghano oppure turkmeno. I messaggi sono apparentemente partiti da un centralino VoIP mal configurato (forse volutamente) poi localizzato in Bulgaria.
Verso la fine del 2008 un altro caso interessante di SPIT è accaduto nella svizzera tedesca. Secondo [21] ben 10.000 telefoni (posti in stazioni od altri luoghi pubblici) hanno iniziato a squillare quasi contemporaneamente dalle 12:30 alle 15:00 del 24/11/2008. Le persone che avessero provato a sganciare la cornetta avrebbero ascoltato lo spot pubblicitario di una canzone hip-hop.
La seguente tabella riassume gli eventi SPIT illustrati nel presente box.

Quando	Dove	Evento
2004/02	Giappone	Messaggi commerciali relativi a siti web per adulti
2004/11	Giappone	Messaggi registrati finalizzati al collezionamento di informazioni di tipo privato
2004/08	Giappone	Squilli di disturbo
2006/03	Olanda	Messaggi religiosi originate dagli USA verso corporate Olandesi
2006/04	USA	Primo attacco SPIT attraverso il servizio di Skype
2006/04	USA	Primo caso di Vishing registrato da Claudmark
2007/11	USA	Attacchi SPIT alla Columbia University
2008/09	Europa/Israele	Attacchi originati dalla rete Freenet
2008/11	Svizzera	Telefonate verso cabine pubbliche

Tabella – Eventi SPIT registrati

Tutti gli addetti ai lavori della sicurezza VoIP sono concordi nel considerare lo SPIT una minaccia imminente molto grave , con forti prospettive di crescita e dal potenziale catastrofico.
Alla luce della breve analisi dei fenomeni SPIT documentati fino ad oggi, sembra che tale previsione, almeno nel breve periodo, non sia realistica. Si condivide comunque l'idea che non convenga sottovalutare la potenziale gravità dello SPIT o sopravvalutare la capacita attuale delle NGN di evitarne la diffusione, rischiando così di ripetere gli stessi errori commessi con la posta elettronica ("entro due anni da ora, lo spam sarà eliminato", Bill Gates, World Economic Forum, 2004).

rappresentato rispecchia una situazione molto realistica (molti degli attacchi SPIT finora registrati sono di questo tipo): un utente malevolo, attraverso un sistema automatico, genera massivamente una serie di tentativi di chiamata verso molti, diversi utenti, riproducendo quindi un messaggio registrato (ad esempio pubblicitario) alla risposta sempre della stessa durata. Riuscire ad identificare e a bloccare un attacco di questo genere è sicuramente una necessità per l'operatore, che deve salvaguardare sia la privacy dei clienti che l'integrità della rete. È ovvio che lo SPIT può trasformarsi in un vero e proprio attacco DOS in quanto vengono impegnate risorse che non possono quindi essere utilizzate per le normali esigenze operative.

2.1.3 Problemi distribuiti

Una famiglia di problematiche relative alle reti di telecomunicazioni sono i cosiddetti fenomeni distribuiti. Essi si possono verificare quando uno specifico problema si presenta su più punti della rete contemporaneamente, come ad esempio la presenza di un baco software all'interno della versione di un software largamente utilizzato o una determinata configurazione errata riportata su tutti gli apparati di un certa rete. In questo caso il problema non è puntuale (come nel caso di un singolo apparato malfunzionante), ma è comune a tutti i punti della rete. Il classico esempio è un bug in un sistema operativo. Tutte le piattaforme dello stesso tipo soffrono dello stesso problema. Nel mondo VoIP questo rischio si può concretizzare nei residential gateway, quegli apparati, cioè, che connettono l'utenza alla rete dell'ISP oppure sui terminali telefonici (softphone o telefoni VoIP). A causa della loro elevata numerosità se si manifestasse un problema su una certa release di software a bordo di tutti o molti degli apparati utente, è facile immaginare che la rilevanza del problema sarebbe enorme.

Due scenari di sfruttamento malevolo di questo problema possono essere le botnet e i worm. Nel primo caso un attaccante potrebbe sfruttare il bug per far compiere azioni illecite agli apparati. Essi sono in tutto e per tutto dei sistemi general purpose, eventualmente dalle funzionalità ridotte. Molti di essi sono equipaggiati con il sistema Linux (o analogo) e quindi possono svolgere alcune delle operazioni che un pc standard può fare. Nel caso del worm, invece, la rete potrebbe essere colpita da un codice autoreplicante (virus worm) che si diffonde apparato dopo apparato a macchia d'olio [11]. In questi due contesti un bug largamente diffuso, potrebbe portare a problemi molto gravi coinvolgendo grandi porzioni di rete. In questi casi la turbolenza nella rete sarebbe molto forte e quindi un sistema di anomaly detection come SAD solleverebbe un'elevata quantità di allarmi, dando così all'operatore la possibilità di accorgersi del problema.

2.2 *Requisiti di un sistema di detection per un large ISP*

Nel momento in cui il sistema SAD è stato progettato uno dei requisiti fondamentali è stato capire come collezionare le informazioni necessarie per le analisi. L'approccio della cattura di pacchetti tramite sonde non sembrava potesse essere adottato sul breve termine; infatti, la generale razionalizzazione dei sistemi di monitoraggio portava ad una loro riduzione piuttosto che ad un aumento. È stata quindi condotta un'indagine per capire dove potessero essere presenti in azienda delle informazioni utili ai fini della detection. L'indagine ha portato ad individuare nel sistema di assurance della piattaforma VoIP un buon candidato. Questo sistema dispone, infatti, di una serie di sonde disseminate in rete e di un sistema di collezione dei cartellini o CDR (Call Detail Record). Questi cartellini contengono i dati analitici di ogni chiamata telefonica VoIP e forniscono tutte le informazioni ritenute utili ai fini dell'analisi: nello specifico il mittente ed il destinatario, l'ora di inizio, l'ora di fine e l'esito della chiamata. Tali informazioni sono estratte dalle sonde direttamente dai messaggi di segnalazione SIP catturati in rete e rese accessibili al sistema SAD.

2.3 *L'architettura generale di SAD*

In *figura 2* è visualizzata l'architettura generale del sistema SAD. Come si può notare, il sistema è costituito dal VoIP Security Subsystem che, a sua volta, s'interfaccia con il VoIP Assurance Subsystem, in precedenza descritto, per l'acquisizione dei dati.

Il VoIP Security Subsystem è composto da 4 componenti: il **prad-scanner** ed il **prad-analyzer**, costituiscono l'Engine del sistema SAD che si fa carico della manipolazione dei dati fornendo le informazioni richieste, la graphical user interface (**GUI**) e il **database**.

Il componente **prad-scanner** fa parte del sottosistema di analisi protocollare (PRAD o Protocol Anomaly Detection), da cui prende il nome. Lo scanner è un software che esegue una scansione di file contenenti cartellini, già presenti sul file system, e li carica all'interno del database. In questa fase vengono implementati alcuni semplici controlli tipo pattern-matching: per esempio sulla correttezza sintattica nei campi "from" e "to" del cartellino (cioè mittente e destinatario della chiamata). Nel caso in cui rileva dei problemi (per es. errori nelle uri) lo scanner riporta sul database specifici allarmi. Il lavoro dello scanner è piuttosto oneroso perché esso deve processare dato per dato, senza possibilità di aggregazione da parte di altri componenti. All'interno dei cartellini ricevuti sono presenti le chiamate originate o destinate alla rete VoIP (oltre che, naturalmente, quelle originate e destinate internamente ad essa). Le chiamate che non coinvolgono almeno un utente VoIP (per es. quelle tra un utente PSTN e un utente mobile) non attraversano le sonde VoIP e quindi non sono visibili in questi cartellini.

Il componente chiamato **analyzer** utilizza i dati presenti nel database per rilevare anomalie comportamentali degli utenti. L'algoritmo che viene utilizzato per questi calcoli è l'entropia: questo approccio innovativo (coperto da brevetto) consiste nel calcolare un valore simbolico relativo che "misura il polso" dell'intera rete e che permette di capire se qualcosa si sta modificando rispetto ad un comportamento rilevato nel passato e considerato "normale" o standard. Il discostamento rispetto al comportamento normale viene verificato tramite il calcolo della media e della deviazione standard, che permette di definire un intervallo entro il quale eventuali variazioni sono considerate accettabili. L'entropia viene trattata con un approfondimento tematico nel box relativo.

Al fine di perseguire l'obiettivo di identificare situazioni anomale che interessano la rete SIP nella sua interezza, è stato necessario costruire un modello che potesse gestire l'intera rete telefonica VoIP in modo da poterne osservare i macro-comportamenti.

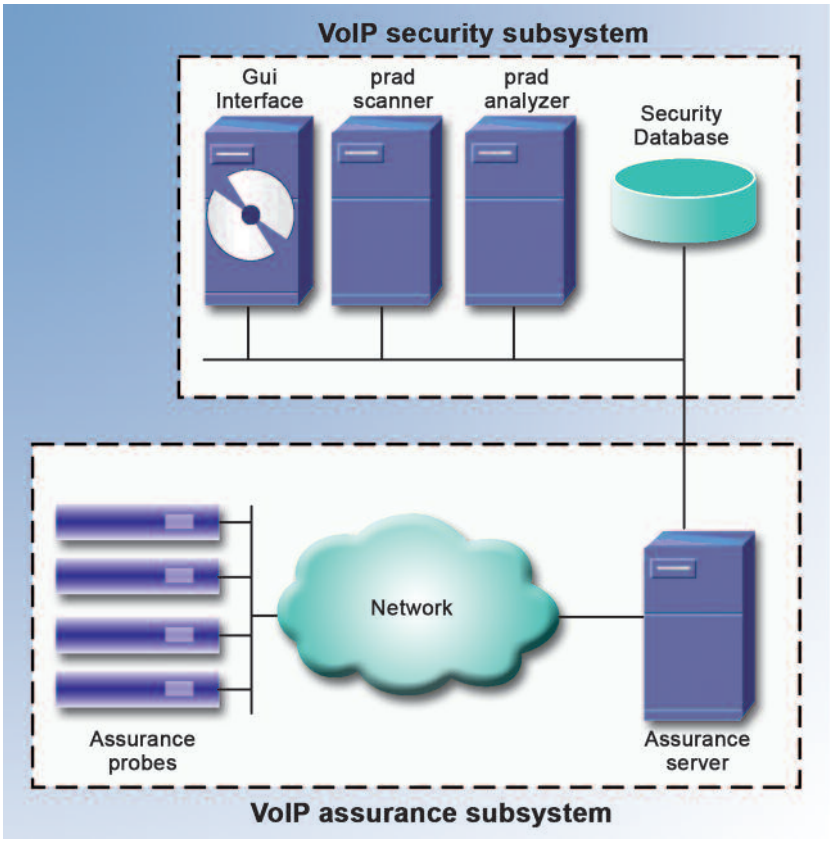


Figura 2 – Architettura generale

Il modello utilizzato è basato sul concetto di grafo orientato i cui nodi sono le URI (identificativi) degli utenti, mentre gli archi sono le relazioni di chiamata. Questo grafo rappresenta la rete telefonica SIP osservata in un determinato intervallo temporale [T1,T2]. Ogni arco ha poi una serie di proprietà che descrivono le relazioni. Attualmente le proprietà prese in esame sono:

- **totBytes**: il numero totale di byte scambiati tra i due utenti;
- **totDuration**: la durata totale delle chiamate tra i due utenti:
- **conversations**: il numero di conversazioni effettuate.

Quindi il seguente grafo:

Indica che l'utente 1 ha chiamato 2 volte l'utente 2, per un totale di 120 secondi di chiamata e 200 byte di segnalazione scambiati.

Su questo modello è possibile fare delle operazioni come confronto tra reti, calcolo di attributi (entropia, numero archi/nodi, ecc). Ad ogni iterazione si accede al database dei dati e si caricano in memoria N reti, con N specificabile in configurazione. Le reti 1..N-1 formano lo storico, la rete N è la rete "attuale", cioè quella sottoposta ad esame.

L'analisi procede seguendo l'approccio *sliding window* (*figura 3*). In questo approccio si effettua un'analisi su un certo lasso di tempo (finestra), poi si procede a far "scorrere" la finestra in avanti di intervalli di tempo prefissati di 5 minuti. L'analisi viene quindi eseguita nuovamente sulla nuova finestra, cosicchè, col passare del tempo, i dati su cui si sono effettuate le prime analisi diventano i dati storici.

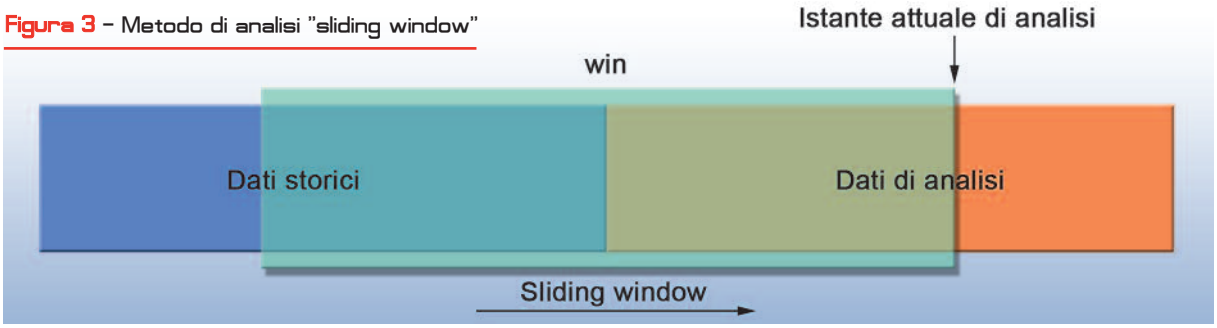
La GUI di SAD è essenzialmente il punto di consultazione degli allarmi. Essa permette di visualizzare gli allarmi sollevati dal prad-analyzer e di esaminarne i dettagli.

Permette inoltre di avere delle viste aggregate sui dati grezzi (per esempio numero di chiamate e numero di errori aggregati per giorno). Permette inoltre di gestire le utenze (locali e LDAP) gestendone l'inserimento e la scadenza ed implementa l'autenticazione forte (tramite token crittografico) degli operatori del sistema, funzionalità necessaria in ambienti operativi che trattano dati di traffico.

2.4 *Prad-scanner: dettagli*

Lo scanner è stato sviluppato in C++ e pensato per funzionare su un sistema Linux. Utilizza, infatti, una interfaccia del kernel chiamata INotify [19] che consente di creare un *listener* di eventi sul file system. In questo caso il listener creato sorveglia la creazione di file. L'interfaccia è molto più efficiente di un sistema di *polling* perché in caso di inattività non vengono sprecati cicli di cpu, utilizzabili da altri componenti del sistema operativo. Lo scanner effettua una verifica sintattica sulle URI per vedere che esse siano conformi agli standard di riferimento (rispetto ai principali RFC IETF) dopodiché carica i dati pre elaborati nel database di SAD. Sebbene i cartel-

Figura 3 – Metodo di analisi "sliding window"



lini originariamente contengono un timestamp che tiene conto anche di frazioni di secondo, in base alle statistiche calcolate in SAD si è ritenuto che la granularità fino al secondo fosse sufficiente, per cui le frazioni vengono ignorate.

2.5 *Prad-analyzer: dettagli*

Il prad-analyzer si occupa dell'analisi comportamentale degli utenti. È un componente scritto in C++ ed utilizza molte delle funzionalità della libreria Boost [16]. In particolare tale libreria viene utilizzata per costruire in memoria un modello della rete VoIP sotto forma di grafi.

Su questo modello è possibile fare delle operazioni come confronto tra reti, calcolo di attributi (entropia, numero archi/nodi, ecc). Ad ogni iterazione si accede al database dei dati e si caricano in memoria N reti, con N specificabile in configurazione. Le reti 1..N-1 formano lo storico, la rete N è la rete “attuale”, cioè quella sottoposta ad esame. Il processo di caricamento dei dati è dipendente dal database, ed è in generale la parte più onerosa. La fase di calcolo invece dura pochi secondi.

Durante la sperimentazione, il sistema è riuscito ad effettuare il caricamento di 5 finestre temporali da 1 ora (indicativamente 200.000 chiamate ciascuna) in circa 4 minuti, mentre l'analisi è stata effettuata in pochi secondi.

È quindi possibile effettuare un'iterazione ogni 5 minuti circa. Questo valore (chiamato risoluzione del sistema) indica in pratica la sua velocità: se si verifica un'anomalia essa viene notata al più tardi dopo 5 minuti.

2.6 *Database: dettagli*

Il database di SAD è il componente più critico dell'intera infrastruttura. Un database lento impedisce di fatto l'esecuzione in tempi accettabili dei componenti di SAD. Il prototipo utilizza un database MySQL su un server HP dotato di 16GB di RAM. La presenza di molta RAM consente una configurazione del server in cui molta memoria viene allocata staticamente per operazioni di indicizzazione e caching. Con le configurazioni utilizzate nel prototipo si è ottenuto un database di circa 60 GB con una velocità di risposta molto alta. Inoltre l'elevata quantità di operazioni in cache permette una fruizione delle GUI molto rapida.

2.7 *Gui: dettagli*

L'ultimo componente di SAD è la graphical user interface (GUI), un'interfaccia utente accessibile via web. Questa componente è stata sviluppata utilizzando l'ambiente RubyOnRails (ROR), un ambiente di sviluppo per applicazioni web particolarmente innovativo. La principale caratteristica di RoR è quella di consentire uno sviluppo molto veloce e agile, utilizzando una serie di paradigmi della programmazione tradizionale mescolati con altri più moderni. L'esperienza con questo ambiente è stata positiva perché ha permesso al team di progetto di ottenere ottimi risultati grafici e funzionali in breve tempo e con un dispendio di energie contenuto. Il risparmio di energie ha permesso di concentrarsi maggiormente sugli aspetti algoritmici e logici. L'utilizzo di questo tipo di approccio si colloca in un contesto più ampio di ottimizzazione dei processi prototipali che si sta sviluppando in Security Innovation. L'obiettivo è quello di ridurre i tempi di sviluppo, mantenendo la stessa qualità o di aumentare la capacità di sviluppo mantenendo il tempo costante. La metodologia utilizzata è particolarmente adatta per gli ambienti prototipali, consentendo ai team di sviluppo una gestione agile, alleggerendo alcuni aspetti formali che per un prototipo sarebbero ridondanti.

RoR implementa il modello MVC (Model-View-Controller) per la separazione delle competenze dei componenti ed il paradigma Convention-Over-Configuration per ridurre la quantità di codice di servizio in favore di alcune convenzioni implicite nel framework. Sebbene non sia ancora chiaro quale sarà il futuro di questo ambiente (per

adesso non è ancora utilizzato in grossi ambienti di produzione), per lo sviluppo di prototipi si è rivelato molto utile, efficiente e rapido.

3 SAD: risultati sperimentali

Il prototipo del progetto SAD è stato testato attraverso due trial in campo con dati provenienti dalla rete di esercizio VoIP Alice di Telecom Italia. Il primo trial si è svolto nei mesi di ottobre/novembre 2007, il secondo nei mesi di aprile/giugno 2008. Il primo trial aveva l'obiettivo di provare il funzionamento della piattaforma, mentre il secondo aveva l'obiettivo di confermare la validità delle segnalazioni sollevate.

Per questo motivo il team SAD ha lavorato in collaborazione con la struttura di esercizio della piattaforma VoIP con cui ha potuto attuare le necessarie verifiche.

3.1 *Contesto di esercizio*

La piattaforma di surveillance per il VoIP è composta da 32 sonde distribuite sul territorio nazionale (una per ogni PoP). Questi apparati hanno il compito di monitorare la rete e si posizionano vicino agli utenti finali. Ogni 15 minuti vengono generati dei cartellini relativi alle ultime chiamate e questi vengono inviati ad un sistema di collezione centrale detto TPM (Traffic Performance Monitoring).

Nelle tabelle successive (*Tabella 1*) sono riportati alcuni dati relativi ai dati di esercizio, secondo quanto collezionato nel secondo periodo di trial (aprile/giugno 2008).

Eseguendo l'analisi con i parametri di configurazione opportuni, si ottengono dati sul numero di nodi (gli utenti), sul numero di archi (le chiamate) per ciascun intervallo temporale, riportati in *Tabella 2*.

3.2 *Architettura di rete e processo di anonimizzazione*

In *Figura 4* si può vedere l'architettura generale della piattaforma VoIP e del suo interfacciamento verso SAD in occasione dei trial.

Le informazioni relative alle chiamate degli utenti sono dati di traffico e quindi sensibili per la normativa Italiana e per le policy interne Telecom Italia. Non è quindi in alcun modo possibile utilizzarle all'interno di un prototipo di laboratorio. È necessario procedere ad una loro anonimizzazione, che le rende non sensibili, ed a quel punto ne è possibile l'uso. I dati sono stati quindi elaborati da un sistema di anonimizzazione (Data Mask) che li ha poi veicolati sul server di collezione di SAD, generando quindi dei dati del tutto analoghi ai dati reali (con stesso formalismo e pari attributi statistici), ma non riconducibili agli utenti, e quindi utilizzabili in ambiente di laboratorio.

Il database che ospita tutti i dati è un componente particolarmente critico in quanto contiene

Tabella 1 – Quantita' di dati utilizzati per il trial

Utenti coinvolti	Oltre 20 milioni
Chiamate totali	Oltre 100 milioni
Chiamate giornaliere medie	Circa 2 milioni
Dimensione del database (MySQL)	Circa 60 Gb
Contratti totali	2 milioni circa
Utenti registrati solo una volta	1.5 milioni circa
Utenti che utilizzano il voip regolarmente	800.000 circa

Tabella 2 – Valori medi per intervalli temporali

Utenti		Chiamate		Chiamate per utente	
Media	82602.002	Media	50434.268	Media	0.61
Minimo	294	Minimo	163	Minimo	0.55
Massimo	201158	Massimo	127351	Massimo	0.63

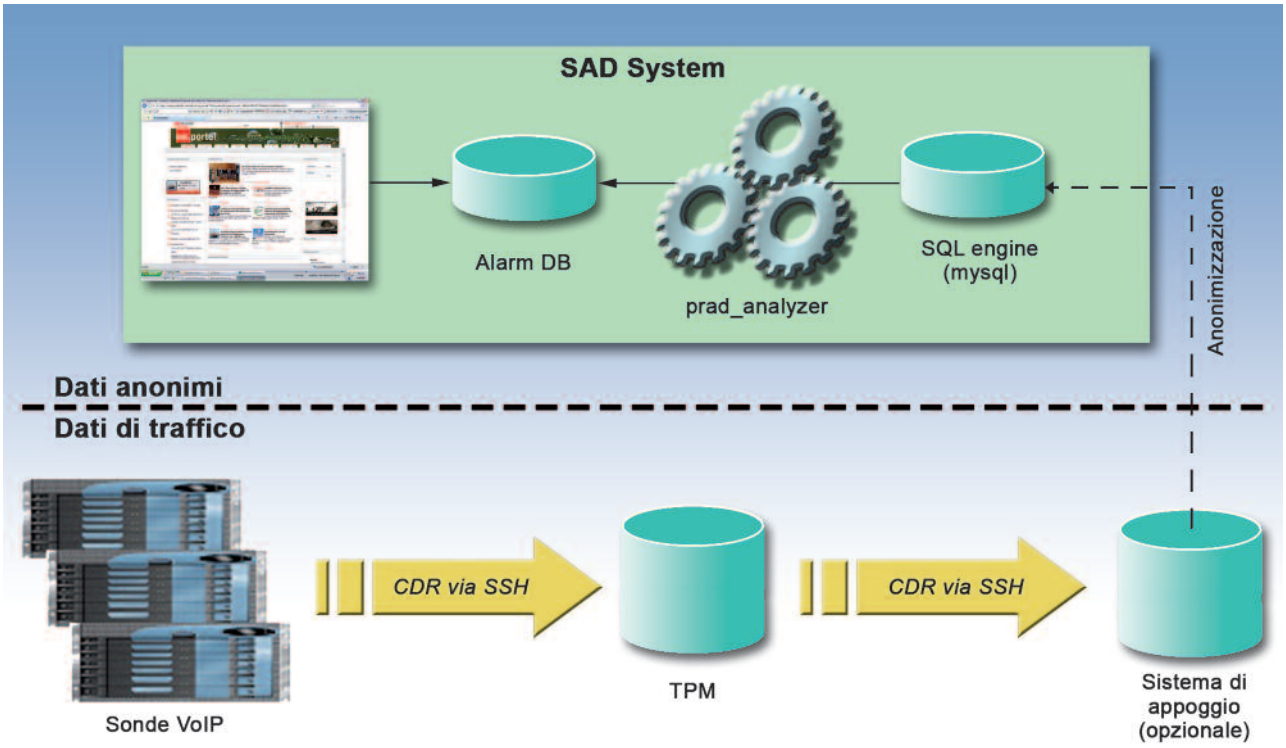


Figura 4 – Architettura di SAD durante il trial

un elevato numero di dati e quindi non può essere trattato come un database normale (la velocità di risposta sarebbe inaccettabile).

Grazie ad ottimizzazioni molto precise di caching e ad un hardware potente (server con 16Gb di RAM) è stato possibile ottenere un sistema la cui velocità di risposta è ridotta all'ordine dei secondi e che ha permesso l'esecuzione del motore di detection con prestazioni accettabili.

3.3 *I risultati del trial*

L'analisi dei dati da parte del sistema SAD ha portato ad evidenziare alcune situazioni interessanti che sono state approfondite con successive fasi di analisi al fine di dare una spiegazione ai fenomeni individuati, non necessariamente legati a problematiche di sicurezza, ma utili per capire il comportamento del servizio e dei clienti VoIP. Quello che è stato ottenuto è una serie di casi di interesse che vedremo approfonditi nei seguenti paragrafi.

3.3.1 Comportamento della rete VoIP

In *figura 6* è possibile vedere una visualizzazione dell'andamento delle chiamate per un periodo di 11 giorni. È da notare l'andamento a "doppia gobba" tipico delle reti dove è presente una componente umana. La prima gobba corrisponde circa alle 11 del mattino, orario di picco delle attività, mentre la seconda circa alle 3 del pomeriggio, altro picco.

La sella centrale corrisponde ad un momento di ridotta attività (12:30 circa), mentre i minimi a margine sono relativi alle ore notturne. La somiglianza dell'andamento dell'entropia (*Figura 5*) con l'andamento del traffico vero e proprio (*Figura 6*) indica che il modello rappresenta in maniera efficace il traffico. Si noti anche che Sabato e Domenica presentano un traffico inferiore rispetto ai giorni feriali. Per capire se ci sono anomalie, il valore attuale dell'entropia viene confrontato con i valori calcolati in precedenza.

Una discordanza rilevante è indice del fatto che qualcosa è avvenuto in rete, ma trattandosi

Trattazione matematica dell'entropia

La tecnica dell'entropia è una nuova tecnica per la rilevazione delle anomalie statistiche nel traffico nella rete VoIP causato da fallimenti del funzionamento della rete o da attacchi generali come (Distributed) Denial of Service ((D)DoS) oppure SPIT (SPam over Internet Telephony). Le caratteristiche simboliche estratte dal traffico della rete SIP sono SIP URI della sorgente e del destinatario (denotate come *sURI* e *dURI*, rispettivamente) prese dai Call Data Record (CDR) corrispondenti alle sessioni SIP. In generale l'entropia è una misura della dispersione statistica associata ai valori simbolici dei dati in un intervallo di tempo. Questa misura raggiunge il suo minimo se tutti i valori sono identici tra loro e raggiunge il suo massimo se i valori sono distribuiti uniformemente nella gamma dei valori simbolici ammissibili. Nel nostro approccio viene selezionato un intervallo di tempo di lunghezza fissa T che viene fatto scorrere nel tempo avanzando con un passo discreto pari a ΔT. L'entropia utilizzata è l'entropia quadratica associata alle frequenze relative dei valori simbolici osservati. L'essenza della tecnica consiste nel computare l'entropia quadratica per ogni finestra e trovare i punti nel tempo dove l'entropia varia improvvisamente rispetto ad un'appropriata differenza relativa. Ci si aspetta che il cambiamento relativo dell'entropia tra due finestre consecutive possa essere molto più piccolo nel traffico normale rispetto a quando c'è una transizione dal traffico normale al traffico anomalo.
Più precisamente, per la k-esima finestra, poniamo che $F_i(k)$ denoti il numero delle volte in cui un valore simbolico a_i si raggiunge, per esempio la frequenza assoluta di questo valore, e che $f_i(k)=F_i(k)/n_k$ denoti la corrispondente frequenza relativa, dove nk denota il numero totale dei valori osservati nella finestra. L'entropia quadratica per la k-esima finestra poi va computata come

$$Q(k) = \sum_i f_i(k)(1-f_i(k)) = 1 - \sum_i f_i(k)^2 = 1 - C(k)$$

dove $C(k)$ è la corrispondente misura quadratica della concentrazione. In pratica, solo le frequenze relative più alte possono essere messe in conto, trascurando quelle più basse. La differenza relativa va poi computata come

$$\delta(k) = \frac{(C(k) - C(k-1))^2}{\sqrt{(C(k)C(k-1)(1 - C(k))\ (1 - C(k-1))}}$$

L'allarme va annunciato se il valore $\delta(k)$ eccede una soglia definita in modo tale che la quantità dei falsi allarmi sia sufficientemente bassa nel traffico normale.
In uno scenario di incondizionato, le variabili simboliche osservate sono sURI, dURI, e la variabile congiunta (sURI, dURI). In questo caso vengono considerate le misure della concentrazione delle >

SIP URI della sorgente, SIP URI del destinatario, oppure dei SIP URI della sorgente e del destinatario congiunti. Nello scenario condizionato, le variabili simboliche osservate sono sURI | dURI o dURI | sURI. In questo caso si tratta delle misure della concentrazione delle SIP URIs della sorgente condizionate ai SIP URIs del destinatario o viceversa. Se il valore della variabile che condiziona non è fisso, potrebbe essere vantaggioso calcolare la misura media condizionata della concentrazione. Se fi,j (k) denota la frequenza relativa del valore (ai, bj) della variabile simbolica congiunta (x, y), la frequenza relativa del valore simbolico ai condizionata al valore simbolico bj si definisce come

$$f_{i|j}(k) = \frac{f_{i,j}(k)}{f_j(k)}$$

dove $f_j(k)=\sum_i f_{i,j}(k)>0$ è la frequenza relativa del valore simbolico bj. La misura media condizionata quadratica della concentrazione del x dato y va poi computata come

$$\overline{C}_{x|y}(k) = \sum_i f_i(k) \sum_i f_{i|j}(k)^2 = \sum_{i,j} \frac{f_{i,j}(k)^2}{f_j(k)}$$

In questo caso si valuta la misura media della concentrazione delle SIP URI della sorgente condizionata alle SIP URI del destinatario o viceversa.
In caso di attacco DDoS che colpisce un numero piccolo di destinatari da un numero grande di destinazioni (scelte casualmente) ci si aspetta che la misura quadratica della concentrazione del *sURI* decrementi e che la misura media quadratica della concentrazione del *sURI* condizionata al *dURI* decrementi ancora più significativamente. Similarmente, in caso di attacco SPIT verso un numero grande di destinatari casualmente scelti verso un numero piccolo di SIP URI sorgente, ci si aspetta che la misura media quadratica della concentrazione delle *URI destinazione* condizionata alle sorgenti decrementi significativamente. Equivalentemente, ci si aspetta che le corrispondenti entropie quadratiche incrementino significativamente in tutti due casi.

jovan.golic@telecomitalia.it

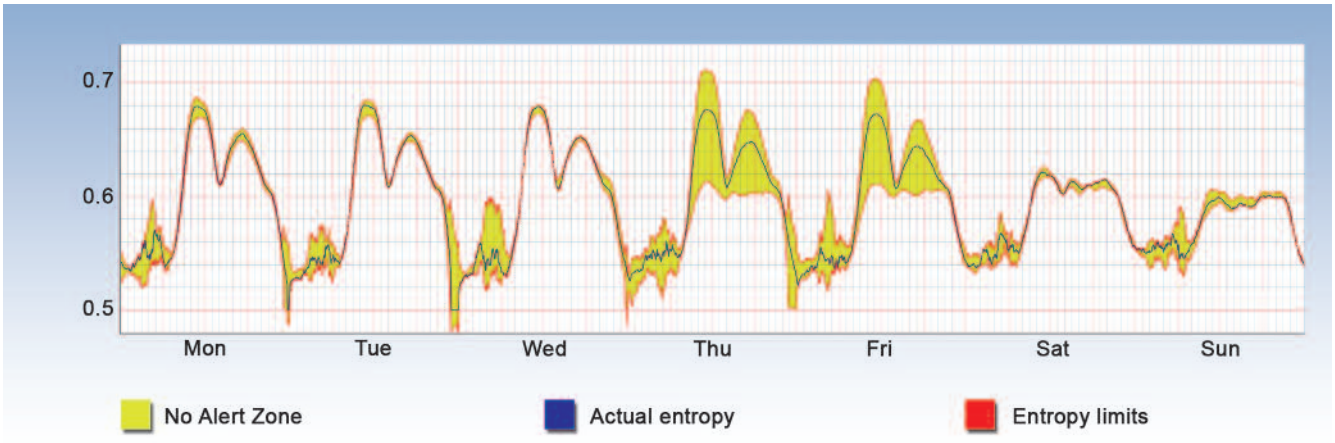


Figura 5 – Grafico dell'entropia in una settimana di traffico

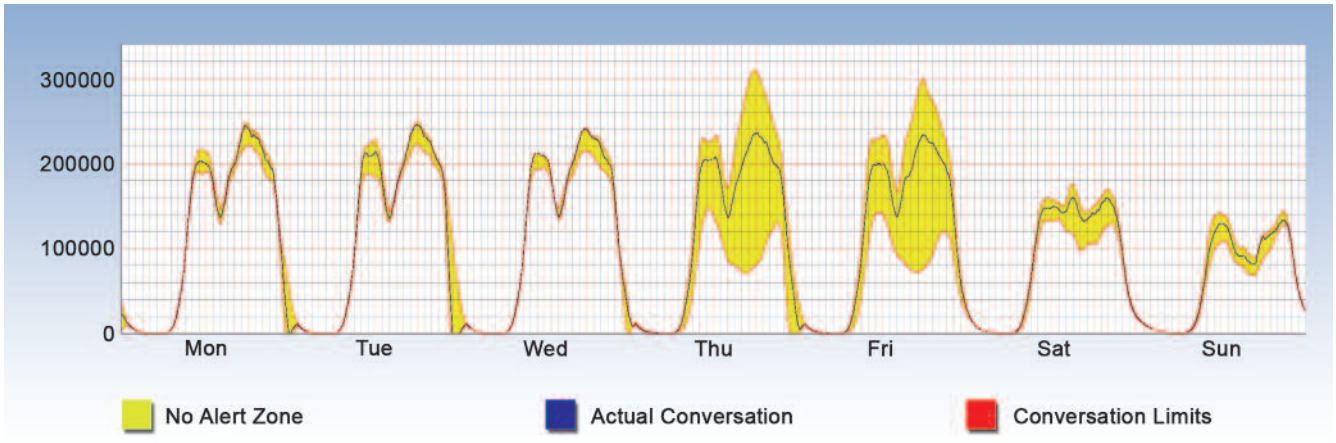


Figura 6 – Grafico delle chiamate

di un valore aggregato non consente, in generale, di individuare lo specifico parametro (utente per esempio) responsabile di tale anomalia.

Per avere un maggior precisione l'analyzer avvia un processo di approfondimento al fine di far emergere quali utenti hanno contribuito a far variare l'entropia.

Viene quindi analizzato il comportamento di ogni singolo utente scartando quelli che hanno bassa rilevanza (poco traffico): quello che si ottiene è una lista di utenti sospetti.

3.3.2 Grafi relazionali

Elaborando i dati collezionati da SAD è possibile creare dei grafi analoghi a quelli che il sistema si crea in memoria. Questi grafi permettono all'analista di avere una rappresentazione grafica dello stato della rete, il che consente di evidenziare situazioni altrimenti difficili da individuare. In *Figura 7* è riportato un grafo relativo ad un comportamento della rete VoIP rilevato durante la sperimentazione . Nell'immagine è possibile individuare una grande quantità di utenti poco rilevanti (a bordo immagine), una grande quantità di utenti di media rilevanza (visibili spostandosi verso il centro) ed un ridotto numero di grossi cluster che rappresentano utenze molto attive. In generale, un elevato numero di archi entranti suggerisce che il nodo possa essere un call-center, un elevato numero di archi uscenti indica invece un contact-center oppure un ele-

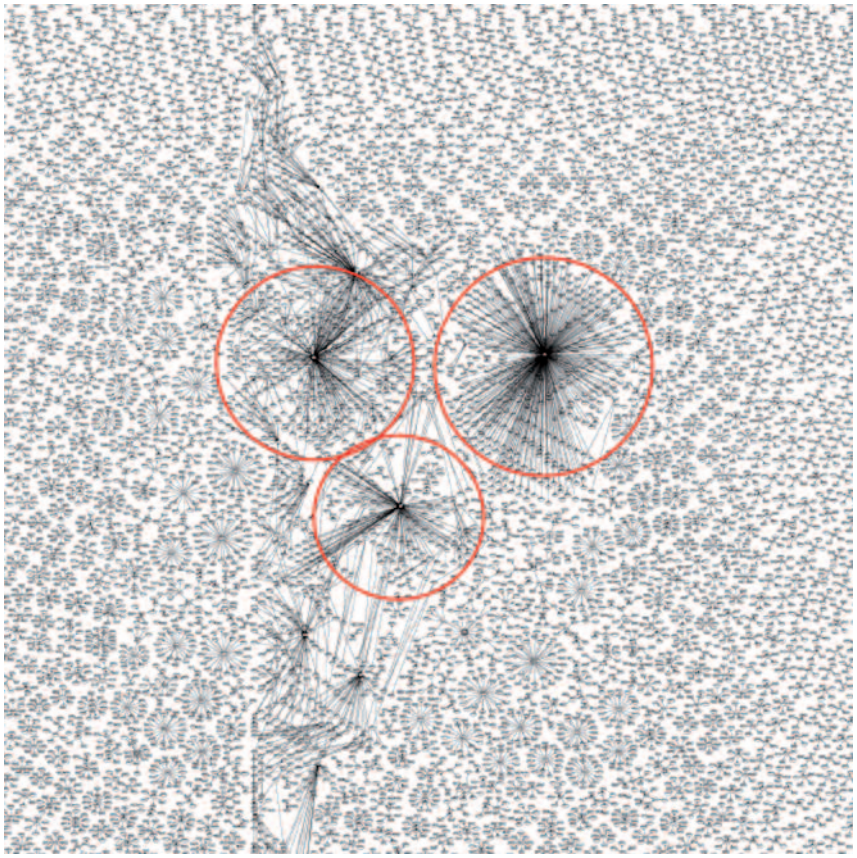


Figura 7 – Grafo delle chiamate costruito nel motore di SAD

mento di rete che presenta un unico numero per le chiamate uscenti ma che gestisce più numeri per le chiamate in ingresso; infine un nodo con diversi archi sia uscenti che entranti è che molto probabilmente un centralino.

Durante il normale funzionamento del sistema, vengono creati tanti grafi relazionali come quello in *Figura 7*. Essi vengono trattati come istantanee che rappresentano lo stato della rete in determinati momenti. Le "immagini" precedenti all'istante attuale vengono trattate come dato storico e contribuiscono ad effettuare l'addestramento (training) del sistema. Questo metodo consente di rilevare anomalie derivanti da discostamenti rispetto a ciò che è successo in precedenza. Per esempio la presenza di un grosso cluster è anomala solo se questo cluster non era già presente. Infatti un importante call center (es. Il 187) è presente sin dall'inizio dell'analisi con una quantità di archi costante nel tempo (a parità di orari). Il sistema quindi non solleva un falso allarme (es. attacco DoS verso il cliente 187) grazie alla presenza dello storico che consente di rilevare tale comportamento come normale. Se un volume di chiamate analogo si rilevasse su un'utenza che non ha manifestato in precedenza attività analoghe, allora saremmo in presenza di un probabile attacco.

3.3.3 Rilevazione di sistemi di test

Durante l'analisi dei dati, il sistema ha rilevato anomalie comportamentali relativamente a specifiche utenze rivelatesi in seguito come sistemi di testing. Approfondendo l'analisi il sistema ha notato che queste utenze effettuavano un numero di chiamate talmente elevato tale da implicare l'utilizzo di sistemi automatici di chiamata. In particolare è stato notato che le chiamate effettuate erano tutte non andate a buon fine ma terminate immediatamente dal mittente (487 Request Terminated), e la loro durata era di conseguenza inferiore ad un secondo. Questa situazione poteva essere realizzata da un tool di attacco (phone scanner, DoS, ecc) o essere causata da una qualche anomalia non intenzionale. Dal punto di vista di un operatore un attacco intenzionale, un malfunzionamento o qualunque fenomeno che può inficiare il corretto funzionamento del servizio è ugualmente pericoloso e da mantenere sotto controllo. In *Figura 8* ed in *Figura 9* si può vedere la rappresentazione grafica del comportamento di uno degli utenti sospetti. Nella figura "calls per day" si può osservare che l'utente normalmente non fa traffico mentre in certi giorni genera dei picchi di chiamate. Questi picchi non hanno il codice

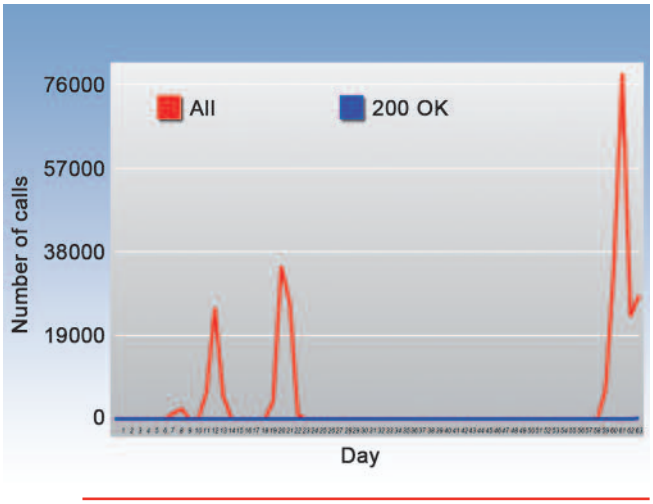


Figura 8 – Chiamate al giorno per un utente sospetto

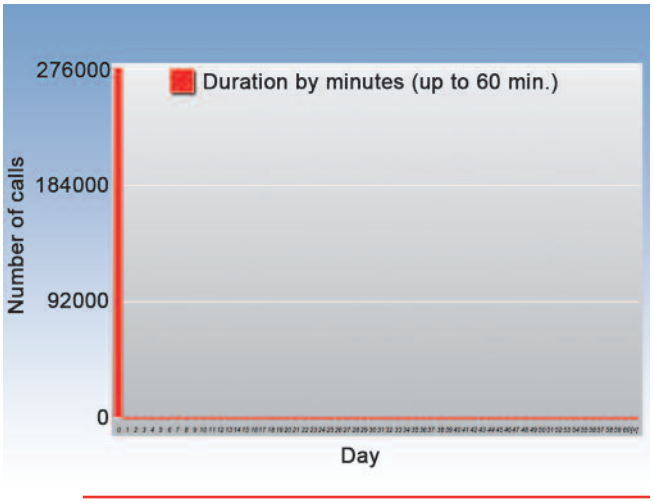


Figura 9 – Distribuzione delle chiamate per un utente sospetto

200OK (chiamata andata a buon fine), quindi tutte le chiamate che tenta di instaurare sono fallite. Si ha riscontro di questo guardando la "call duration distribution", cioè il conteggio di quante chiamate sono state fatte suddivise sulla base della loro durata (quante lunghe meno di 1 minuto, meno di 2, meno di 3 e così via).

L'analisi svolta sui dati ha spiegato che questi fenomeni erano riconducibili a sistemi di test impiegati direttamente in rete. Il comportamento di questi sistemi era comunque simile ad un attacco e quindi il sistema ha correttamente evidenziato il fenomeno, dimostrando l'efficacia del metodo di analisi adottato.

3.3.4 Anomalie legate alle festività infrasettimanali

Il sistema di monitoraggio effettua dei confronti tra giorni uguali all'interno della settimana, quindi va a confrontare i lunedì con i lunedì, i martedì con i martedì e così via. Questo modello funziona finché i giorni trattati sono effettivamente simili.

Se in una settimana cade un giorno di festa il comportamento muta. Confrontare un lunedì' qualunque con il giorno di Natale (se esso cadesse di lunedì) porta inevitabilmente a rilevare dei falsi positivi. Questo è un difetto fisiologico di tutti i sistemi di monitoraggio che può essere evitato solo ampliando enormemente il periodo di osservazione. È stato verificato sperimentalmente come sia ottimale, l'utilizzo di un periodo di training di 1 mese.

Qualora si attraversassero dei giorni di festa è noto che verranno sollevate delle anomalie: in questo caso l'operatore, dopo averle verificate può ignorarle. È questo il caso che si è verificato durante il trial che si è svolto tra metà aprile e metà giugno, attraversando la festività del 2 Giugno. In quella data sono state rilevate anomalie relativamente a importanti call-center che hanno manifestato un inferiore numero di chiamate entranti. Ciò è normale, per quanto detto sopra, ed è spiegabile con il fatto che meno persone chiamano per segnalare problemi.

4 Conclusioni

Per effettuare il trial è stato necessario sottoporre i dati ad un processo di anonimizzazione, e per questo motivo il trasferimento dei dati è stato effettuato in modalità batch. Il modo ideale di analizzare i dati è invece in maniera automatica e continuativa, in modo che il motore di analisi possa elaborare i dati continuamente.

Sebbene i dati non possano essere convogliati sul server SAD nel momento stesso in cui sono

creati, la riduzione al minimo dei tempi di trasferimento può portare ad avere un effetto "near-real-time" che consentirebbe all'operatore di avere delle segnalazioni dalla rete il più tempestive possibili: la previsione è di avere un tempo di reazione inferiore ad 1 ora dal momento di inizio del fenomeno.

L'entropia risulta essere una tra le grandezze ideali da controllare per verificare la presenza di anomalie. La sua espressività e la sua sintesi sono due elementi chiave nel suo utilizzo. Il discostamento dai comportamenti normali viene poi verificato tramite il calcolo della media e della deviazione standard, che permettono la creazione di una "fascia di quiete" o "intervallo di confidenza", una zona cioè in cui, rispetto all'entropia, il comportamento della rete è considerato normale. SAD, essendo un software di anomaly detection, non solleva allarmi per fenomeni puntuali, ma li solleva in tutti quei casi che hanno un forte impatto sulla rete, per cui l'ordine di grandezza della variazione si avvicina agli ordini di grandezza delle proprietà sotto esame. Il prototipo di SAD ha superato con successo i trial e si è dimostrato utile ai fini di monitoraggio e di security. Ha dato inoltre evidenza di come un efficiente riutilizzo delle informazioni già presenti in azienda possa portare a risultati di notevole interesse.

Una riduzione della ridondanza delle informazioni a vantaggio di un utilizzo molteplice delle stesse può essere un *driver* per lo sviluppo di sistemi innovativi che sfruttino l'alto valore di certe informazioni per ottenerne delle altre di valore ancora superiore. Una efficace analisi delle fonti di dati provenienti dai sistemi di esercizio e dei loro punti di stoccaggio può portare ad una utile razionalizzazione dei dati.

Attualmente si sta pianificando un utilizzo del sistema SAD in esercizio.

ACRONIMI

DNS Domain Name System
DOS Denial of service
GUI Graphical User Interface
IMS IP Multimedia Subsystem
IP Internet Protocol
LDAP Lightweight Directory Access Protocol
MVC Model View Controller
NGN Next Generation Network
PoP Point of Presence
PSTN Public Switched Telephone Network
RoR Ruby on Rails
SAD Sip Anomaly Detection
SBC Session Border Controller
SIP Session Initiation Protocol
SPIT Spam Over Internet Telephony
URI Uniform Resource Identifier
VoIP Voice over IP

BIBLIOGRAFIA

[1] Misitano, Pasquinucci "VoIP: una interessante novità con molte problematiche di sicurezza"

[2] Aurobindo Sundaram, "An Introduction to Intrusion Detection", *http://www.acm.org/crossroads/xrds2-4/intrus.html*

[3] Università degli Studi di Genova, Tesi di Laurea "Studio e sviluppo di metodologie per la rilevazione di attacchi Denial of Service basati sul protocollo SIP", Ferreri Fabia

[4] Canal "SIP: una tecnologia per reti e servizi di prossima generazione", ISBN 88-85404-37-5

[5] Minerva, Manzalini, Moiso "I dati di un Operatore: quali sono, come usarli e perché "esporli". Cod. Doc. TFC0900006

[6] IETF, RFC5039, "The Session Initiation Protocol (SIP) and Spam", J. Rosenberg C. Jennings (Cisco), January 2008

[7] ETSI TR 187 009 "Feasibility study of prevention of unsolicited communication in the NGN"

[8] 3rd Generation Partnership Project; Technical Report Group Services and System Aspects; Protection against SMS, MMS and IMS SPAM; Study of Different SPAM Protection Mechanisms Release 8

[9] ITU-T Study Group 17, X.ocsip – Overview of countering SPAM for IP multimedia application - TD 2499 Rev.1
[10] VoIP-IRC bot, Mohamed Nassar, Radu State, Olivier Festor "VoIP security : a myth or a fact ? let's try this bot and we'll get the response later" *http://www.loria.fr/~nassar/readme.html*
[11] Storm botnet, *http://en.wikipedia.org/wiki/Storm_botnet*
[12] February 14, 2008: Blue Box #76: Cisco, Skype and BT vulnerabilities, when SIP looks like SPIT, VoIP security threat predictions and the FBI forgets to pay their bills, plus listener comments and more... *http://www.blueboxpodcast.com/2008/02/blue-box-76-cis.html*
[13] Morello, Repubblica Affari&Finanza, 04 febbraio 2008
[14] MSNbc, "'Vishing' scams use your telephone to hook you", *http://www.msnbc.msn.com/id/14138614/from/ET/*
[15] Il sole 24 ore, "Gli IT manager italiani non temono il VoIp", 04 Febbraio 2008
[16], Spam Monthly Report, May 2008
[17] *http://www.voipsa.org/pipermail/voipsec_voipsa.org/2006-March/001324.html*
[18] *http://www.guardian.co.uk/technology/2007/nov/01/news.hacking*
[19] *http://inotify.aiken.cz/*
[20] *http://www.boost.org/*
[21] *http://www.20min.ch/news/zuerich/story/12093602*

dario.lombardo@telecomitalia.it
paolo.delutiis@telecomitalia.it

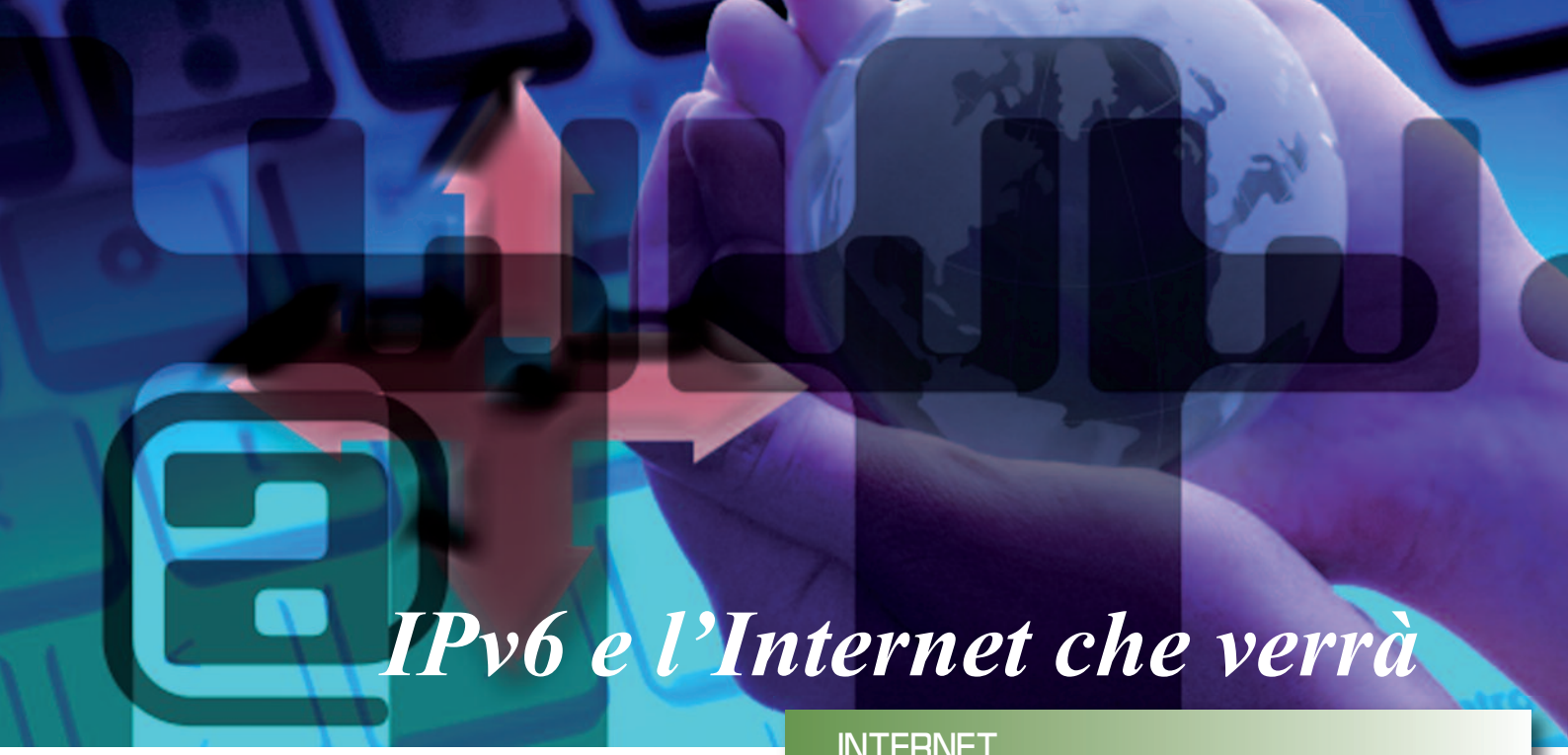
AUTORI

Dario Lombardo

laureato nel 2001 in Ingegneria è entrato in Tilab dopo la laurea e da allora ha lavorato nel gruppo di Security. Le prime esperienze sono state di vulnerability assessment sia per attività interne che verso clienti esterni. Le esperienze successive si sono rivolte principalmente allo sviluppo di sistemi innovativi in ambito Security con particolare attenzione verso i protocolli di routing e il VoIP. Oggi è responsabile di progetto per lo sviluppo di prototipi in ambito network security con particolare interesse verso la protezione dell'infrastruttura DNS ■

Paolo De Lutiis

laureato in Scienze dell'informazione, entra in Telecom Italia nel 2000, prendendo parte da subito a progetti relativi alla sicurezza ICT. Attualmente in Security Innovation, partecipa a progetti relativi alla sicurezza della NGN2 e segue i lavori di standardizzazione TISPAN nel Working Group 7, in qualità di rapporteur di specifici work Item sulla sicurezza NGN ■



IPv6 e l'Internet che verrà

INTERNET

Paolo Fasano, Domenico Marocco, Maurizio Siviero

Obiettivo di questo lavoro è fornire un aggiornamento sulla necessità di affrontare l'introduzione del nuovo protocollo in rete e analizzare cosa voglia dire compiere questo passo nella realtà della rete di un Service Provider come Telecom Italia. Inoltre ci si propone di gettare uno sguardo su ciò che potrebbe diventare Internet nel prossimo futuro per inquadare l'esigenza di IPv6 nel più generale scenario di evoluzione della rete.

1 Introduzione

Sono passati alcuni anni da quando il protocollo IPv6 [1] è stato specificato. Anche la fase di sperimentazione che ha visto l'operatività della rete 6Bone [2] si è conclusa. In questi anni l'attuale versione del protocollo IP ha mostrato capacità di sopravvivenza inimmaginabili, pur tuttavia le ragioni per un nuovo protocollo non sono venute meno e l'esaurimento degli indirizzi IP sta diventando una preoccupazione concreta per gli Internet Service Provider.

IPv6 offre uno spazio di indirizzamento davvero ampio e propone nuovi meccanismi per la configurazione automatica dei terminali; queste novità sono sufficienti a fare di IPv6 un protocollo incompatibile con IPv4 e quindi a rendere il problema della migrazione alquanto complesso. I Service Provider stanno studiando le modalità più opportune per affiancare il trasporto di IPv6 a quello di IPv4 con impatti su tutte le componenti tecnologiche, infrastrutture di rete, piattaforme di controllo, servizio e gestione e terminali d'utente.

Si prospetta all'orizzonte uno scenario in cui i due protocolli coesisteranno a lungo e anche altri segnali, ad esempio i più recenti sviluppi nel contesto dell'Home Networking, fanno intravedere la prospettiva di una rete sempre più eterogenea. La nuova frontiera del networking potrebbe proprio essere quella di gestire secondo nuovi paradigmi questa realtà diversificata.

2 Perché IPv6

2.1 *Stato delle assegnazioni IPv4*

Alla fine degli anni '70 quando fu definito il protocollo IPv4 ancora oggi in uso [3], un campo indirizzo di 32 bit, corrispondente ad uno spazio di 2^32 (circa 4,3 miliardi) di indirizzi, sembrò sufficiente a soddisfare ogni possibile espansione di Internet.

Tuttavia già nella metà degli anni '90 apparve chiaro che la limitazione dello spazio di indirizzamento, anche a causa delle soluzioni tecniche e delle politiche di assegnazione adottate, sarebbe diventato a breve un vincolo all'espansione della rete. Furono quindi adottate soluzioni tecniche per contrastare l'esaurimento degli indirizzi (principalmente CIDR e NAT), e fu operata una prima revisione delle politiche di assegnazione.

Il CIDR (Classless Inter-Domain Routing) [4] costituisce per Internet il passaggio da routing Classfull a routing Classless: con il routing Classfull si assume che una classe di indirizzi non possa essere partizionata in due domini separati. Con il routing classless viene superata questa rigidità, pemettendo quindi di partizionare classi di indirizzi tra organizzazioni differenti e aumentando l'efficienza di utilizzo.

Il Network Address Translation (NAT) [5] e la sua estensione Network Address Port Translation (NAPT anche noto come PAT) [6] permettono ad apparati appartenenti ad una rete con indirizzamento privato di utilizzare un numero limitato di indirizzi pubblici per accedere all'esterno della propria organizzazione. L'introduzione del NAT costituisce una radicale modifica dell'architettura di Internet: viene meno il principio end-to-end che costituiva uno dei capisaldi del progetto iniziale della rete, principio che voleva assicurare che la rete fornisse puro trasporto. Con i meccanismi di NAT gli apparati intermedi si interpongono sempre almeno a livello TCP o UDP, ma in alcuni casi devono intervenire anche nei messaggi di protocolli applicativi (ad esempio SIP) che possono trasportare informazioni su indirizzi di livello 3 e/o 4. Una rete in cui si utilizzi NAT richiede quindi anche Application Layer Gateway (ALG) per specifici protocolli. Si risparmiano quindi indirizzi ma i costi crescono.

Non sempre inoltre l'introduzione di un ALG è sufficiente a garantire un corretto funzionamento dei software applicativi: applicativi basati sul paradigma peer-to-peer, in generale non funzionano correttamente in presenza di NAT. Il NAT introduce inoltre problemi nella gestione di meccanismi di mobilità basati su Mobile IP e di meccanismi di sicurezza end-to-end basati su IPsec: 'rompere' il principio 'end-to-end', vista l'attuale architettura di servizi ed applicativi di Internet, è quindi critico.

Le soluzioni tecniche introdotte e le nuove politiche di assegnazione hanno determinato nella seconda metà degli anni '90 un sensibile rallentamento nella corsa agli indirizzi, come evidente dall'andamento della curva di *Figura 1*.

A partire dal 2001, in corrispondenza dell'espansione degli accessi Broadband fissi, dell'espansione della rete in tutti i continenti ed ultimamente dell'accesso alla rete dati anche parte dei terminali mobili, la crescita è ripresa. Vari tentativi sono stati fatti negli anni di interpolare la curva e fare delle previsioni realistiche sulla possibile data di esaurimento: tra gli altri sono da citare gli studi di G. Huston a cui si rimanda [7][8]. Dal gennaio 2004 al gennaio 2009 sono state assegnate 54 classi A, circa 10 all'anno, con un picco di 13 nel 2007. Ne rimangono 31 da allocare; a questi ritmi, potranno bastare per 24/36 mesi. IANA (Internet Assigned Number Authority) assegna gli indirizzi agli enti di assegnazione regionali (come il RIPE NCC responsabile per Europa, Medio Oriente e Asia Centrale); trascorreranno quindi probabilmente altri 12 mesi prima che un'organizzazione che ne faccia richiesta non riceva indirizzi. Quindi se nulla di nuovo succede, entro il 2012 non saranno più disponibili indirizzi IPv4.

Ma può succedere ancora qualcosa? Ci sono 16 classi A riservate per usi futuri: utilizzarle potrebbe essere complicato perchè alcuni router potrebbero non instradare i pacchetti verso que-

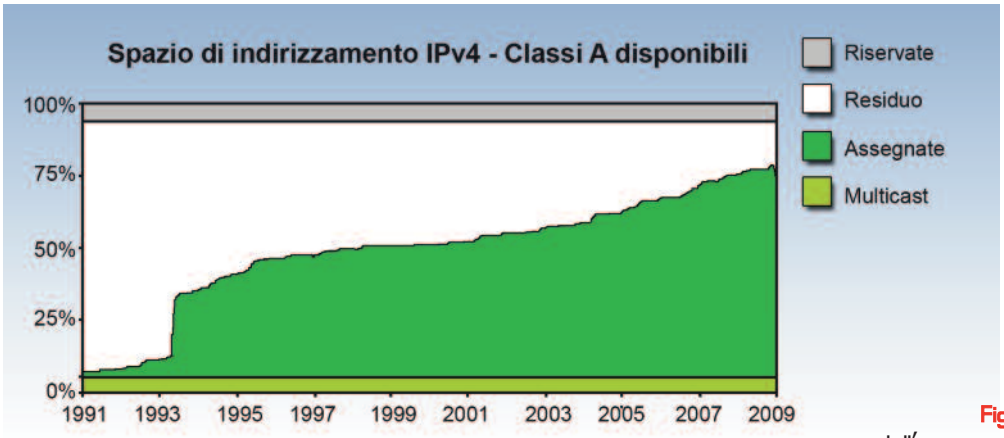


Figura 1 – Andamento dell'assegnazione degli indirizzi IPv4 da parte IANA

sti indirizzi (un router conforme allo standard dovrebbe scartarli). Esistono poi molte classi di indirizzi assegnate nei primi anni '90 che sono sotto utilizzate. Ad oggi sono difficilmente recuperabili: chi dovrebbe rilasciare gli indirizzi, deve sicuramente sostenere costi e disservizi per rinumerare le proprie reti.

Un punto di vista interessante è stato recentemente espresso da B. Edelman [9], economista di Harvard: gli indirizzi IPv4 sono stati sino ad oggi concessi ad un costo irrisorio, fuori da una politica di mercato. Inoltre è espressamente vietato trasferire indirizzi da un'organizzazione ad un'altra. Gli indirizzi non utilizzati possono solo essere restituiti agli enti che li hanno assegnati. Rivedere questa politica, permettendo il trasferimento a prezzi di mercato, potrebbe favorire un uso più efficiente, il rilascio degli indirizzi non utilizzati e quindi prolungare la vita di IPv4. Ma la comunità di Internet è pronta a questo passo?

2.2 *Problemi pratici per gli ISP*

L'esaurimento degli indirizzi IPv4 sta progressivamente trasformandosi da un tema di studio, ad un problema pratico per gli ISP: ad oggi le richieste di indirizzi da parte degli ISP verso gli enti di assegnazione regionali vengono ancora soddifatte, ma sono aumentati i tempi burocratici per ottenere gli spazi di inidirizzamento richiesti. Per evitare una corsa all'accaparramento di indirizzi viene effettuato un monitoring dell'efficienza di utilizzo degli indirizzi già assegnati e vengono assegnati nuovi indirizzi solo quando il richiedente dimostri di avere esaurito, con la migliore efficienza di utilizzo possibile, gli spazi di indirizzi già assegnati. Si vedano ad esempio le norme vigenti all'interno che RIPE NCC [10].

Naturalmente l'adozione di queste norme porta ad un carico operativo più elevato nella gestione della rete da parte dell'ISP. Frequente è anche la tentazione di utilizzare indirizzi privati per numerare terminali che nel breve termine non hanno necessità di accesso ad Internet o per i quali l'impiego di NAT non costituiscano nell'immediato un problema rilevante.

Le tecniche di NAT, se da una parte sono sempre state osteggiate da parte degli architetti di Internet, hanno sempre trovato terreno abbastanza fertile tra i Service Provider. Vi sono almeno tre ragioni che giustificano questo atteggiamento. La prima ragione è che il NAT introduce un incremento di costi che spesso è a carico dell'utente stesso: attualmente il NAT viene realizzato ai bordi della rete, su CPE o Network Termination, i cui costi sono a carico dell'utente finale o facilmente ribaltabili sullo stesso. La seconda ragione è che il NAT è percepito come un meccanismo che permette un maggior controllo da parte dell' ISP del traffico in rete: limita le applicazioni end-to-end (almeno per gli utenti non sufficientemente esperti) e quindi fa in modo che per l'utente finale

sia più semplice utilizzare i servizi forniti dagli ISP rispetto a quelli forniti da terze parti. Da sempre, infine, è opinione comune che i NAT costituiscano un meccanismo di sicurezza a basso costo. Naturalmente è possibile confutare queste affermazioni, tutte molto deboli in una prospettiva di medio/lungo periodo, ma più difficili da scalzare nel breve periodo, tanto è vero che proprio per l'interesse degli ISP vengono oggi riproposte in IETF (*Internet Engineering Task Force*) soluzioni di NAT a più livelli (si veda ad esempio [22]), anche note come Carrier Grade NAT o Large Scale NAT [21].

IPv6 dal punto di vista dei Service Provider è sicuramente una scelta molto più controversa. Il regime di concorrenza impone ai Service Provider attenzione alle richieste del mercato, alla stabilità delle reti e ai costi delle soluzioni fornite. IPv6 sembra andare contro tutti questi principi basilari della fase commerciale di Internet. Non vi è richiesta da parte del mercato perchè non vi sono ad oggi nuove prestazioni abilitate in IPv6 che non siano già disponibili in IPv4. Dal punto di vista della stabilità della rete, anche se negli ultimi dieci anni si sono moltiplicate le sperimentazioni, vale la massima generale, che ogni cambiamento introduce necessariamente transitori in cui si creano disservizi. Da ultimo i costi: per partire occorre intervenire riprogettando la rete, formando il personale, aggiornando le release sulle macchine; nel transitorio in cui la rete sarà dual stack, occorre disporre di macchine con prestazioni maggiori in grado di gestire entrambi i protocolli; per completare la migrazione infine occorre sostituire gli apparati più vecchi, non in grado di evolvere ad IPv6 e migrare conseguentemente l'utenza che da questi è servita.

Nei Service Provider è, quindi, ancor oggi radicata la convinzione che, anche se il passaggio ad IPv6 sarà inevitabile, saranno necessariamente coloro che si muoveranno per primi a sostenere i maggiori costi e ad avere maggiori ritorni negativi: maggiori costi e minor stabilità della rete non potranno che risultare in una minore competitività sul mercato per gli early adopter. Secondo questa logica è quindi conveniente procrastinare l'introduzione di IPv6 il più a lungo possibile. Ma è davvero questa la logica corretta?

Il dubbio è più che legittimo; la migrazione verso IPv6 come detto è un processo molto lungo e chi parte troppo tardi rischia di trovarsi in difficoltà, esattamente come chi inizia la migrazione troppo presto.

Occorre notare che in passato sono state studiate soluzioni per permettere a traffico IPv6 di attraversare isole IPv4 only: queste soluzioni basate su tunneling sono presenti oggi all'interno di diversi sistemi operativi, ad esempio XP e Vista. Non offrire connettività IPv6 nativa da parte di un ISP potrebbe lasciare spazio alla realizzazione di reti overlay basate su questi meccanismi.

Dal punto di vista degli ISP quindi l'unica strategia vincente, o forse non perdente, è quella di un progressivo adeguamento dell'infrastruttura, con una realistica attenzione ai costi: progettazione della rete IPv6, scelta di apparati e di release IPv6 ready, accensione delle funzionalità, individuazione degli apparati obsoleti e delle procedure per sostituirli. Acquisire quindi la capacità di fornire connettività IPv6 in tutti i punti della propria rete e la libertà di cominciare a farlo quando sia più opportuno; ma non necessariamente sostenere i costi di fornire IPv6 a tutta la clientela a partire da domani.

3 IPv6: cosa cambia

Nel progettare il nuovo protocollo ci si pose l'obiettivo di espandere lo spazio di indirizzamento e di migliorare alcuni aspetti di IPv4 che si erano rivelati critici, quali la sicurezza, la semplicità di configurazione, la gestione della mobilità, il multicast, la qualità di servizio. Le innovazioni introdotte in IPv6 relativamente a questi ultimi aspetti sono state successivamente recepite in opportune estensioni di IPv4 e quindi oggi per questi aspetti non vi sono particolari differenze di funzionalità e prestazioni in IPv6 rispetto ad IPv4.

Le differenze fondamentali si limitano pertanto al formato del *Header* dei pacchetti, alla struttura degli indirizzi ed ai meccanismi per assegnare ed

utilizzare gli indirizzi all'interno della rete. Queste differenze sono sufficienti a fare di IPv6 un protocollo simile ma incompatibile con IPv4 e quindi a determinare i conseguenti problemi di migrazione.

L'*Header* IPv6 [11] ha lunghezza fissa pari a 40 byte; per contro l'*Header* IPv4 (*Figura 2*) può avere lunghezza variabile per la presenza di *Option* (raramente utilizzate). Eventuali informazioni aggiuntive sono inserite in IPv6 in *Extension Header* che di norma vengono analizzati solo agli estremi della comunicazione: la presenza di *Extension Header* viene segnalata nel campo *Next Header* che di norma indicherà il codice del protocollo di livello superiore (ad esempio TCP o UDP) già presente nel campo Protocol del pacchetto IPv4. In IPv6 inoltre non è possibile la frammentazione del pacchetto da parte di router intermedi: un pacchetto di lunghezza incompatibile con l'MTU (*Maximum Transfer Unit*) di una data tratta viene semplicemente scartato ed una segnalazione di errore viene inviata alla sorgente che dovrà adeguare da quel momento in poi la dimensione dei pacchetti inviati.

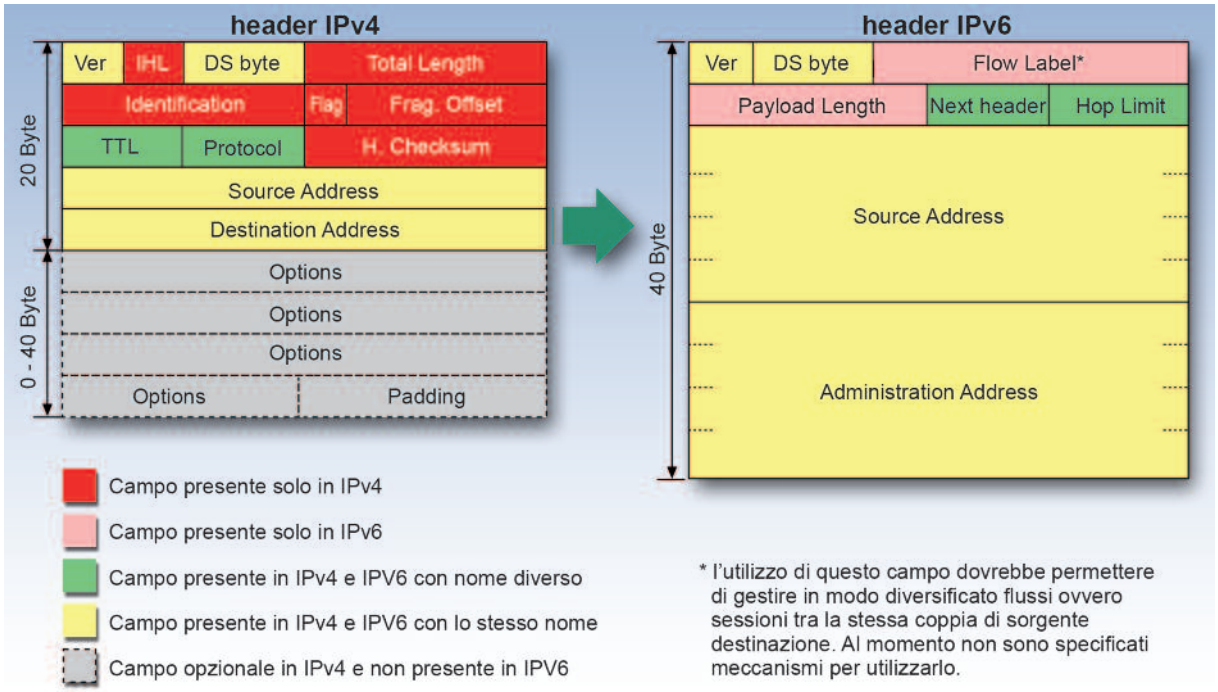
Nel *Header* sono inseriti i campi per gli indirizzi sorgente e destinazione che in IPv6 hanno lunghezza pari a 128 bit. Lo spazio di indirizzamento è strutturato in modo da ricavare delle classi di indirizzi dedicate a specifici utilizzi; in particolare sono definiti gli indirizzi *Global Unicast, Link Local, Unique Local Addresses* e *Multicast* [43].

Il formato di questi indirizzi è riportato in *Figura 3*. Si osservi che le prime tre classi di indirizzi sono strutturate secondo uno schema generale del tipo prefisso+identificativo di interfaccia; l'identificativo di interfaccia ha una lunghezza pari a 64 bit, per permettere di ricavare in modo automatico dall'indirizzo MAC (*Medium Access Control*) dell'interfaccia fisica questa parte dell'indirizzo e quindi semplificare le operazioni di configurazione di una sottorete.

Va notato che con IPv6 un'interfaccia di rete può avere più indirizzi che utilizzerà per scopi differenti; questo rende anche molto semplice la rinumerazione delle reti IPv6, in quanto nel transitorio viene semplicemente aumentato il numero degli indirizzi assegnati e quindi si può gestire la rinumerazione senza perdita di operatività.

In particolare un'interfaccia ha sempre un indirizzo Link Local, ricavato dal prefisso FE80::/64

Figura 2 – Confronto tra il formato di *Header* IPv4 ed IPv6



e dall'indirizzo MAC dell'interfaccia fisica; questo indirizzo permette la comunicazione di tutti gli apparati di una stessa LAN in modo automatico, senza nessuna operazione di configurazione sulle relative interfacce. Per comunicazioni tra apparati remoti è invece richiesto un indirizzo Global Unicast, che può essere configurato in modo esplicito o ottenuto attraverso alcuni meccanismi di tipo automatico: IPv6 *Stateless Address Autoconfiguration* [44], DHCPv6 [45] o DHCPv6 con *prefix delegation* [46].

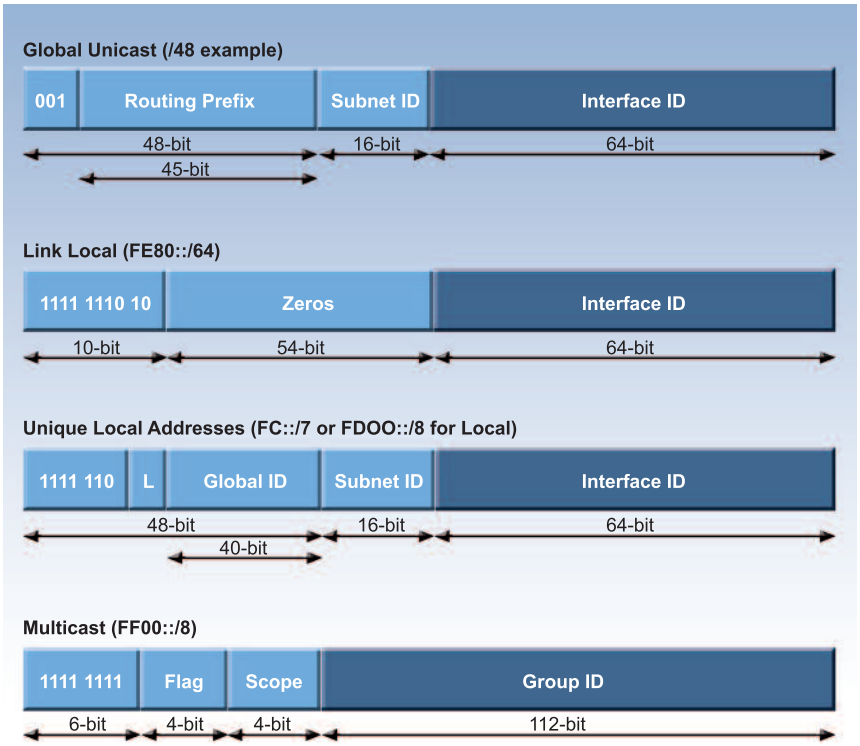


Figura 3 – Formato degli indirizzi IPv6

I meccanismi di autoconfigurazione sono particolarmente potenti e permettono di semplificare notevolmente le configurazioni di rete: hanno tuttavia il costo di utilizzare in modo non ottimale lo spazio di indirizzamento, in particolare riservare 64 bit per l'indirizzo di interfaccia evita la configurazione manuale, ma, di fatto, "spreca" la metà dello spazio di indirizzamento.

4 L'introduzione in rete

4.1 *4.1 Abilitare l'infrastruttura*

In ambito IETF sono state fatte negli anni ripetute analisi sulle modalità di introduzione di IPv6 in una rete IPv4, con l'obiettivo di preservare il più possibile gli investimenti, ridurre i disservizi e procedere in modo graduale all'abilitazione del nuovo protocollo. In particolare da lungo tempo viene proposto l'approccio così detto *dual-stack*: tale approccio è basato sulla capacità di un router di instradare pacchetti appartenenti a protocolli differenti; ad esempio un router può instradare indifferentemente pacchetti IP e pacchetti MPLS.

Introdurre IPv6 in modalità *dual stack* significa in estrema sintesi assegnare ad ogni interfaccia fisica di ogni router della rete, già numerata con IPv4, anche una numerazione IPv6; distribuire tra i diversi router della rete le informazioni di raggiungibilità delle sottoreti IPv6 attraverso protocolli di routing adeguati e costruire all'interno dei router le tabelle di routing e di forwarding specifiche. Inoltre sui router devono essere abilitati una serie di protocolli e di procedure, specifici di IPv6 (ad esempio *Router Advertisment*, DHCPv6,…), come dettagliato in [12].

I protocolli di routing utilizzati per IPv6 sono opportune estensioni dei protocolli di routing utilizzati per IPv4. In particolare le estensioni per MP-BGP [16] e per IS-IS [18] discendono naturalmente dalla capacità di questi protocolli di tra-

sportare informazioni di raggiungibilità per più classi di indirizzi diversi. Forse meno naturali sono le estensioni specificate per OSPF [17], che era stato progettato pensando esplicitamente ad IPv4.

Inserire queste funzionalità sui router può risultare in alcuni casi critico: in primo luogo si ha un maggior carico sul piano di controllo, che non dipende, come detto, dall'inserimento di nuovi protocolli di routing, quanto dalla necessità di elaborare oltre alle informazioni sulla raggiungibilità delle reti IPv4 anche quella delle reti IPv6. Questo carico addizionale è presumibilmente modesto nella fase iniziale, in quanto IPv6 per sua natura si presta meglio all'aggregazione delle informazioni di routing ed inoltre la Full Internet table IPv6 è ad oggi limitata ad alcune decine di migliaia di rotte. Un secondo problema è legato alla crescita della memoria occupata per gestire le varie tabelle IPv6: a causa della lunghezza degli indirizzi, anche un numero modesto di rotte, occupa uno spazio di memoria significativo. Un ultimo problema rilevante è legato al piano di forwarding; ad oggi questa operazione viene realizzata per IPv4 in HW sulle macchine di classe più elevata: non tutti i router sono tuttavia predisposti per il forwarding in HW di IPv6. Una valutazione della rete da migrare ad IPv6, deve in primo luogo identificare quelle porzioni critiche rispetto a questi parametri, ovvero gli apparati che non potranno a regime sostenere traffico IPv6 e che quindi non sarà conveniente aggiornare ma si dovranno sostituire al termine del percorso di migrazione.

In [13] viene proposto un approccio graduale per attuare la migrazione:

- Passo 1 "*launch*": prevede l'acquisizione degli indirizzi IPv6 dai registri ed il collegamento agli upstream provider, ovvero l'aggiornamento dei router di bordo verso altri domini;
- Passo 2a "*backbone*": aggiornamento a dual stack dei router di backbone;
- Passo 2b "*customer connection*": aggiornamento dei router di edge per permettere il collegamento nativo della clientela in IPv6;
- Passo 3 "*complete*": è la fase in cui tutta la rete supporta IPv6, ovvero quella in cui si sostituiscono gli eventuali apparati che si ritiene conveniente non migrare.

Rispetto a questo approccio, si osserva che il passo 2a può risultare particolarmente oneroso, in quanto richiede un aggiornamento del protocollo di routing IGP utilizzato nel backbone.

Per questo motivo e traendo beneficio dal fatto che negli ultimi anni si è affermato l'utilizzo di MPLS nel backbone per il trasporto di IP, in luogo del forwarding di IPv6 nativo è stata proposta la soluzione 6PE [19]: in questa soluzione, come avviene per le VPN MPLS, il backbone trasporta solo traffico MPLS (e quindi non viene aggiornato ad IPv6), ed il protocollo MP-iBGP viene utilizzato per scambiare le informazioni (rotte IPv6 ed associate *label* MPLS) di una *address family* dedicata. Come per MPLS VPN è richiesto che tutti i router di bordo condividano l'intera tabella di routing e che incapsulino il pacchetto IPv6 con una doppia etichetta MPLS, un'etichetta esterna che instrada il pacchetto sul circuito virtuale (LSP, *Label Switched Path*) infrastrutturale verso il PE destinazione e una label più interna che indirizza la tabella delle rotte IPv6. Questa soluzione ha il principale inconveniente di non supportare traffico multicast IPv6, ma allo stato attuale questa limitazione viene comunemente accettata.

Adottando l'architettura 6PE (*Figura 4*) occorre quindi attivare la funzionalità sia sui router di Edge, sia sui router di bordo del dominio. Si noti che il protocollo di routing MP-iBGP richiede in alternativa o una maglia completa di sessioni iBGP tra tutti i PE o una sessione iBGP da ogni PE verso un *Route Reflector Server* che ridistribuisca le informazioni di routing a tutti i 6PE. Quest'ultima soluzione è normalmente preferita perchè più scalabile; il *Route Reflector Server* può essere sia un apparato dedicato all'*address family* IPv6, sia un apparato già utilizzato per rotte IPv4 o VPNv4.

L'impatto della funzionalità 6PE sugli apparati di edge, va sempre valutato in termini di incremento di carico sul piano di controllo e di occupazione di memoria (almeno tutte le rotte interne al dominio devono essere distribuite su tutti i 6PE). L'incremento percentuale di entrambi i pa-

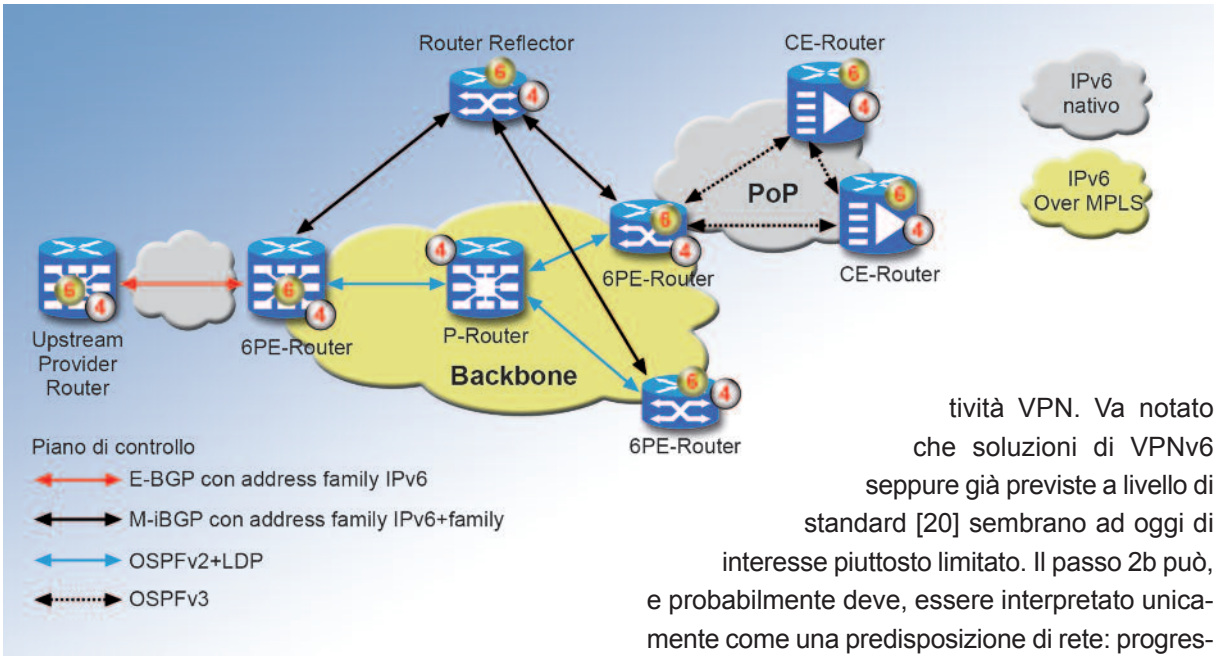


Figura 4 – Possibile architettura di rete con soluzione 6PE

rametri può essere quasi trascurabile per apparati utilizzati per utenza business che abbiano già oggi funzionalità di PE MPLS attive. Può risultare al contrario significativo invece per apparati dedicati all'utenza residenziale per i quali non sarebbe altrimenti previsto l'utilizzo di MPLS. Il punto esatto in questo segmento di rete in cui abilitare tale funzionalità è quindi un parametro di progetto che può variare in funzione del contesto specifico.

Per quanto riguarda il passo 2b, l'obiettivo di un ISP è sicuramente quello di proporre IPv6 all'utenza nel modo meno invasivo possibile, con le stesse modalità con cui è offerto oggi IPv4, per favorire la migrazione graduale del traffico da IPv4 ad IPv6, grazie all'adeguamento progressivo dei terminali. Per questo motivo sono ad oggi di particolare interesse per la clientela residenziale xDSL le soluzioni basate su PPP [14][15]. È quindi necessaria l'introduzione sugli apparati di Edge di una serie di funzionalità necessarie al supporto di queste soluzioni. Più semplice risulta la realizzazione di collegamenti nativi IPv6 verso utenza business: in questo caso tuttavia occorre garantire la coesistenza sia con i servizi di accesso ad Internet IPv4, sia con i servizi di connettività VPN. Va notato che soluzioni di VPNv6 seppure già previste a livello di standard [20] sembrano ad oggi di interesse piuttosto limitato. Il passo 2b può, e probabilmente deve, essere interpretato unicamente come una predisposizione di rete: progressivamente l'ISP predispone ogni apparato di bordo in modo che sia possibile connettere in *dual stack* la clientela a questa collegata, ma fornisce la connettività solo a chi ne faccia esplicita richiesta.

Il passo 3 è il deployment dei servizi, che quindi prevede il progressivo passaggio a connettività dual stack della clientela. Nel passo 3 sono inoltre da affrontare i problemi tecnologici rimasti aperti nelle fasi precedenti, quale ad esempio il multicast se è stata scelta un'architettura 6PE.

4.2 *Abilitare le piattaforme*

L'abilitazione delle piattaforme di controllo e servizio è il passo necessario per fare in modo che il protocollo IPv6 possa essere effettivamente adottato e fruito dagli utenti. Si tratta quindi di un passo indispensabile per poter andare verso l'adozione generale di IPv6 e, nel lungo termine, l'abbandono di IPv4.

Una mappa dei sistemi di controllo e servizio rilevante nell'evoluzione verso IPv6, include (*figura 5*):

- Domain Name System (DNS);
- sistemi di per l'autenticazione e la gestione della presenza di rete;

- sistemi per la configurazione remota degli apparati a casa dell'utente;
- sistemi per la fornitura dei servizi ToIP residenziali e business;
- portali Internet di servizio e di intrattenimento (e.g. nel contesto TI alice.it, 187.it, yalp!,…)
- piattaforme per servizi IPTV;
- sistemi per l'interfacciamento tra le reti ToIP e la PSTN/PLMN e l'SBC (Session Border Controller) per l'interfacciamento con le reti degli altri operatori (OLO);
- eventuali SoftSwitch e nodi di accesso multiservizio.

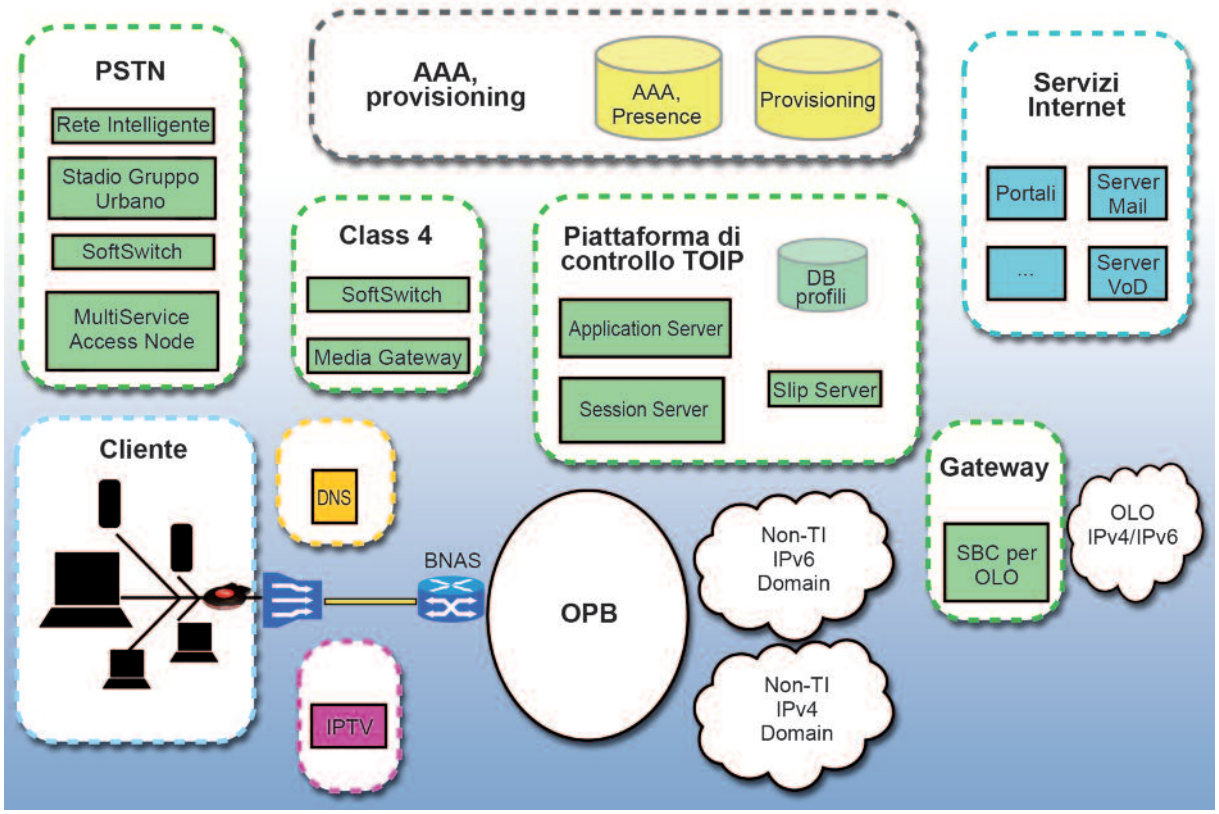
Su alcune piattaforme in esercizio l'introduzione di IPv6 potrebbe essere difficoltosa (per obolescenza tecnologica) o molto onerosa.

È quindi opportuno dividere la migrazione in due fasi: la prima, che coinvolge il DNS ed i sistemi di autenticazione e di *presence*, da svolgere in concomitanza con la migrazione dell'infrastruttura; la seconda, che prevede la migrazione delle restanti piattaforme, che può essere posticipata.

In particolare, il DNS deve essere aggiornato già nella prima fase per contenere, oltre agli indirizzi IPv4, anche i nuovi indirizzi IPv6, che di loro natura sono estremamente poco mnemonici. Questa esigenza deriva dal fatto che, anche per effettuare semplici operazioni di debugging in rete, l'utilizzo dei nomi logici è praticamente obbligatorio vista la lunghezza degli indirizzi IPv6. Il coinvolgimento del sistema di autenticazione e di presence nella prima fase è indispensabile per garantire, anche in presenza di IPv6, la funzione di AAA (Authentication, Authorization, Accounting), ossia l'accesso alla rete degli utenti e dei terminali.

Un requisito generale per la migrazione delle piattaforme è la garanzia di continuità del funzionamento di tutti i servizi per la clientela che rimane su IPv4. Questo requisito è reso ancora più indispensabile dal fatto che, nelle fasi iniziali di questo processo, la clientela IPv4 costituirà in pratica la totalità del parco clienti. Un altro impor-

Figura 5 – Principali piattaforme di controllo/servizio



IPv6 nella rete Mobile

L'evoluzione nella tecnologia di accesso radio, in particolare il dispiegamento della tecnologia HSPA (High Speed Packet Access)[1], la disponibilità di terminali mobili sempre più evoluti e di "chiavette" USB per l'accesso alla rete mobile tramite laptop, stanno determinando una vera e propria esplosione del traffico dati sulla rete mobile, cioè sulla rete GPRS (General Packet Radio Service). In prospettiva, la significativa diffusione di terminali per applicazioni machine-to-machine basati su SIM solo dati (sistemi di monitoraggio e telegestione, etc.), e la crescente popolarità di servizi che necessitano di raggiungibilità globale e connettività always-on, eroderanno significativamente la già scarsa quota di indirizzi IPv4 ancora disponibili, rendendo ben presto impraticabile l'assegnazione di indirizzi IPv4 pubblici agli utenti mobili.
Per questo motivo, molti operatori mobili stanno già iniziando a pensare all'introduzione di IPv6 nelle proprie reti.
Nella rete GPRS il GGSN (Gateway GPRS Support Node) costituisce il gateway di accesso verso le reti IP e come tale svolge tutte quelle funzioni correlate alla connettività IP, come l'assegnazione degli indirizzi IP ai terminali, il routing dei pacchetti e l'enforcement di policy sul traffico utente (e.g. policy di QoS, gating e/o charging). Il GGSN è il primo nodo di rete che su cui è terminato il traffico d'utente, che invece attraversa trasparentemente l'SGSN (Serving GPRS Support Node) tramite tunneling GTP.
Per accedere ad un servizio, l'utente richiede l'attivazione di uno specifico APN (Access Point Name), che si traduce nella richiesta di connettività verso un'opportuna PDN (Packet Data Network). Ad esempio, nella rete di Telecom Italia l'APN "ibox.tim.it" viene utilizzato per la navigazione in Internet. L'APN richiesto dall'utente arriva all'SGSN, che lo utilizza per selezionare un GGSN in grado di offrire la connettività desiderata, dopodiché il terminale resta ancorato a tale GGSN per tutta la durata della sessione di comunicazione.
Di conseguenza l'introduzione del supporto per IPv6 in rete GPRS impatta primariamente terminali e GGSN, che devono essere in grado di tramettere e ricevere traffico IPv6 sul piano d'utente.
Nella fase di transizione, specialmente per la navigazione in Internet, i terminali dovranno essere in grado di raggiungere destinazioni IPv4-only e/o IPv6-capable. Assumendo di disporre di GGSN e terminali dual-stack, cioè capaci di supportare simultaneamente IPv4 e IPv6, ciò potrà essere realizzato in una delle seguenti modalità:

a) All'attivazione della connessione a PDN, il terminale ottiene un indirizzo IPv4 ed un indirizzo IPv6. In questo caso IPv6 verrebbe utilizzato per comunicare con le destinazioni IPv6-capable, mentre IPv4 verrebbe utilizzato per comunicare con le destinazioni IPv4-only. Per ovviare alla scarsità di indirizzi IPv4 sono possibili due alternative:
   - utilizzo di indirizzi IPv4 privati e NAT;
   - l'indirizzo IPv4 è pubblico ma viene rilasciato nel caso in cui non venga utilizzato per un certo periodo di tempo (per poi aprirlo nuovamente in caso di necessità). Viceversa la connettività IPv6 può essere completamente always-on. Questa è una soluzione che è stata introdotta dal 3GPP in Rel-8.

b) Il terminale configura solo l'indirizzo IPv6. In questa configurazione il terminale possiede sol-

1 I protocolli HSPA sono un'evoluzione dell'UMTS che, in termini di prestazioni, apporta migliorie analoghe a quelle introdotte dall'EDGE sullo standard GPRS. High Speed Downlink Packet Access (HSDPA) permette di aumentare la velocità di trasmissione raggiungendo la velocità teorica di 14,4 Mbit/s; gli sviluppi attuali hanno raggiunto velocità di 7,2Mbit/s, con un massimo di 384 Kbit/s in up-link. High Speed Uplink Packet Access (HSUPA) viceversa permette di migliorare le performance di up-link fino a 5,76Mbit/s teorici.
Recentemente l'HSPA è stato ulteriormente migliorato, introducendo nuove versioni indicate come HSPA Evolution (HSPA+), ed in grado di offrire velocità di accesso fino a circa 50 Mbps.

tante requisito è che i servizi che prevedono una comunicazione tra due o più clienti (come i servizi ToIP) possano essere fruiti indipendentemente dalla versione del protocollo adottata dai clienti. Questo implica che le piattaforme che realizzano tali servizi dovranno garantire l'interoperabilità tra clientele eterogenee.

I principali scenari di servizio che dovranno essere affrontati sono i seguenti:

- Scenario navigazione/servizi web-based: per questi servizi si prevede che la clientela IPv6-enabled debba continuare ad utilizzare per lungo tempo anche IPv4, al fine di accedere ai siti e alle applicazioni sul web che sono ospitati su server IPv4. In sostanza, per supportare tali applicazioni la clientela IPv6 sarà dotata di dual stack. I principali web browser in uso (Internet Explorer, Mozilla Firefox) supportano nativamente il dual stack, ossia sono in grado di utilizzare lo stack corretto per accedere ad un sito, in funzione del tipo di indirizzo che è stato restuito come risultato della query DNS.

  Per incrementare l'utilizzo e la diffusione di IPv6, sarebbe auspicabile che i portali di servizio dell'ISP fossero abilitati a funzionare anche in IPv6.

  Chiaramente, l'approccio *dual stack* non implica un significativo risparmio di indirizzi IPv4 per l'operatore di rete. Va però notato che approcci basati sull'assegnazione di soli indirizzi IPv6 (quindi con risparmio di indirizzi IPv4) e traduzione effettuata in rete, sia a livello rete-trasporto (NAT-PT) sia a livello applicativo (ALG), non appaiono percorribili in quanto soffrono dei già citati problemi di mancanza di trasparenza rispetto alle applicazioni e di costi aggiuntivi.
- Scenario servizi ToIP: per questi servizi si prevede di poter introdurre, da un certo momento, nuovi AG (*Access Gateway*) e nuovi terminali IPv6-enabled. Dovrà essere quindi garantita sia la compatibilità di tali AG/terminali con le piattaforme di controllo legacy, che potrebbero avere tempi di migrazione diversi

> tanto un indirizzo IPv6 e quindi la comunicazione verso destinazioni IPv4 è possibile solo attraverso Application Level Gateway (ALG) o traduttori di protocollo. Inoltre può rendersi necessaria un'ulteriore funzione di interlavoro sul terminale nel caso in cui l'utente stia utilizzando un'applicazione che supporti solo IPv4.

Oltre che su terminali e GGSN, l'introduzione di IPv6 in rete mobile ha anche impatti sui tradizionali nodi che sono alla base dell'operatività di una qualunque rete IP quali: server DNS, estesi per risolvere richieste di risoluzione relative ad indirizzi IPv6, server di Autenticazione Autorizzazione ed Accounting (AAA), piattaforme di tariffazione, estese per poter gestire Charging Data Record (CDR) contenenti indirizzi IPv6, piattaforme di policy control, sistemi di per l'intercetto legale, etc.

Meccanismi di transizione quali ad esempio il dual-stack andrebbero inoltre considerati per i portali ed altre piattaforme di servzio, quali portali WAP (Wireless Application Protocol), server di posta, server MMS (Multimedia Messaging Service), piattaforme IMS (IP Multimedia Subsystem), etc.

Guardando infine ai terminali, ad oggi il protocollo IPv6 è già supportato da vari sistemi operativi, come ad esempio Symbian OS, Windows Mobile, Android e Mac OS X. Di conseguenza è immaginabile che i costruttori di terminali mobili che implementano quei sistemi operativi sarebbero già in grado di utilizzare IPv6 per l'accesso alle reti dati. Tuttavia, questa funzionalità non è ancora esposta da vari terminali, anche molto evoluti (e.g. iPhone e BlackBerry), probabilmente perché non ancora richiesta dagli operatori e per non complicare i test di interoperabilità necessari per la validazione e certificazione del modello.

Da questa analisi emerge che l'eventuale attivazione di IPv6 in rete mobile ha impatti su molti nodi di rete ed il meccanismo di transizione è un processo non banale che va pianificato in modo molto accurato.

da quelli dei terminali, sia la possibilità di comunicare tra i nuovi terminali IPv6 e terminali legacy IPv4.

Per realizzare entrambe le funzioni, un ruolo molto importante sarà svolto dai Session Director (SD), che già oggi implementano funzioni di ALG. In futuro, le funzioni potrebbero essere estese per disaccoppiare l'introduzione dell'IPv6 nei terminali ToIP dall'evoluzione del Core di controllo/servizio.

La migrazione ad IPv6 delle connessioni di gestione e/o dedicate ai servizi ToIP su AG, permetterebbe di risparmiare un elevato numero di indirizzi IPv4. Quindi anche se ha un impatto in rete significativo, può risultare prioritaria.

- Scenario servizi IPTV: le prime analisi indicano che un vincolo rilevante per la migrazione del servizio riguarda il supporto del multicast IPv6 da parte dell'infrastruttura di rete, in quanto, come già osservato, attualmente non previsto dall'architettura 6PE. Un primo possibile passo potrebbe essere la gestione con dual stack: i contenuti Video on Demand (VoD) potrebbero essere erogati su IPv6, mentre i contenuti mutlicast continuerebbero ad essere distribuiti su IPv4, in attesa di una soluzione di rete che consenta l'utilizzo del solo protocollo IPv6.

Infine, sarà necessario aggiornare per tutti i servizi per i quali via via si intende predisporre l'offerta, i sistemi di gestione automatizzata e i sistemi per la tariffazione.

4.3 *Abilitare i terminali*

Per quanto riguarda i terminali occorre distinguere tra quelli forniti e gestiti dal ISP e quelli che gli utenti acquistano sul mercato.

Per quanto riguarda i terminali forniti da ISP (quali terminali ToIP, Set Top Box, Access Gateway,…) dovrà essere definito il piano di abilitazione coerentemente con le scelte e le tempistiche decise a livello di piattaforma. In generale, essendo la maggior parte dei dispositivi basati su Linux, il supporto di IPv6 dovrebbe essere garantito per quanto riguarda il S.O., mentre dovrà essere realizzato il porting delle applicazioni.

Per quanto riguarda i terminali non distribuiti da ISP, la velocità di adozione di IPv6 non è controllabile direttamente dall'operatore, ma è determinata dalle scelte dei clienti nonché dalle implementazioni decise dai fornitori dei device, dei Sistemi Operativi e degli applicativi.

Analizzando il panorama dei sistemi operativi, è possibile affermare che la maggioranza dei S.O. in uso oggigiorno supporta IPv6, anche se non sempre tale protocollo è automaticamente attivato sulle macchine; in particolare:

- Linux supporta IPv6 già dalla versione 2.2 del kernel, rilasciata nel 1999. Tuttavia, i kernel 2.2 e 2.4 non sono più stati aggiornati per supportare gli ultimi RFC, pertanto al momento l'utilizzo di IPv6 è consigliato solo con kernel dell'ultima generazione (2.6.x). Il Linux IPv6 User Group e mantiene una lista di applicazioni portate su IPv6 ed effettua il porting di nuove applicazioni.
- UNIX BSD è il sistema operativo che ha la più lunga tradizione di supporto di IPv6. Free-BSD e Net-BSD hanno gruppi di utenti che mantengono liste di applicazioni portate su IPv6.
- Microsoft ha rilasciato versioni "trial" dello stack IPv6 già per Windows 95, 98 e 2000 (IPv6 Technology Preview). Solo a partire da Windows XP Service Pack 1 (settembre 2002) lo stack IPv6 è stato ufficialmente rilasciato e supportato. Si noti che con XP lo stack è già installato, ma deve essere attivato. Soltanto con Vista IPv6 è abilitato già di default.
- MAC OS X supporta IPv6 dalla versione 10.2 "Jaguar". Tuttavia, non è previsto alcun supporto in MAC OS Release 9. Apple mantiene una lista delle applicazioni portate su IPv6.

Per quanto riguarda le applicazioni, il panorama appare sicuramente articolato, ma sono già numerose le applicazioni abilitate ad utilizzare IPv6 anche in modalità dual-stack. A titolo di esempio, indichiamo:

- Web browser: supporto in Microsoft Internet Explorer (ottimale solo dalla release 7), Mo-

zilla Firefox (dalla rel. 1.5), Opera (dalla rel. 7.20), Google Chrome;
- Client e-mail: supporto in Microsoft Outlook 2007, Windows Mail (embedded in Vista), Apple Mail, Mozilla Thunderbird;
- Media Player: supporto in Windows Media Player (dalla 9.0), Video Lan Client (VLC), WinAmp (dalla 5.34);
- File sharing: l'implementazione di IPv6 nei protocolli Kad e ED2K è in corso. Tuttavia, l'implementazione dei protocolli e delle funzionalità interne dei programmi peer-to-peer (come eMule), nonchè la diffusione nelle reti peer-to-peer delle nuove versioni dei protocolli potrebbe richiedere alcuni anni.

5 L'Internet che verrà

5.1 *Una rete eterogenea*

Quando fu proposta in IETF per la prima volta la soluzione *dual stack* IPv4/IPv6 l'obiettivo era una migrazione tra i due protocolli del tutto trasparente: partendo con un sufficiente anticipo vi era la possibilità di avere il 100% dell'utenza Internet connessa in IPv6 prima di giungere ad un esaurimento degli indirizzi IPv4; in una situazione di questo tipo ogni applicativo, poteva essere migrato ad IPv6 in modo del tutto scorrelato dall'evoluzione di rete. Se a livello applicativo si fosse a questo punto privilegiato IPv6 rispetto a IPv4 in tutti i casi in cui questo fosse stato possibile, il traffico IPv4 sarebbe gradatamente diminuito in rete, sino a rendere di fatto inutili nuove assegnazioni di indirizzi IPv4, se non per esigenze molto specifiche.

È molto probabile che questa situazione ottimale non si possa più verificare neanche con un'accelerazione improvvisa dell'introduzione di IPv6 [7]. È quindi molto probabile che l'esaurimento degli indirizzi IPv4 si verifichi prima che IPv6 abbia raggiunto una penetrazione significativa. Da quel momento in poi a nuovi utenti potranno essere assegnati solo indirizzi pubblici IPv6. Internet sarà di fatto partizionata in tre categorie di utenza, IPv4 only, IPv4/IPv6 ed IPv6 only. In questo scenario solo gli utenti dual stack potranno nativamente accedere a tutte le risorse di rete. Per i rimanenti dovranno essere previste opportune funzionalità di adattamento di protocollo. Il problema è rilevante e le attività in IETF si stanno concentrando su questo aspetto.

In realtà un meccanismo base detto "Network Address Translator – Protocol Translator" (NAT-PT) per permettere la comunicazione tra apparati IPv4 ed apparati IPv6 è stato definito da tempo [23] ma successivamente deprecato [24] perchè percepito come un elemento che poteva ritardare la migrazione ad IPv6. Le nuove soluzioni allo studio di fatto sono varianti migliorate di tale meccanismo base.

Un esempio è la recente proposta del meccanismo denominato NAT64 [25] che ha lo scopo di permettere le comunicazioni, iniziate dal client, tra un client IPv6 only ed un server IPv4 only, o comunicazioni tra un peer IPv6 only and un peer IPv4 only, sempre a condizione che queste siano iniziate dal terminale IPv6 only. Per il funzionamento di questa soluzione è necessario che il DNS sia adeguato in conformità a quanto previsto in [26].

Emerge chiaramente la considerazione che se da un lato la possibilità di gestire la coesistenza di reti eterogenee rappresenta un indubbio vantaggio (non è richiesto l'aggiornamento immediato o veloce dei terminali e delle piattaforme di servizio legacy) la promessa che IPv6 fa di una rete più semplice e meno costosa sarà effettivamente realizzabile solo con la prospettiva di lungo termine di una sostituzione completa di IPv4 in Internet.

5.2 *L'Internet delle cose*

Una delle motivazioni che portarono alla definizione di IPv6 è sicuramente la visione di una Internet del futuro che a partire dal Web e dai

La sperimentazione di Telecom Italia

Considerato l'elevato impatto dell'introduzione di IPv6 in rete, Telecom Italia ha deciso di intraprendere da diverso tempo delle sperimentazioni mirate a raggiungere seguenti obiettivi:

a) monitorare il grado di evoluzione della tecnologia che viene progressivamente resa disponibile dai fornitori, al fine di poter migliorare le sinergie con essi;
b) pianificare delle strategie di introduzione di IPv6 in modo coerente con le roadmap di crescita tecnologica e con lo scopo di abbattere o ridurre sensibilmente il fabbisogno di indirizzi IPv4 nel più breve arco temporale.

Superare la fase delle sperimentazioni di laboratorio e verificare IPv6 nella rete di produzione ha lo scopo di sperimentare non solo la connettività IPv6 di base che apparati e terminali possono offrire, ma anche il grado di adeguatezza e di compatibilità delle varie funzionalità che costituiscono la struttura portante dei servizi erogati dagli ISP quali ad esempio Quality of Service, Multicast, AAA, Security (antispoofing, Lawful Intercept), protocolli di gestione.
I servizi di accesso ad Internet dual-stack IPv4/IPv6 per clientela residenziale e business costituiscono un banco di prova adeguato a questo scopo e sono realizzabili nel breve termine, richiedendo limitati adattamenti a livello di piattaforme. Per questo motivo Telecom Italia ha deciso di focalizzarsi in tal senso nell'attuale fase di trial in campo, rimandando ad un prossimo futuro la sperimentazione della migrazione ad IPv6 di servizi quali ad esempio VoIP o IPTV, che essendo caratterizzati da relazioni di traffico che si richiudono prevalentemente all'interno della propria rete, potrebbero permettere una fase di transizione più rapida, ma che utilizzano terminali e piattaforme non ancora predisposti per IPv6.
In tale ottica, Telecom Italia ha intrapreso un'attività sperimentale nell'ambito del progetto CVIS (Cooperative Vehicle-Infrastructures System) [48] finanziato dall'Unione Europea e che vede la partecipazione consorziata di ISP, Vendor di TLC e delle principali case automobilistiche. Il progetto è finalizzato ad offrire servizi avanzati di infomobilità e prevede l'utilizzo di speciali dispositivi installati a bordo dei veicoli che scambiano dati in connettività IPv6 sia tra di loro che verso un centro servizi.
Il compito specifico di Telecom Italia è costituito dalla fornitura di infrastruttura di rete IP dual stack al centro servizi mediante accesso ADSL.
Lo schema su cui è stata basata l'architettura è il seguente (***figura A***):

a) Router ADSL presso il centro servizi di Infomobility con modalità di accesso PPPoE dual stack (IPv4 ed IPv6);
b) Rete ADSL pubblica di Telecom Italia: Access Node, Access Network e BroadBand Network Access Server (BNAS);
c) Piattaforma di autenticazione AAA
d) Trasporto del traffico IPv6 su OPB (Optical Packet Backbone) in modalità 6PE, ovvero all'interno di un tunnel MPLS;
e) Connettività verso Big Internet IPv6 attraverso Gateway Internazionale
f) DNS Server "IPv6 aware", ovvero in grado di risolvere URL associate ad indirizzi IPv6 basandosi su DNS query IPv4.

Il router ADSL è dotato di un client PPP (Point-to-Point Protocol) dual-stack sull'interfaccia WAN che consente di poter dialogare con il BNAS attraverso i protocolli per la gestione dell'indirizzamento IPv4 (Link Address Control Protocol - LCP, Network Control Provocol - NCP, PPP IP Control Protocol IPCP, Password Authentication Protocol - PAP) ed IPv6 (DHCPv6, ICMPv6).
La rete di accesso è completamente trasparente alla modalità di indirizzamento utilizzata (IPv4 o >

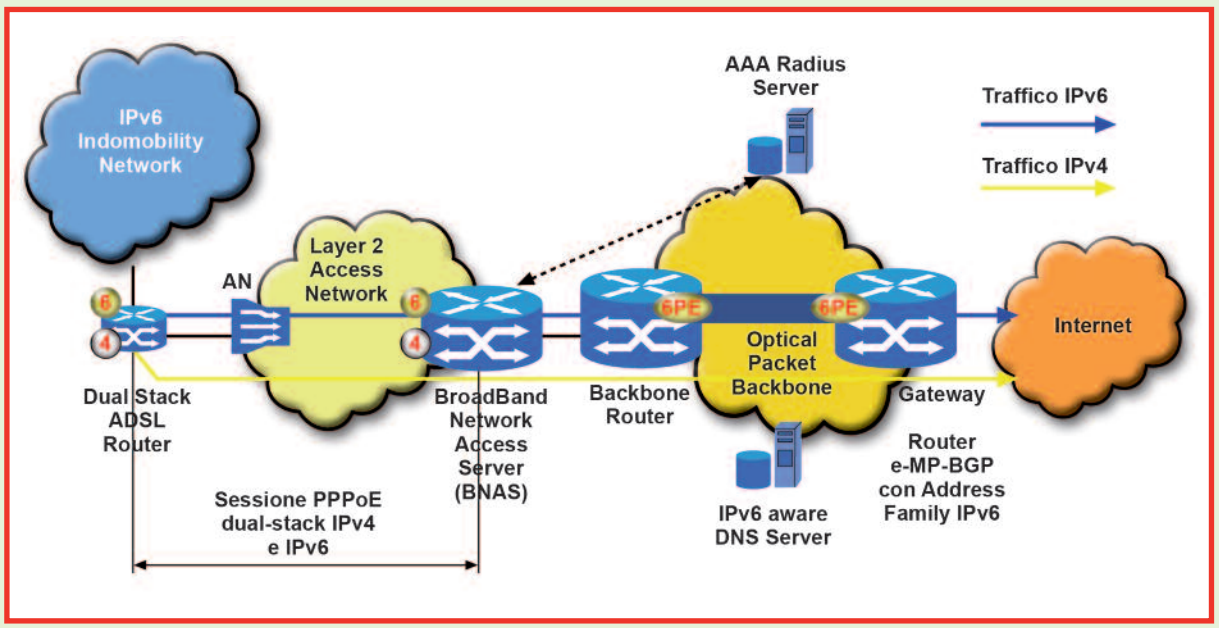


Figura A – Architettura di riferimento del trial realizzato nell'ambito del progetto CVIS

IPv6) ed ha il solo scopo di aggregare a livello 2 gli Access Node della rete ADSL.
Il BNAS svolge una funzione chiave nell'architettura in questione, poiché è l'elemento di rete che svolge le seguenti funzioni:

- negoziazione a livello logico lato client del router ADSL dei parametri di connettività IPv4 ed IPv6,
- interazione con la piattaforma AAA (Authentication, Authorization, Accounting) per l'autenticazione del cliente, l'applicazione di alcuni parametri relativi al profilo di utenza (residenti nel DB della piattaforma AAA) e l'emissione dei dati di accounting.

Quando la sessione PPP dual-stack è instaurata, nel BNAS sono presenti due interfacce virtuali punto-punto verso il cliente, afferenti a due istanze di routing differenziate, rispettivamente IPv4 ed IPv6.
Il routing verso il BackBone prevede la consegna del traffico IPv6 ad un nodo con funzionalità 6PE per il trasporto verso l'end-point 6PE del nodo Gateway verso Internet, ove sono presenti sessioni e-MP-BGP con address family IPV6 verso l'upstream provider.
Il server DNS (Domain Name Server) "IPv6 aware" supporta la risoluzione di URL associate ad entrambe le tipologie IPv4 ed IPv6: la query DNS è basata sul protocollo IPv4.
In conclusione, l'esperienza CVIS sta rappresentando una occasione ottimale per Telecom Italia di sperimentazione IPv6 per le seguenti ragioni:

- la rete è inserita in uno scenario "green-field" di terminali ed apparati: non si hanno quindi stringenti requisiti di retro-compatibilità con il protocollo IPv4 poiché le relazioni di traffico si racchiudono prevalentemente all'interno della rete CVIS.
- Permette di minimizzare l'impiego di indirizzi IPv4 che sono riservati esclusivamente alla connettività del centro servizi connesso in modalità dual-stack che rappresenta l'unico elemento centralizzato IPv4 compatibile. L'utilizzo di indirizzamento IPv4 sugli autoveicoli non risulterebbe comunque adeguato poiché non permetterebbe di scalare ad un potenziale bacino di milioni di utenze rappresentato dagli autoveicoli attualmente in circolazione.

servizi telematici, si espande prima all'ambito delle telecomunicazioni in generale (telefoni IP, cellulari, sistemi di videoconferenza, …), per diventare infine pervasiva e permettere la comunicazione ed il controllo remoto di un'ampia varietà di oggetti all'interno e all'esterno degli edifici.

In futuro gli ambienti (le case, gli uffici, le automobili, le città) diventeranno sempre più intelligenti e informatizzati e i nuovi dispositivi saranno in grado di interagire e sfruttare l'intelligenza largamente distribuita in questi ambienti (servizi di Ambient Intelligence). In futuro si tenderà ad utilizzare sempre più applicazioni che, a partire dalle funzionalità di localizzazione, permetteranno ai terminali di "capire il contesto" (Context Awareness) in cui operano consentendo alle applicazioni di adattarsi alla situazione.

Le nuove applicazioni non comporteranno solo interazione fra esseri umani e servizi, ma anche fra macchine e macchine (servizi "Machine to Machine"). I servizi possibili spaziano dalle applicazioni relative alla sanità e al wellness a quelle relative al controllo remotizzato di apparati e sistemi di produzione e controllo, dai sistemi di infomobilità, ad applicazioni complesse di smistamento merci. L'Internet delle cose forse non consumerà grandi quantità di banda, ma avrà sicuramente dei requisiti stringenti su come i dati saranno trasportati e inviati agli estremi della rete.

Occorre tuttavia notare che questi scenari, pur se da tempo dibattuti, tardano a concretizzarsi: in passato, ed in certa misura ancora oggi, lo spazio di indirizzamento non ha costituito una limitazione al collegamento ad Internet di nuovi terminali, ma non per questo vi è stato un pervasivo proliferare di dispositivi IP. Alcune iniziative stanno avendo un successo contrastato; si veda ad esempio lo standard Universal Plug and Play (UPnP) [36] e le correlate specifiche della Digital Living Network Alliance (DLNA) [37] che appoggiandosi su la suite di protocolli TCP/IP permettono la condivisione di contenuti (in particolare audio/video) tra diversi dispositivi all'interno della casa quali Personal Computer, palmari, cellulari, video registratori o video Player.

Per collegare apparecchi di consumer electronics a corto raggio è più comune l'impiego di standard industriali wired e wireless alternativi all'IP: basti pensare alle tecnologie Bluetooth [38], Infrared Data Association (IrDA) [40], Firewire [39], USB [41], ZigBee [42]. PC ed in futuro probabilmente Access Gateway evoluti hanno il compito di costituire il Gateway tra le sottoreti domestiche 'specializzate' e la rete IP. Il successo di queste soluzioni è sicuramente determinato anche da cartelli industriali dei costruttori degli apparati di consumer electronics, che cercano attraverso la specificità delle soluzioni di proteggere il loro mercato dall'invadenza della rete IP. Ma è innegabile che per collegare tra loro un numero limitato di device queste soluzioni siano molto semplici ed efficaci.

Queste soluzioni, accanto a soluzioni proprietarie, sono molto impiegate anche nelle reti di sensori, all'interno delle quali un dispositivo master che ha interfaccia verso la rete Internet è normalmente utilizzato per raccogliere le informazioni dai singoli sensori e convogliarle verso opportuni centri di controllo.

In questo quadro, non si può quindi affermare che l'industria sia compatta nell'attesa della diffusione di IPv6 per dare nuovo impulso alle applicazioni di tipo machine-to-machine. Tuttavia IPv6, con la sua abbondanza di indirizzi, costituisce un elemento importante affinchè i fornitori di tecnologia IP e i Service Provider possano aspirare ad un ruolo in questo nuovo potenziale mercato.

5.3 *Oltre l'IP*

Un punto di vista originale sull'evoluzione di Internet è presentato in alcuni studi recenti [27][28]: secondo questo punto di vista, è necessario non tanto rivedere il protocollo di trasporto dei dati, quanto il paradigma stesso su cui Internet è stata costruita. Di fatto le comunicazioni su Internet sono costruite su due pilastri: il principio end-to-end ed il modello client-server. Entrambi questi pilastri sono mutuati dalla rete telefonica, di cui Internet si è mostrata la naturale evoluzione. Ul-

timamente però sta emergendo un modello di comunicazione diverso, che viene comunemente detto Information Centric Networking, che si sta sviluppando in reti overlay, quali le reti peer-to-peer (e.g. eMule/eDonkey [29], BitTorrent [30], Kazaa [31], Skype [32]] o di distribuzione dei contenuti (Akamai [33], etc.).

Il modello client-server presenta alcuni problemi intrinseci a cui negli anni si è riusciti a dare solo soluzioni parziali. Ad esempio il problema di sicurezza dei dati, per il quale sono stati proposti meccanismi per realizzare canali sicuri tra il client ed i server, senza riuscire a risolvere il problema di fondo di garantire l'autenticità e l'integrità del dato. O i problemi di congestione, legati a malfunzionamenti dei nodi di rete, colli di bottiglia intrinseci dei server, o attacchi di tipo DDOS. O anche la gestione di multicast o di mobilità, che nelle reti IP attuali presentano complessitità realizzative e problemi di scalabilità.

Nell'approccio Information Centric Networking si riconosce la centralità del dato, della singola unità di informazione, e si costruisce la rete sulla base di un'interfaccia cliente di tipo *publish/subscribe*: i dati vengono pubblicati in rete ed è la rete che si preoccupa di preservarne l'integrità, tracciarne le copie autentiche disponibili ed individuare il meccanismo più idoneo a distribuirli, a singoli utenti o a gruppi di utenti che ne facciano richiesta. L'Information Centric Networking supera il principio di trasparenza end-to-end, in quanto pone la rete come mediatore delle comunicazioni; implicitamente supera quindi anche la necessità di utilizzare tecnologia unificante per Internet: non essendovi necessità di gestire comunicazioni end-to-end, Internet può essere partizionata in più sottoreti, alcune delle quali utilizzeranno ad esempio IPv4 per le comunicazioni tra i vari nodi; altre potranno utilizzare IPv6 o altre soluzioni ancora, ad esempio quelle tipiche della home network o delle reti di sensori.

Ciò che deve essere preservato e reso univoco, nonostante l'eterogeneità tecnologica dei nodi che costituiscono tale rete, è il 'nome' del dato che che ne costituisce la chiave unica di ricerca e di instradamento. L'impiego di tecnologie di auto-certificazione dei dati [34] potrà permettere di raggiungere un elevato livello di sicurezza delle informazioni prescindendo dall'impiego di canali di comunicazione sicuri end-to-end. Da notare che, in questo contesto, dato è sinonimo di informazione digitale nel senso più generale del termine, ovvero pagine web, filmati, registrazioni audio o programmi eseguibili, sono istanze di 'dati', caratterizzati da opportuni attributi.

La rete Information Centric mantiene tabelle di instradamento dei 'nomi', ovvero tabelle che risolvono un nome nell'indirizzo fisico di in uno o più repository. Un meccanismo di indirezione nella tabella dei nomi può essere utilizzato per gestire in modo semplice la mobilità di un terminale (e quindi dei dati in questo memorizzati) attraverso Internet. Il problema tecnologico da affrontare è proprio quello di gestione delle tabelle dei nomi: in linea di principio uno spazio dei nomi (anche detto dizionario) piatto è preferibile; è questa la soluzione adottata in alcuni overlay peer-to-peer, che dal punto di vista tecnologico utilizzano soluzioni basate su Distributed Hash Table (DHT). Ma questi meccanismi sono accettabili da un punto di vista amministrativo/commerciale (è accettabile che parte dei nomi di dati registrati da utenti di un dato ISP siano gestiti da un ISP concorrente o da un ISP che si trova agli antipodi?), e sono sufficientemente potenti da indirizzare l'intero spazio dei nomi della futura Internet? Ed inoltre, è possibile conciliare questi meccanismi con l'esigenza di trasferire il dato nel modo più efficiente possibile, contrariamente a quanto oggi sembra accadere negli overlay P2P? Questi sono solo alcuni degli innumerevoli punti aperti che devono essere risolti prima che da un'idea potenzialmente molto promettente si possa passare ad una fase realizzativa.

Le potenzialità delle soluzioni Information Centric sono enormi e per questo motivo, importanti progetti di ricerca, tra i quali ad esempio il progetto IST 4WARD [47], stanno investendo un significativo effort in questa direzione. Dal punto di vista dei Service Provider questo approccio è particolarmente interessante, perchè verrebbe ad attribuire agli ISP un ruolo centrale nella società dell'informazione, simile a quello rivestito dai Telco con la rete telefonica: garantire agli utenti

la possibilità di reperire istantaneamente ed in modo affidabile qualunque informazione disponibile semplicemente identificandola con il nome, ad un costo ragionevole e soprattutto con la garanzia di autenticità. Proprio come oggi è possibile entrare in comunicazione con un qualunque utente della rete telefonica ovunque si trovi semplicemente digitandone il numero. Si rimanda a [35] per ulteriori approfondimenti.

6 Conclusioni

L'esaurimento degli indirizzi IP è un problema prossimo venturo e per alcuni versi è già un problema di oggi, in quanto ottenere nuovi indirizzi è diventato difficile e l'amministrazione oculata di quelli a disposizione è costosa.

IPv6 rimane l'unica soluzione identificata dalla comunità internazionale in grado di dare nuova linfa al modello di Internet che tanto successo ha avuto negli ultimi decenni. Le alternative, sostanzialmente basate sull'uso estensivo di indirizzi privati e NAT, introducono architetture di rete complesse, non convenienti dal punto di vista economico e con la prospettiva di ridurre pesantemente la flessibilità nello sviluppo di nuovi servizi.

Questo articolo descrive un approccio al percorso di introduzione di IPv6 in rete che evidenzia come i primi passi siano di tipo infrastrutturale e richiedano attività di industralizzazione su tutti i segmenti di rete e sulle principali piattaforme abilitanti ai servizi di connettività. La disponibilità di servizi di connettività IP *dual stack* abiliterà in seguito la possibilità di scegliere IPv6 come base per l'evoluzione di servizi, sia servizi nuovi, sia servizi per i quali la scarsità di indirizzi IPv4 divenga problematica.

Scostando il velo del futuro vengono evidenziati tre aspetti. Dapprima la considerazione che introducendo IPv6 andremo per un lungo periodo verso una situazione di reti eterogenee, che certamente richiederà di gestire situazioni di interlavoro tra i due protocolli IP. La seconda è che la disponibilità dell'enorme spazio di indirizzamento di IPv6 fornirà alle manifatturiere delle tecnologie IP e ai Service Provider un'opportunità di entrare nel potenziale mercato dell'Internet delle cose con un punto di forza in più, nonostante anche in questo contesto l'orientamento odierno sembra privilegiare uno scenario di reti eterogenee. Da ultimo, i prossimi sviluppi sul modello dell'Information Centric Networking potrebbero rappresentare la prossima frontiera per consentire ai Service Provider una gestione vantaggiosa dell'etorgeneità delle reti.

BIBLIOGRAFIA

[1] F. Iuso, "IPv6: la nuova versione del protocollo Internet", «Notiziario Tecnico Telecom Italia», Anno 7, n. 1, aprile 1998, pp. 47-56.

[2] P. Fasano, G. Girardi e I. Guardini, "IPv6 nell'evoluzione di Internet", «Notiziario Tecnico Telecom Italia», Anno 10, n. 2, Settembre 2001, pp. 60-73.

[3] RFC 791 "Internet Protocol", J. Postel, September 1981.

[4] RFC 1519 "Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy", V. Fuller, T. Li, J. Yu, K. Varadhan, September 1993.

[5] RFC 1631 "The IP Network Address Translator (NAT)", K. Egevang, P. Francis, May 1994.

[6] RFC 3022 "Traditional IP Network Address Translator (Traditional NAT)", P. Srisuresh, K. Egevang, January 2001.

[7] G. Huston "The changing Foundation of the Internet: confronting IPv4 Address Exhaustion", in The Internet Protocol Journal, volume 11, number 3, September 2008.

[8] G. Huston "The IPv4 Internet Report", *http://ipv4.potaroo.net, August 2008*.

[9] Benjamin Edelman "Running Out of Numbers: Scarcity of IP Addresses and What To Do About It", Harvard Business School, Working Paper 09-091, August 2008.

[10] "IPv4 Address Allocation and Assignment Policies for the RIPE NCC Service Region",

ripe-449, Febbruary 2009.
[11] RFC 2460 "Internet Protocol, Version 6 (IPv6) Specification", S. Deering, R. Hinden, December 1998.
[12] RFC 4294 "IPv6 Node Requirements", J. Loughney, April 2006.
[13] RFC 4029 "Scenarios and Analysis for Introducing IPv6 into ISP Networks", M. Lind, V. Ksinant, S. Park, A. Baudot, P. Savola, March 2005.
[14] RFC 4241 "A Model of IPv6/IPv4 Dual Stack Internet Access Service", Y. Shirasaki, S. Miyakawa, T. Yamasaki, A. Takenouchi, December 2005.
[15] RFC 4779 "ISP IPv6 Deployment Scenarios in Broadband Access Networks", S. Asadullah, A. Ahmed, C. Popoviciu, P. Savola, J. Palet, January 2007.
[16] RFC 2545 "Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing", P. Marques, F. Dupont, March 1999.
[17] RFC 5340 "OSPF for IPv6", R. Coltun, D. Ferguson, J. Moy, A. Lindem, July 2008.
[18] RFC 5308 "Routing IPv6 with IS-IS", C. Hopps, October 2008.
[19] RFC 4798 "Connecting IPv6 Islands over IPv4 MPLS Using IPv6 Provider Edge Routers (6PE)", J. De Clercq, D. Ooms, S. Prevost, F. Le Faucheur, February 2007.
[20] RFC 4659 "BGP-MPLS IP Virtual Private Network (VPN) Extension for IPv6 VPN", J. De Clercq, D. Ooms, M. Carugi, F. Le Faucheur, September 2006.
[21] draft-nishitani-cgn-01 "Common Functions of Large Scale NAT (LSN)", T. Nishitani, S. Miyakawa, A. Nakagawa, H. Ashida, November 2008.
[22] draft-shirasaki-nat444-isp-shared-addr-01 "NAT444 with ISP Shared Address", Yasuhiro Shirasaki, Shin Miyakawa, Akira Nakagawa, Jiro Yamaguchi, Hiroyuki Ashida, March 2009.
[23] RFC 2766 "Network Address Translation - Protocol Translation (NAT-PT)", G. Tsirtsis, P. Srisuresh, February 2000.
[24] RFC 4966 "Reasons to Move the Network Address Translator – Protocol Translator (NAT-PT) to Historic Status", C. Aoun, E. Davies, July 2007.
[25] draft-bagnulo-behave-nat64-03 "NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", Marcelo Bagnulo, Philip Matthews, Iljitsch van Beijnum, March 2009.
[26] draft-bagnulo-behave-dns64-02 "DNS64: DNS extensions for Network Address Translation from IPv6 Clients to IPv4 Servers", Marcelo Bagnulo, Andrew Sullivan, Philip Matthews, Iljitsch van Beijnum, Masahito Endo, March 2009.
[27] Van Jacobson, "A New Way to look at Networking ", in Google Tech Talks August 30, 2006, *http://video.google.com/videoplay?docid=-6972678839686672840*.
[28] V. Jacobson, M. Mosko, D. Smetters, J.J. Garcia-Luna-Aceves, "Content-Centric Networking", January 2007, *http://cscbalston.dmeid.org/darpa/meetings/presentations/jQfxH8aC/PARC.pdf*.
[29] Web source: *http://www.emule-project.net/home/perl/general.cgi?l=1*
[30] Web source: *http://www.bittorrent.com/*
[31] Web source: *http://www.kazaa.com/us/index.htm*
[32] Web source: *http://www.skype.com/intl/en/*
[33] Web source: *http://www.akamai.com/*
[34] D. Mazières, M. Kaminsky, M. F. Kaashoek, and E. Witchel "Separating key management from file system security", Operating Systems Review 34(5), pag. 124–139, December 1999
[35] M. D'Ambrosio, P. Fasano, V. Vercellone, M. Ullio "Providing Data Dissemination Services in the Future Internet", GLOBECOM 2008:5606-5611.
[36] Web source: *http://www.upnp.org/*
[37] Web source: *http://www.dlna.org/home*
[38] Web source: *http://www.bluetooth.com/Bluetooth/*
[39] "IEEE Standard for a High-Performance Serial Bus", IEEE Std. 1394-2008, October 2008.
[40] Web source: *http://www.irda.org/*

[41] Web source: *http://www.usb.org/home*
[42] Web source: *http://www.zigbee.org/*
[43] RFC 4291 "IP Version 6 Addressing Architecture", R. Hinden, S. Deering, February 2006.
[44] RFC 4862 "IPv6 Stateless Address Autoconfiguration", S. Thomson, T. Narten, T. Jinmei, September 2007.
[45] RFC 3315 "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", R. Droms, Ed., J. Bound, B. Volz, T. Lemon, C. Perkins, M. Carney, July 2003.
[46] RFC 3769 "Requirements for IPv6 Prefix Delegation", S. Miyakawa, R. Droms, June 2004.
[47] Web source: *http://www.4ward-project.eu/*
[48] Web Source: *http://www.cvisproject.org/en/home.htm*

paolo.fasano@telecomitalia.it
domenico.marocco@telecomitalia.it
maurizio.siviero@telecomitalia.it

AUTORI

Paolo Fasano

dottore di Ricerca in Ingegneria Elettronica, è in Azienda dal 1993 dove si è dedicato all'innovazione delle reti a pacchetto. Si è inizialmente occupato di reti e servizi a larga banda, partecipando alle prime sperimentazioni geografiche a livello europeo di reti in tecnologia ATM (Asynchronous Transfer Mode). Ha spostato successivamente i suoi interessi sui servizi di rete basati sull'Internet Protocol (IP); dal 1995 al 2001 ha partecipato attivamente a numerosi gruppi di lavoro dell'IETF (Internet Engineering Task Force) ed è stato pioniere sul tema IPv6 in Telecom Italia. È oggi il responsabile della funzione Broadband Network Services Innovation che si occupa l'innovazione relativa alle reti a pacchetto (Ethernet, IP, MPLS, etc.) e ai servizi di rete su queste realizzate ■

Domenico Marocco

laureato in Ingegneria Elettronica, entra in Azienda nel 1987, dove si è occupato delle prime soluzioni di trasmissione dati e voce su doppino.
Dal 1990 al 2000 ha seguito lo sviluppo di reti dati nell'ambito della Direzione Business di Telecom Italia, contribuendo al consolidamento delle reti a pacchetto X.25 ed all'evoluzione delle reti Frame Relay, ATM e IP. Successivamente ha contribuito, in ambito Network, all'ingegnerizzazione e sviluppo del backbone IP. Dal 2004 ad oggi segue lo sviluppo dello Strato di Servizio delle reti IP (Edge IP), sia per la componente di progettazione rete che di sviluppo dei servizi per la clientela Consumer e Business. Inoltre segue lo sviluppo funzionale e dimensionale dell'Intranet dell'intero Gruppo Telecom Italia ■

Maurizio Siviero

laureato in Ingegneria Elettronica, in Telecom Italia dal 1991. Inizialmente si è occupato di standard internazionali e progetti Europei di ricerca sul controllo delle reti a larga banda. In seguito ha lavorato sull'evoluzione del controllo delle reti di Telecom Italia, dalle soluzioni per l'accesso commutato ad Internet, all'integrazione di voce e dati sulle reti a pacchetto (prima ATM e poi IP), contribuendo alla ricerca ed allo sviluppo delle soluzioni di controllo della Qualità del Servizio nelle reti a larga banda. È attualmente responsabile della struttura TILAB "Control Layer Innovation" che si occupa delle attività di ricerca e innovazione del controllo di rete fissa e rete mobile, mediante il presidio dei principali gruppi di normativa internazionale, la definizione e la prototipazione di soluzioni innovative ■

Codici a barre bidimensionali: tecnologia e campi applicativi

MOBILE

Cecilia Corbi, Stefania Lisa, Giuseppe Piersantelli

I codici a barre bidimensionali, o mobile code (tag 2d), sono rappresentazioni in grafica matriciale “machine readable” di informazioni e dati. Nati per applicazioni logistiche ed industriali, si stanno progressivamente diffondendo nel mercato consumer, comparendo su giornali, riviste, confezioni di prodotti, biglietti da visita e pubblicità e stanno diventando un veicolo per accedere, promuovere e distribuire contenuti multimediali in mobilità, “one click content”, senza dover inserire manualmente complesse URL sul browser del terminale mobile. Oltre ai supporti fisici, si prevede la diffusione dei codici a barre anche su siti web e contenuti video trasmessi su piattaforme televisive, al fine di abilitare una maggiore interazione ed una convergenza tra servizi fissi e mobili.

La scansione di codici a barre può essere oggi effettuata con un terminale mobile equipaggiato con fotocamera digitale, connessione wireless a larga banda e una delle molte applicazioni, dette “barcode reader” o “mobile code client”…, per la scansione, acquisizione e decodifica dei codici a barre bidimensionali. I codici a barre bidimensionali abilitano un’importante convergenza tra supporti materiali e fisici come la carta stampata, informazioni, contenuti multimediali presenti sul web e applicazioni interattive.

L’utilizzo di codici a barre bidimensionali coinvolge diversi attori: dai content provider ai generatori di tag 2D, dagli operatori mobili agli enti di certificazione, dalle manifatturiere alle agenzie di pubblicità e agli editori. L’interoperabilità è una parola chiave per il successo dell’ecosistema dei servizi legati ai mobile code.

1 Introduzione

I Mobile Code, ovvero i barcode 2d o codici a barre bidimensionali (anche noti come tag 2d) sono rappresentazioni di informazioni interpretabili da una macchina e si presentano come formati grafici impressi su superfici di tipo materiale (carta, prodotti di elettronica) e multimediale (video clip). Nei codici a barre lineari le informazioni sono rappresentate da linee parallele di differente spessore (*figura 1*).

Figura 1 – Codice a barre lineare

Figura 2 – Esempi di codici DataMatrix e QRcode

Nei codici bidimensionali (ovvero tag 2D, "mobile code",…), evoluzioni di quelli lineari, le informazioni sono rappresentate da un insieme di moduli quadrati bianchi e neri (o colorati) che formano figure quadrate o rettangolari (*figura 2*) . La loro capacità di contenere informazioni varia da pochi caratteri ad alcune migliaia. I tag bidimensionali possono contenere oltre 3000 caratteri ASCII e vengono acquisiti e interpretati dai terminali mobili dotati di fotocamera e client di decodifica.

2 Tecnologia

Le componenti tecnologiche basilari per il funzionamento dei mobile code sono la simbologia, il reader (client) di lettura e decodifica a bordo dei cellulari, le architetture di risoluzione degli indirizzi (metodo diretto e indiretto) e le regole di "specification", che vengono applicate al momento della creazione del tag e usate per implementare vari business model e diverse tipologie di servizi.

2.1 *Simbologia*

Il modulo della matrice è la dimensione di una cella, che tipicamente può assumere i valori 0 (cella bianca) o 1 (cella nera), con la possibilità di invertire i colori. Le celle sono organizzate in una matrice di righe e colonne in numero sia pari che dispari. Un insieme di celle forma una regione, organizzata in righe e colonne, con bordi ben delimitati per facilitarne al client l'acquisizione. Il processo di codifica dei dati avviene in due macro-fasi: la *high level encoding* in cui i dati vengono convertiti in piccole unità da 8 bits (CW), e la *low level encoding* in cui i CW sono convertiti in quadratini bianchi o neri.

Generalmente nel tag è inserito un sistema di *detection and error correction* per consentire la ricostruzione dei quadratini stampati male, sbiaditi o cancellati.

I formati di barcode più diffusi nel mondo sono gli standard QRCode (*figura 3*) sviluppati nel 1994 da Denso Wave e popolarissimo in Giappone e altri Paesi asiatici, e Datamatrix (*figura 4*), inventata da RVSI Acuity, utilizzato in Europa (in particolare in Francia). Sono disponibili anche soluzioni proprietarie, alcune molto diffuse, come gli EZCode (Spagna, Italia, USA, Messico).

A seconda dei formati può variare il numero delle informazioni codificabili in un barcode: ad esempio un barcode DataMatrix può contenere circa 3000 caratteri ASCII

All'interno di un codice a barre è possibile codificare informazioni come contatti personali, indirizzi email e di siti internet. Il risultato è una piccola immagine di forma quadrata che può essere inserita in pubblicità, brochure, giornali, riviste o su video (contenuti televisivi, pubblicità, video sul web, videoclip musicali).

I barcode sono solitamente generati con l'ausilio di una piattaforma di encoding, dotata di un'interfaccia web mediante la quale un set di informazioni, tra cui contenuto del barcode, nome, titolo, data di scadenza, è codificato in immagini di piccole dimensioni. Successivamente, la piattaforma genera uno o più barcode (in formato .jpeg o .png) in dimensioni diverse (piccolo, medio, grande) a seconda del supporto di destinazione.

Una volta generati, i barcode possono essere agevolmente decodificati da appositi programmi, barcode o mobile code reader (client), che li acquisiscono sotto forma di immagini e ne interpretano il contenuto.

Figura 3 – QR code codificano dati alfanumerici "a123456789a12..." di lunghezza 40, 100 e 200 caratteri

Figura 4 – Data Matrix code codificano dati alfanumerici "a123456789a12..." di lunghezza 40, 100 e 200 caratteri

Un esempio di servizio offerto da Nokia (*figura 5*) consente ai visitatori di generare barcode partendo dalla URL di un sito web, nonché di scaricare un'applicazione gratuita per terminali mobili che effettua la scansione e la decodifica dei codici a barre.

Una volta installata sul cellulare, l'applicazione pilota la fotocamera digitale che, a questo punto, riconosce il mobile code, lo legge e lo decodifica Se il contenuto decodificato è un indirizzo web, il telefono attiva una connessione dati e il browser del terminale mobile apre la pagina web richiesta.

Figura 5 – Esempio di Nokia per la creazione dei codici a barre

2.2 *Mobile code reader (client)*

Negli ultimi anni sono state sviluppati client da installare su terminale mobile che sfruttano le API di gestione delle fotocamere dei telefoni cellulari per effettuare la scansione del tag e acquisirne, mediante collegamento WAP, le relative informazioni. I barcone reader sono ormai maturi e molti in grado di leggere sia barcode lineari sia bidimensionali. La maggior parte è in grado di leggere più simbologie.

2.3 *Architetture di risoluzione*

Le architetture di riferimento sono di due tipologie: metodo diretto (*figura 6*) e metodo indiretto (*figura 7*). Il primo metodo, adottato in Giappone, si basa sulla diretta decodifica da parte del client dell'informazione contenuta nel tag.

L'architettura del metodo diretto prevede i seguenti passaggi:

- un client su terminale mobile acquisisce, riconosce e decodifica un codice a barre;
- il processo di decodifica si completa sul client;
- il risultato della decodifica è una URI ad un servizio di 3° party che fornisce il servizio richiesto.

Il metodo indiretto aggiunge un passaggio intermedio: il client dopo la lettura, ottiene un identificativo che deve essere risolto "server side" (nel tag viene codificato un identificativo, mai un contenuto). Il vantaggio di questa soluzione, adottato dagli operatori Europei e Americani, è quello di poter associare nel tempo alla stessa tag informazioni diverse, aggiornandole, e gestire regole di business di vario tipo. I modelli di business derivanti possono quindi essere molteplici, e dipen-

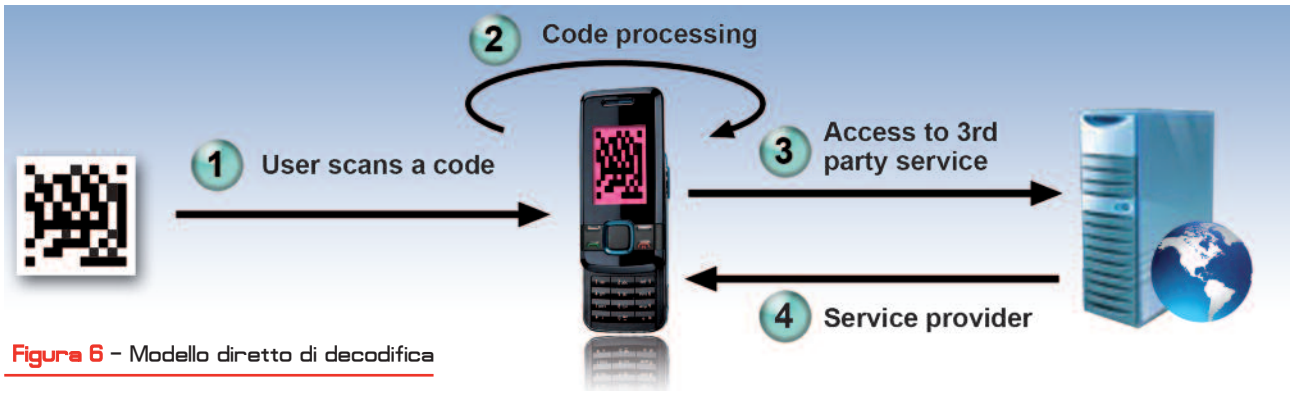


Figura 6 – Modello diretto di decodifica

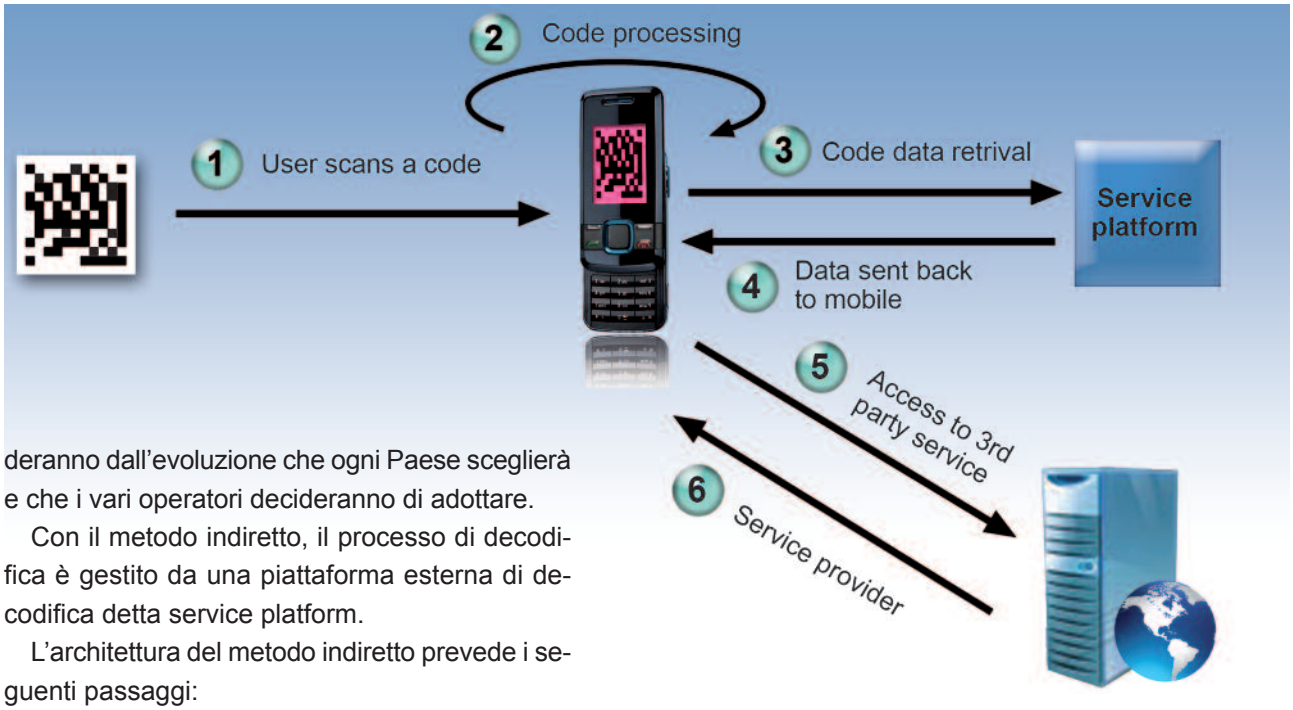


Figura 7 – Modello indiretto di decodifica

deranno dall'evoluzione che ogni Paese sceglierà e che i vari operatori decideranno di adottare.

Con il metodo indiretto, il processo di decodifica è gestito da una piattaforma esterna di decodifica detta service platform.

L'architettura del metodo indiretto prevede i seguenti passaggi:

- un client su terminale mobile acquisisce, riconosce e decodifica un codice a barre;
- il processo di decodifica non si completa sul client: il cliente legge l'identificativo inserito nella tag e quindi si connette ad una service platform;
- la service platform risolve l'identificativo dal codice a barre e invia un contenuto o la URI di un servizio al terminale mobile.

2.4 *Mobile Code Specifications*

Le Mobile Code Specifications, applicabili solo al metodo indiretto, consentono ai gestori dei servizi delle piattaforme di implementare diversi use case di servizi e regole associate.

Gli obiettivi principali delle Mobile code specifications nel modello indiretto sono:

- interoperabilità tra i client;
- implementazione di diversi modelli di servizio;
- le Specifications sono parte delle attività in corso di svolgimento presso il gruppo di standardizzazione OMA.

3 Standard

Oggi per le due simbologie principali dei codici a barre bidimensionali, il QR Code (Quick Re-

sponse) e il Data Matrix Code, esiste uno standard relativo. Al momento non esiste alcuno standard per i modelli funzionali legati alle architetture. Ciò ha creato negli ultimi anni una frammentazione sul mercato di diversi vendor e piattaforme di creazione e gestione dei vari componenti dei mobile code non interoperabili tra di loro. Pertanto OMA, Open Mobile Alliance, sta lavorando per la creazione di uno standard per l'interoperabilità di tutti i componenti dell'ambiente "mobile code", ovvero simbologie, client di lettura e decodifica, funzionalità di registrazione e risoluzione "server side", sia a livello di requisiti, sia a livello di architettura logica. I primi risultati sono attesi per fine 2009.

3.1 *OMA*

L'obiettivo di OMA , creata nel giugno del 2002, è quello di produrre standard aperti e specifiche a partire dalle richieste del mercato e dei clienti, che realizzino un'interoperabilità end-to-end in maniera trasversale rispetto sia ai vari fora di standard, che ai differenti attori della catena del valore (Operatori,Vendors, Content Providers, IT,..) .

Il mandato di OMA non si applica più solo al mondo mobile, ma anche al mondo fisso in ottica di convergenza: OMA non specifica "servizi", ma "Service Enablers"; in tal modo gli attori della catena del valore possono sviluppare "servizi" innovativi e differenziati, mantenendo comunque la massima interoperabilità e riducendo i costi. Con Enabler si intende pertanto un insieme di specifiche che definiscono una determinata funzionalità, che viene offerta al livello applicativo per la realizzazione di servizi per l'utente finale; lo scopo di OMA è definire quindi un framework architetturale comune, indipendente da tecnologie proprietarie e "bearer agnostic".

Telecom Italia è tra gli Operatori più attivi all'interno dell'ente ed ha una consistente ed attiva partecipazione ai vari gruppi tecnici, detenendo alcune chairmanship di rilievo.

All'interno della struttura operativa dei gruppi tecnici di OMA, il gruppo MCE-MC (Mobile Code) sta lavorando attivamente alla specifica completa (ovvero requisiti, architettura, e specifica tecnica di dettaglio) dell'enabler Mobile Code. Partecipano attivamente e con interesse al gruppo sia Operatori (Telefonica, Vodafone, Orange, At&T, T-Mobile, NTTDoCoMo, China Mobile, …), sia manifatturiere (Samsung, Nokia, HP, Fujitsu) sia media vendor (Scanbuy, Neomedia, …). Il primo deliverable pubblico rilasciato dal gruppo OMA Mobile Code è stato un white paper (pubblicato nel 2008 in linea con le linee guida di GSMA).

Obiettivo principale del gruppo tecnico è quello di realizzare un'architettura standard completa per l'ecosistema dei Mobile Codes (utenti, publishers, service providers), che permetta l'interoperabilità fra gli erogatori dei servizi legati ai mobili code, mantenendo una backwards compatibility con esistenti e rilevanti sistemi di 2D barcode. Tale interoperabilità abiliterà la decodifica delle stesse simbologie da client di lettura realizzati da vendor diversi; inoltre sarà realizzata un'architettura di gestione e risoluzione (per il metodo indiretto) che permetterà all'utente finale di usufruire di servizi messi a disposizione dai diversi operatori in modo completamente trasparente. Il gruppo MCE-MC sta lavorando a requisiti per il barcode reader, la simbologia, l'architettura diretta e indiretta di registrazione e risoluzione, le specifiche per la sintassi di decodifica e risoluzione e ad alcune funzionalità a corredo del buon funzionamento dei servizi abilitati dai tag 2d, come la security.

La specifica è ora nella fase di definizione dell'architettura e si prevede la sua chiusura nel primo semestre del 2010.

4 User experience: "One click content"

Tramite barcode bidimensionali è possibile accedere a contenuti multimediali in mobilità con poche e semplici operazioni: inquadrare un barcode, acquisirlo, decodificarlo, accedere ad una pagina web o a un contenuto multimediale.

Il cliente beneficia dell'uso dei codici a barre poiché tale modalità diminuisce il numero di click per accedere ai contenuti e unifica il meccanismo per l'accesso a servizi forniti da Operatori diversi (informazioni turistiche, orari dei mezzi di trasporto, messaggi pubblicitari, streaming audiovideo, chiamate a numeri di pubblica utilità…).

I barcode portano benefici anche ai content provider in quanto tale modalità aumenta la visibilità dei loro contenuti e come conseguenza favorisce gli acquisti di impulso (ad esempio, scaricare una suoneria, effettuando la scansione di un barcode su un video musicale o un DVD).

Lo schema seguente (*figura 8*) esemplifica il passaggio da un codice a barre stampato su supporto fisico o pubblicato su un video ad un contenuto multimediale.

5 Servizi disponibili

L'utilizzo di codici a barre inizia negli anni '40 grazie al lavoro di due ricercatori americani del Drexel Institute of Technology e successivamente dell'azienda Sylvania. Lo scopo primario consisteva nell'identificazione di vagoni ferroviari; dagli anni '60 i codici a barre sono applicati in ambito commerciale per l'identificazione di beni di largo consumo.

Negli ultimi dieci anni si sono diffuse le prime applicazioni consumer, grazie anche all'introduzione di telefoni cellulari con fotocamera. In Giappone), dove i tag e i relativi lettori sono diffusi da una decina d'anni, circa il 90% dei consumatori accede regolarmente alle informazioni relative ai beni di consumo, effettuando scansioni con la fotocamera del telefono cellulare per connettersi ad Internet.

Oggi dall'Estremo Oriente i tag 2D si stanno diffondendo anche in Europa, in particolare in Francia (figura 9): sono stampati su quotidiani, periodici, manifesti pubblicitari, sull'etichetta di indumenti e sui biglietti da visita.

Con il telefono cellulare dotato di una fotocamera e di un'applicazione specifica (sviluppata in Java oppure nativa Symbian, Windows Mobile, …) è possibile effettuare la scansione e la decodifica di un tag, e quindi far compiere automaticamente l'azione desiderata, ad esempio aprire il link Wap o Web contenuto nel codice e visualizzare informazioni aggiuntive relative al prodotto pubblicizzato, oppure acquistare contenuti digitali (sfondi e suonerie) o scaricare applicazioni sul terminale. Tutto ciò inquadrando semplicemente un'immagine, ovvero un tag anziché inserendo

Figura 8 – Il passaggio da un codice a barre a un contenuto multimediale

Figura 9 – Alcuni esempi di giornali e pubblicita` con barcode

un indirizzo Internet. Si tratta di un sistema innovativo per la promozione di prodotti, servizi e contenuti.

L'utilizzo dei tag diminuisce quindi fortemente il numero di click per la ricerca di informazioni in mobilità e al contempo consente di aumentare la visibilità dei contenuti e favorire gli acquisti di impulso. Ecco alcuni esempi applicativi: la memorizzazione di un contatto mediante la lettura del tag stampato su un biglietto da visita; l'interazione con servizi interattivi multimediali, quali il televoto; l'advertising multimediale e il one-click content, ovvero il reperimento facile e l'invio veloce di informazioni su terminali mobili. Per esempio in Giappone i tag sono usati ovunque, persino sulle etichette del vino (*figura 10*) per ottenere maggiori informazioni sulla qualità, gli abbinamenti, la gradazione alcolica del contenuto di una bottiglia…

Anche in Italia da qualche tempo il servizio di scansione e decodifica tag si sta diffondendo, soprattutto sulla carta stampata e la cartellonistica.

Nel maggio 2008 il quotidiano La Gazzetta dello Sport (gruppo RCS) ha lanciato gazza&play (*figura 11*), un servizio mobile multimediale basato sui codici a barre bidimensionali, che consente ai lettori di approfondire le notizie pubblicate sul quotidiano mediante la fruizione di contenuti mobili multimediali (audio e video); per accedere a tali contenuti, è necessario disporre

di un terminale mobile con fotocamera, installare il lettore di barcode (se non preinstallato sul terminale) ed effettuare una scansione del barcode associato al singolo articolo. I barcode utilizzano il formato QRCodes (Quick Response Codes), attualmente molto popolare in Giappone. Se il terminale è privo di fotocamera, è possibile inviare ad un centro servizi un SMS con un codice numerico associato al barcode. A termine di questa operazione, si riceve un link che punta ad un contenuto multimediale (una pagina del sito WAP della Gazzetta, un video o un commento audio). Il servizio è gratuito ad eccezione dei costi di connessione per ricevere i contenuti multimediali.

PIU' CLICK, PIU' INFO →

Inquadra la TAG con la fotocamera del tuo cellulare, per visionare l'intervista di approfondimento sul tema dei codici a barre; utilizza il software Flash Me presente nel menu delle applicazioni.
Se non hai Flash Me scaricalo gratuitamente dal sito Tim>119>I miei Players, oppure da *http://app.scanlife.com/appdownload/dl*
Costi di navigazione e download del video secondo il tuo piano tariffario.

Figura 10 – Esempio delle tag sulle etichette dei vini

Inoltre l'edizione italiana della rivista Wired, apprezzata dagli appassionati di tecnologie, ospita regolarmente inserti pubblicitari corredati di codici a barre. In fondo ad ogni messaggio pubblicitario sono normalmente presenti le istruzioni per scaricare ed installare una delle molte applicazioni barcode reader gratuite disponibili sul web.

Ed ancora: la rivista Quattroruote offre un servizio di approfondimento dei propri articoli mediante l'utilizzo di QRCode: effettuando la scansione, si accede a pagine web con contenuti extra relativi all'articolo in oggetto.

Se passiamo poi al campo delle affissioni murali, su una delle pareti della stazione FFSS Santa Lucia di Venezia, oggi campeggia un

Figura 11 – I barcode de "La Gazzetta dello sport"

enorme poster giallo che raffigura un codice QRCode da cui è possibile accedere ad informazioni turistiche sulla città.

Sovente, quando si frequentano conferenze, incontri e riunioni all'estero, professionisti e manager di aziende straniere porgono i loro biglietti da visita corredati da piccoli barcode cui è associata generalmente una v-card, ovvero un file con le informazioni di contatto (nome, cognome, email, indirizzo, numero di telefono), leggibile da molti programmi di posta elettronica e dalle rubriche dei principali terminali mobili.

Alcuni siti web che distribuiscono applicazioni per terminali mobili consentono, poi, di effettuare i download mediante la scansione di codici a barre, anziché l'inserimento di lunghe e complesse URL nella barra degli indirizzi del browser mobile: accedendo con un computer a tali siti web, accanto a ciascuna applicazione è visualizzato un barcode: per accedere alla URL specifica e scaricare l'applicazione desiderata, è sufficiente lanciare un barcode reader dal cellulare ed effettuare la scansione di quel codice a barre.

Quando si acquista un biglietto per un volo aereo o per un viaggio in treno, è inoltre sempre più usuale ricevere per email la relativa documentazione in formato elettronico, spesso corredata da un codice a barre. Alcune compagnie aree consentono di effettuare l'imbarco mediante la scansione di un barcode ricevuto per email su un terminale mobile. Il barcode, visualizzato sul

display del terminale, è acquisito da uno scanner installato al varco verso l'aeromobile. Tale modalità semplifica e accelera la procedura di imbarco.

Telecom Italia ha recentemente lanciato il servizio TIM FlashMe, disponibile per telefoni con sistemi operativi Java, Windows CE e Symbian...

L'applicazione FlashMe è stata anche oggetto di un trial di servizio condotto presso il Telecom Italia Future Center a Venezia. Nel settembre 2008, presso il secondo chiostro al piano terreno del convento è stata organizzata una mostra di fotografie intitolata "Vapore d'acqua" con il patrocinio del Comune di Venezia e della società di trasporti pubblici ACTV.

In tale occasione, è stato predisposto un sistema sperimentale di guide multimediali basato sulla scansione di codici a barre associati ad alcune delle fotografie esposte. Terminali Apple iPhone equipaggiati con il client per la lettura di tag e connessi ad una rete wireless ad alta velocità, appositamente predisposta, sono stati consegnati ai visitatori della mostra che, inquadrando i barcode, hanno così potuto accedere direttamente sull'iPhone a molti contenuti multimediali: video interviste all'autore, descrizioni testuali, immagini correlate e strumenti interattivi di votazione e commento.

Le immagini seguenti (*figura 12*) sono alcune delle pagine web visualizzate dal browser del ter-

Figura 12 – Informazioni aggiuntive reperite tramite scansione di barcode

minale iPhone pochi istanti dopo la scansione di un barcode associato ad un'opera esposta. La prima schermata mostra i contenuti disponibili: testo, immagini e una video intervista fruibile in progressive download. La seconda schermata mostra gli strumenti interattivi messi a disposizione dei visitatori.

6 Conclusioni

È facile prevedere un aumento della diffusione e dell'utilizzo dei codici a barre bidimensionali per facilitare l'accesso alle informazioni in mobilità o anche solo per gestire operazioni semplici, come aggiungere un numero alla rubrica del cellulare, o chiamare un taxi, inquadrando il barcode associato al numero di telefono della compagnia. I barcode non saranno ospitati esclusivamente su supporti fisici e cartacei, come giornali, cartelloni, copertine di CD, o su pagine di siti web, ma troveranno spazio anche su supporti immateriali come video trasmessi in televisione per abilitare, ad esempio, l'interattività dello spettatore con programmi televisivi o l'acquisto di contenuti correlati con il programma fruito (ad esempio, la suoneria di un video musicale visto in TV).

Anche l'acquisto di contenuti mobili o la sottoscrizione a servizi mobili a valore aggiunto, al momento gestita mediante l'invio di comandi testuali a numeri large account, potrà essere effettuata mediante la scansione e l'acquisizione di codici a barre associati a particolari servizi.

ACRONIMI

QRCode	Quick Response Code
URL	Universal Resource Locator
WAP	Wireless Application Protocol
3G	Third Generation Network
URI	Uniform Resource Identifier
OMA	Open Mobile Alliance

BIBLIOGRAFIA

[DATAMATRIX] "Information technology - International symbology specification - Data Matrix", ISO/IEC 16022:2000.

[EAN/UPC] "Information technology -Automatic identification and data capture techniques - Bar code symbology specification - EAN/UPC", ISO/IEC 15420.

[FLASHCODE] "Flashcode Reader International Specification", Version 1.0 http://www.mobiletag.com/beta/en/contactspecification.html

[FLASHME] Il client, se non già presente sul temilane mobile TIM, può essere scaricato dal sito WAP: http://FlashMe.getscanlife.com ed installato su molti telefoni cellulari compatibili con l'applicazione. TIM FlashMe è anche disponibile tra le applicazioni della community NextInnovation (www.nextinnovation.it)

[GIAPPONE] http://onlypunjab.com/fullstory2k7-insight-Brings+Barcode-status-19-newsID-21380.html

[MIME] "Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types", RFC 2046 http://www.ietf.org/rfc/rfc2046.txt

[NDEF] "NFC Data Exchange Format (NDEF) Technical Specification", NFC Forum http://www.nfc-forum.org/specs/

[NFCRTD] "NFC Record Type Definition (RTD) Technical Specification", NFC Forum

[NOKIA] "Mobile Codes" http://mobilecodes.nokia.com

[NTTDOCOMOGUIDE] "Guidelines and criteria for creating QR codes compatible with all terminals", NTT DoCoMo, http://www.nttdocomo.co.jp/english/service/imode/make/content/barcode/about/#p02

[NTTDOCOMOFUNC] "Bar Code Function", NTT DoCoMo, http://www.nttdocomo.co.jp/english/service/imode/make/content/barcode/function/

[OMAURI] "URI Schemes for the Mobile Applications Environment", Version 1.0, Open Mobile Alliance™, OMA-TS-URI_Schemes-V1_0-20070718-D, URL:http://www.openmobilealliance.org/.

[QR] "Information technology - Automatic identification and data capture techniques -QR Code 2005 bar code symbology specification", ISO/IEC 18004:2006.
[SPRTD] "NFC Smart Poster RTD Technical Specification", NFC Forum
[TAGURI] "RFC 4151. The 'tag' URI Scheme", IETF, http://www.faqs.org/rfcs/rfc4151.html.
[TEXTRTD] "NFC Text RTD Technical Specification", NFC Forum
[URI] "RFC 3986. Uniform Resource Identifier (URI): Generic Syntax", IETF, http://www.ietf.org/rfc/rfc3986.txt.
[URIRTD] "NFC URI RTD Technical Specification", NFC Forum
[URNRES] "RFC 2169 - A Trivial Convention for using HTTP in URN Resolution", IETF, http://www.faqs.org/rfcs/rfc2169.html.

ceciliamaria.corbi@telecomitalia.it
stefania.lisa@telecomitalia.it
giuseppe.piersantelli@telecomitalia.it

AUTORI

Cecilia Corbi

laureata in Matematica, è entrata in azienda nel 1989. Responsabile di diversi progetti sull'area dello sviluppo dei servizi sia per la rete fissa che per la rete mobile, ha ricoperto per diversi anni il ruolo di Client Manager, coordinando progetti relativi allo sviluppo e messa in campo di architetture e servizi VAS innovativi per gli Operatori Mobili delle consociate estere.
Dal 2006 è Technical Area Coordinator per Telecom Italia Lab per quanto riguarda le tematiche di standardizzazione che afferiscono al Service Layer ed in particolare coordina la partecipazione di Telecom Italia all'ente OMA (Open Mobile Alliance). È presente come membro Telecom Italia nel Board of Director OMA, nonché chairman del comitato operativo del Board 'Strategic Plan' ■

Stefania Lisa

laureata in Scienze dell'Informazione, entrando nel 1994 a far parte del Gruppo Telecom Italia. Ha lavorato alla progettazione e design di servizi multimediali e per anni si è occupata di e-learning, coordinando progetti con l'obiettivo sia di prototipazione e realizzazione di nuove applicazioni e di servizi, sia di studio e messa a punto di nuove modalità formative e di apprendimento, legate alle opportunità del mondo Web e mobile. Negli ultimi ha lavorato nell'ambito dei servizi multimedia broadband e broadcast (IPTV, DVB-T, DVB-H) e delle applicazioni a valor aggiunto legate al multimedia (rich media languages, 3d, mobile code). Attualmente è responsabile di un progetto di innovazione sul tema del "Visual Search" evoluto in Telecom Italia Lab ■

Giuseppe Piersantelli

ha lavorato per l'Università di Genova e Xerox. È entrato in Telecom Italia nel 2001 all'interno del progetto Sviluppo Professionalità Internet. Presso Telecom Italia Lab si è occupato di comunità virtuali, home network, scouting di dispositivi multimediali e nuove piattaforme per la distribuzione di contenuti digitali multimediali. È stato più volte keynote speaker in conferenze internazionali sui temi della mobile TV e della convergenza broadband-broadcast.
Attualmente è responsabile di un progetto sulla fotografia digitale e scrive di digital imaging sul blog Business Ecosystems. del Future Centre di Venezia ■

Analisi contestuale in ambito automobilistico

INNOVAZIONE

Alessio Dore, Francesco Pasini, Carlo Regazzoni

Negli ultimi anni la tecnologia applicata in ambito automobilistico ha assunto un ruolo sempre più rilevante per il miglioramento della sicurezza e del confort di guida. L'estrazione di dati contestuali tramite l'analisi di flussi video acquisiti da videocamere può avere ricadute in molteplici applicazioni sia orientate alla prevenzione di incidenti, sia per fornire informazioni utili alla guida. Nell'ambito del laboratorio congiunto Telecom Italia Lab - Università di Genova, il gruppo di ricerca ISIP40 (vIdeo and SIgnal Processing for Telecommunications), in collaboarazione con la divisione di TILab "Research & Trends", si è focalizzato sullo sviluppo di tecniche per l'analisi di sequenze video acquisite da telecamere orientate in modo tale da inquadrare la porzione di strada davanti al veicolo e telecamere posizionate all'interno dell'abitacolo puntate sul guidatore. Le tecniche di elaborazione video per le telecamere esterne permettono l'estrazione automatica di dati contestuali relativi alla strada e al traffico rilevato; in particolare è possibile estrarre il numero di corsie di marcia, la corsia corrente, un indicatore del livello di traffico che varia da "blocco" a "strada libera" e il numero di persone rilevate dal software nel campo di vista del sensore. L'elaborazione delle immagini acquisite dalla telecamera interna è invece utile per l'estrazione automatica di dati relativi al livello di attenzione del guidatore (posizione del volto, direzione dello sguardo, occhi chiusi, ecc). Nel seguito di questo articolo è descritta l'architettura del sistema e sono presentati alcuni esempi di risultati di estrazione del contesto ottenuti su sequenze reali acquisite da veicoli in movimento.

1 Architettura del Sistema

Il sistema realizzato è composto, dal punto di vista dell'hardware impiegato, da due webcam collegate ad un laptop PC presente all'interno dell'autovettura. Il PC ha una connessione internet attraverso una PC card UMTS per la pubblicazione delle informazioni di contesto sul server remoto e per la fruizione dei filmati registrati sul veicolo.

In *figura 1* viene rappresentata in maniera molto semplice l'architettura del sistema realizzato, visualizzando le connessioni tra i differenti componenti hardware (2 webcam, 1 laptop pc e 1 scheda UMTS).

La *figura 2* rappresenta invece l'architettura logica del sistema, visualizzando i moduli software che sono stati utilizzati per realizzare il prototipo.

I moduli 1 (TLC1) e 2 (TLC2) sono i thread di acquisizione immagini dalle 2 webcam utilizzate. Le immagini acquisite vengono inviate ai moduli software di registrazione che generano filmati temporizzati dal modulo TIMER, che è un thread in grado di inviare ai i segnali di inizio/fine filmato.

Parallelamente le immagini acquisite vengono rese disponibili anche ai moduli di elaborazione che sono in grado di estrarre i dati di contesto (interno ed esterno all'abitacolo) e inviarli al modulo XML_POST che in maniera asincrona e a intervalli prestabiliti trasmette le informazioni contestuali opportunamente formattate in un file XML ad un server remoto, utilizzando la connettività garantita dalla scheda UMTS.

L'acquisizione e l'elaborazione dei dati è realizzata quindi tramite una vettura equipaggiata con un PC portatile a cui sono collegate due webcam opportunamente posizionate per inquadrare

Figura 1 – Architettura fisica del sistema realizzato

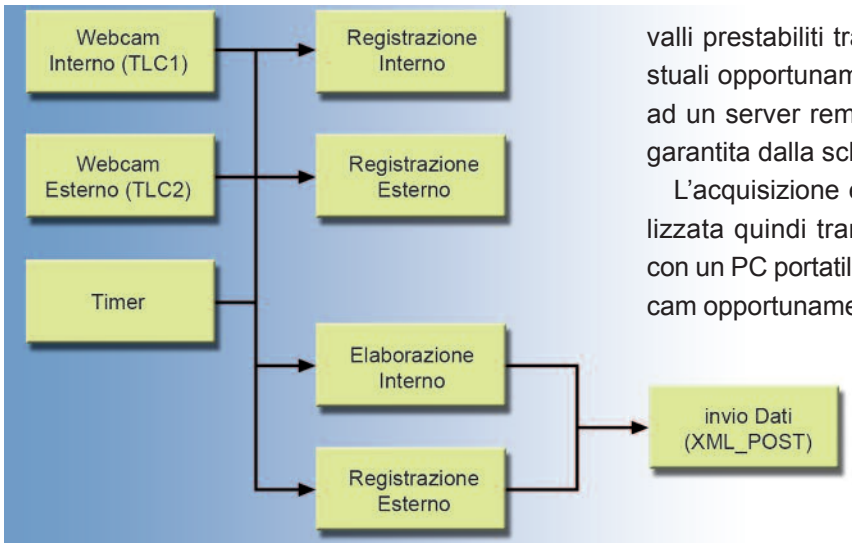


Figura 2 – Architettura logica del sistema realizzato

l'esterno dell'abitacolo e il guidatore. I sensori ottici montati a bordo del veicolo in questa fase del progetto sono:

- telecamera-Webcam a colori con vista verso l'esterno del veicolo, puntata nello spazio di strada di fronte al veicolo. Tale sensore permette di acquisire ad esempio immagini di altri veicoli che si trovano sulla carreggiata e della carreggiata stessa;
- telecamera-Webcam a colori con vista all'interno del veicolo, puntata sul guidatore. Tale sensore permette di acquisire immagini sullo stato del conducente.

Un'interfaccia è stata sviluppata per l'acquisizione e la memorizzazione real-time dei diversi dati sincronizzati tra di loro. Il modulo di acquisizione permette pertanto di ottenere fino ad un massimo di 4 flussi video a 25 frame/sec, acquisiti da telecamere anche con caratteristiche diverse, e di sincronizzarli temporalmente fra loro in modo da permettere l'associazione coerente dei dati elaborati in ciascuno di essi.

2 Analisi video del contesto stradale

Lo scopo degli algoritmi per l'analisi di contesto da telecamere esterne è quello di estrarre informazioni su ciò che accade all'esterno del veicolo stesso. I principali elementi di interesse presenti nello spazio intorno al veicolo sono le linee della strada, gli altri veicoli, gli ostacoli ed i pedoni.

In questo progetto l'attenzione è stata incentrata sulle seguenti caratteristiche:

- posizione delle linee della carreggiate, numero di carreggiate;
- stato del traffico caratterizzato da tre livelli di intensità, ovvero, scorrevole, moderato, intenso;
- posizione del veicolo nella strada;
- altri veicoli presenti sulla strada con particolare attenzione a quelli di fronte che potrebbero essere causa di tamponamenti. In questo caso è utile conoscere le posizioni nel tempo degli altri veicoli per poter meglio valutare potenziali situazioni pericolose.

L'elaborazione dei flussi video è stata realizzata attraverso moduli interconnessi, in modo da sfruttare le informazioni ottenute ad ogni passo. A partire dal flusso video ottenuto dalla telecamera orientata all'esterno del veicolo si operano le seguenti operazioni:

- rilevazione di corsie: in questa fase si effettua la ricerca delle linee che delimitano le carreggiate e vengono identificate le corsie di marcia;
- valutazione posizione del veicolo sulla strada: viene effettuata la valutazione della posizione relativa del veicolo tra le carreggiate rilevate al passo precedente. In questo modo è possibile stimare il comportamento di guida tramite analisi temporale dei cambi di carreggiata;
- rilevazione di veicoli nelle prossimità: in questo modulo vengono ricercati veicoli che si trovino nelle vicinanze. Questa elaborazione permette di valutare lo stato del traffico e rilevare possibili situazioni di pericolo causate dagli altri automobilisti;
- inseguimento veicoli estratti: i veicoli rilevati ad ogni istante sono quindi coerentemente associati per momenti successivi in modo da valutarne il movimento all'interno della strada.

Nel seguito verranno descritti singolarmente i sottomoduli del sistema, specificando per ognuno le informazioni che vengono estratte, le caratteristiche che sono già state implementate e quelle da migliorare/implementare.

2.1 *Estrazione contorni con algoritmo di Canny*

L'algoritmo di Canny è un operatore per l'estrazione dei contorni (edge detection) ideato nel 1986 da John F. Canny ampiamente utilizzato in elaborazione delle immagini. Canny ha presentato una teoria del riconoscimento dei contorni [3] che si propone di spiegare i fondamenti di questa tecnica. Questo algoritmo si basa su un metodo di calcolo multi-stadio per individuare i contorni

presenti nelle immagini. In particolare si utilizza il calcolo delle variazioni, una tecnica basata sulla ricerca della funzione che ottimizza un dato funzionale. La funzione ottimale è definita come somma di quattro termini esponenziali, ma può essere approssimata dalla derivata prima di una funzione gaussiana. Un esempio di estrazione dei contorni da un'immagine della strada acquisita da un veicolo in movimento è presentata in *figura 3 (a)*.

2.2 *Estrazione delle linee con algoritmo di Hough*

Dopo aver estratto i contorni l'immagine binaria risultante viene utilizzata per rilevare le linee presenti nell'immagine utilizzando l'algoritmo di Hough [4, 5, 6]. Prima di iniziare con la ricerca delle linee sull'immagine binaria viene applicata una maschera per escludere le zone di non interesse. Nel caso di immagini da veicolo, per esempio, la parte in alto si può considerare inutile ai fine della ricerca delle linee della strada. Naturalmente nel nostro caso verranno ricercate le linee che hanno caratteristiche che possono essere associate alle linee della carreggiata.

Nell'elaborazione delle immagini la trasformata di Hough viene utilizzata per individuare forme diverse definite analiticamente (ad esempio linee, cerchi ellissi…).

L'algoritmo riceve come ingresso le coordinate dei punti dell'immagine (binaria) e fornisce in uscita una descrizione parametrica dell'insieme delle curve riconosciute, appartenenti ad una fissata figura analitica (nel nostro caso le linee). L'algoritmo di Hough, applicato sull'immagine dei contorni che rappresenta il contesto esterno al veicolo, estrae un numero abbastanza alto di linee come si può vedere da *figura 3 (b)*. Per l'estrazione delle linee nel sistema in realtà vengono utilizzati due estrattori in parallelo uno per le linee a destra del veicolo (nella figura disegnate di rosso) e uno per le linee a sinistra del veicolo (nella figura disegna te di verde).

Il modulo successivo si occupa di prendere la decisione su quali sono, tra tutte le linee, quelle che identificano al meglio la strada che il veicolo sta percorrendo.

2.3 *Estrazione informazioni strada e posizione veicolo*

I moduli successivi si occupano di trovare le informazioni che riguardano la strada e il veicolo a partire dalle linee estratte ai passi precedenti.

Le informazioni interessanti che possono essere estratte sono:

- numero di carreggiate;
- stato del traffico;
- posizione veicolo.

L'algoritmo di ricerca delle linee si fonda sul fatto che nella strada le uniche linee che dovrebbero essere presenti sono quelle relative alle linee laterali che la delimitano e la linea di mezzeria se presenti.

Il primo passo da eseguire è la ricerca della linea a destra e della linea a sinistra prossime al veicolo. Esse sono la prima linea destra e la prima sinistra tra quelle trovate partendo dal basso dell'immagine. Una volta estratte le due linee si focalizza l'attenzione su una zona all'interno del triangolo formato da esse. Su tale zona viene eseguita una statistica dei pixel appartenenti alla strada nel tempo allo scopo di creare un modello della strada. Tale statistica si effettua all'interno di una finestra temporale che viene spostata avanti nel tempo per considerare possibili cambiamenti delle condizioni della strada (come ad esempio variazioni dell'illuminazione).

Il valore dei pixel della zona considerata (strada) vengono utilizzati per creare in modo iterativo un modello gaussiano [7] $N(\mu,\Sigma)$ con media μ e matrice di covarianza Σ. Da notare che si può assume l'indipendenza dei tre canali colore, pertanto la matrice di covarianza può essere considerata diagonale con le varianze sul rosso, sul blu e sul verde sulla diagonale.

In seguito alla creazione del modello della strada basato sul colore, sono considerati tutti i

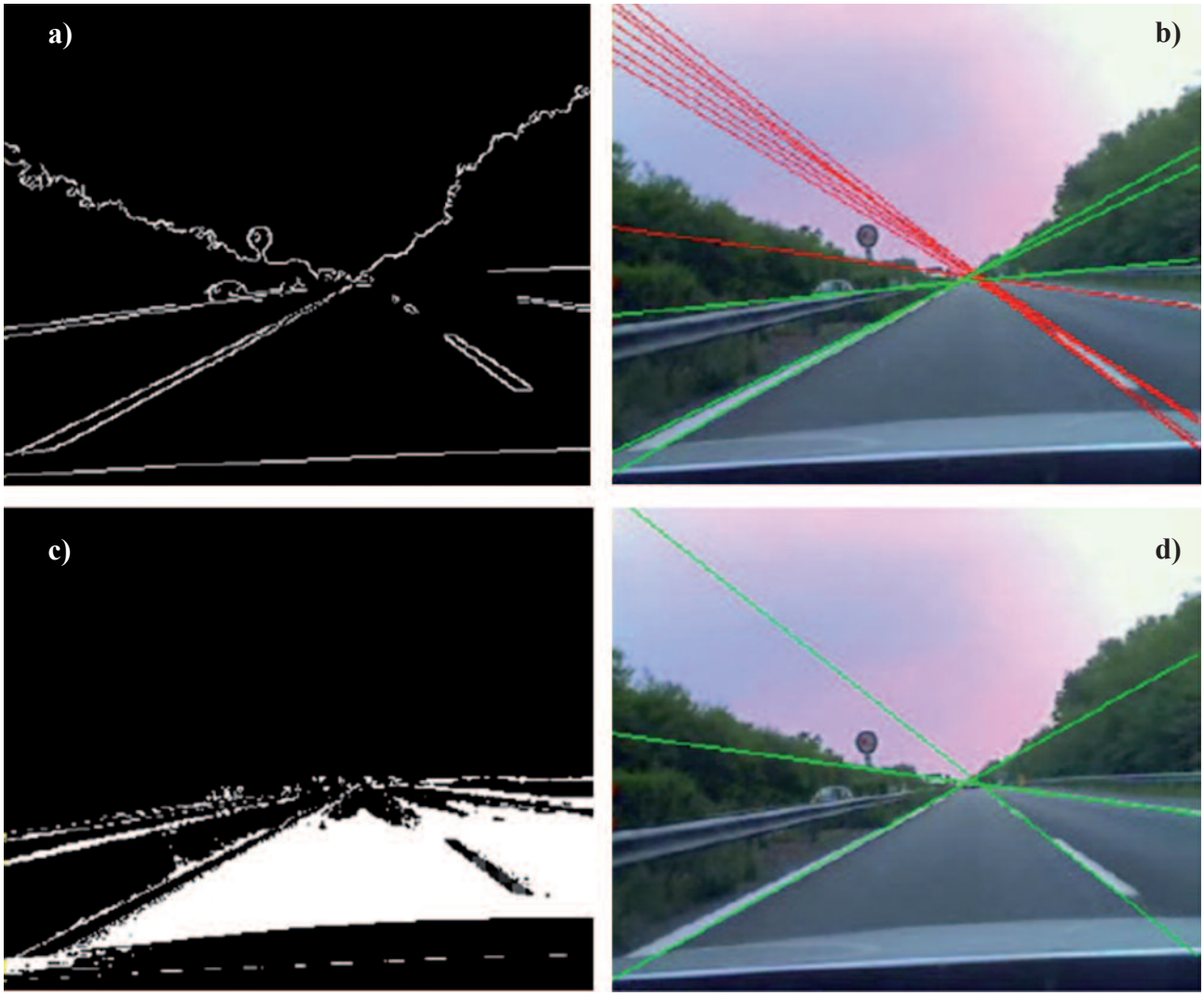
pixel dell'immagine al di sotto del punto di intersezione tra le due linee trovate e ognuno viene confrontato con il modello gaussiano della strada, cercando quelli che sono più simili al modello. Un esempio può essere visto in *figura 3 (c)*.

Il passo successivo è capire se la strada è a una o due corsie e la posizione del veicolo all'interno di essa. Le altre linee vengono cercate spostandosi verso i lati dell'immagine, le successive linee a destra e a sinistra sufficientemente distanti dalle prime trovate sono le candidate. Se le linee sono presenti vengono create due zone una tra due linee a destra e una tra le due linee a sinistra e per ogni zona vengono calcolati il numero di punti appartenenti alla strada. Se una delle due zone ha un numero sufficiente di pixel appartenenti alla strada questa viene considerata a due corsie e la nuova linea come linea di delimitazione della strada. Un esempio di linee estratte è mostrato in *figura 3 (d)*.

Le situazioni che possono capitare dopo aver estratto le prime due linee sono le seguenti:

- nessuna nuova linea trovata. La strada è ad una corsia;

Figura 3 – Esempio di passi di elaborazione per l'estrazione di dati contestuali della strada, a] estrazione dei contorni tramite l'algoritmo di Canny; b] rilevazione delle linee mediante calcolo della trasformata di Hough; c] estrazione del modello della strada; d] rilevazione delle linee di carreggiata.



- trovata nuova linea a destra. La strada è a due corsie e il veicolo si trova a sinistra;
- trovata nuova linea a sinistra. La strada è a due corsie e il veicolo si trova a destra.

Successivamente conoscendo posizione della strada nell'immagine e la posizione del veicolo, sarà possibile restringere il campo di ricerca dei veicoli esterni solo ad alcune parti dell'intera immagine con notevole risparmio sotto il punto di vista computazionale.

Sono stati eseguiti alcuni test per valutare la bontà del modulo di estrazione delle informazioni che riguardano la strada. È necessario, tuttavia, precisare che il modulo deve essere utilizzato in zone stradali dove sono ben visibili le linee appartenenti alla carreggiata.

Per queste prove son state utilizzate tre sequenze video riprese da telecamera con vista frontale. Le sequenze sono state riprese nel tratto autostradale Genova – Chiavari.

Durante i test vengono estratte due informazioni che sono:

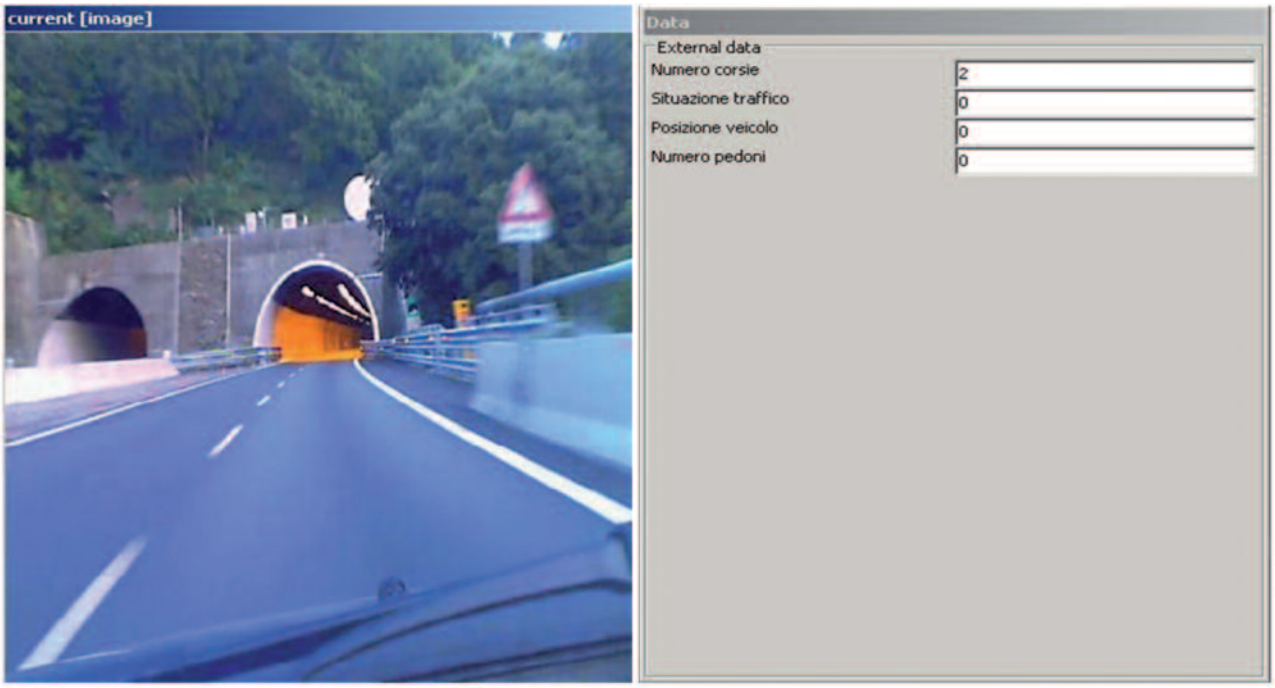
1. numero di carreggiate della strada;
2. posizione del veicolo.

Queste informazioni sono estratte per ogni frame del filmato, ma siccome sia il numero di carreggiata che la posizione del veicolo non cambiano in modo istantaneo l'uscita (risultato) viene mediata su una finestra temporale più lunga. Le estrazioni istantanee possono essere limitate grazie al processo di media temporale. Il numero di carreggiate è calcolato su di una finestra di 2 secondi, mentre per la posizione del veicolo che ha maggiore variabilità viene utilizzata finestra di 1 secondo.

Per realizzare queste prove è stata implementata un'applicazione che riceve le informazioni correnti oltre che a visualizzare il video; è possibile vedere come si presenta l'interfaccia in *figura 4*.

Per quanto riguarda la posizione del veicolo viene indicato un numero che è zero se il veicolo si trova sulla corsia più a destra e viene incrementato di 1 mano a mano che il veicolo si sposta di una corsia più a sinistra. In *tabella 1* e *tabella 2* vengono mostrati i risultati.

Figura 4 – Esempio estrazione dati contesto strada



Sequenza	Rilevamenti	Corretti	Errati	% corretti
1 (30 sec)	30	25	5	83
2(39 sec)	39	29	10	74
3(11 sec)	11	11	0	100

Tabella 1 – Statistiche posizione veicolo su strada

Sequenza	Rilevamenti	Corretti	Errati	% corretti
1(30 sec)	15	11	4	74
2(39 sec)	19	19	0	100
3(11 sec)	5	5	0	100

Tabella 2 – Statistiche su estrazione numero carreggiate strada

È necessario far notare che per come è stato implementato il modulo la parte di estrazione della posizione del veicolo è strettamente correlata al numero esatto di corsie; infatti nel caso di due corsie dove una viene riconosciuta correttamente se il veicolo è a sinistra sarà sicuramente un errore, mentre se veicolo è a destra sarà sempre corretto.

Invece quando il numero di corsie è corretto ma la posizione del veicolo è calcolata in modo sbagliato (ad esempio in sequenza 2), significa che le linee estratte non sono corrette (non sono quelle che identificano la strada).

2.4 *Rilevamento di veicoli in prossimità*

Per lo sviluppo di un efficiente sistema intelligente che permetta il miglioramento della sicurezza è necessario poter estrarre nella maniera più accurata possibile le informazione che riguardano lo spazio intorno al veicolo stesso. In questo spazio potranno essere presenti oggetti fissi come ad esempio edifici o alberi e/o oggetti in movimento che sono rappresentati principalmente da tutti gli altri veicoli (motocicli, autovetture, camion, ecc.) presenti in prossimità dell'automobile.

Un accurato sistema di riconoscimento di veicoli da telecamera mobile si basa su due passi consecutivi:

1. generazione delle ipotesi. Si intende la generazione delle ipotesi di dove potrebbe essere posizionato il possibile veicolo/oggetto all'interno dell'immagine;
2. verifica delle ipotesi. Si intende la verifica della presenza dei veicoli/oggetti all'interno dell'immagine partendo dalle ipotesi generate nel precedente passo (ad esempio prima di inizializzare il modulo di inseguimento).

Una soluzione robusta di questo tipo di problematiche non è semplice, infatti un sistema deve soddisfare principalmente due requisiti:

1. numero di mancati allarmi;
2. numero di falsi allarmi.

Bisogna quindi raggiungere un buon compromesso sulle due caratteristiche richieste. Inoltre è necessario studiare le strategie di decisione (verifica delle ipotesi) per migliorare la detezione. La più semplice decisione è quella di considerare un veicolo presente appena viene riconosciuto; sicuramente in questo caso c'è il rischio che accadano alcune situazioni di falso allarme, se l'algoritmo di ricerca non è preciso.

Per risolvere questo problema è stato studiato ed implementato un algoritmo basato sull'estrazione e l'inseguimento delle caratteristiche (*feature*) KLT (Kanade Lucas Tomasi) [8], ovvero dei punti ad elevata curvatura (*corner o punti angolosi*).

Con questo metodo si cerca di estrarre esclusivamente il veicolo di fronte nel momento in cui appare. Partendo dalle informazioni delle due

linee della strada che “prossime al” veicolo, estratte dai moduli precedenti, si restringe il campo di ricerca ad una zona intorno al punto di intersezione di tali linee. Successivamente vengono estratte in tale zona le *feature* di Kanade Lucas Tomasi per più frame consecutivi considerando solo quelle che:

- sono stabili (cioè presenti per più frame consecutivi);
- presentano un movimento superiore ad una certa soglia prefissata.

Raggruppando tali caratteristiche si cerca la “forma” che appare di fronte. L’algoritmo di raggruppamento (*clustering*) che è stato implementato ed utilizzato è piuttosto semplice e si basa su due caratteristiche delle feature che sono state estratte:

- posizione;
- movimento.

Quindi dopo questa operazione sono raggruppate le feature che sono vicine fra loro e che hanno movimento simile. Molto importante è quindi la scelta di due soglie di vicinanza e di movimento. Da sottolineare inoltre che, essendo la zona di ricerca ristretta, al massimo può essere presente un solo veicolo. Pertanto si suppone che nella zona di interesse sia presente un veicolo se la forma trovata dall’algoritmo di raggruppamento delle caratteristiche KLT è composta da un numero sufficiente di feature e presenta dimensioni e proporzioni appropriate, ovvero che possono essere assimilate ad un veicolo che si trova in posizione frontale rispetto alla vettura. Un esempio di questa procedura è mostrato in *figura 5*.

In questo esempio la zona di ricerca è disegnata in verde, le *feature* che sono state raggruppate sono associate al colore rosso mentre le rimanenti sono in verde. Infine la forma dell’oggetto estratto è colorato in viola.

Quindi per utilizzare questo metodo per ogni frame della sequenza video devono essere

Figura 5 – Esempio estrazione veicoli tramite il metodo KLT

estratte e inseguite le caratteristiche KLT (*feature tracking*) che si presentano nella zona di interesse. Nel seguito si dà una breve descrizione di cosa sia intenda per *feature tracking*.

Per capire cosa sia il *feature tracking* è sufficiente dare una definizione dei due termini chiave: *feature* (caratteristica) e *tracking* (inseguimento). Una *feature* è una caratteristica particolare dell'immagine. Quali siano queste caratteristiche dipende dall'algoritmo utilizzato per l'inseguimento, ma in generale si richiede che la caratteristica sia facilmente individuabile in un certo intorno, ad esempio una porzione dell'immagine dall'intensità luminosa uniforme generalmente non contiene *feature*, perché è difficile distinguere al suo interno tra un punto e l'altro. Invece, un pixel con intensità diversa da quelli confinanti può essere identificato con maggiore semplicità nei frame successivi. Tuttavia è difficile osservare in maniera affidabile le proprietà dei singoli punti di un'immagine a causa del rumore introdotto dalla telecamera per cui solitamente una *feature* non è un punto, ma un'intera zona.

L'inseguimento consiste nel cercare la corrispondenza tra le *feature* trovate in due frame successivi, in modo da poter valutare lo spostamento avvenuto nell'intervallo tra questi due frame.

Il *feature tracking* dipende quindi dalla scelta di buone (ovvero in grado di essere discriminative) caratteristiche dell'immagine e dalla loro associazione per i frame successivi al fine di osservarne lo spostamento. Quello del *feature tracking* è stato un problema ampiamente studiato negli anni. Uno degli algoritmi proposti è quello sviluppato da Lucas, Kanade e Tomasi. L'algoritmo su cui si basa si fonda su un primo lavoro del 1981 di Lucas e Kanade, successivamente sviluppato e completato da Tomasi e Kanade. Una descrizione dettagliata dell'algoritmo si trova in un articolo di Shi e Tomasi [8].

2.5 *Inseguimento dei veicoli estratti*

Dopo aver estratto i veicoli nello spazio esterno è sicuramente utile tenerne traccia e conoscerne il loro spostamento nel tempo allo scopo di avere una completa analisi del contesto esterno e di tutti i suoi "protagonisti".

A questo scopo sono stati adattati due algoritmi di inseguimento sviluppati dal gruppo di ricerca in seguito ad attività precedenti: uno basato sul Meanshift [9] e l'altro basato sui corner e sulla GHT [10]. Al fine di migliorare le prestazioni di inseguimento e di ottenere algoritmi più robusti sono state esplorate le seguenti possibilità:

1. modifica e miglioramento dell'algoritmo basato su Meanshift;
2. adattamento alle esigenze del progetto dell'algoritmo sperimentale denominato MAPT [11].

3 Analisi video dello stato di attenzione del conducente

L'analisi di ciò che avviene all'interno del veicolo è una delle sorgenti di informazione di maggiore interesse per le applicazioni *automotive*. In particolare dall'analisi di contesto a bordo di veicoli si cercano di estrarre considerazioni sul comportamento del guidatore al fine di identificare possibili situazioni di pericolo (colpi di sonno, cambi di corsia senza aver verificato la presenza di altri veicoli che sopraggiungono, ecc.) o al fine di determinare se lo stile di guida rispetti le regole del codice della strada. Per ottenere ciò, i dati più significativi che possono essere estratti da una camera che monitorizza il guidatore sono la direzione dello sguardo, la posizione della faccia, la frequenza di battito di ciglia (*blink*) e il fatto che la bocca sia aperta o chiusa.

Nell'articolo presentato in [12] gli autori presentano un sistema che mira ad utilizzare algoritmi di visione artificiale al fine di sviluppare metodi di assistenza al guidatore per prevenire situazioni di pericolo. L'approccio utilizzato si distingue rispetto a quelli presenti in molti altri lavori dello stato dell'arte dove l'analisi è limitata a ciò che

accade all'esterno dell'auto (presenza pedoni, ostacoli, avvicinamento ad altri veicoli, ecc.). Infatti gli autori considerano tre componenti del sistema, ovvero il veicolo, l'ambiente esterno e il guidatore e costruiscono dei modelli e delle tecniche per analizzare i dati relativi a questi tre elementi al fine di favorire la guida sicura. Per quanto riguarda l'ambiente esterno gli elementi da estrarre dai flussi video acquisiti sono la presenza di altri veicoli o pedoni, la posizione rispetto alle linea della carreggiata, mentre dai dati presenti sul can bus del veicolo possono essere ottenute informazioni sull'angolo di sterzo delle ruote, la velocità, l'utilizzo del freno, ecc. Tuttavia al fine di definire se il comportamento del guidatore sia veramente rischioso e per predire le azioni che saranno compiute devono essere derivati lo stato fisico e mentale del guidatore e i suoi comportamenti. A questo fine possono essere utilizzati diversi tipi di algoritmi di video processing finalizzati a questo tipo di analisi.

Nella letteratura di applicazioni *automotive* si trovano diversi lavori che sono incentrati su algoritmi per l'analisi del guidatore. Molti approcci sono classificati in base alla tipologia sottolineando la condizione di utilizzo, le ipotesi e le prestazioni ottenibili. Alcuni metodi si basano sullo studio del colore caratteristico delle componenti della faccia (occhi , bocca) rispetto alla pelle, oppure analizzano il movimento di qualche *feature* rilevante nella zona degli occhi e della bocca. Altri lavori cercano di far fronte alle difficili condizioni ambientali delle applicazioni *automotive* (vale a dire frequenti e rilevanti modifiche nell'illuminazione) individuando gli occhi con l'ausilio di telecamere ad infrarossi (ad esempio si veda [12]). Questi approcci sono usualmente molto robusti e possono operare anche con bassa luminosità, ma il costo delle camere infrarossi è molto più alto rispetto a quello delle webcam tradizionali.

È possibile categorizzare gli algoritmi utilizzati in letteratura per l'analisi de guidatore in tre classi principali:

1. rilevamento e inseguimento del volto;
2. stima della posa della testa;
3. rilevazione degli occhi legata all'analisi della direzione dello sguardo e del battito di ciglia.

Nel seguito sono descritti alcuni degli approcci di maggiore interesse per queste tre categorie.

3.1 *Rilevamento e inseguimento di volti*

Il rilevamento e l'inseguimento di volti è uno dei problemi che ha raccolto maggiore interesse nella comunità scientifica della visione artificiale. Infatti, in molte applicazioni (p.e. ambient intelligence, videosorveglianza, ecc.) è richiesto di comprendere dove si trovino i volti delle persone all'interno del frame acquisito da una telecamera. Il problema del rilevamento del volto, ovvero identificare la posizione del volto all'interno del frame della sequenza video, è il passo fondamentale di tutti questi tipi di applicazioni. In letteratura sono stati proposti molti approcci che possono essere categorizzati (vedi la monografia sull'argomento in [13]) nelle seguenti classi:

- tecniche di confronto del modello (template matching) ;
- tecniche basate sull'analisi e la rappresentazione dei dati (feature invariant) ;
- tecniche basate sulla rappresentazione (appearance based);
- AdaBoost con Haar feature [14].

Gli algoritmi possono poi essere classificati anche in base al dominio in cui vengono effettuate le operazioni, se in quello spaziale o in quello compresso. Da sottolineare il fatto che in molte delle principali tecniche di face detection viene effettuato un prefiltraggio in cui vengono ricercate le zone di lavoro tramite metodi di segmentazione della pelle basti sull'analisi del colore. Nel seguito viene descritto più in dettaglio il metodo utilizzato nel nostro lavoro ovvero l'AdaBoost con Haar feature.

Il metodo proposto da Viola e Jones [14] è stato largamente utilizzato negli ultimi anni per applicazioni di rilevamento di oggetti in generale e di facce in particolare data la sua velocità che ne permette l'utilizzo in tempo reale e data l'invarianza alla scala.

L'algoritmo si basa sull'utilizzo di *feature* tipo-Haar calcolate in modo veloce tramite il metodo dell'*Integral Image*. L'algoritmo dell'AdaBoost è utilizzato per selezionare un sottoinsieme di caratteristiche da utilizzare per effettuare il rilevamento. Infine i classificatori in grado di distinguere gli oggetti in base alle caratteristiche selezionate sono combinati in una struttura a cascata (cascade classifier) che permette di eliminare nei primi passi le zone di sfondo in modo da concentrarsi sulle zone con maggior probabilità di presenza di oggetti e, quindi, di limitare i tempi di elaborazione.

Il metodo dell'*Integral Image* permette di ridurre molto il peso computazionale del calcolo delle caratteristiche tipo-Haar che sono impiegate in questo approccio. In particolare nell'algoritmo di Viola-Jones sono scelte delle caratteristiche rettangolari in grado di rappresentare strutture geometriche dell'immagine ispirate dalle *feature* di Haar. Queste sono caratterizzate da un numero variabile, da due a quattro, di rettangoli di dimensione uguale affiancate una all'altra. A differenza delle *feature* di Haar quelle utilizzate rappresentano un insieme sovra-completo in quanto sono considerate anche quelle linearmente dipendenti. Nel sistema in esame dove la risoluzione di base del rilevatore (ovvero la dimensione della sottofinestra dell'immagine dove effettuare la ricerca) è 24x24 pixel, il numero di *feature* è 45.396.

Nel nostro lavoro utilizziamo questo approccio per determinare la posizione iniziale del viso, degli occhi, del naso e della bocca. Per questo motivo per ogni componente della faccia è impiegato un rilevatore di tipo Viola-Jones. Quindi abbiamo bisogno di cinque rilevatori per identificare ogni componente della faccia, questo fa aumentare la complessità temporale dell'approccio. Per ridurlo la rilevazione *bottom-up* della bocca è sostituita da un'inizializzazione top-down attraverso vincoli geometrici.

3.2 *Stima della posa della testa*

Nell'articolo di Murphy-Chutorian e Trivedi [15] è proposto una completa descrizione dei metodi presenti nello stato dell'arte utilizzati per la stima della posa della testa. Questi approcci permettono di ottenere informazioni rilevanti riguardanti la direzione dello sguardo senza necessità di dover individuare gli occhi e le pupille, operazione questa spesso molto complessa.

I metodi di stima della posa della testa possono essere classificati nelle seguenti categorie:

- metodi basati sull'apparenza della sagoma (Appearance Template Methods): si confronta una nuova immagine della testa con un insieme di esempi ognuno dei quali etichettati con un valore di posa al fine di trovare quello più simile;
- metodi basati sulla griglia di detettori (Detector Array Methods): si addestrano una serie di detettori di teste ognuno dei quali specifico per un valore discreto di posa;
- metodi basati sulla regressione non-lineare (Non-linear Regression Methods): si utilizzano strumenti di regressione non lineare per mappare le immagini o le relative feature a dei valori di posa della testa;
- metodi basati sugli iperspazi inclusi (Manifold Embedding Methods): si cercano i sottospazi di dimensione minore che modellano la variazione della posa della testa. Nuove immagini possono essere inclusi in questi iperspazi e utilizzati per template matching o regression;
- modelli flessibili (Flexible Models): si adatta un modello non rigido alla struttura della faccia di ogni individuo nel piano immagine. La stima della posa si ottiene tramite un confronto a livello di feature o dai parametri del modello;
- metodi geometrici (Geometric Methods): Si utilizzano le posizioni di occhi, bocca, naso, per determinare la posa a partire dalla loro configurazione;
- metodi basati sull'inseguimento (Tracking Methods): si valuta la posa della testa in base ai movimenti osservati tra istanti successivi;
- metodi Ibridi (Hybrid Methods): si combinano due o più dei metodi precedenti per superare le limitazioni inerenti in ciascun singolo approccio.

In questo lavoro si è deciso di usare un approccio *tracking based* per calcolare la direzione di vista,

perché esso fornisce un buon rapporto tra precisione e complessità computazionale. I metodi di tracking operano seguendo il movimento relativo della testa tra frame consecutivi durante una sequenza video [12]. Dopo una fase di inizializzazione, effettuata tramite la ricerca della faccia all'interno di un frame, con l'algoritmo AdaBoost con caratteristiche di Haar, viene utilizzato un algoritmo di tracking per localizzare la posizione della faccia nei frame successivi e la posizione relativa di ogni componente che costituisce il viso come naso, bocca e occhi. Quando la stima della posizione della faccia è ottenuta dentro un frame video, vengono calcolati l'angolo di vista e le altre informazioni di interesse.

Per ogni componente della faccia viene utilizzata un'istanza dell'algoritmo di inseguimento basato sulle *feature* di Kanade-Lucas-Tomani (KLT) [8]. Questo algoritmo è basato su un modello di cambiamento di immagine affine, che usa un metodo specifico per la selezione delle *feature* e una tecnica di monitoraggio delle *feature* durante l'inseguimento. La selezione è studiata appositamente per massimizzare la qualità dell'inseguimento ed è quindi ottimale per costruzione, al contrario di misure basate sulle caratteristiche locali (*texture*). Il monitoraggio è computazionalmente poco oneroso e aiuta a discriminare tra *feature* adatte o meno sulla base di misure di diseguaglianza basate sul movimento affine legate al cambiamento del modello dell'immagine sottostante. Grazie a questa scelta si ottiene una precisione sulla stima della posizione degli occhi e la possibilità di riuscire a valutarla anche quando si verifica una rilevante rotazione della testa. Inoltre l'uso delle caratteristiche KLT permette di avere informazioni aggiuntive sui cambiamenti dell'orientazione del volto, che possono essere utilizzati per determinarne la posa come descritto più dettagliatamente in seguito.

3.2.1 Calcolo angolo di vista

L'angolo di vista è una delle più importanti informazioni per valutare lo stato del guidatore. Può essere suddiviso in angolo di *yaw* (angolo di rotazione rispetto al piano orizzontale), *roll* (angolo di rollio ovvero di rotazione longitudinale al movimento) e *pitch* (angolo di rotazione verticale). In questo lavoro è proposto un metodo che mira ad ottenere un soddisfacente compromesso tra velocità di esecuzione e precisione della stima. Per questo fine lo sguardo è stato caratterizzato da n possibili valori discreti. Per stimare l'angolo di *yaw* vengono utilizzate le informazioni sulla posizione delle *feature* KLT, lo spostamento degli occhi durante la fase di inseguimento e il triangolo formato tra i due occhi e il naso. Un primo valore ε dell'angolo di *yaw* può essere calcolato dalle feature KLT nel modo seguente:

$$\varepsilon = \frac{\mathrm{var}\,(R_x)}{\mathrm{var}\,(L_x)}$$

dove $\mathrm{var}(R_x)$ è la varianza sull'asse delle x dell'occhio destro e $\mathrm{var}(L_x)$ è la varianza sull'asse delle x dell'occhio sinistro.

Lo spostamento degli occhi durante la fase di inseguimento associato allo studio della posizione relativa tra gli occhi e il naso garantiscono un miglioramento della stima dell'angolo di *yaw*. Naturalmente se non c'è una conoscenza sulla posizione del naso, il triangolo costruito tra naso e occhi non può essere creato quindi si utilizzano solo le informazioni delle *feature* KLT e dello spostamento. In questo modo l'angolo di *yaw* è meno preciso, ma ugualmente è possibile una stima.

L'angolo di *roll* è calcolato grazie all'analisi della posizione del centro degli occhi con il centro del viso, più la differenza tra il valore della y tra gli occhi. Con questi due parametri è possibile giungere ad una stima accettabile dell'angolo.

3.2.2 Risultati

I video utilizzati per testare il metodo sono stati tutti ottenuti da una semplice webcam a 320x240 di risoluzione. Per i video di interno sono state utilizzate sequenze di tre persone diverse che girano la testa in diverse direzioni. Per i video a bordo la stessa webcam è stata installata su una macchina ed è stata utilizzata per analiz-

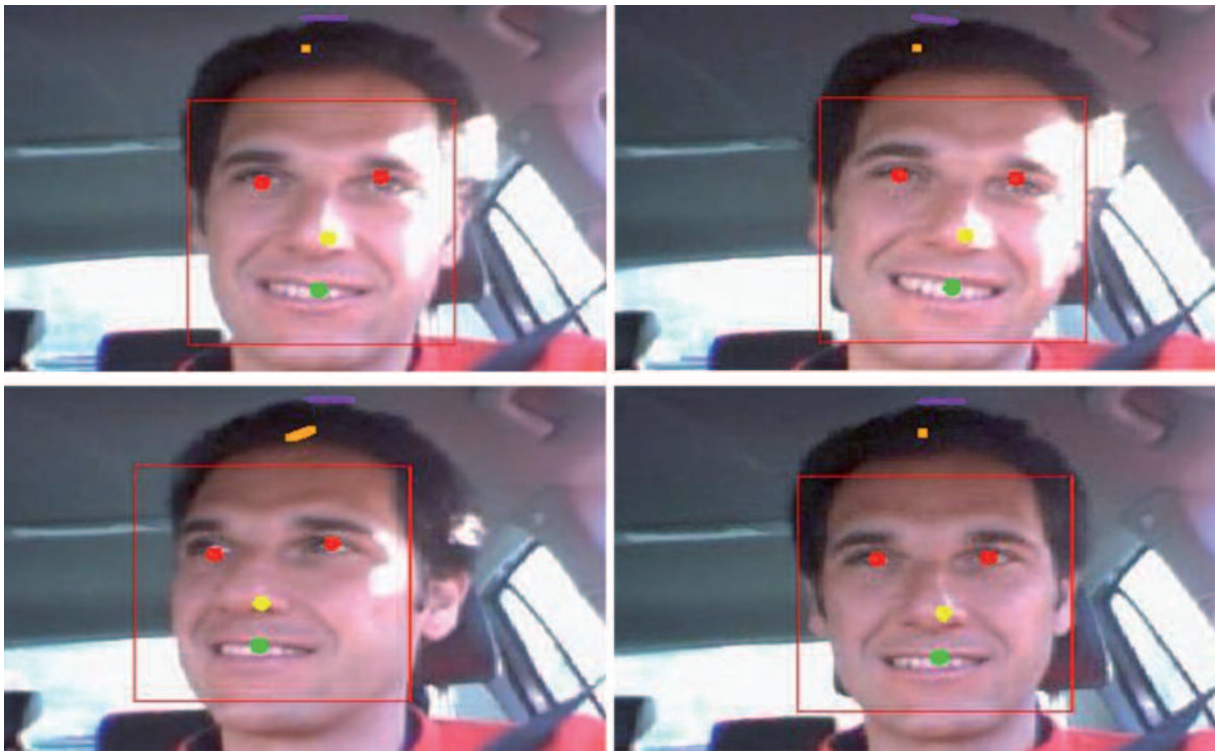
zare due guidatori, diversi dalle persone analizzate nei video di interno, durante un viaggio di mezz'ora ciascuno. Naturalmente ci sono alcune differenze tra le due tipologie di video che sono state prese in considerazione. Il primo tipo di esperimenti non è soggetto a cambi di luminosità che rendono più complessa la fase di inseguimento della seconda tipologia. Per questo, nel caso interno, l'inseguimento è robusto a meno di grosse occlusioni o di rotazioni molto significative, invece nel caso della macchina, il *traking* necessita di essere inizializzato più frequentemente proprio a causa dei cambi di luminosità. Utilizzando una metodologia di re-inizializzazione totale o parziale si riesce a mitigare questo problema.

La stima degli angoli di *roll* e di *yaw* è abbastanza buona. Si può verificare che la maggior parte degli errori si osservano quando il movimento della testa è determinato dalla modifica di due o più angoli di rotazione.

In *figura 6* è mostrato un esempio di video in macchina; i risultati non sono molto differenti rispetto a quelli del caso indoor, in quanto il vero problema è il cambio di luminosità che ci obbliga a re-inizializzare il modulo di inseguimento. In questo modo possiamo seguire la faccia del guidatore e dare una stima del suo angolo di vista senza molte differenze rispetto al caso interno. In *tabella 3* sono analizzati i risultati sperimentali ottenuti sia nel caso interno e sia nel caso esterno di rispettivamente 10:06 minuti e 45:49 minuti a 25 frame al secondo.

La percentuale di frame con errori sono state ricavate confrontando i risultati dell'algoritmo con le osservazioni. I video acquisiti a bordo del veicolo hanno complessivamente più errori rispetto a quelli di interno, in particolare l'errore è maggiore nella fase di rilevamento e inseguimento, piuttosto che nel calcolo dell'angolo di vista. Questo è facilmente spiegabile, gli errori sono dovuti proprio al cambio di luminosità che ci porta a re-inizializzare l'algoritmo, o che porta errore nel *tracking*, in particolare si può notare che nella maggior parte dei casi solo una componente sbaglia. L'angolo di *roll* soffre maggiormente degli errori sul *tracking*, per questo motivo le prestazioni sono migliori nei casi di interno, mentre l'angolo di *yaw* è più robusto, ma presenta più errori nel

Figura 6 – Esperimenti a bordo auto



	Video in Interno	Video in Esterno
No Rilevamento Faccia	0.5%	3.1%
No Rilevamento Naso/Bocca	0.7%	0.8%
Errore Inseguimento Occhi/Naso	5.5%	11.4%
Errore Inseguimento Bocca	5%	7.8%
Errore angolo di Yaw	6.8%	4.7%
Errore angeolo di Roll	12%	16.1%

Tabella 3 – Percentuale di frame con errori

caso indoor in quanto sono presenti rotazioni della testa più consistenti.

4 Integrazione dati per comprensione di relazioni causa-effetto

A partire dai dati contestuali derivati dall'analisi dei flussi video riguardanti l'interno e l'esterno dell'autovettura è di notevole interesse realizzare algoritmi in grado comprendere le relazioni tra situazione del traffico e comportamento del guidatore. Un esempio di questo tipo di applicazioni è presentato da McCall [16] che descrive la possibilità di combinare informazioni riguardanti l'angolo di vista del conducente, la posizione del veicolo rispetto alle carreggiate e tutti i dati provenienti dal CAN bus della vettura, ovvero velocità, angolo di sterzo, pressione sul pedale del freno,, al fine di predire le intenzioni di cambio di carreggiata. Lo scopo di questa applicazione è quello di inviare un allarme preventivo nel caso in cui si rilevi l'intento di effettuare un cambio di carreggiata in presenza di una macchina che sta sopraggiungendo. In questo contesto, inoltre, l'apprendimento dello stile di guida del singolo guidatore permette di migliorare la capacità di rilevazione di tali comportamenti potenzialmente rischiosi tramite la possibilità di adattarsi al conducente e la conseguente personalizzazione di tale funzionalità. In un altro lavoro [17], gli autori mostrano come questi concetti di apprendimento possono essere utili, al fine di visualizzare messaggi personalizzati per il rispetto dei limiti di velocità. L'esperimento condotto mostra come alcune modalità di visualizzazione di messaggi di superamento del limite di velocità siano più efficaci rispetto ad altre. In questo caso quindi sono elaborate congiuntamente le informazioni di posizione dell'automobile ottenute con il GPS, tramite cui viene derivato il limite di velocità nel relativo tratto di strada e quelle di velocità dell'automobile ottenute dal CAN bus.

Pertanto un possibile sviluppo della ricerca presentata può riguardare l'introduzione di moduli di elaborazione che realizzino funzionalità di integrazione dei dati relativi al traffico e quelli riguardanti il conducente, in modo da determinare le relazioni causa-effetto tra essi. Per esempio può essere utile apprendere il comportamento del guidatore in presenza di traffico intenso per determinarne la pericolosità. Inoltre anche lo stato di attenzione del conducente può essere associato alla presenza di altri veicoli o di prossimità di curve al fine di comunicare messaggi di attenzione. Tramite questa analisi sarebbe possibile, inoltre, determinare le reazioni del guidatore ai diversi messaggi di allerta o di informazione, in modo da ottimizzare l'efficacia dell'interazione tra il sistema e l'utente.

5 Conclusioni

In questo articolo è stata descritta l'attività di ricerca svolta all'interno del laboratorio congiunto Telecom Italia Lab – Università di Genova. Il gruppo di ricerca ISIP40, in collaboarazione con la divisione di TILab "Research & Trends", si è

occupato di realizzare moduli software per l'analisi di contesto in ambito automobilistico. Sono stati realizzati algoritmi in grado di fornire elementi utili alla comprensione della situazione di guida sia riferiti al veicolo, cioè posizione rispetto alla carreggiata, cambi di corsia e presenza di altri veicoli in prossimità, sia riguardanti lo stato di attenzione del guidatore, ovvero direzione dello sguardo. Questi dati possono essere ottenuti in tempo reale e inviati, tramite una scheda wireless UMTS, ad un server centrale, che si occupa del loro utilizzo. Le possibili applicazioni, per cui queste informazioni possono risultare di interesse, sono molteplici, dalla prevenzione di incidenti tramite segnali di allerta in caso di situazioni di pericolo, alle funzionalità tipo "scatola nera" per la registrazione di video in caso di incidente, ai servizi per il miglioramento di confort di guida sfruttando l'analisi congiunta dei dati provenienti da più auto in modo da segnalare agli utenti lo stato di traffico.

BIBLIOGRAFIA

[1] http://cvrr.ucsd.edu/

[2] http://www.roadscan.co.uk/

[3] Canny, J.,"*A Computational Approach To Edge Detection*", IEEE Trans. Pattern Analysis and Machine Intelligence, 8:679-714, 1986.

[4] P.V.C. Hough, "*Machine Analysis of Bubble Chamber Pictures*", Proc. Int. Conf. High Energy Accelerators and Instrumentation, 1959.

[5] Duda, R. O. and P. E. Hart, "*Use of the Hough Transformation to Detect Lines and Curves in Pictures*". Comm. ACM, Vol. 15, pp. 11–15 (January, 1972).

[6] Shapiro, Linda and Stockman, George. "*Computer Vision*", Prentice-Hall, Inc. 2001

[7] John Aldrich. Earliest "*Uses of Symbols in Probability and Statistics*". Electronic document, retrieved March 20, 2005.

[8] J. Shi and C. Tomasi, "*Good features to track*". Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on, pp. 593–600, 1994.

[9] D. Comaniciu, V. Ramesh and P. Meer, "Real-time tracking of non-rigid objects using Mean Shif", Proceedings of 2000 IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head, SC, volume II, June 2000.

[10] M. Asadi, A. Beoldo, A. Dore, and C.S. Regazzoni, "*Tracking by Using Dynamic Shape Model Learning in the Presence of Occlusion*", IEEE AVSS, London, UK, 5-7 September 2007.

[11] A.Dore, A. Beoldo, and C.S. Regazzoni, "*Multiple Cue Adaptive Tracking of Deformable Objects with Particle Filter*", IEEE International Conference on Image Processing, ICIP 2008, San Diego, CA, USA, 12 - 15 October 2008.

[12] M. Trivedi, T. Gandhi, and J. McCall, "*Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety*", IEEE Transactions on Intelligent Transportation Systems, vol. 8, no. 1, pp. 108–120, 2007.

[13] M. Yang, D. J. Kriegman, and N. Ahuja, "*Detecting faces in images: a survey*", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 24, pp. 34–58, 2002.

[14] P. Viola and M. Jones, "*Robust real-time object detection*", International Journal of Computer Vision, 2002.

[15] E. Murphy-Chutorian and M. M. Trivedi, "*Head pose estimation in computer vision: A survey*", IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 4, pp. 607–626, 2009.

[16] Joel McCall, David Wipf, Mohan M. Trivedi, Bhaskar Rao, "*Lane Change Intent Analysis Using Robust Operators and Sparse Bayesian Learning*", IEEE Transactions on Intelligent Transportation Systems, Sept 2007

[17] Shinko Y. Cheng, Anup Doshi, Mohan M. Trivedi, "*Active Heads-up Display based Speed Compliance Aid for Driver Assistance: A Novel Interfac e and Comparative Experimental Studies*", IEEE Intelligent Vehicles Symposium, June 2007.

carlo@dibe.unige.it
dore@dibe.unige.it
pasini@ginevra.dibe.unige.it

AUTORI

Alessio Dore

laureato in Ingegneria delle Telecomunicazioni presso l'Università di Genova, ha svolto parte della tesi di laurea presso la Kingston University – London, UK. Nel 2006 ha collaborato col gruppo di ricerca video and Signal Processing for Telecomunications (ISIP40) nel Dipartimento di Ingegneria Biofisica ed Elettronica, Università di Genova. Dal Gennaio 2007 è studente di dottorato presso il Dipartimento di Ingegneria Biofisica ed Elettronica dell'Università di Genova. Nel 2008 è stato per due mesi visiting student presso l'University of Illinois at Chicaco (UIC), Chicago, USA. I campi di ricerca in cui è coinvolto sono: metodi Bayesiani per visual tracking, algoritmi bio-inspired per l'apprendimento di interazioni e sorveglianza cognitiva. Ha pubblicato 11 articoli su conferenze internazionali e 1 capitolo di libro ■

Francesco Pasini

ha conseguito la Laurea Triennale in Informatica e la Laurea Specialistica in Informatica curriculum Grafica e Immagini presso l'Università di Genova. Nel 2008 ha lavorato per la MEEO s.r.l (Meteorological and Environmental Earth Observation) su immagini satellitari multispettrali. Dal novembre del 2008 all'aprile del 2009 ha collaborato col gruppo di ricerca video and Signal Processing for Telecomunications (ISIP40) nel Dipartimento di Ingegneria Biofisica ed Elettronica, Università di Genova. Dal maggio del 2009 è assegnista presso il Dipartimento di Ingegneria Biofisica ed Elettronica dell' Università di Genova. I campi di ricerca in cui è coinvolto sono: computer vision, sistemi di intelligenza d'ambiente, tracking ■

Carlo Regazzoni

ingegnere elettronico è Dottore di Ricerca in Telecomunicazioni dell'Università di Genova. Dal 1990 è responsabile dell'area Industrial Signal and Image Processing (ISIP) del gruppo di ricerca in Elaborazione dei Segnali e Telecomunicazioni (SP&T) nel Dipartimento di Ingegneria Biofisica ed Elettronica (DIBE). Dal 1999 è responsabile del Gruppo SP&T ed è Professore Ordinario dal 2008. È stato organizzatore e co-organizzatore di diverse conferenze e workshop internazionali. È stato Guest Editor di diversi Special Issue su argomenti collegati all'elaborazione e interpretazione dati per la video sorveglianza (Real Time Imaging 2001, Proccedings of the IEEE 2001). Attualmente è Associate Editor delle seguenti riviste internazionali, IEEE Signal Processing Letters, IEEE Transactions on Mobile e International Journal on Image and Graphics. È stato responsabile di progetti europei di ricerca e sviluppo e in molti contratti di ricerca con industrie italiane. È stato revisore per conto di numerose riviste internazionali e valutatore per svariati programmi di ricerca europei. Ha vinto il Best paper award 2002 IEEE Transactions on Vehicular Technologies sezione Vehicular Electronics ed è autore e co-autore di più di 70 articoli su riviste internazionali e 250 articoli presentati a conferenze internazionali. Gli argomenti di ricerca di cui si occupa riguardano applicazioni di video sorveglianza e ambienti interattivi intelligenti ■

Cloud Computing: stato dell'arte e opportunità

INNOVAZIONE

Antonio Manzalini, Corrado Moiso, Elisabetta Morandin

Il "Cloud Computing" è un insieme di tecnologie informatiche che permettono di fornire come servizi "on-demand" l'accesso e l'utilizzo di risorse di elaborazione, memoria ed applicative, distribuite e virtualizzate in rete. Tale paradigma abilita nuovi modelli di distribuzione del software e di offerta di applicativi, permettendo di rendere disponibili come servizi, accessibili tramite interfacce web, applicazioni e piattaforme informatiche. Attraverso l'analisi dello stato dell'arte e delle caratteristiche del Cloud Computing, l'articolo si propone di identificare le potenziali opportunità che questo paradigma potrebbe offrire agli Operatori. L'articolo descrive inoltre alcuni scenari applicativi e l'attuale posizionamento di Telecom Italia. Infine si presentano alcuni spunti di riflessione ed una visione tecnologica sulle future evoluzioni del Cloud Computing.

1 Introduzione

Il termine "Cloud Computing" è ormai familiare a molti utilizzatori di servizi Internet: l'espressione sta ad indicare un insieme di tecnologie informatiche che permettono l'utilizzo di risorse, ad esempio server o storage, distribuite e virtualizzate in rete, al fine di fornire servizi all'utente finale. Tipicamente l'utente finale accede a tali servizi mediante un browser. L'interesse per questo paradigma sta crescendo giorno per giorno, come evidenzia Google Trends (*figura 1*). Del resto, come fa notare il Cloud Computing Journal [13], sono numerosi i fattori che sostengono questo interesse da parte delle aziende, tra i quali vi è la necessità di disporre di sempre maggiori risorse computazionali fornite on-demand.

Nonostante la semplice definizione, il Cloud Computing implica diverse realtà tecnologiche,

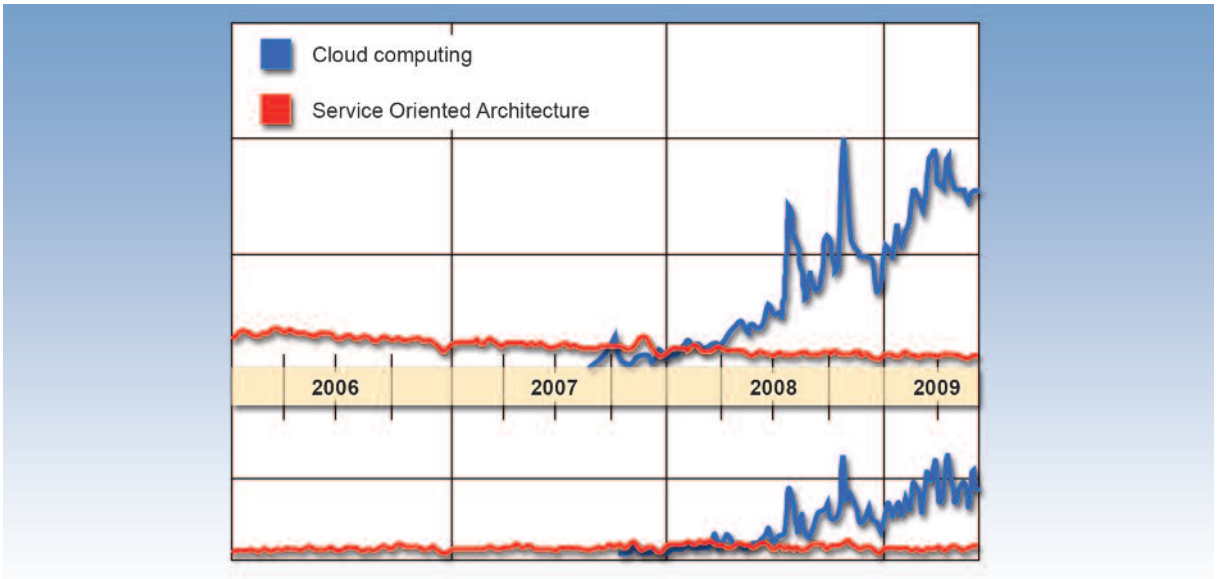


Figura 1 – Confronto delle ricerche in rete dei termini "Cloud Computing" e "Service Oriented Architecture"[1]

economiche e di business. Un esempio della complessità del Cloud Computing e degli interessi che lo circondano è la vicenda della redazione dell'"Open Cloud Manifesto" [9]. Questo doveva essere un documento aperto che indicasse alcuni principi di riferimento delle soluzioni di Cloud Computing valide per tutti i fornitori di funzionalità e in particolare che indirizzasse le problematiche tecnologiche di sicurezza, interoperabilità, portabilità delle Cloud attraverso l'impiego di standard aperti. La stesura del documento è stata molto travagliata e alcune aziende fra cui Google e Microsoft non l'hanno ratificato. Microsoft ha criticato il modo stesso in cui il documento era stato segretamente redatto ben lontano ad una filosofia aperta e partecipativa. Amazon ha invece sostenuto che i Web Services sviluppati sono già utilizzabili via multi-piattaforma, multi-linguaggi di programmazione, multi-sistemi operativi [19].

Tale conflitto dimostra che saranno in molti a competere su questo terreno: Google, con la sua infrastruttura parallela e distribuita di computers low-cost ed il cui "cuore pulsante" è rappresentato da MapReduce; Yahoo!, che sulla piattaforma Apache Hadoop ha lanciato la Yahoo! Search Webmap, utilizzando un Cloud di 10.000 CPUs; Microsoft con i Cloud Computing tools di Azure; IBM, che ha lanciato Blue Cloud; Amazon, con i suoi già noti servizi EC2 e AWS, Sun Microsystems, con Network.com; oltre ad una serie di aziende più piccole, tra le quali spicca 3Tera, con il suo AppLogic. Ma ci sono anche Dell, HP, Oracle, ed EMC, la quale ha comprato PI Corporation, una start-up che sviluppa software per il Cloud Computing e moltissimi altri *competitor* sicuramente arriveranno.

Dal punto di vista tecnologico le fondamenta del Cloud Computing sono radicate nell'evoluzione e nella storia dell'elaborazione distribuita. Il Cloud Computing può essere, allora, visto come una soluzione che nasce da tecnologie come "Grid Computing" [10], "Autonomic Computing" [1] ed "Software as a Service"[2], tutto integrato e reso fruibile attraverso le tecnologie proprie del Web 2.0, a partire dalla centralità del browser e dall'approccio "web as a platform".

Infatti, un "Cloud computazionale" è un sistema distribuito complesso che deve essere regolato ed ottimizzato per dare le prestazioni elaborative desiderate: è necessario identificare e allocare varie funzionalità distribuite in rete, allocarle a di-

1 http://www.google.com/trends?q=cloud+computing

2 http://en.wikipedia.org/wiki/Software_as_a_service

versi sistemi di calcolo, integrare le funzioni in una logica di servizio e rappresentare in maniera significativa i risultati della computazione verso il cliente.

Nel Cloud Computing confluiscono più filoni tecnologici: soluzioni per l'esposizione e la composizione di funzionalità distribuite, quali le Application Programming Interface (APIs) e relativi meccanismi per le chiamate remote (basate su SOAP, REST, CORBA, o altri protocolli), linguaggi di programmazione, con prevalenza di quelli di alto livello come PHP e Python, per permettere una programmazione agile e aperta ad un vasto repertorio di programmatori, soluzioni per la rappresentazione e la condivisione dei dati (ad esempio, mediante XML o JSON) e meccanismi per il loro reperimento e memorizzazione (ad esempio database come MySQL), meccanismi di monitoraggio e di accounting delle risorse utilizzate.

Tra le conoscenze informatiche più interessanti che contribuiscono alla realizzazione di Cloud computazionali si possono citare:

- la tecnologia sviluppata per i web e gli application server: si possono riconoscere in un Cloud computazionale le diverse funzioni attribuite ad un application server, quali la presentation, la business e la data logic, con la differenza che esse sono ora distribuite su sistemi cooperanti;
- le soluzioni innovative per gestione di grandi moli di dati: il Cloud Computing utilizza le funzioni di successo dei search engine per identificare, indicizzare e memorizzare le informazioni; queste sono in fase di ulteriore sviluppo per integrare capacità semantiche e per favorire il data mining intelligente;
- le Rich Internet Application (RIA) che hanno portato in dote le tecnologie per permettere di accedere ai servizi informatici mediante interfacce utente molto ricche ed accattivanti usufruite attraverso browser; esse si sono recentemente arricchite anche con la possibilità di operare anche quando l'utente non è connesso, come proposto, ad esempio, da Google Gears o da Adobe Air;
- i principi delle Service Oriented Architecture che hanno permesso di organizzare le componenti computazionali distribuite, in modo tale da poterle identificare, allocare, orchestrare e ottimizzare per fornire funzionalità agli utenti;
- l'autonomic computing definito da IBM come il sistema nervoso di un ambiente distribuito: permette di arricchire i server e le funzionalità con la capacità di autogestirsi, in modo da limitare il bisogno di intervento umano nelle aree proprie della gestione (Fault, Configuration, Accouting, Performance e Security).

Queste tecnologie, proprie dei sistemi altamente distribuiti, possono essere affiancate con soluzioni specifiche per gestire i clienti e fornire sicurezza delle transazioni, quali, ad esempio:

- Identity Management, per la gestione dell'identità e dei profili d'utente, ad esempio mediante l'adozione di soluzioni quali OpenId, Liberty Alliance o CardSpace;
- accounting delle risorse utilizzate, anche considerando prestazioni per gestire micro pagamenti: esempi sono forniti da Amazon per l'accesso al Simple DB, oppure da Gomez [3] per contabilizzare transazioni su macchine degli utenti valorizzate a 0,0005 $ per minuto di utilizzo di CPU;
- sicurezza della trasmissione e accesso ai dati, come meccanismi di autenticazione sicura, crittazione e di single sign on.

Un aspetto rilevante per l'Operatore è che il Cloud Computing, per funzionare adeguatamente, ha bisogno di connettività di rete, sia tra i nodi del Cloud, sia tra i client (degli utenti) e i server. Questa connettività prevedrà dei requisiti di qualità (tra cui transazioni, sicurezza, ritardo contenuto, VPN) che potrebbero essere la base per la fornitura di servizi a valore aggiunto da parte dell'Operatore. Inoltre, se i servizi saranno caratterizzati da un'elevata multimedialità, aumenteranno considerevolmente anche i requisiti di banda necessaria.

Dal punto di vista economico il Cloud Computing trae la sua giustificazione dalla possibilità of-

3 http://www.gomezpeerzone.com/ViewCurrentRates.aspx

ferta a molte aziende di non investire sulle infrastrutture informatiche, ma di "comprare on-demand" da un'azienda che ha realizzato un'infrastruttura di Cloud Computing potenza di calcolo, storage e servizi informatici: in questo modo si converte CAPEX in OPEX. In un recente articolo di ZDNet si legge lo slogan "start your company with a credit card and a cloud" [14], per sottolineare i potenziali risvolti economici del Cloud Computing. I servizi e le infrastrutture informatiche diventano quindi delle "utility" da pagare a consumo, ribaltando così un modello di business consolidato negli anni e, secondo alcuni, riproponendo un modello simile a quello dei mainframe. Un vantaggio di tale approccio è innegabilmente quello di poter disporre (lato utente) di un sistema sempre aggiornato dal punto di vista tecnologico e di poter rilassare la necessità di rinnovare il parco macchine interno, per essere al passo con la tecnologia e le esigenze di calcolo.

Si torna quindi a parlare di "thin client", ossia macchine con limitate capacità e prestazioni, ma in grado di accedere efficacemente ad Internet per usufruire dei servizi in rete.

Ovviamente c'è qualcuno che argomenta i possibili "lati oscuri" del Cloud Computing. Alcuni fautori dei sistemi aperti (ad esempio Stallman [11]) vedono, però, nel Cloud Computing un ritorno al passato verso ambienti chiusi e proprietari, che a lungo termine porteranno al fenomeno del lock-in degli utenti su una specifica piattaforma proprietaria: possibili rischi riguardano, ad esempio, la perdita di controllo sui dati sensibili aziendali memorizzati nell'infrastruttura di un fornitore di Cloud Computing, la concentrazione dei servizi di Cloud Computing nelle mani di pochi grossi player, i disservizi, in alcuni casi già sperimentati dai grossi utenti di alcune attuali piattaforme, fatti che hanno contribuito a creare un certo scetticismo sull'affidabilità delle attuali soluzioni.

Dal punto di vista dei modelli di business la situazione è estremamente variegata. Le offerte sono molto ampie e al momento non si è palesato un chiaro modello di business. Gli attori di questo mondo tendono a coprire delle nicchie. Ad esempio Amazon offre un servizio di database in rete, dove la memorizzazione dei dati è gratuita, mentre il recupero dei record è a pagamento. StrikeIron invece ha creato un ambiente per l'hosting delle API, mettendo insieme un ecosistema composto da fornitori di tecnologia di base (gli application server), i fornitori di servizi ed i programmatori. Un tentativo interessante è quello di utilizzare anche tecniche di social network, per fornire servizi di Cloud Computing: ad esempio Cucku, Crashplan e Zoomgo, seppur in modalità diverse, propongono un servizio che oltre allo storage centralizzato, utilizza la memorizzazione distribuita delle informazioni e dei file degli utenti su macchine e dispositivi di amici o conoscenti. In ogni caso queste tecnologie e approcci sono e saranno sempre più rilevanti per definire uno strato di servizi componibili e programmabili. È necessario, pertanto, identificare le funzionalità di valore e le strategie che un Operatore può perseguire per essere parte attiva in questi nuovi modelli di business.

In *figura 2* sono rappresentati la varietà delle soluzioni e lo spettro dell'applicabilità del Cloud Computing.

2 Software/Platform/Infrastructure as a Service: cosa sono?

Il diffondersi delle tecnologie di virtualizzazione e di Cloud Computing, insieme ad una crescente esigenza di abbattere i costi di gestione delle applicazioni e dei sistemi nel mondo IT, hanno portato alla diffusione di politiche di delivery dei servizi IT in modalità on-demand, così da permettere la diffusione di nuovi modelli, o l'estensione di quelli esistenti, relativi alla distribuzione del software ed all'accesso degli applicativi software. Ecco, quindi, che si parla di software, piattaforme o infrastrutture che sono resi disponibili come servizi [17]. Questi servizi possono essere considerati come le componenti di base per lo sviluppo del Cloud Computing.

Andando nel dettaglio per ciascuna tipologia di servizio:

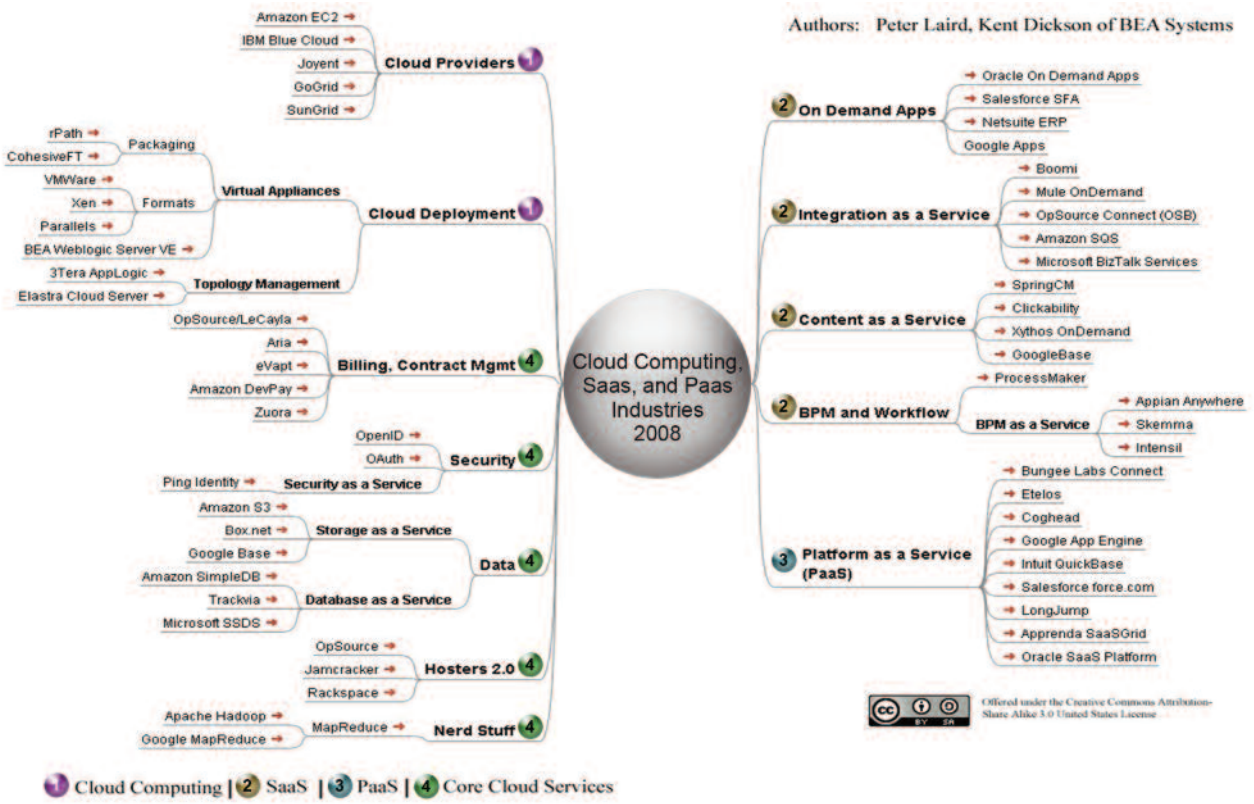


Figura 2 – La mappa dell'industria del Cloud Computing [4]

- **Software as a Service** (SaaS): il termine si riferisce alla fornitura di un applicativo in modalità centralizzata e accessibile via Web. La caratteristica comune di questa tipologia di servizio è quella di fornire un applicativo che sia condiviso tra tutti i clienti a meno di funzionalità opzionali ed eventualmente abilitate in base ad una configurazione di amministrazione. Nella maggioranza dei casi, una singola istanza dell'applicativo gestisce clienti diversi, pur garantendo la separazione logica dei dati di ciascun cliente. Tramite un'architettura hardware virtualizzata o un applicativo virtualizzato (virtual appliance), il servizio applicativo può essere reso disponibile in modalità multi istanza, secondo cui ogni istanza applicativa è dedicata ad un singolo cliente.

Normalmente, le politiche di pagamento di questo servizio applicativo dipendono dal numero di utenti abilitati all'accesso, con eventuali costi extra in base alle funzionalità applicative opzionali abilitate, alla capacità di banda utilizzata, o alle risorse aggiuntive di memoria utilizzate (extra-storage).
Spesso i servizi applicativi offrono un'interfaccia via Web Services, che permette l'integrazione e l'interoperabilità con altri applicativi e che quindi consente di sviluppare nuove applicazioni, seguendo i principi SOA (Service Oriented Architecture).
Tipici servizi che sono già resi disponibili al mercato, secondo la modalità SaaS, sono i servizi applicativi di base come CRM, ERP, back-up dei dati centralizzati, documentali, mail e archiviazione.

- **Platform as a Service** (PaaS): il termine si riferisce alla modalità tramite cui una piattaforma rende disponibile, via Web, tutti quegli strumenti e prodotti utilizzati nello sviluppo e delivery di nuovi servizi applicativi. Ad esem-

4 http://saaslink.googlepages.com/saasindustrymap (creative commons)

pio, sono piattaforme che rendono disponibili strumenti di sviluppo come workflow, di creazione di interfacce web, database integration, storage, integrazione di web-service. In alternativa, rendono disponibili degli strumenti per la gestione del software, come la gestione delle versioni, la configurazione di delivery, strumenti di controllo delle performance dell'applicativo e quindi di configurazione della scalabilità.

Alcuni esempi di PaaS sono gli ambienti di sviluppo Azure Services Platform di Microsoft, gli ambienti di sviluppo via Web Services di Amazon Web Services (AWS), oppure la piattaforma BlueCloud annunciata da IBM. Inoltre, alcuni produttori forniscono delle soluzioni per sviluppare dei marketplace di soluzioni applicative disponibili in modalità SaaS: tra questi possiamo citare JameCraker, SalesForce.com, o NEC.

Un'altra soluzione interessante è Google App Engine, una piattaforma che permette di sviluppare e ospitare applicazioni web ed è gratuita fino a certi livelli di consumi di risorse.

Un'interessante analisi di comparazione dei servizi offerti da Amazon e Google PaaS è riportata in [18].

- **Infrastructure as a Service (IaaS)**: il termine si riferisce alla modalità di offrire come servizi infrastrutturali, risorse di elaborazione, memoria e comunicazione, come macchine virtuali, CPU, memoria, schede LAN, apparati di rete e loro configurazioni, servizi di backup. Normalmente i servizi offrono una visione virtualizzata delle piattaforme. I clienti accedono ai servizi tramite Internet, e questi servizi sono pagati in base al loro effettivo consumo o utilizzooro effettivo consumo. Il provider normalmente supporta il cliente con strumenti per la configurazione e il monitoraggio delle risorse necessarie da utilizzare, come ad esempio per la configurazione dei firewall, oppure per la configurazione delle politiche di backup, o di scale up/down in base all'effettivo utilizzo delle risorse.

  Molti sono i provider che si stanno aprendo al mercato per supportare questa tipologia di servizi: tra i primi e di maggior successo c'è Amazon EC2 (Elastic Cloud Computing). Successivamente, si sono aggiunti i maggiori fornitori di hardware o software come IBM, SUN, Microsoft, Oracle. Non mancano anche le piattaforme Open Source come Eucalyptus e quelle di altri fornitori come GoGrid che possono offrire servizi i cui costi sono competitivi rispetto a quelli offerti da Amazon.

In *figura 3* è schematizzata l'architettura concettuale del paradigma "Software/Platform/Infrastructure as a Service".

Sicuramente tutte queste tre tipologie di servizi presentano e offrono delle caratteristiche comuni che possono essere così sintetizzate:

- **facile disponibilità di accesso al servizio**, con semplici interfacce web, che non richiedono la disponibilità di elevate capacità hardware da parte del cliente, è possibile gestire una **auto-configurazione del servizio richiesto** (self-provisioning);

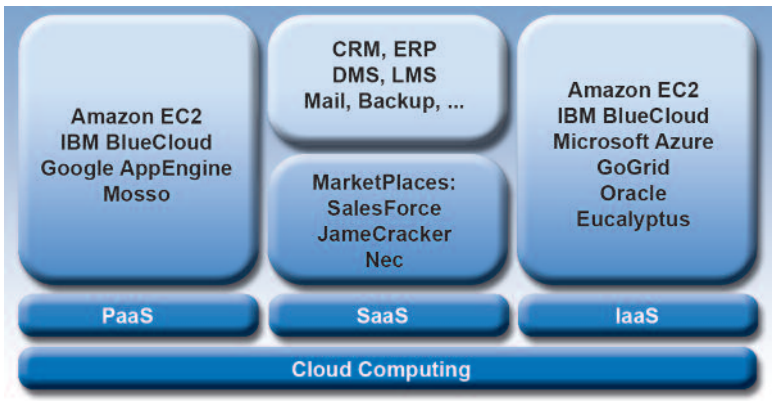


Figura 3 – Architettura concettuale del paradigma "Software/ Platform/ Infrastructure as a Service" [5]

5 adattamento da: http://ivanov.wordpress.com/2008/05/01/cloud-computing/

- **la disponibilità di un servizio fruibile da clienti diversi (multitenancy)**, poiché i clienti condividono le stesse risorse hardware o software, pur mantenendo una separazione logica o fisica dei dati;
- **minori spese di investimento**, in quanto le infrastrutture (hardware o software) sono rese disponibili da chi fornisce il servizio (service provider);
- accesso ai servizi da remoto mediante interfacce che **sono indipendenti dal tipo di device** utilizzato e che rendono **trasparente l'effettiva locazione** del punto di erogazione del servizio stesso;
- erogazione di un **servizio disponibile in modalità anche estemporanea ma comunque immediata (on-demand), con forme di pagamento che sono legate strettamente al reale utilizzo delle risorse (pay-as-you-go)**. Ad esempio, Amazon può garantire la disponibilità di una decina di macchine virtuali in una o due ore e il pagamento avviene con semplice addebito in carta di credito, in base all'effettiva "accensione" della virtual machine;
- **aggiornamenti software o nuove funzionalità sono resi disponibili a tutti i clienti del servizio**, in maniera trasparente e normalmente senza ulteriori aggravi di spesa da parte degli utenti, facendo quindi non gravare sui clienti tutte le politiche di gestione delle patch e di nuovi aggiornamenti;
- possono essere resi disponibili dei servizi che sono il risultato di un'**integrazione di differenti servizi disponibili nelle varie Cloud**, servizi che possono essere configurati anche dinamicamente in base ad esempio a criteri di scalabilità, di disponibilità delle risorse, o ai costi di erogazione dei servizi stessi.

3 Alcuni messaggi da "Cloud Computing Conference & Expo"

La partecipazione alla conferenza "2nd International Cloud Computing Conference and Expo" [6] tenutasi nel marzo 2009 a New York (USA) ha permesso di avere una panoramica aggiornata sullo stato dell'arte, delle tecnologie e dei servizi del Cloud Computing. Questa è stata, infatti, l'occasione per i più importanti fornitori di hardware o software come IBM, Microsoft, Sun, o società di servizi. di presentare le loro soluzioni già operative o di annunciare le prossime iniziative nel settore.

Tra le prime società che hanno iniziato ad erogare servizi infrastrutturali di Cloud Computing possiamo citare Amazon.com, che ha lanciato il servizio nel 2002, utilizzando piattaforme virtualizzate con il prodotto Xen. Amazon.com ha poi arricchito ulteriormente la sua offerta di servizi, resi disponibili in internet come web services (Amazon Web Services: AWS), tramite HTTP e i protocolli REST o SOAP.

Via Web Services sono quindi disponibili servizi di configurazione delle risorse hardware virtualizzate (Elastic Compute Cloud: EC2), di memorizzazione (simple storage service: S3), persistenza in DBMS (SimpleDB Services), di gestione delle code (Simple Queue Service), di autenticazione, di gestione pagamenti e fatturazione (DevPay), e molti altri. Tutti questi servizi sono sempre erogati e fruiti in modalità on-demand, e permettono di sviluppare applicazioni completamente distribuite in internet.

Vari sono i partner che possono vantare di avere collaborato con Amazon.com per la gestione della piattaforma, come IBM, Sun, Rightscale e 3Tera.

Amazon.com può vantare diversi casi di successo di utilizzo dei propri servizi. Ad esempio: Animoto, Facebook, o per l'accesso ai dati storici, NASDAQ, sono applicativi che sono ospitati nell'infrastruttura di Amazon.com. Un esempio efficace di utilizzo estemporaneo dei servizi Amazon è invece il caso del Washington Post, che ha utilizzato i servizi di Amazon EC2 per indicizzare e rendere disponibili al pubblico oltre 17 mila pagine di documenti pdf, che riportavano le attività di Hillary Clinton al tempo della presidenza del marito dal 1993 al 2001. Per realizzare questa

6 http://cloudcomputing.sys-con.com/

elaborazione, sono stati utilizzati circa 200 virtual machine per circa 24 ore per un totale di circa 1400 ore (tempo utilizzo delle virtual machine), con una spesa totale di soli $144,62!

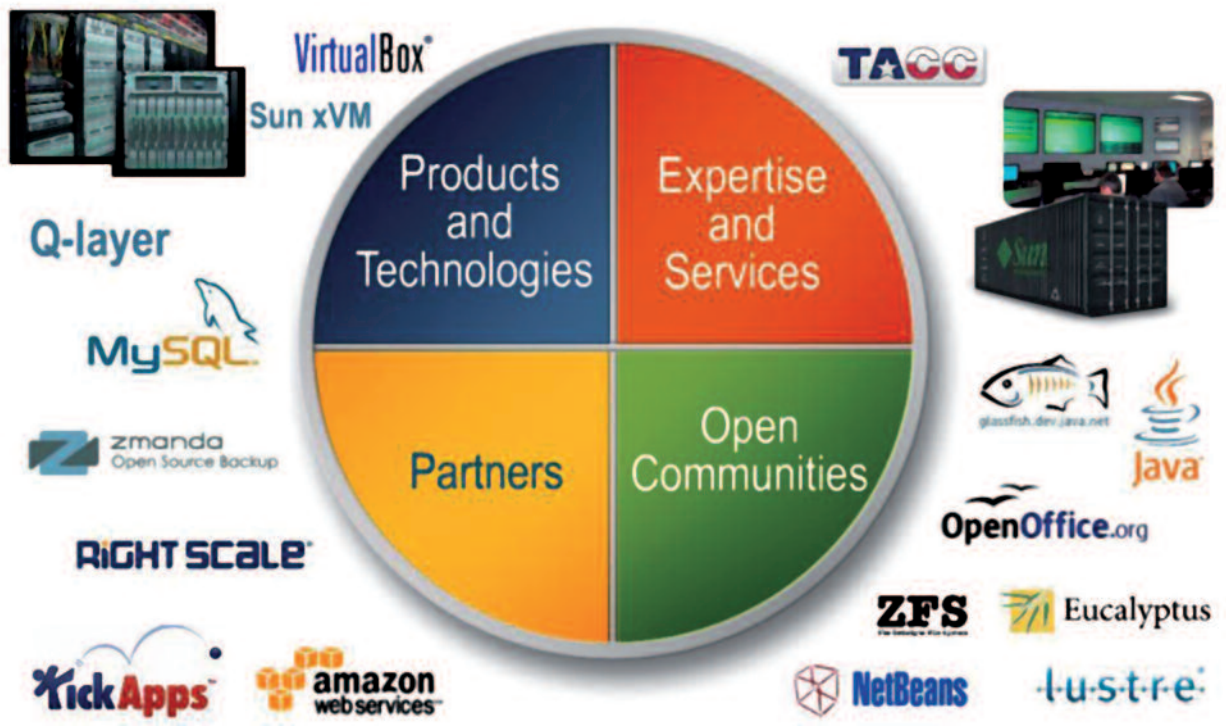
La soluzione **IBM**, invece, si sta caratterizzando soprattutto per la gestione di infrastrutture dinamiche che permettono la gestione di Cloud private, pubbliche o ibride (parte pubbliche e private) e propone una piattaforma Blue Cloud di servizi infrastrutturali e applicativi che sono integrati in una piattaforma (basata su Tivoli) di gestione del provisioning, del performance management, di security, accounting,…

Durante la conferenza **Microsoft** ha poi presentato la sua piattaforma Windows Azure Platform, che offre alcuni dei servizi proposti da Amazon, ma in sola tecnologia virtualizzata basata su prodotti Microsoft e nei data center di Microsoft. Offrono servizi di memorizzazione massiva di dati (blob), servizi di memorizzazione di tabelle e di gestione di code di dati/messaggi. Le applicazioni che utilizzano i servizi Windows Azure utilizzano protocollo .NET e linguaggi come Visual Basic, WCF, C#, C++. Microsoft ha comunque dichiarato che supporterà anche applicativi sviluppati con altri linguaggi, come php, Perl, Phyton, java,… I servizi sono per ora offerti gratuitamente e in futuro saranno contabilizzati in base al consumo effettivo delle risorse.

Il fornitore che invece sta puntando alla predisposizione di una suite Open Source per la gestione dei servizi di Cloud Computing, è **SUN** (*figura 4*), che ha presentato alcuni strumenti prototipali di integrazione dei tool Eucalyptus, ZFS, Lustre, MySQL, NetBeans,…

Altri esempi interessanti che sono stati presentati riguardano poi la fornitura di applicativi che sono di supporto per la gestione di macchine fi-

Figura 4 – La proposta SUN di piattaforma OpenSource per il CloudComputing [*fonte SUN*]



siche e virtualizzate in differenti data center e quindi sono di supporto per i provider dei servizi IaaS, SaaS o PaaS. Tra i più interessanti ci sono i prodotti **RightScale** e **3Tera**, che come citato sopra sono anche partner di Amazon.com.

Altri provider, come **The Rackspace Cloud** e **GoGrid**, si stanno, invece, proponendo al mercato con prezzi e funzionalità competitive rispetto ad Amazon.com: ad esempio i servizi di load balancing e clustering sono gratuiti. Questi provider sono in grado di offrire servizi infrastrutturali, memorizzazione massiva di dati, gestione di applicativi web ed email hosting.

Un'altra interessante soluzione è stata quella proposta da **AppZero** che, sfruttando i meccanismo di Virtual Application Appliance (VAA), permette di costruire il packaging di un applicativo con tutte le sue dipendenze, ma separando totalmente le dipendenze dal sistema operativo, Linux, Unix o Windows. In questo modo il delivery di un applicativo, anche se implementato su più livelli, può essere facilmente gestito e distribuito su server fisici, virtuali o distribuiti nella Cloud. Questo meccanismo permette ad esempio di spostare facilmente un applicativo, o un layer dell'applicativo, da un Cloud Provider all'altro, in base a eventuali problemi di sovraccarico della rete o a politiche di sconto vantaggiose.

La conferenza di New York voleva essere anche un'occasione per finalizzare l'**Open Cloud Manifesto**, identificando i punti cardine su cui basare i futuri sviluppi di servizi e tecnologie, in modo da garantire che i servizi possano interoperare in Cloud private e pubbliche e non essere legati a standard proprietari. A questo manifesto hanno aderito più di 150 aziende che operano in questo ambito, ma purtroppo non hanno per ora aderito i grandi player come Amazon.com e Microsoft!

4 Le soluzioni di Cloud Computing di Telecom Italia

Questa sezione offre una panoramica delle soluzioni di Cloud Computing e dei servizi di IaaS e SaaS che Telecom Italia ha realizzato e rende disponibili al mercato.

I servizi di IaaS sono offerti ai clienti PMI e Top Client tramite la piattaforma NetComputing, che permette di gestire servizi infrastrutturali su hardware virtualizzato e utilizzando diverse piattaforme, come IBM AIX, VMware, Solaris, Hp-UX. La piattaforma, tramite la componente di Service Provisioning, permette la gestione semi-automatica delle richieste di provisioning delle risorse hardware, attivate tramite la componente NIS, e dei servizi applicativi erogati in modalità SaaS, attivati automaticamente tramite l'attivatore NAS (*figura 5*).

La piattaforma è, inoltre, integrata nella catena di delivery dei servizi di Telecom Italia e quindi nei sistemi di provisioning e billing della catena OSS/BSS aziendale.

La piattaforma è dotata di un portale, a cui i clienti registrati possono accedere alla componente di self-provisioning, per l'eventuale acquisto delle soluzioni applicative fornite in modalità SaaS[7].

Tra i servizi erogati in modalità SaaS, Telecom Italia ha già messo a disposizione diverse soluzioni applicative che si basano sia su prodotti commerciali, sia su prodotti Open Source e il cui ambito rientra in quei servizi ICT offerti alla clientela TOP e Business. I domini applicativi spaziano dai documentali, ai CRM, ai sistemi ERP, alle soluzioni di eLearning per la formazione, o ai servizi base di gestione del backup, o di mail. Alcuni esempi di servizi SaaS sono: CRM Open source SugarCRM (offerta MyCustomerEasy), CRM Dynamics di Microsoft (offerta MyCustomer), ERP opensource OpenBravo (offerta MyCompanyEasy), ERP IBM ACG (offerta MyCompany), documentale IBM Document Management e documentale Kelyan, piattaforma di eLearning basata sul prodotto opensource Moodle.

Una delle possibili evoluzioni funzionali della piattaforma NetComputing e dei servizi applicativi SaaS, consiste nell'estendere la componente di

7 Il portale è accedibile tramite il link: http://www.pmi.telecomitalia.it/pmi/

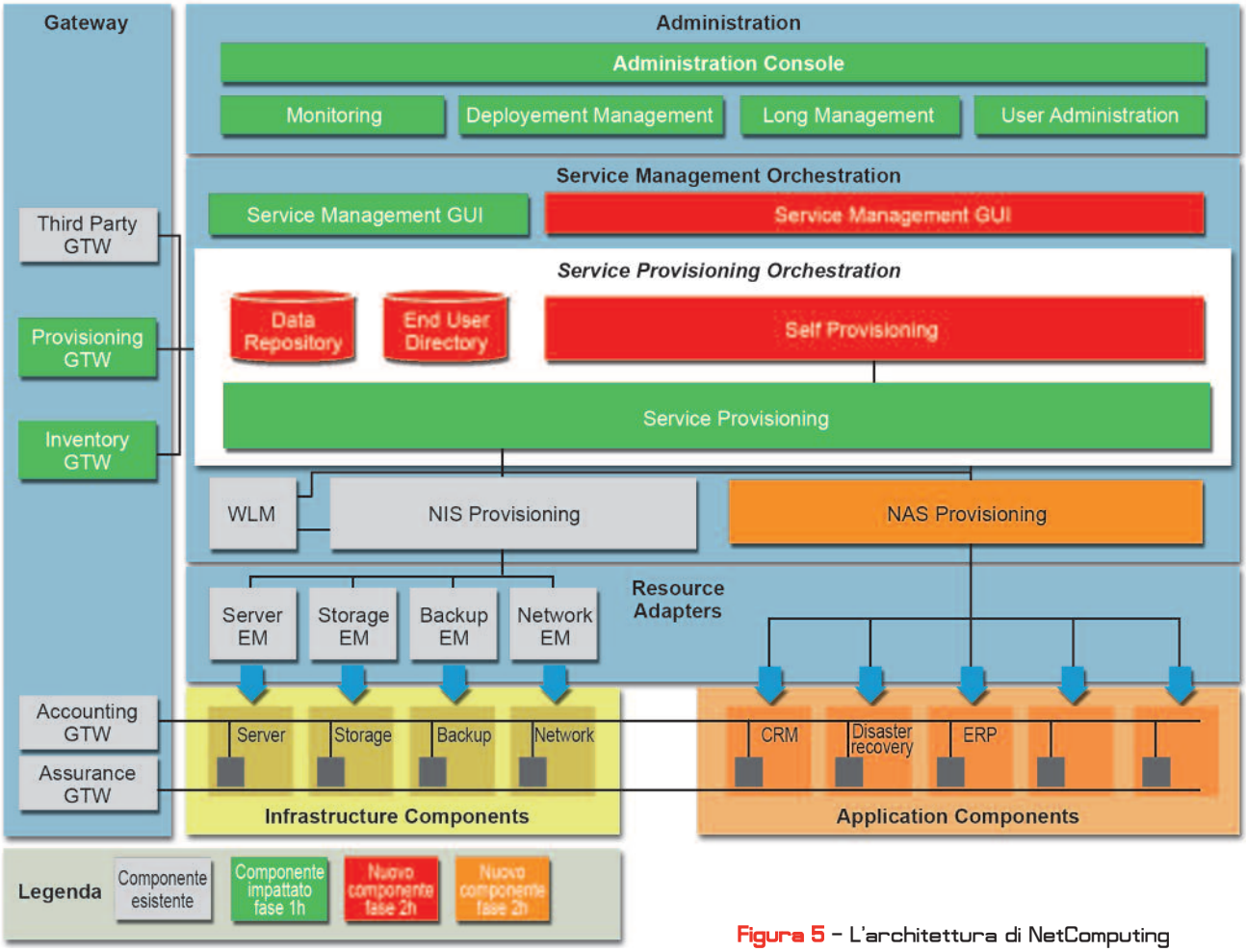


Figura 5 – L'architettura di NetComputing

self-provisioning per rendere il processo di erogazione dei servizi IaaS e Saas completamente automatico e mettere a disposizione dell'acquirente forme di pagamento on-line come carta di credito, paypal, o con possibilità di addebito alla bolletta telefonica fissa o conto mobile (*figura 6*).

Gli obiettivi di questa evoluzione sono diversi:

- integrare, riusare e mettere a fattor comune processi e strumenti sviluppati nello sviluppo dei servizi VAS nell'ambito mobile per l'evoluzione dell'offerta dei servizi ICT offerti da TI (es. piattaforma dei servizi VAS – BAAS, forme di pagamento on-line o conto mobile, gestione provisioning pacchetti SMS, ecc.);
- permettere di allargare il bacino di clienti potenziali, di ridurre i costi e i tempi di provisioning e di gestione dei servizi stessi con un processo di self-provisioning completamente automatico;
- estendere i modelli di business dal puro SaaS verso nuovi modelli di Reselling e Revenue Sharing, allo scopo di consolidare un ecosistema di clienti e partner che abbia Telecom Italia come focal point per il portafoglio dei servizi e l'infrastruttura di erogazione.

Nella *figura 6* è illustrata l'evoluzione di NetComputing verso la piattaforma ICT&VAS che integra ed estende i portali di self-provisioning, portali di marketplace o i canali di provider esterni per l'erogazione di nuovi servizi ICT con le funzionalità già erogate nel mondo Mobile VAS. La piattaforma NAS di Netcomputing è quindi estesa con la componente di DataCollection ed integrata con i servizi di pagamento on-line offerti dalla piattaforma VAS.

La componente BAAS, Business As As Service, ha il ruolo centrale di orchestrazione dei servizi ICT e Mobile VAS, e di collegamento con

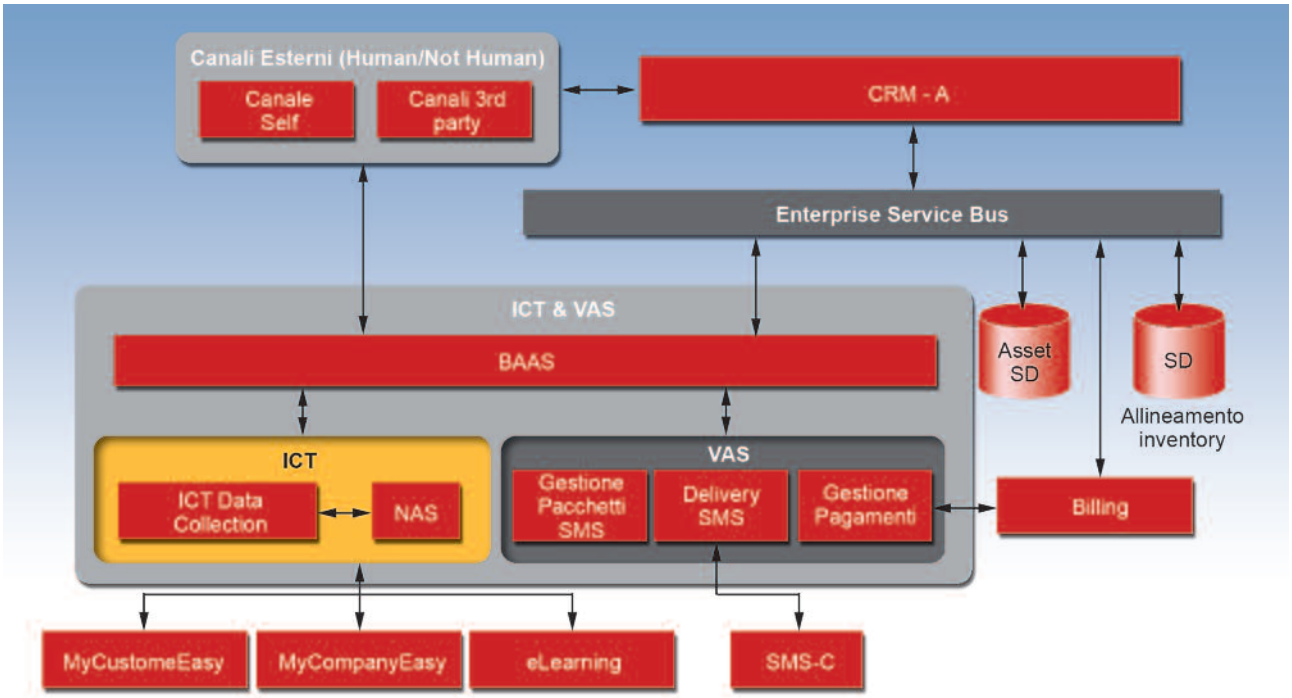


Figura 6 – Evoluzione dell'architettura di NetComputing per i servizi ICT&VAS

la catena standard di delivery dei servizi, e principalmente con i sistemi CRM-Affari, Asset Inventory, Service Inventory, e sistemi di Fatturazione.

5 Ulteriori scenari di servizio abilitati dal cloud computing

Questa sezione introduce ulteriori scenari applicativi abilitati dalle tecnologie del Cloud Computing.

5.1 *Il modello Software-plus-Service*

Sono in molti a credere che la rivoluzione del Cloud Computing si attuerà attraverso il pressoché completo trasferimento di applicazioni e dati nel Cloud (secondo il modello SaaS). C'è chi sostiene invece che un modello ibrido possa essere più efficace: le applicazioni ed i dati "in the cloud" si affiancheranno, integrandosi, con applicazioni in esecuzione sui client e sui server interni all'azienda (secondo il modello Software-plus-Service, S+S).

In effetti questo secondo modello potrebbe consentire una maggiore flessibilità nella scelta del device di accesso alle applicazioni (PC, telefono, ecc.), del tipo di connessione, del luogo dove eseguire una certa logica di servizio "lato server" (in-the-cloud/on-premise), e può tenere conto degli sviluppi della tecnologia dei dispositivi di Utente, che, nel prossimo futuro, segnerà notevoli aumenti della capacità di esecuzione e memorizzazione.

Un interessante esempio di questo modello ibrido è fornito dall'idea di clonare il cellulare nel cloud [15], così da consentire da un lato di sfruttare appieno le potenzialità di elaborazione e memoria dei cellulari di futura generazione e dall'altro di facilitare la reperibilità dei propri contenuti e servizi e la loro condivisione con altri utenti, ad esempio all'interno di una rete sociale.

5.2 *Rich Internet Application e servizi di Network PC*

I servizi Web 2.0 offrono un ampio insieme di funzionalità e capacità. Essi tendono a replicare, estendendolo e migliorandolo, il modello delle applicazioni per il desktop. Il "caso" delle applicazioni di tipo "Microsoft Office" in rete è paradigmatico: le funzionalità e servizi offerti permettono agli utenti di gestire documenti, calendari, fogli di calcolo, posta elettronica ed altro direttamente tramite il browser. I dati degli utenti sono memorizzati in rete e recuperati indipendentemente dal tipo di accesso alla rete, inoltre essi possono essere condivisi, gettando le basi per nuove applicazioni di lavoro cooperativo. Questo approccio sembra essere una possibile minaccia per il consolidato mercato delle applicazioni desktop.

In sostanza la tecnologia Web 2.0 coniugata con l'accesso a larga banda abilitano una nuova classe di servizi denominata Rich Internet Applications (RIA). Esse sono fruite per mezzo di un browser, sono localmente eseguite in un ambiente sicuro, detto "sandbox" (cioè, una macchina virtuale che implementa controlli di sicurezza sull'uso delle risorse locali), mentre il carico computazionale pesante (cambio di stati, gestione e memorizzazione dei dati) è eseguito remotamente da server specializzati realizzati in un Cloud computazionale. Siccome la logica di servizio e i dati sono mmorizzati nel Cloud computazionale, i servizi sono acceduti mediante una connessione broadband anche in modalità nomadica e mobile, se si utilizza un accesso adeguato. Esempi di questa nuova classe di servizi sono: Google's Docs & Spreadsheets suite [8] o l'ambiente collaborativo zimbra [9].

Un esempio di applicazioni RIA è il ricorrente interesse per i servizi che mimano il "network PC". La classe di servizi che va sotto il nome di WebOS si riferisce a servizi Web 2.0 che utilizzano le capacità della rete per virtualizzare un ambiente operativo in rete, mettendo a disposizione del singolo utente un disco di rete, potenza di calcolo, un insieme di applicativi di interesse (ad esempio gli applicativi di tipo Office), strumenti di comunicazione personale (instant messager, posta elettronica ed altro) e in più la possibilità di estendere le funzionalità, accedendo ad un sistema di repository di applicazioni di terze parti. Tipicamente i servizi di tipo WebOS sono basati sulle seguenti assunzioni:

- disponibilità in rete di un'elevata capacità di calcolo;
- disponibilità in rete di capacità di memorizzazione elevate;
- possibilità di essere sempre connessi;
- esigenza di dover accedere ai servizi di network PC da diversi tipi di terminali (da casa, dall'ufficio, in viaggio).

I primi due requisiti sono naturalmente soddisfatti da un'infrastruttura di Cloud Computing.

Un capostipite di questo approccio è stato YouOS [10], un "web operating system" sperimentale che offriva prestazioni di un moderno sistema operativo. Gli utenti potevano accedere, tramite un browser, ad un sistema che emulava un moderno sistema operativo, simile a Linux, che offriva le applicazioni tipiche di un desktop e che permetteva di crearne di nuove, mediante un proprio sistema di sviluppo delle applicazioni integrato. Tale approccio è ora perseguito da eyeOs, da i-cloud e da varie distribuzioni Linux: in particolare "One", una nuova versione di Ubuntu, è stata progettata per potere offrire questo tipo di servizi.

Sebbene il servizio di Network PC sia potenzialmente interessante, non è stato finora declinato in maniera vincente. Inoltre, tale mercato è frammentato in varie iniziative a tal punto che si avverte la necessità di sistematizzare le problematiche dei WebOS mediante la standardizzazione delle API.

8 http://www.google.com/google-d-s/tour1.html

9 http://www.zimbra.com/products/desktop.html

10 http://en.wikipedia.org/wiki/YouOS

5.3 *Il Data Web*

Il Web tradizionale è una rete di documenti collegati tra loro tramite dei link (gli URL), così da realizzare un "enorme" ipertesto. Accanto a questo, si sta affermando un'altra rete di informazioni, costituita da un Web di dati, in cui i dati sono descritti e messi in relazione tramite linguaggi formali (es. RDF), possibilmente arricchiti con informazioni semantiche.

La motivazione di un Web di dati nasce dalla necessità di organizzare i dati presenti in Internet non solo come una rete di documenti HTML, come nel Web, o come un insieme di tabelle, come nei Data Base tradizionali, ma come una rete di dati associati ed aggregabili. Il Web di dati è la base per abilitare applicazioni di mash-up e di data mining, in quanto è molto più facile reperire, estrarre e combinare le informazioni di interesse da singoli dati e dalle loro relazioni, piuttosto che da documenti.

Questa visione è perseguita da due iniziative internazionali: Dataweb [8], sviluppato da OASIS, ed il progetto Linked Data sotto l'egida del W3C [11].

Esempi di tali Web di dati sono quelli generati a supporto di servizi altamente dinamici ed adattativi, come, ad esempio, quelli elaborati nel contesto dell'iniziativa Wiki-city del MIT [12]. Un tale Web dei dati è caratterizzato da grossi moli di dati, raccolte da diverse fonti (es. sensori, contenti/informazioni prodotte dagli utenti, informazioni generate da piattaforme di servizi interne e/o esterne all'operatore). La validità di tali dati può dipendere da condizioni spazio-temporali e il loro accesso e visibilità possono essere soggetti a complessi vincoli d'autorizzazioni. Tali dati devono essere opportunamente collezionati, memorizzati, correlati, resi accessibili tramite strumenti di ricerca e di navigazione. Inoltre, i nuovi dati potrebbero dover essere opportunamente distribuiti alle applicazioni ed utenti che hanno precedentemente espresso un interesse.

La dinamicità, la quantità, il ritmo d'acquisizione e di distribuzione dei dati richiedono elevate capacità di elaborazione e di memorizzazione che possono essere messe a disposizione da un'infrastruttura di Cloud Computing. Inoltre, i meccanismi insiti in una tale infrastruttura possono essere in grado di assorbire adeguatamente ai picchi di traffico e di scalare la soluzione al crescere degli utenti e delle quantità di dati. A partire dalle informazioni strutturate secondo l'approccio del Data Web si potrebbe creare un "Data Cloud", da intendersi come una rete che mette in relazione tali dati. La sua realizzazione potrebbe utilizzare soluzioni che offrono una gestione semplice e con una vista unitaria, i dati memorizzati su diversi sistemi di Cloud Computing, anche di fornitori diversi [16].

6 Possibili evoluzioni verso Cloud pervasivi e adattativi

L'aumento della penetrazione delle connessioni a banda larga, il contemporaneo emergere del mobile broadband, la riduzione dei costi delle memorie di massa e dei server, stanno creando già oggi le condizioni per lo sviluppo di sistemi decentralizzati, che permettono di accedere ad una svariata gamma di applicazioni, come la diffusione e la condivisione di contenuti, lo storage online di dati e backup, Network PC, le applicazioni "Software as a Service", ecc.

È prevedibile, inoltre, che nei prossimi cinquedieci anni lo sviluppo tecnologico dei terminali di utente porterà ad un aumento sensibile delle capacità computazionali, mentre le memorie raggiungeranno dimensioni dell'ordine dei Terabyte. Inoltre, lo sviluppo dell'Internet delle Cose, permetterà di mettere in rete una vasta gamma di oggetti digitali, dai sensori, agli elettrodomestici, che si potranno raggiungere anche utilizzando informazioni "associate", ad esempio descrizioni di tipo semantico o le cosiddette tag. L'introduzione di queste nuove tipologie di terminali e la loro integrazione e sinergia con i sistemi di rete porte-

11 http://www.w3.org/DesignIssues/LinkedData.html

ranno alla creazione di Cloud di risorse che si espandono in maniera pervasiva e senza soluzione di continuità dai terminali all'Edge della rete, ai server dei data center in rete o dei content/service provider.

La conseguente crescente distribuzione e complessità richiederà soluzioni capaciti di auto-adattarsi alla forte dinamicità del contesto servizi e al tempo stesso di auto-gestirsi (con un limitato intervento umano). Infatti, le caratteristiche di questi "Pervasive Cloud" di risorse, quali, ad esempio, la distribuzione geografica, la dinamicità con cui le risorse entrano nel Cloud o lo lasciano (churn rate), impediscono di introdurre sistemi di supervisione logicamente centralizzati, anche se distribuiti/distribuibili su sistemi differenti per ragioni prestazionali o di resistenza ai guasti, per implementare funzioni di monitoring ed ottimizzazione. Infatti:

- l'elaborazione di soluzioni ottime globali a livello del "pervasive cloud" (ad esempio, per il load balancing) potrebbero essere inutili, siccome la soluzione computata potrebbe essere obsoleta o inutile a causa della sua dinamicità;
- un sistema di gestione centralizzato potrebbe non essere in grado di avere un quadro aggiornato dello stato del sistema, tramite la collezione degli eventi, log, allarmi di interesse, a causa della distribuzione geografica delle risorse, dell'evoluzione dinamica delle risorse disponibile e dalla possibile, temporanea, disconnessione di alcune risorse o di sottoreti.

Un'alternativa è quella di introdurre soluzioni altamente decentralizzate, realizzate come sistemi distribuiti basati su tecnologie di Autonomic Computing e Communication. Le tecnologie autonomiche sono nate proprio con l'obiettivo di sviluppare soluzioni di elaborazione distribuita, in grado di rispondere ai requisiti della crescente complessità ed eterogeneità delle reti del futuro. L'Autonomic Computing è stato introdotto nel 2001 da IBM [7]: nel suo manifesto introduttivo alla tecnologia, IBM ha argomentato che, a causa della complessità crescente dei sistemi informativi, i computer e le applicazioni dovevano imparare come gestire se stesse, guidate solamente da policy di alto livello fornite dagli operatori umani. La visione prendeva spunto dalla metafora del sistema nervoso neuromotorio, in particolare alla sua funzione di regolare l'omeostasi dell'organismo attraverso meccanismi autonomi, la cui complessità è mascherata alla volontà cosciente. I sistemi autonomici, pertanto, sono in grado di prendere decisioni autonomamente, controllando costantemente il loro stato e modificando il proprio comportamento adattandolo al variare delle condizioni interne ed esterne di esecuzione, ad esempio per migliorare le prestazioni, oppure per recuperare situazioni critiche o di errore.

L'introduzione di prestazioni autonomiche nei modelli di comunicazione ha permesso la realizzazione di sistemi autonomici distribuiti. Tali sistemi sono tipicamente organizzati come un insieme cooperante di componenti autonomiche dotate di prestazioni di auto-adattamento del proprio comportamento, come quelle denominate self-CHOP (self-configuration, self-healing, self-optimizing, self-protection).

Un esempio di componente autonomica è l'ACE (Autonomic Communitation Element) definito dal Progetto CASCADAS [5]. Sebbene i dettagli del modello di comunicazione possano variare, la distanza "massima" a cui due componenti autonomiche possono comunicare e interagire è in generale piccola, se comparata alla dimensione dell'intero sistema. Inoltre, le comunicazioni sono realizzate attraverso un'Overlay Network, che interconnette tutte le componenti. Tali Overlay Network sono create, mantenute ed ottimizzate, attraverso algoritmi di auto-organizzazione [6], i quali possono anche essere utilizzati per creare raggruppamenti di componenti secondo le loro proprietà (es. il tipo di risorsa gestita).

L'auto-organizzazzione è una delle caratteristiche particolarmente interessanti delle reti di componenti autonomi che, in quanto offre la capacità di far emergere spontaneamente, dalle interazioni locali dei singoli elementi costituitivi, pattern e/o comportamenti adattativi organizzativi a livello dell'intero sistema.

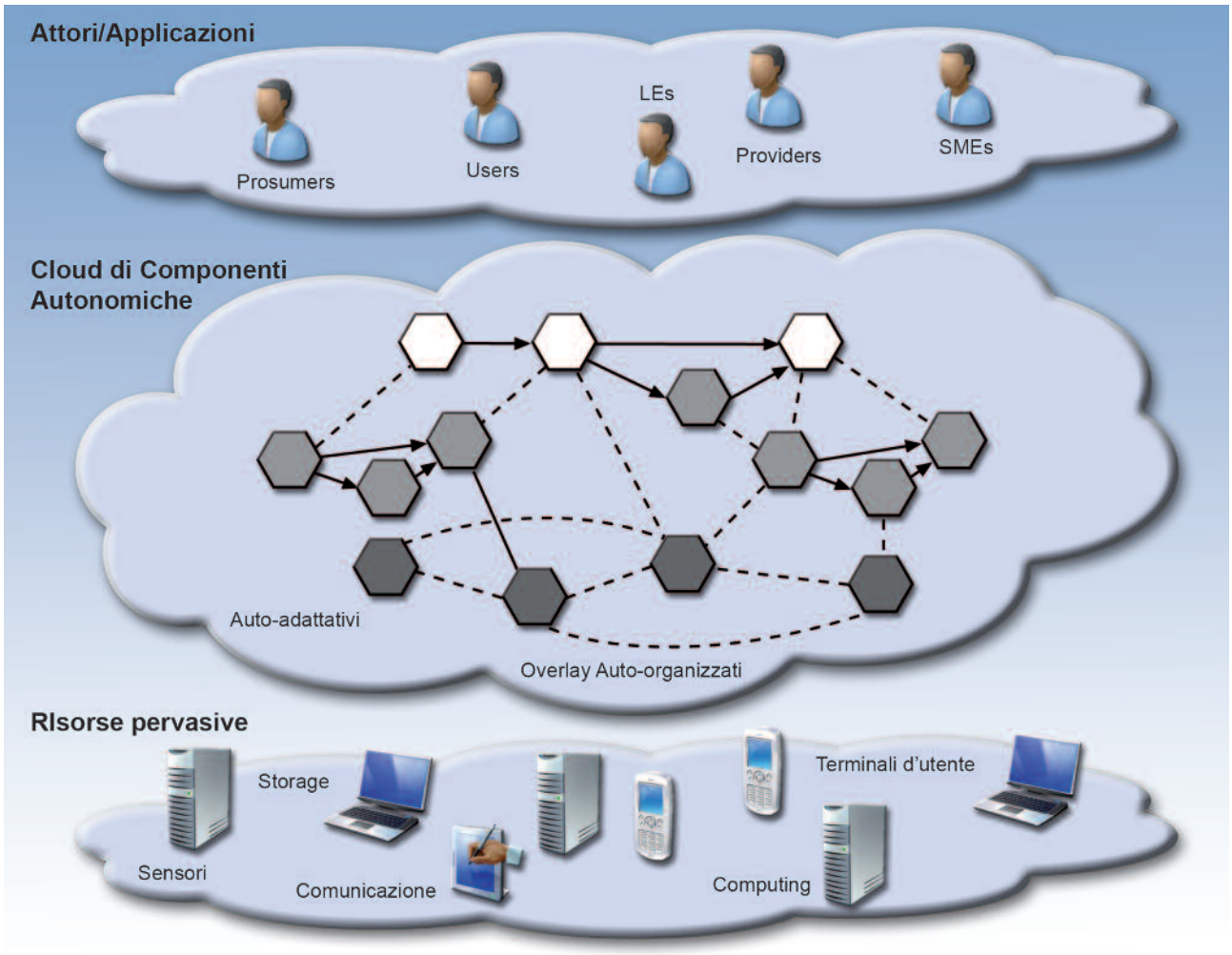


Figura 7 – Pervasive Cloud di risorse

I sistemi di supervisione di un "Pervasive Cloud" di risorse possono essere, pertanto, strutturati come un insieme di componenti autonomiche co-operanti tramite un'Overlay Network auto-organizzata: ogni componente ha il computo di monitorare e gestire una singola risorsa, o un gruppo di risorse, del Cloud, e possono, ad esempio, essere dispiegate sulla risorsa stessa. Le componenti adattano il proprio comportamento e quello della risorsa controllata mediante prestazioni autonomiche, secondo logiche locali di supervisione. Tali logiche locali possono elaborare eventi locali alla componente/risorsa e combinarli con le informazioni (ad esempio, relative a guasti, dati di carico, banda disponibile/utilizzata, ecc.) scambiate con i nodi vicini nell'Overlay Network (*figura 7*).

Tramite lo scambio di informazioni con i vicini, i singoli nodi si creano un'approssimazione dello stato del "Pervasive Cloud". Infatti, i dati ottenuti da uno dei vicini nell'Overlay non sono solo relativi allo stato della risorsa associata a questo, ma rappresentano una combinazione delle informazioni che questo ha ricevuto a sua volta. In questa maniera, le informazioni aggregate sullo stato degli elementi nel "Pervasive Cloud" si diffondono attraverso questo, permettendo alle singole logiche locali di supervisione di avere una visione dello stato complessivo. Tale visione, seppur parziale, è, però, sufficiente per fare emergere, dalle singole logiche locali, un comportamento di supervisione complessiva del "Pervasive Cloud".

Questi meccanismi di diffusione ripetuta d'informazioni sono detti di *gossiping*: ad ogni passo,

un nodo comunica una piccola quantità di dati con un limitato sottoinsieme di nodi nel sistema, in generale, tra i vicini in un'Overlay Network.

Tramite tali meccanismi si può bilanciare il carico elaborativo, ad esempio per soddisfare le condizioni negoziate sulla qualità di servizio, recuperare situazioni d'errore e realizzare politiche d'ottimizzazione nell'utilizzo delle risorse [2].

A titolo di esempio, l'approccio può essere illustrato da un algoritmo completamente decentralizzato per la realizzazione di politiche di risparmio energetico. La logica si basa sulla considerazione che un server attivo, ma in Idle, consuma circa 20 volte di più di un server in stand-by, mentre l'energia consumata da un nodo attivo si incrementa di poco all'aumentare del carico. Pertanto, un gruppo di server con un basso utilizzo causa uno spreco d'energia, in quanto lo stesso carico potrebbe essere gestito da un numero inferiore di server. Attraverso lo scambio d'informazione, tramite *gossiping*, sul carico elaborativo dei server, le logiche locali di supervisione possono decidere di mettere in stand-by la risorsa controllata, o di "risvegliarne" una vicina, nel caso d'informazioni di sovraccarico.

Valutazioni fatte tramite ambienti di simulazione hanno dimostrato che, tramite quest'algoritmo completamente decentralizzato, arricchito con una logica di bilanciamento del carico elaborativo, è possibile ottenere risparmi dell'ordine del 7%-10%, con un peggioramento marginale dei tempi d'esecuzione [3].

Un secondo punto d'attenzione è la realizzazione di politiche di allocazione dinamica delle risorse, per soddisfare le richieste di un servizio.

Anche in questo caso le caratteristiche di elevata distribuzione e dinamicità del "Pervasive Cloud" richiedono di realizzare soluzioni decentralizzate. Un possibile approccio è quello di dotare i servizi/applicazioni di prestazioni per richiedere alle risorse le capacità richieste dalla loro esecuzione e di arricchire i gestori delle risorse con le corrispettive funzioni per l'assegnazione dinamica di tali capacità, secondo opportune politiche di allocazione.

Uno degli approcci di maggiore interesse è quello basato sulle "aste elettroniche" [4]: i gestori delle risorse nel "Pervasive Cloud", anch'essi implementabili tramite componenti atomiche, sono organizzati in un'Overlay Network utilizzata dalle applicazioni sia per inoltrare richieste di "discovery" di risorse, sia per negoziare l'allocazione di loro capacità.

Ad esempio, quando un'applicazione ha bisogno di una certa capacità di una risorsa, registra la sua richiesta come un'asta. Questa informazione è distribuita, attraverso l'Overlay Network, a tutti i gestori delle risorse del tipo richiesto. Tali entità, che hanno il compito di regolare l'allocazione delle capacità delle risorse gestite, rispondono con un'offerta. La formulazione delle offerte è realizzata considerando lo stato attuale di allocazione delle capacità della risorsa gestita e le politiche di allocazione e di "pricing". Tale approccio ha il vantaggio di essere completamente distribuito, sia nella negoziazione che nell'allocazione delle risorse, sia nell'auto-organizzazione dell'Overlay Network che collega le applicazioni ai gestori delle risorse.

In un quadro evolutivo di questo tipo è assolutamente necessario ricercare nuovi modelli di business, basati sul paradigma degli ecosistemi. Ciò che caratterizza un ecosistema, ispirandosi alla metafora biologica, è proprio l'insieme di relazioni, anche caotiche, che in qualche modo condizionano il comportamento e l'evoluzione del sistema stessa.

Gli ecosistemi, che si creano intorno ad un'innovazione tecnologica, possono rappresentare potenziali opportunità di sviluppo per gli Operatori, nella misura in cui questi riescano a ritagliarsi un ruolo e inserirsi in un modello di business vantaggioso.

Negli ecosistemi il valore non è tanto legato ad una specifica applicazione, servizio o informazione, ma alla totalità dell'offerta. Si passa da un valore puntuale (il servizio che utilizzo), ad uno potenziale (la varietà di servizi che potrei utilizzare). Proprio per questo la persona diventa un partecipante all'ecosistema, proprio come accade con il Web: nessun sito è talmente interessante da motivare l'acquisto di un collegamento ad Internet, ma l'enorme varietà del Web, fa percepire al singolo l'utilità dell'acquisto.

Quello dunque che potrebbe emergere dall'evoluzione del Parvasive Cloud è un vero e proprio ecosistema per Telecomunicazioni, ICT e Internet: una miriade di risorse dati e/o servizi (autonomici) che interagiscono direttamente o indirettamente condizionandosi a vicenda, adattandosi e facendo emergere strutture organizzate.

7 Conclusioni

Il Cloud Computing potrebbe richiedere dei servizi di comunicazione che vanno oltre alla semplice connettività. L'Operatore dovrebbe comprendere quali funzionalità potrebbe fornire per entrare nell'eco-sistema del Cloud Computing: ad esempio, le soluzioni a supporto delle VPN dati o i servizi di sicurezza e transazionali.

A livello di servizio il Cloud Computing è un mezzo per permettere all'Operatore per entrare ancora più prepotentemente nel mercato dei servizi IT, valorizzando i propri asset unici, come i data center e i servizi di connettività "intelligente". Inoltre, la flessibilità del Cloud Computing potrebbe permettere agli Operatori di aggregarsi dinamicamente tra loro, al fine di costituire delle piattaforme elaborative, in grado di competere con quelle dei migliori attori del mondo Web, come Google ed Amazon.

Dal punto di vista delle reti sociali, inoltre, l'Operatore potrebbe costruire anche delle Cloud miste che comprendono sia risorse dell'Operatore, sia risorse messe a disposizione dagli utenti finali, per eseguire servizi ed applicazioni a supporto dei social network e di applicazioni con implicazioni socio-economiche, come, ad esempio, "Emergent City" [12], infrastrutture su richiesta per associazioni di volontariato o per situazioni di emergenza.

A tendere questa visione di aggregazione di risorse distribuite in rete e sui device d'utente potrebbe essere un'alternativa, percorribile sia dal punto di vista tecnico sia da quello economico, al paradigma client-server adottato dai grandi attori del Web e a quello della Network Intelligence che centralizza tutte le funzioni nella rete dell'Operatore.

[12] http://www.stanza.co.uk/emergentcity/?page_id=6

8 Ringraziamenti

Gli autori ringraziano Fabrizio Broccolini e Roberto Minerva per l'attenta supervisione e i preziosi suggerimenti forniti all'arricchimento del presente articolo.

BIBLIOGRAFIA

[1] S. Dobson, S. Denazis, A. Fernández, D. Gaïti, E. Gelenbe, F. Massacci, P. Nixon, F. Saffre, N. Schmidt, F. Zambonelli, A survey of autonomic communications, ACM Trans. Auton. Adapt. Syst., 1(2): 223-259, 2006.

[2] P. Deussen, L. Ferrari, A. Manzalini, C. Moiso, "Highly Distributed Supervision for Autonomic Networks and Services", Proceedings of the Fifth Advanced International Conference on Telecommunications (AICT2009).

[3] S. Sahuquillo, L. Ferrari, A. Manzalini, C. Moiso, J. Solé Pareta, B. Otero, "Self-Organized Server Farms for Energy Savings", in Proc. 6th International Conference on Autonomic Computing and Communications (ICAC09).

[4] S. Magrath, F. Chiang, S. Markovits, R. Braun, F. Cuervo, Autonomics in telecommunications service activation. In *Autonomous Decentralized Systems 2005* (pp. 731- 737).

[5] A. Manzalini, F. Zambonelli, L. Baresi, A. Di Ferdinando, The CASCADAS Framework for Autonomic Communications, in: "Autonomic Communication" (Eds.: A. Vasilakos, M.

Parashar, S. Karnouskos, W. Pedrycz), Springer book (2009).

[6] F. Saffre, R. Tateson, J. Halloy, M. Shackleton, J.L. Deneubourg, Aggregation Dynamics in Overlay Networks and Their Implications for Self-Organized Distributed Applications, The Computer Journal (2008).

[7] IBM, An architectural blueprint for autonomic computing. *http://www-01.ibm.com/software/tivoli/autonomic/pdfs/AC_Blueprint_White_Paper_4th.pdf, June 2006.*

[8] OASIS, The Dataweb: An Introduction to XDI, *http://xml.coverpages.org/XDI-IntroWhitePaper20040120.pdf.*

[9] Open Cloud manifesto, *http://www.opencloudmanifesto.org/Open%20Cloud%20Manifesto.pdf*

[10] I. Foster, C. Kesselman, J. M. Nick, S. Tuecke, The Physiology of the Grid - An Open Grid Services Architecture for Distributed Systems Integration, *http://www.globus.org/alliance/publications/papers/ogsa.pdf*

[11] The Guardian, Cloud Computing is a trap, warns GNU founder Richard Stallman, *http://www.guardian.co.uk/technology/2008/sep/29/cloud.computing.richard.stallman*

[12] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, C. Ratti, Real-Time Urban Monitoring Using Cellular Phones: A Case-Study in Rome. MIT SENSEable City Laboratory Working Papers, *http://senseable.mit.edu/ (2007)*.

[13] A. Chakraborty, Ten Key Reasons Enterprise Cloud Computing Is Going Places, Cloud Computing Journal, *http://cloudcomputing.sys-con.com/node/782055* (Dicembre 2009).

[14] R. Harris, Start your company with a credit card and a cloud, ZDNet, *http://blogs.zdnet.com/storage/?p=410*, (Marzo 2009).

[15] C. Mims, Sending Cell Phones into the Cloud, MIT Technology Review, *http://www.technologyreview.com/communications/22571/?nlid=1994* (Maggio 2009).

[16] K. Green, Moving Data around the Clouds, MIT Technology Review, *http://www.technologyreview.com/computing/22586/?nlid=2001* (Maggio 2009).

[17] Wikipedia.org, Everything as a service, *http://en.wikipedia.org/wiki/Everything_as_a_service*

[18] D. Hinchcliffe, Comparing Amazon's and Google's Platform-as-a-Service (PaaS) Offerings, *http://blogs.zdnet.com/Hinchcliffe/?p=166&tag=btxcsim* (Aprile 2008).

[19] L. Dignan, Amazon Web Services: No Open Cloud Manifesto for us, *http://blogs.zdnet.com/BTL/?p=15341* (Marzo 2009).

antonio.manzalini@telecomitalia.it
corrado.moiso@telecomitalia.it
elisabetta.morandin@telecomitalia.it

AUTORI

Antonio Manzalini

laureato in Ingegneria Elettronica, è entrato in Telecom Italia nel 1990 e ha partecipato a diversi progetti di ricerca finanziati dalla Comunità Europea, riguardanti infrastrutture di networking e trasporto ottico, e piattaforme innovative di servizio, occupando, inoltre, varie posizioni di responsabilità. Ha partecipato in ITU-T ed ETSI a molte attività di standardizzazione nelle telecomunicazioni. Attualmente fa parte dell'area Future Centre & Technical Communication di Telecom Italia dove si occupa di architetture distribuite (quali Autonomic e Cloud Computing), abilitanti ecosistemi di TLC e l'Internet delle Cose. Nel 2008 ha conseguito la certificazione internazionale PMI come Project Manager. È autore di molte pubblicazioni, di un libro e di cinque brevetti su soluzioni di reti e servizi ■

Corrado Moiso

laureato in Scienze dell'Informazione, nel 1984 entra in Azienda. Inizialmente ha studiato linguaggi logici e funzionali, l'elaborazione distribuita ad oggetti ed il loro uso in TMN. Dal 1994, con diversi ruoli di responsabilità tecnica, ha investigato l'introduzione di IT nell'Intelligenza di Rete, contribuendo alla sperimentazione di TINA, allo standard Parlay ed all'introduzione di SOA in piattaforme di servizio. Attualmente, nel contesto delle attività del Future Centre & Technical Communication di Telecom Italia, investiga l'adozione di architetture decentralizzate e di tecnologie autonomiche nelle infrastrutture di rete. Ha collaborato a progetti finanziati da EC ed Eurescom; è autore d'articoli in conferenze e riviste e di 7 brevetti su sistemi e metodi per servizi ■

Elisabetta Morandin

laureata in Matematica, entra in Sodalia S.p.A nel 1995 (ora Software Factory di Trento) e partecipa come Task Leader ai progetti europei ed interni all'Azienda nell'area del riuso del software, nella direzione Ricerca e Sviluppo. Dal 2000 partecipa ai progetti di sviluppo della Factory, assumendo nel 2002 l'incarico di Project Manager per il progetto di sviluppo del sistema di provisioning dei servizi xDSL (OM). Nel 2003 inizia a collaborare con l'incarico di Chief Architect nel gruppo di IT Architectures ed è principalmente coinvolta nelle attività di sviluppo del progetto e-Foundation. Nel 2006, con il ruolo di responsabile dell'Architettura e Analisi, è coinvolta nelle attività di ingegnerizzazione dei progetti IT per la Sanità (Piattaforma eHCP, Telemonitoraggio, ArchivingGateway). Dal 2008 è Delivery Manager, nell'Ingegneria IT delle soluzioni verticali per il Mercato (T.IT.MS.VP), delle Business Solution erogate in modalità SaaS (CRM, ERP, Piattaforma di eLearning, Homeland Security, ecc.) e del sistema di provisioning NetComputing ■

Analisi di mercato delle tecnologie e dei servizi della Presenza

INTERNET

Gianluca Zaffiro

Questo articolo si concentra sui risultati di un´analisi di mercato compiuta da Telecom Italia nell'ambito dell'azione di coordinamento Peach, un progetto fondato dall'unità delle Tecnologie Future ed Emergenti della Comunità Europea. L'analisi ha identificato, classificato ed elaborato ulteriormente le aree applicative, o mercati, che attualmente traggono beneficio dalle tecnologie della Presenza, la scienza dell'interazione digitale mediata. In questo articolo viene derivata e presentata una tassonomia di questi mercati, che descrive in che modo e per quale motivo la Presenza è rilevante in ognuno di essi. Infine dall'estrapolazione di alcuni dati, si mostrano quali sono i mercati maggiormente indirizzati e quali le tecnologie della Presenza, a cui si fa il ricorso più ricorrente. I risultati qui riportati sono basati su di un'ampia ricerca condotta sulle aziende che attualmente risultano aver adottato, o stanno fornendo soluzioni di Presenza. In una sezione dell'articolo ci si concentra specificamente sui trend di innovazione che la Presenza porterà alle Telecomunicazioni.

1 Introduzione

La Presenza è stata per anni oggetto di studio presso i laboratori di università e i centri di ricerca, tuttavia a causa dell'elevata capacità di calcolo richiesta e degli alti costi associati, le applicazioni sono migrate dal mondo della ricerca a quello industriale molto lentamente. I problemi

tecnici ed i costi non sono però l'ostacolo principale: la chiave è rendere questa tecnologia adeguata agli impieghi della vita reale.

Continui miglioramenti nel rapporto prezzo/prestazioni dei relativi dispositivi, comunque, hanno reso le tecnologie per la Presenza più economiche, tanto che molte applicazioni della Presenza sono oggi citate in articoli e su giornali non solo a livello sperimentale e di ricerca, ma anche industriale e commerciale.

Per sua natura la Presenza è un campo profondamente interdisciplinare, che copre un ampio numero di aree: dalle neuroscienze e scienze cognitive, all'intelligenza artificiale, sensoristica e sistemistica. Questa caratteristica orizzontale rende la Presenza un terreno affascinante e fertile, ma costituisce anche un impedimento al suo sviluppo, data la distribuzione dei ricercatori su disparate discipline e gruppi in tutto il mondo.

Un'attività di coordinamento intesa a promuovere il dialogo tra le discipline, a costruire un'identità e a favorire l'integrazione nel processo dell'individuazione del futuro della ricerca e dell'indirizzamento delle linee guida, è stata pertanto promossa nell'ambito del programma quadro Peach [1]. Peach è un progetto fondato dall'unità delle Tecnologie Future ed Emergenti della Comunità Europea della durata di tre anni, iniziato nel maggio 2006 e concluso a maggio 2009.

Il primo obiettivo di Peach è stato quello di stimolare, strutturare e supportare la comunità di ricerca, con un'attenzione speciale alle sfide associate all'inter-disciplinarietà del campo e alla produzione di uno scenario futuro e di una mappa evolutiva a supporto della costruzione della Presenza. In secondo luogo, dato che la ricerca sulla Presenza è orientata a produrre tecnologie dirompenti che possono avere un impatto sociale profondo e sollevare serie questioni etiche, Peach ha avuto l'obiettivo di analizzare la relazione tra le tecnologie della Presenza e la società (tendenze, etica, aspetti legali), stimolare il contatto dei ricercatori con il mercato e migliorare la comprensione pubblica sulla ricerca e le tecnologie della Presenza.

In questo articolo si presenta uno studio delle aree di applicazione della Presenza che possono essere individuate nel mercato. Si segnala che questa ricerca non intende costituire una lista esaustiva in tutte le aree, ma è stata piuttosto pensata per metter in luce prodotti e servizi interessanti e rappresentativi. L'analisi è stata coordinata da Telecom Italia in quanto partner di Peach e in qualità di responsabile delle attività di *Market Interaction* di Peach.

L'analisi è partita dalla realizzazione di un elenco di aziende che operano con la Presenza, condotta attraverso una ricerca e selezione delle stesse effettuate tra aprile 2007 e febbraio 2009, utilizzando varie fonti indicate in *tabella 1*. In relazione ai risultati precedenti sono quindi state classificate le aree di applicazione industriale della Presenza.

Per ogni azienda è stata compilata, o su indicazione diretta della stessa, o dall'analisi del profilo e del sito *web*, una scheda con le seguenti informazioni:

- una breve descrizione dell'azienda;
- una breve descrizione dei prodotti e servizi;
- l'area di competenza, con più scelte possibili (es. interfacce acustiche, intelligenza artificiale, processamento dei segnali);

Fonte	Lista iniziale Peach	Lista ISPR	Contatti IST2006	Peach who is who 2006	Ricerche web	Totale
Accettati	34	2	15	3	98	152
Respinti	2	37	9	4	7	59
Totale analizzati	36	39	24	7	105	211

Tabella 1 – Principali fonti e relativi i dati numerici per la costruzione della lista delle aziende che operano con e/o nella Presenza

- l'area applicativa, con più scelte possibili (es. telecomunicazioni, medicina, intrattenimento).

La lista delle aziende è stata infine pubblicata sul sito di Peach [2].

Dall'analisi della lista di aziende sono stati derivati progressivi cambiamenti e aggiustamenti alla tassonomia.

Inoltre la lista delle aziende ha fornito la base dati per l'analisi statistica delle caratteristiche del mercato della Presenza, consentendo per esempio di estrarne la distribuzione per paese e per competenze.

Le idee iniziali sulle aziende da includere nella lista sono state basate su alcuni contatti già noti al consorzio Peach, su una lista pubblica della Società Internazionale per la Ricerca sulla Presenza (ISPR) [3], associazione questa che supporta la ricerca accademica in particolare relativa alla tele-presenza, su un indirizzario di ricercatori ed esperti di Presenza, denominato "Peach Who is Who" e redatto dal consorzio Peach [4], ed infine dai contatti stabiliti durante una sessione di *networking* organizzata da Peach a Helsinki durante l'evento IST 2006 supportato dalla Comunità Europea [5], durante il quale sono stati raccolti un certo numero di riferimenti di persone.

Successivamente la ricerca via *web* è stata focalizzata e basata sull'identificazione delle aree di applicazione e tecnologiche, così come su ulteriori contatti personali ed informazioni da altre fonti. In totale sono state analizzate 211 aziende da tutte le fonti citate, di cui 152 aziende sono state finalmente mantenute nell'attuale lista.

2 Mercati per la Presenza

La Presenza è un campo dedito allo studio della scienza, della tecnologia e degli impatti sociali dell'interazione digitale mediata. Consiste di un insieme di aree di ricerca che studiano come produrre esperienze apparentemente "reali" e l'impatto delle nuove tecnologie di interazione sulle reti sociali.

La scienza della Presenza [6] studia come il cervello umano costruisce il modello della realtà e di sé stesso sostituendo, o incrementando l'informazione sensoriale e d'interazione ad esso indirizzata. Essa appartiene ad una più ampia classe di campi di ricerca che studiano come i sistemi cognitivi costruiscono modelli del loro ambiente e interagiscono con esso.

Il principale obiettivo è quello di sviluppare una scienza e tecnologia per raggiungere un efficace livello di sostituzione e interazione ed offrire accesso ad un gran numero di potenti applicazioni.

Il campo di ricerca della Presenza si può separare in tre aree principali:

- **scienze cognitive umane e sociali**: da considerarsi nel senso più ampio, includendo sia la parte di intelligenza che di azione, le emozioni ed i processi volitivi. Tra le discipline che vi ricadono vi sono la psicologia e le neuroscienze, le scienze sociali e cognitive;
- **interfacce uomo-macchina**: tecnologie per inviare e ricevere informazioni dagli esseri umani agli agenti informatici. In pratica sono dei sistemi di comunicazione bi-direzionali tra uomo e macchina: schermi, telecamere, microfoni, altoparlanti, sensori elettro-fisiologici, stimolatori vestibolari o di altro tipo, sintetizzatori di odori, stimolatori magnetici trans-cranici ecc;
- **scienza cognitiva applicata alla macchina**: questo campo comprende l'intelligenza artificiale (nel senso più ampio possibile), l'intelligenza computazionale (inclusa la *fuzzy logic*, l'apprendimento statistico ecc.) così come i sistemi di gestione di grandi moli di dati, la classificazione automatica, l'analisi statistica e il processamento dei segnali. Questo aspetto si rende probabilmente meno rilevante nello scenario di comunicazione uomo-uomo mediato dalla tecnologia, ma diventa critico nella interazione uomo-macchina, fornendo l'essenziale "*ghost in the machine*".

Tra le tecnologie che abilitano la Presenza, le interfacce uomo-macchina sono probabilmente le più importanti. Gli agenti umani e artificiali sono entrambi dotati di attuatori e sensori e sono in

grado d'interagire anche con interfacce dirette.

Una lista di tecnologie importanti per la Presenza, anche se non esaustiva, è stata presa in considerazione durante la ricerca e la selezione della lista di aziende. Queste tecnologie possono essere associate alle seguenti principali aree di competenza:

- agenti umani virtuali;
- computer grafica;
- intelligenza artificiale;
- interazione uomo-calcolatore;
- interfacce acustiche;
- interfacce cervello-calcolatore;
- interfacce tattili;
- misure e rappresentazione di immagini mediche;
- processamento dei segnali;
- realtà aumentata o mista;
- realtà virtuale;
- tecnologie di comunicazione;
- visione artificiale.

3 Quali aree di applicazione?

Analizzando la lista delle aziende attive sulla Presenza, è stata derivata una tassonomia che identifica otto aree di applicazione, all'interno delle quali le aziende selezionate operano nel mercato utilizzando le tecnologie della Presenza. Tali aree di applicazione, qui chiamate "mercati", sono:

- intrattenimento;
- educazione & formazione;
- medicina & psicologia;
- telecomunicazioni;
- marketing;
- militare;
- manifatturiero & design;
- architettura & costruzioni.

Nelle sezioni che seguono sono descritti, per ogni mercato, quali sono i principali vantaggi che la Presenza porta alle attività svolte dalle aziende attive nelle corrispondenti aree applicative.

3.1 *Intrattenimento*

È uno dei campi di applicazione della Presenza maggiormente indirizzati attualmente in termini di aziende che vi operano. L'intrattenimento beneficia della Presenza, rivoluzionando il modo di vedere un film attraverso l'introduzione della interattività (rispetto alla trama o ai personaggi per esempio) ed immersività, ed offrendo gli strumenti per incontrare personaggi di fantasia in un ambiente virtuale. Un altro vantaggio che le tecnologie della Presenza aggiungono è quello di far sì che le persone interagiscano con il contesto di gioco e si comportino come se fossero realmente calate al suo interno. Inoltre va registrata la grande diffusione dell'intrattenimento con ambienti simulati dal calcolatore, i cosiddetti mondi virtuali, costruiti per essere abitati e vissuti in modo interattivo dagli utenti attraverso i loro avatar. Esempi noti sono *Second Life*, *There.com*, o *Club Penguin*: ogni giorno milioni di utenti visitano questi mondi virtuali che esistono solo come mondi di fantasia, di divertimento e di gioco di ruolo.

Infine alcune tecnologie della Presenza, come le nuove interfacce cervello-calcolatore, che consentono la comunicazione diretta da uomo a macchina attraverso la lettura ed interpretazione dell'elettro-encefalo-gramma (EEG), sono oggi applicate nell'ambito dei videogiochi e dei giocattoli.

3.2 *Educazione e Formazione*

È un'area estremamente interessante ed indirizzata quasi quanto quella dell'intrattenimento. La Presenza è usata in questo caso per sviluppare strumenti di apprendimento ed addestramento o formazione altamente immersivi (ad esempio con interfacce di simulazione, di realtà virtuale e con giochi di ruolo), che consentono ai discenti e ai formandi di provare ciò che in un'esperienza reale sarebbe difficile da gestire e generare. Un aspetto che va sottolineato è che è scientificamente provato che le reazioni psicofi-

siche e comportamentali indotte dall'esperienza virtuale nel soggetto sono confrontabili con quelle che si avrebbero nell'ambiente reale corrispondente, nonostante il soggetto sappia che l'ambiente in cui si trova è una mera simulazione. Conseguentemente applicazioni come le esercitazioni in caso di incendio o per altre situazioni di pericolo assumono notevole valore. Inoltre la Presenza può essere impiegata nell'educazione scolastica per visualizzare concetti astratti (e non).

3.3 *Medicina e Psicologia*

Si tratta di una delle prime e più note aree di applicazione per la Presenza, come risulta sia dall'analisi della lista delle aziende che vi si dedicano, sia dalla numerosità di articoli e ricerche scientifiche che vi fanno riferimento. La medicina e la psicologia beneficiano dalla Presenza in numerosi modi: tecnologicamente la visualizzazione aumentata supporta il chirurgo durante un'operazione espandendo il contenuto informativo disponibile e consentendo di vedere gli organi attraverso i tessuti [7]; un ambiente immersivo 3-D consente il trattamento di fobie ed altre problematiche mentali, guidando il paziente in un'esperienza virtuale controllata dal terapeuta [8], proprio in virtù del fatto che l'esperienza virtuale può generare reazioni psicofisiche e cognitive confrontabili con quelle determinate da una esperienza reale corrispondente; l'impiego di simulatori di parti anatomiche e di strumenti medicali, con *feedback* anche tattili, offre al medico la possibilità di far pratica, apprendere tecniche e procedure complesse in un contesto equivalente al reale ma in condizioni prive di rischio e con una efficace gestione dei costi [9].

3.4 *Telecomunicazioni*

La video-teleconferenza o tele-presenza (vedi *figura 1*) oltre a rappresentare la tipica applicazione di ambiente immersivo è anche un esempio evidente di come la scienza della Presenza sia la base per l'illusione di comunicare tra persone, come se ci si trovasse effettivamente nello stesso luogo.

I dati dei venditori e degli utilizzatori dimostrano che il ricorso allo strumento di tele-presenza nelle organizzazioni che se ne sono dotate si attesta al 30 o 40%, mentre il tasso d'uso dei sistemi di videoconferenza tradizionale si limita al 6 o 7%. Fino al 15% del mercato di tele-presenza sarà legato ai servizi pubblici, come ad esempio i sistemi installati nei centri business e negli alberghi. Oggi un sistema di tele-presenza costa tra 150.000 e 250.000 euro a postazione e

Figura 1 – Sistema di tele-presenza di HP

tra 7.000 e 20.000 euro al mese per postazione completamente gestita [10].

Un altro vantaggio portato dalla Presenza è quello di ricreare e potenziare l'esperienza sociale di comunicazione e collaborazione: applicazioni di questo tipo si trovano nell'ambito della collaborazione mediata, a supporto di gruppi di lavoro distribuiti o per la gestione e condivisione della conoscenza. Le telecomunicazioni possono anche beneficiare della co-presenza o *social presence* [7][12][13], introducendo nelle comunicazioni mediate tradizionali gli elementi informativi non-verbali [14][15], gli aspetti emozionali e le informazioni di contesto. Infine è possibile dotare gli agenti virtuali [16] dedicati all'interazione con il pubblico, come ad esempio gli assistenti virtuali diffusi su Internet, di intelligenza artificiale sociale, rendendoli capaci di relazionarsi in modo più naturale con gli esseri umani.

3.5 *Marketing*

La Presenza offre al *marketing* strumenti per la visualizzazione avanzata di oggetti, la virtualizzazione e l'"*augmentation*". Un prodotto, anche se non ancora realizzato, può essere visto e manipolato con un'animazione 3-D a 360°. È possibile far indossare ai clienti scarpe o altri accessori in camerini virtuali, facendosi un'idea del prodotto prima di acquistarlo e decidendo come personalizzarlo sulla base delle proprie preferenze rispecchiate dal modello virtuale. I pubblicitari possono adattare la pubblicità ad uno specifico contesto usando la virtualizzazione di immagini. Infine nei mondi virtuali gli avatar di persone reali possono vedere i prodotti reali (o virtuali) di marche note, per poi acquistarli nel mondo virtuale (o reale) con denaro virtuale che alla fine, quasi un gioco di parole, è denaro reale. Secondo C. E. Hudson della società americana Serious Business, il mercato globale dei beni e/o regali virtuali nel 2008 ha generato ricavi per 2 miliardi di dollari (di cui il 10% negli USA).

3.6 *Militare*

Questa area trae speciale vantaggio dal potenziamento delle capacità umane in azioni di guerra, ottenibile dall'impiego delle tecnologie di Presenza. Per esempio si realizzano visori con sistemi a realtà aumentata che aiutano a fornire informazioni aggiuntive ai soldati. Inoltre le forze militari possono effettuare un addestramento simulato di impiego di armi e veicoli militari in condizioni controllate e di sicurezza, ed infine impiegare ambienti virtuali e la tecnologia dei video-giochi per l'addestramento tattico e il reclutamento [17].

3.7 *Manifatturiero e Design*

La Presenza in questo settore consente di migliorare la visualizzazione. Impiegare la realtà virtuale è molto vantaggioso per costruire prototipi virtuali o presentare lavori di *design* (per esempio per organizzare una sfilata di moda virtuale), o per studiare virtualmente l'ergonomia di un prodotto o, infine, per migliorare la sicurezza ed efficienza di un posto di lavoro in fase di progettazione. La visualizzazione di dati complessi, come ad esempio giacimenti petroliferi profondi, è anche un altro modo efficace di contenere i costi industriali [18].

3.8 *Architettura e Costruzioni*

Anche questa area gode dei benefici legati alle caratteristiche di visualizzazione ed immersività offerti dalla Presenza. L'applicazione *architectural walkthrough* consente, attraverso l'uso del calcolatore e di un sistema interattivo, di simulare un percorso all'interno di un modello architettonico tridimensionale, per esempio aiutando a verificare l'impatto di un progetto urbano proposto [19]. Un'altra applicazione è il modello di costruzione 4-D, che consente di vedere su schermo

tridimensionale come un progetto edilizio evolve nelle sue varie fasi temporali, permettendo di rivedere la pianificazione o di verificare lo stato di avanzamento.

4 Scenari futuri nelle Telecomunicazioni

In questa sezione si affronta il tema degli scenari futuri per le Telecomunicazioni basati sulle tecnologie della Presenza [20]. Come è stato detto queste tecnologie possono potenziare la comunicazione tra persona e persona grazie soprattutto all'impiego di interfacce uomo-macchina (HMI) in grado di creare la sensazione credibile di trovarsi in un luogo o con qualcuno.

In questo modo è possibile aumentare il contenuto informativo scambiato durante la comunicazione mediata.

L'impiego di un'interfaccia consente non solo di riprodurre un luogo o la co-presenza di altre persone in modo altamente verosimile oppure adatto allo svolgimento di un compito specifico; consente anche di aggiungere indicazioni che sono utili per comunicazione stessa o la collaborazione degli intervenuti, e pertanto incrementare le informazioni che sarebbero altrimenti disponibili nella corrispondente (o no) situazione reale.

Una comunicazione potenziata può tenersi in un luogo che riproduce fedelmente la realtà oppure in un luogo virtuale generato dal calcolatore.

Al primo caso appartengono le applicazioni di video tele-presenza, l'uso di interface tattili, la trasmissione di stati mentali attraverso interfacce cervello-calcolatore, le applicazioni di realtà mista o aumentata soprapposte al mondo reale, le stanze stereoscopiche immersive (come per esempio il "Blue-C Portal" presentato alla conferenza Siggraph 2003 [21]).

Al secondo caso invece sono da ricondurre i mondi virtuali alla Second Life e le applicazioni immersive virtuali generate dal calcolatore nei CAVE (*Computer Automatic Virtual Environment*).

Nel primo caso l'obiettivo è di ricostruire il mondo reale in una copia virtuale esattamente identica, creata sensorialmente dalla macchina, mentre nel secondo caso l'obiettivo è di sostituire il mondo reale con uno virtuale dove è ancora possibile interagire in mondo naturale, ma anche in deroga alle regole del mondo fisico (si pensi per esempio alla possibilità di spostarsi volando offerta da Second Life ai suoi abitanti).

La video tele-presenza è l'esempio più evidente di comunicazione potenziata, per la quale sono già disponibili prodotti commerciali, per esempio di Cisco e HP. Altre soluzioni immersive più avanzate sono per ora state solo dimostrate, come il "Blue-C Portal". Eventi "olografici" sono diventati realtà negli studi televisivi, come per la trasmissione su CNN durante le elezioni della presidenza americana, dove ben 35 telecamere ad alta definizione sincronizzate hanno riprodotto un'immagine 3-D video di una giornalista [22] (si veda la *figura 2*).

Componenti significativi per la comunicazione aumentata sono:

- i dispositivi di visualizzazione (schermi stereoscopici, visori 3-D, caschetti per *augmented reality* ecc.);

Figura 2 – la corrispondente di CNN Jessica Yellin mentre appare in diretta nello studio in una videoproiezione 3-D

- strumenti d'interazione naturale (tattile, vocale, gestuale ecc.);
- strumenti di interpretazione degli stati umani psico-fisici, cognitivi ed emotivi (onde cerebrali, emozioni ecc.);
- metodi di rappresentazione (*avatar*, agenti virtuali, mondi e ambienti virtuali ecc.);
- soluzioni di confluenza uomo-macchina (non invasive, come nella realtà mista o aumentata e alcune interfacce BCI oppure invasive come nelle protesi, interfacce neurali, interfacce BCI impiantate).

La combinazione di queste componenti potrà dare origine a soluzioni di comunicazione mediata estremamente diverse, dove l'elemento comune è l'adozione di una tecnologia che potenzia e amplifica la capacità umana di scambio di informazioni o di collaborazione a obiettivi comuni.

Di seguito si descrivono nel dettaglio le componenti citate.

4.1 *Dispositivi di visualizzazione*

Gli schermi stereoscopici si diffonderanno nel mercato nel medio termine. I nuovi schermi "olografici" recentemente sviluppati superano alcuni limiti di quelli lenticolari: ogni punto dello schermo di questi nuovi dispositivi emette raggi di luce di colore ed intensità diversi e in direzioni multiple, in questo modo l'immagine 3-D può essere vista da più persone contemporaneamente e cambiando l'angolo di vista in modo continuo. Inoltre sistemi di proiezione sulla retina offriranno per la visualizzazione personale soluzioni molto più snelle di quelle attuali [10].

La **realtà aumentata** (AR) sta emergendo e conquisterà la massima diffusione di mercato probabilmente nel lungo periodo [10]. Questa tecnologia può essere totalmente immersiva, come nel caso d'impiego di visori che schermano il campo visivo, o semi-occlusivi, per esempio come nel caso dei dispositivi commerciali "MyVu", che lasciano libero parte del campo visivo, o che sono semitrasparenti. Alcuni visori possono proiettare dati alfanumerici per aggiungere informazioni sul contesto (come nel caso di applicazioni militari o per squadre di manutenzione o soccorso). Oggi gli inconvenienti maggiori legati a questi sistemi sono il consumo e la durata della batteria, le dimensioni ingombranti e la scarsa visibilità o luminosità quando usati in luce solare. I telefoni mobili stanno guadagnando sempre più importanza come interfacce per AR, specialmente per supportare la comprensione e interazione delle persone in una situazione contestuale. Quando i cellulari adotteranno soluzioni per scoprire non solo la loro posizione ma anche il loro orientamento nello spazio e si doteranno di soluzioni di riconoscimento dei gesti basati su telecamera, diventeranno allora una potente interfaccia per i contenuti virtuali (vedi *figura 3*).

La *Natural Feature Tracking* è una tecnologia che potrà essere impiegata per sovraimporre ed orientare correttamente nello spazio oggetti virtuali o altri dati su flussi video e immagini, senza utilizzare *marker*. In questo modo per esempio

Figura 3 – Samsung Lab propone il concept del Mobile Immersive Virtual Outreach Navigator: i cellulari diventano una finestra interattiva di massa per mescolare realta' e virtuale [23]

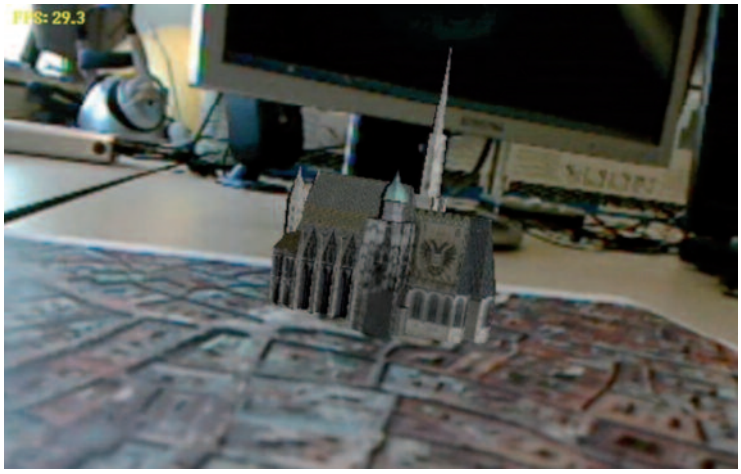


Figura 4 – Con la realta' aumentata e il Natural Feature Tracking si puo' visitare un monumento virtuale sulla cartina di una citta' attraverso la telecamera del cellulare [realizzazione della Graz University of Technology]

ad una mappa geografica, vista attraverso la telecamera di un cellulare, si potranno sovrapporre monumenti ed edifici virtuali 3-D, consentendo di esplorare virtualmente una città. Una tale caratteristica, usata con un cellulare, trasformerà lo stesso in un potente dispositivo in grado di mescolare in ogni istante oggetti virtuali alla realtà circostante (vedi *figura 4*).

4.2 *Strumenti d'interazione naturale*

Un numero sempre maggiore di dispositivi e applicazioni si doteranno di accelerometri ed altri sensori, o implementeranno sistemi di analisi dei flussi video o della voce, o sensori fisiologici: conseguentemente l'interazione tra uomo e macchina diverrà sempre più naturale e trasparente. Attraverso il **riconoscimento dei gesti** con telecamere 3-D (in grado di riconoscere la profondità) è possibile interagire con uno schermo (*figura 5*), così come con il controller della WiiMote dotato di **sensori di movimento** sì può giocare a un videogioco: sono esempi di come questi strumenti potranno rendere l'interazione più naturale e finalmente semplificare la comunicazione.

Le **interfacce tattili** invece non solo daranno "volume" agli oggetti virtuali, ma potranno anche essere impiegate per trasmettere emozioni e supportare la cooperazione [26][27].

4.3 *Strumenti di interpretazione degli stati umani (psico-fisici, cognitivi, emotivi)*

Le BCI, **interfacce cervello-calcolatore**, sono attualmente un tema molto discusso, non solo nell'ambito della comunità scientifica, ma anche dagli analisti di mercato [10]. Le BCI possono identificare gli stati mentali umani, come i livelli di attenzione e meditazione, e tradurli in comandi per videogiochi o giocattoli robotizzati.

Figura 5 – Il riconoscimento dei gesti 3-D per l'interazione user-computer naturale [tecnologia di Softkinetic e Orange Vallee, CES 2009 [25]]

Figura 6 – Un lettore di EEG realizzato da un partner del progetto Peach

I **lettori EEG** possono riconoscere gli stati emotivi o di concentrazione e quindi essere utilizzati per aggiungere informazioni non-verbali altrimenti perse alla comunicazione tradizionale (*figura 6*). Applicazioni non mediche di queste interfacce si ritrovano a supporto di compiti di formazione, o di attività critiche (autisti di autotreni, controllori di volo ecc.).

Il **riconoscimento delle emozioni** in soluzioni commerciali si basa oggi principalmente su analisi delle caratteristiche del segnale vocale [10], mentre si studiano tecniche di video analisi in tempo reale dell'espressività del viso [28].

4.4 *Metodi di rappresentazione*

La comunicazione può avvenire attraverso specifici ambienti di mediazione o supportata da mediatori. Per metodi di rappresentazione qui intendiamo mondi virtuali, realtà virtuale immersiva, avatar, agenti virtuali dotati di aspetto umano e caratteristiche di intelligenza sociale artificiale.

I **mondi virtuali** possono essere sia visualizzati su *desktop* o attraverso CAVE immersivi con visione stereoscopica. I mondi virtuali per *desktop*, oltre al noto successo già avuto per l'intrattenimento, indirizzano il mercato delle imprese in due modi:

- addestramento virtuale in ambienti che riproducono in modo fedele un ambiente di lavoro, offrendo la possibilità di formare il personale senza la necessità di viaggiare o di esporsi a rischi;
- cooperazione mediata dagli *avatar* in ambienti virtuali sia in contesti di riunione che in ambiti di lavoro collaborativo su documenti o altri oggetti, con la possibilità di utilizzare interfacce 3-D interattive e suono spaziale.

La **realtà virtuale immersiva** è uno strumento potente per addestrare e ripassare procedure operative, così come per la prototipazione sicura, dal momento che le persone reagiscono alla realtà simulata come se fosse vera

Gli **avatar** possono intervenire nella comunicazione all'interno di un ambiente virtuale e trasmettere i contenuti non-verbali della stessa, attraverso la loro mimica (anche in modo automatico integrando sistemi di riconoscimento delle emozioni) o possono supportare visivamente l'attuazione di compiti collaborativi complessi.

Molte aziende offrono e sviluppano **agenti virtuali**; la comunicazione aumentata farà leva sull'intelligenza artificiale sociale di questi interlocutori per aiutare le persone che cercano informazioni (assistenza ai clienti ecc).

4.5 *Soluzioni di confluenza uomo-macchina*

La confluenza tra uomo e macchina, in alcuni casi chiamata anche *Human 2.0*, è un campo di ricerca da poco sviluppato, che intende amplificare le capacità fisiche e mentali umane attraverso lo sviluppo di un livello di comunicazione diretta tra uomo e macchina, che non necessa-

riamente si esprime in forma cibernetica. Questa ricerca porterà all'estensione delle capacità sensoriale e percettive, potenzierà la memoria personale o collettiva e accrescerà le prestazioni mentali e cognitive del singolo o distribuite. I primi vantaggi giungeranno in ambito medico alle persone che necessitano assistenza per varie forme di disabilità. Infine si arriverà, come prevedibile conseguenza dell'evoluzione delle interfacce cervello-calcolatore-cervello, alla comunicazione "telepatica" o "empatica" di stati mentali.

5 Il valore aggiunto portato dalla Presenza

Effettuando un'analisi dei benefici portati delle tecnologie della Presenza alle applicazioni commerciali, è possibile individuare quattro elementi principali, che emergono trasversalmente su tutte le aree applicative:

- **visualizzazione 3-D o aumentata**: impatta principalmente su chirurgia, intrattenimento, militare, architettura e costruzioni, educazione. Questa caratteristica attiene alla capacità di mostrare la realtà in modo più efficace, aggiungendo informazioni ed elementi utili, o semplicemente permettendo di rendere visibile ciò che diversamente non lo sarebbe, o lo sarebbe con difficoltà;
- **ambienti immersivi**: sono utili soprattutto nel contesto della terapia psicologica virtuale, delle telecomunicazioni e sistemi di collaborazione, della formazione, dei mondi virtuali, del manifatturiero e design, dell'architettura e costruzioni, dell'educazione. Consente di ricreare completamente la sensazione di essere in un luogo, da soli o con qualcun altro;
- **applicazioni tattili**: portano vantaggio soprattutto alle simulazioni mediche, ai giochi interattivi, alla robotica (ad esempio tele-operazione in chirurgia). Estendono la percezione della virtualità al di là della sola visualizzazione, abbracciando il senso del tatto o dando la sensazione di agire fisicamente su qualcosa di virtuale (attraverso l'uso di una forza retroattiva);
- **co-presenza**: porta vantaggio soprattutto alle telecomunicazioni (e più precisamente alle comunicazioni mediate, ai sistemi di collaborazione e alle applicazioni con umani virtuali) e alla robotica. Riguarda la sensazione di stare con qualcun altro, rendendo semplice e naturale comunicare come se ci si trovasse faccia a faccia, collaborare con qualcun altro, comprendere meglio ed efficacemente la dinamica di un gruppo, ed infine dare ad un agente umano virtuale o un *robot* la capacità di comportarsi in modo più simile a quello umano.

6 Risultati dell'analisi di mercato

I risultati riportati in questo lavoro fanno riferimento alla classificazione di 152 aziende, individuate in quanto produttrici o consumatrici di tecnologie relative alla Presenza. La lista delle aziende è stata analizzata considerando le aree di competenza delle aziende, i mercati all'interno dei quali operano e la collocazione geografica.

La distribuzione delle aziende rispetto alle aree di competenza è presentata nella *tabella 2*.

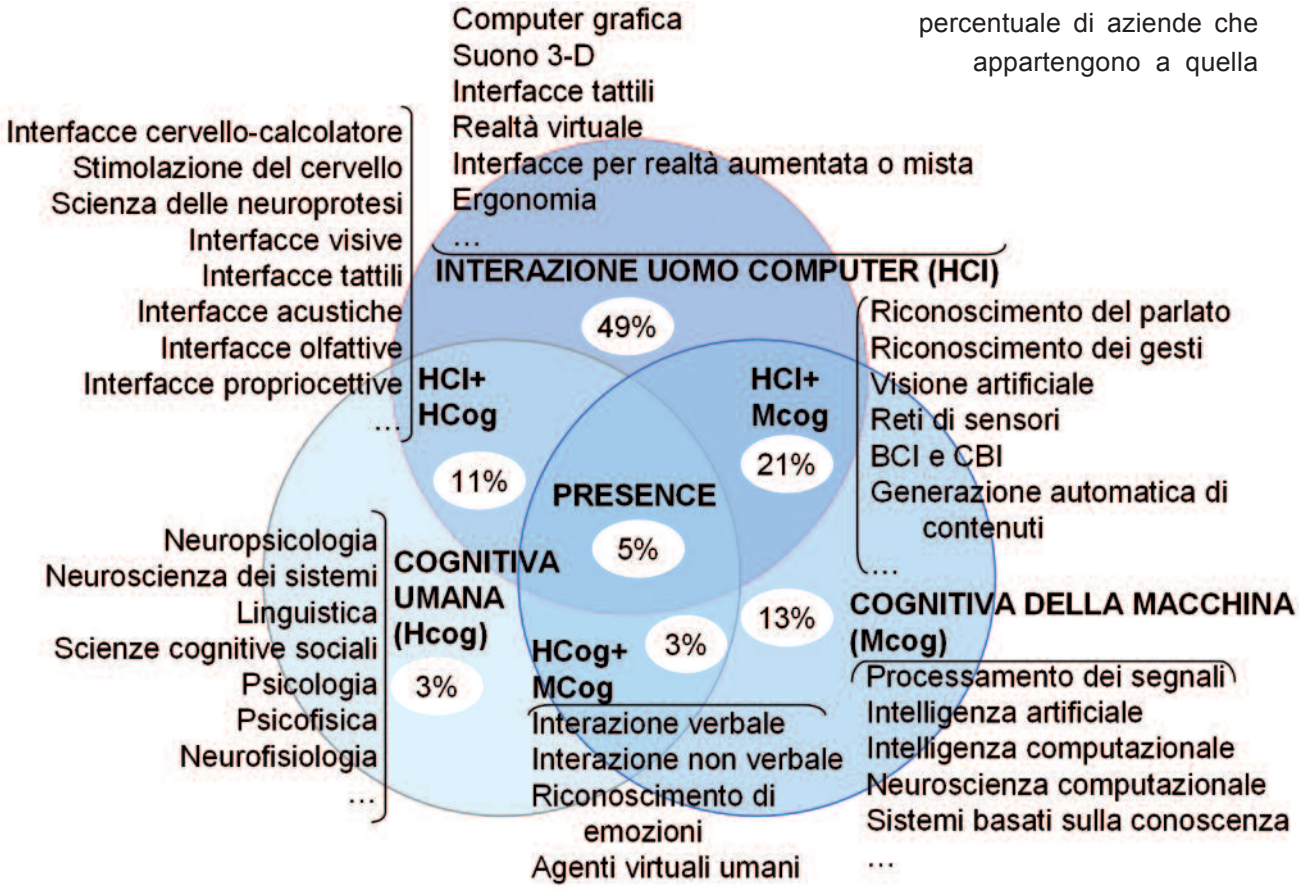
Tabella 2 – Distribuzione numerica delle aziende della Presenza per area di competenza

Area di competenza	Aziende
Realtà virtuale	65
Computer grafica	39
Tecnologie di comunicazione	23
Misure e immagini mediche	19
Interazione uomo-macchina	18
Agenti umani virtuali	13
Realtà mista e aumentata	12
Visione artificiale	12
Interfacce tattili (haptics)	10
Interfacce cervello-calcolatore	8
Intelligenza artificiale	6
Processamento dei segnali	6

La prevalenza di realtà virtuale e computer grafica è il risultato della visione tradizionale di queste tecnologie come elemento centrale per fare Presenza, ma in qualche misura anche del fatto che queste aziende sono quelle più note alla comunità di ricerca.

Basandosi sulla distribuzione delle aziende per area di competenza, le stesse sono state collocate sui tre pilastri disciplinari della Presenza [6] (in *figura 7* il numero nei tondi bianchi mostra la quantità percentuale di aziende che appartengono a quella

Figura 7 – Distribuzione percentuale delle aziende tra i tre pilastri disciplinari della Presenza



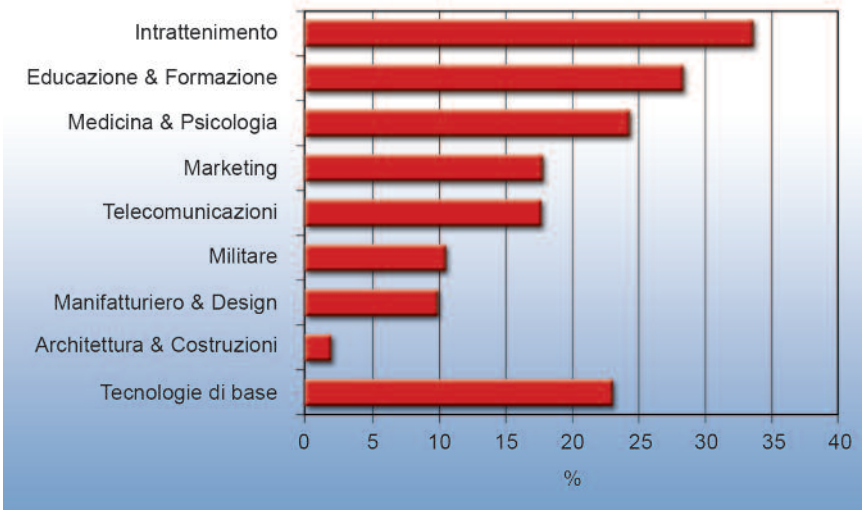


Figura 8 – Distribuzione percentuale delle aziende della Presenza in relazione al mercato in cui operano

area). Per la collocazione di un elevato numero di aziende con competenza nella computer grafica e realtà virtuale, il campo dell'interazione uomo-macchina è il più popolato. Si osserva anche una cerca debolezza nell'impiego in ambito industriale delle competenze maturate dalle scienze cognitive umane.

La distribuzione delle aziende rispetto ai mercati in cui sono attive è invece riportata nella *figura 8*. Si nota che i campi dell'intrattenimento, educazione e formazione, medicina e psicologia sono attualmente i più attivi in termini di aziende che vi operano (questi mercanti non sono tuttavia stati valutati rispetto al volume di affari che generano).

La distribuzione geografica delle aziende è rappresentata in *figura 9*. Il maggior numero totale di aziende nel singolo paese si trova negli Stati Uniti. Comunque il numero totale di aziende nel continente europeo è circa pari a quello americano. Nelle restanti aree sono state identificate aziende in Israele, Cina, Sud Corea e Giappone.

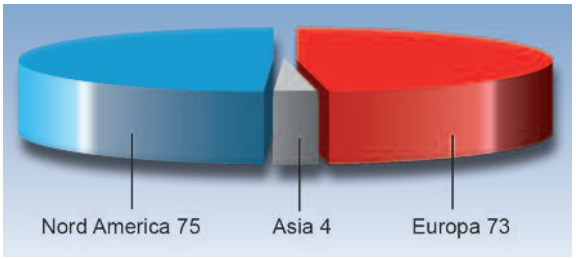


Figura 9 – Distribuzione numerica delle aziende della Presenza per aree geografiche di appartenenza

7 Conclusioni

In questo articolo è stata presentata un'analisi attiva dell'industria mirata a individuare aziende che hanno attualmente adottato, o stanno producendo soluzioni che impiegano tecnologie o applicazioni legate alla Presenza. Sono state identificate 152 aziende, che, a loro volta, possono essere classificate all'interno di otto mercati o aree applicative. I mercati più affollati sono quelli dell'intrattenimento, dell'educazione e formazione e della medicina e psicologia. La maggior parte delle aziende dimostra di avere competenze nelle tecnologie di interazione uomo-macchina, come la computer grafica, la realtà virtuale, la realtà mista e aumentata. Oltre alla realtà virtuale, che è naturalmente la "tecnologia" più prodotta e utilizzata nei mercati della Presenza, molte aziende sono specializzate in computer grafica, centrale nella visualizzazione 3-D e nella costruzione di ambienti immersivi. In modo simile la visualizzazione di immagini mediche è un settore tecnologico molto diffuso. Nella lista di aziende è stato identificato un debole ricorso alle competenze delle scienze cognitive umane, fatto che suggerisce come questo sia un campo attualmente diffuso soprattutto in ambito accademico e di laboratorio. È anche emerso che l'industria è geograficamente ben rappresentata nel Nord America e in Europa.

I trend di innovazione lasciano presagire che la comunicazione e collaborazione tra persone verrà potenziata oltre i limiti fisici comuni, come conseguenza della combinazione di alcune tecnologie di Presenza attualmente ancora in ambito di ricerca. Si può ipotizzare un crescente impiego di tecnologie basate sulle neuroscienze e sul riconoscimento e trasmissione delle emozioni: saranno adottate nuove interfacce capaci di interpretare direttamente i segnali del cervello e di controllare i calcolatori o di riconoscere le intenzioni o gli stati mentali umani.

Questa evoluzione ridurrà i confini tra uomo e macchina, creando un *mix* di realtà e virtualità, e metterà le persone nelle condizioni di manipolare la realtà attraverso applicazioni mediate dalle macchine.

Ringraziamenti

Si ringraziano in particolare il prof. Igor S. Pandzic della Università di Zagabria, Facoltà di Ingegneria Elettrica e Calcolo (Croazia), e i dottori Giulio Ruffini e Cristina Martin-Puig del diparti-

mento di ricerca dell'azienda Starlab di Barcellona (Spagna), per la loro collaborazione nel lavoro di analisi di mercato.

ACRONIMI

AR	Augmented Reality
BCI	Brain to Computer Interface
CAVE	Computer Automatic Virtual Environment
CBI	Computer to Brain Interface
EEG	Elettro-encefalo-gramma
HMI	Human Machine Interface
IST	Information Society Technology
PEACH	Presence Research in Action

BIBLIOGRAFIA

[1] Peach FP6 Coordination Action No 33909. URL: http://peachbit.org

[2] I. Pandzic, G. Zaffiro, *Deliverable D47 Peach Future Markets Issue 2 Annex I*. Peach FP6 Coordination Action No 33909. Apr 2009.

[3] ISPR: International Society for Presence Research. URL: www.temple.edu/ispr

[4] Peach Who is *Who electronic magazine*. URL: http://peachbit.org

[5] IST Event 2006, Networking Session *A common marketplace for Presence, Virtual Reality, Sound and Sense research*, Helsinki, 21-23 Nov 2006. URL: ec.europa.eu/information_society/istevent/2006/cf/network-detail.cfm?id=922

[6] G. Ruffini et al., *Deliverable D31 Visions, Roadmaps, the ERA [Issue 3]*. Peach FP6 Coordination Action No 33909. Apr 2009

[7] K. Kania. (R&D Horizons) *Virtual Reality Moves into the Medical Mainstream*. Medical Device & Diagnostic Industry Magazine. May 2000. URL: www.devicelink.com/mddi/archive/00/05/004.html

[8] S. Rizzo, *Comparing Mental Health Applications Using Individually Administered Virtual Reality and Second Life: Conceptual and Ethical Issues*, AAAS, San Francisco, Feb 2007

[9] Medicine Meets Virtual Reality conference. URL: http://www.nextmed.com

[10] Hype Cycle for Human-Computer Interaction, 2008. ID Number: G00164139. Gartner, Dec 2008.

[11] F. Biocca, C. Harms, J.Gregg, *The Networked Minds Measure of Social Presence: Pilot Test of the Factor Structure and Concurrent Validity*, Presence 2001, 4th Annual International Workshop, Philadelphia, 21-23 May 2001

[12] J. Hauber, H. Regenbrecht, A. Hills, A. Cockburn, M. Billinghurst, *Social Presence in Two- and Three-dimensional Videoconferencing*, Presence 2005, 8th Annual International Workshop on Presence, London, 21-23 Sep 2005

[13] G. Riva, R. Schroeder, G. Zaffiro, *Deliverable D47 Peach Future Markets Annex II*. Peach FP6 Coordination Action No 33909. Apr 2008.

[14] F. Martino, A. Miotto, F. Davide and L. Gamberini, *Exploring Social Network Indices As Cues To Augment Communication and to Improve Social Practices. 1st International Workshop on Maps Based Interaction in Social Networks*, MapISNet '07, Rio de Janeiro, Brasil, 10 Sep 2007

[15] M. C. Brugnoli, F. Morabito, R. Walzer and F. Davide, (2006). *The PASION Project: Psychologically Augmented Social Interaction Over Networks*. PsychNology Journal, 4(1), 103-116. URL: www.psychnology.org/358.php

[16] M. Garau, *The Impact of Avatar Fidelity on Social Interaction in Virtual Environments*, Department of Computer Science University College London, 13 Oct 2003

[17] J. White. It's a Video Game, and an Army

Recruiter. Washington Post. 27 May 2005. URL: www.washingtonpost.com/wp-dyn/content/article/2005/05/26/AR2005052601505.html

[18] Ching-Rong Lin, et al. *Virtual reality for geosciences visualization. Computer Human Interaction*, 1998. Proceedings. 3rd Asia Pacific

[19] J. White, *Virtual Reality and the Built Environment* (2002) Architectural Press, Oxford

[20] G. Zaffiro, G. Ruffini, I. Pandzic, C. Martin-Puig, *Augmenting Human Communication*, Proceedings Poster Session FET09, Praga, 21-23 Apr 2009.

[21] Blue-C Portal URL: http://blue-c.ethz.ch

[22] J. Lite, Star Wars comes to CNN: Network uses holographic journo, Scientific American 5 Nov 2008, URL: www.sciam.com/blog/60-second-science/post.cfm?id=star-wars-comes-to-cnn-network-uses-2008-11-05

[23] Mobile 3D Interface Concept Video from Samsung. URL: www.youtube.com/watch?v=5zzNZGggA_Y

[24] Robust High Speed Natural Feature Tracking on PC. URL: www.youtube.com/watch?v=2Vwt4q8vJ5g

[25] Demo of Orange Vallee's gesture-controlled TV, with Softkinetics technology. URL: www.youtube.com/watch?v=K0-4-FObaRU

[26] M. Monroy, M. Oyarzabal, M. Ferre, A. Campos, J. Barrio, "MasterFinger: Multi-finger Haptic Interface for Collaborative Environments", Lecture notes in Computer Science 2008, Springer Berlin / Heidelberg

[27] J. N. Bailenson, Nick Yee, "Virtual interpersonal touch: Haptic interaction and copre sence in collaborative virtual environments." Multimedia Tools Appl. 37(1): 5-14 (2008)

[28] J. N. Bailenson, E. D. Pontikakis, I. B. Mauss, J.J. Gross, M. E. Jabone, C. A.C. Hutchersond, C. Nassa, O. Johnf "Real-time classification of evoked emotions using facial feature tracking and physiological" International Journal of Human-Computer Studies, 2007, Elsevier

gianluca.zaffiro@telecomitalia.it

AUTORI

Gianluca Zaffiro

laureato in Ingegneria Elettronica presso il Politecnico di Torino nel 1992, con Master in Telecomunicazioni presso COREP/SSGRR nel 1995. Entra in Telecom Italia nel 1994 dove si è occupato di reti ottiche, partecipando all'IEC per gli standard e la pubblicazione di numerosi articoli. Da qualche anno opera nel gruppo Innovation Trends di Telecom Italia Lab. È responsabile per Telecom Italia dell'azione di coordinamento IST FP6 Peach per la ricerca sulla Presenza, sul cui tema ha pubblicato alcuni articoli. Si occupa di elaborare scenari innovativi di medio/lungo periodo di interesse per le telecomunicazioni. Nel 2004-2005 ha collaborato a numerose attività sulla Convergenza Fisso-Mobile. Nel 2003 ha collaborato a dare supporto strategico per l'innovazione tecnologica dell'area Mobile Services di TIM. Nel 2001-2002 ha partecipato al lancio del servizio di Mobile Instant Messaging, TIMCafè, focalizzandosi su aspetti di marketing ■

Tecnologie Powerline e fibre ottiche plastiche: l'esempio Smart Inclusion

TECNOLOGIE

Andrea Bergaglio, Mariano Giunta, Angelantonio Gnazzo

Negli ultimi anni, si è investito nella ricerca di nuove soluzioni di connettività sia wired che wireless, specialmente in ambito residenziale e per edifici già esistenti.

Tra queste, l'utilizzo della rete elettrica quale portante fisico pre-esistente è parso particolarmente attraente. Considerata però la topologia della rete elettrica, non sempre è possibile conoscere a priori le prestazioni offerte dalla tecnologia (powerline). Per ovviare a ciò, nuove soluzioni basate su fibre ottiche in plastica (POF) risultano di interesse per applicazioni in vari ambiti.

Nel presente articolo sono descritte le caratteristiche principali delle tecnologie powerline e POF, e i relativi ambiti applicativi. Viene inoltre fornito un approfondimento su uno specifico esempio relativo al progetto Smart Inclusion e riguardante gli ambiti ospedalieri e scolastici.

1 Introduzione

Le tecnologie di connettività per la realizzazione di reti dati richiedono in genere nuove infrastrutture.

La guida CEI 306-2 [1] prevede la realizzazione di un cablaggio strutturato ed è generalmente applicabile solo nelle nuove lottizzazioni o in edifici in profonda ristrutturazione. Tale soluzione propone un cablaggio a stella indipendente dalle applicazioni per gestire tutti i servizi (multiplay), garantendo un'adeguata flessibilità ed espandibilità.

Anche l'attuale tecnologia Wi-Fi non sempre risulta adeguata in termini di prestazioni (copertura radio, robustezza agli interferenti, massimo throughput, ecc.) per scenari di servizio multiplay.

Risulta quindi fondamentale individuare tecnologie di home network alternative da applicare non solo in ambito residenziale. I macro-requisiti che tipicamente sono presi in considerazione nell'analisi delle soluzioni sono i seguenti:

1) Semplicità di installazione;
2) Bassa invasività della soluzione;
3) Prestazioni adeguate in termini di throughput e copertura;
4) Massima coesistenza e robustezza agli interferenti:
5) Basso costo.

Purtroppo, nessuna delle tecnologie ad oggi disponibili risponde a tutti i requisiti ma, al tempo stesso, è possibile affermare che almeno una delle tecnologie sia sempre applicabile.

2 Tecnologia Powerline

Ormai da più di 10 anni, è in continua crescita l'interesse ad utilizzare la rete elettrica per la trasmissione di voce e dati a larga banda. Alcuni progetti europei del VII° Programma Quadro stanno dedicando risorse all'analisi e alla sperimentazione di soluzione evolute basate sull'utilizzo della rete elettrica [2].

Questo interesse d'altra parte è supportato dalla costante e rapida evoluzione tecnologica della tecnologia Power Line Telecommunication (PLT) [3]. È noto che il principale ostacolo alla realizzazione della società dell'informazione è costituito dagli investimenti legati alle infrastrutture. Ciò vale sia per la rete di accesso, sia, a livello residenziale, per la necessità di appropriati cablaggi, necessari per rendere fruibile i nuovi servizi a larga banda, come ad esempio l'IPTV e altri servizi identificati come "smart home". Ad oggi, le varie evoluzioni delle tecnologie radio (es. Wi-Fi MIMO) ed i classici cablaggi non sono riuscite a fornire una soluzione adatta per tutte le necessità ed in modo particolare, nel caso di edifici esistenti non è agevole installare nuovi cablaggi adatti per la trasmissione di dati ad alta velocità, basati sulla tecnologia Ethernet.

In questo contesto, la tecnologia PLT presenta notevoli vantaggi in quanto permette di:

- utilizzare la rete elettrica esistente;
- distribuire i nuovi servizi all'interno degli ambienti (es. IPTV), eventualmente in modo complementare a tecnologie radio;
- evitare nuovi investimenti per la realizzazione di infrastrutture.

La tecnologia PLT è utilizzata da parecchio tempo anche dalle "utilities" per trasmettere dati o per servizi di comando e telemetria a bassa velocità, come ad esempio la telelettura dei contatori elettrici o la telegestione della rete elettrica ad alta tensione. Questi servizi, che utilizzano la banda di frequenza da 3 a 148,5 kHz, richiedono una bassa velocità trasmissiva e sono regolamentati da un'apposita normativa CENELEC [4].

Nella *figura 1* sono riportate alcune applicazioni abilitate dall'impiego della tecnologia PLT.

L'impiego delle linee elettriche per trasmettere dati con un bit rate molto elevato richiede l'utilizzo di una banda di frequenze che abbia un'estensione da 1 a 30 MHz. Tale porzione di spettro potrebbe presentare problemi di compatibilità elettromagnetica, che includono aspetti legati sia all'immunità ai disturbi, sia alle emissioni di campi elettromagnetici indesiderati. Quest'ultimo risulta quello più difficile da gestire, in quanto l'emissione dei campi elettromagnetici può interferire con vari servizi radio operanti nella stessa banda di frequenze tra cui comunicazioni militari, radioamatori, servizi di radiodiffusione, ecc..

Per prevenire i potenziali problemi di interferenze, dovuti ai segnali PLT, è necessario che gli enti normativi adottino appositi standard con dei limiti di campo elettromagnetico accettabili. Sfortunatamente, lo sviluppo di questi limiti di emissione si sta rivelando un compito molto arduo e controverso, tanto che al momento costituisce un elemento frenante per la diffusione dei sistemi PLT a larga banda.

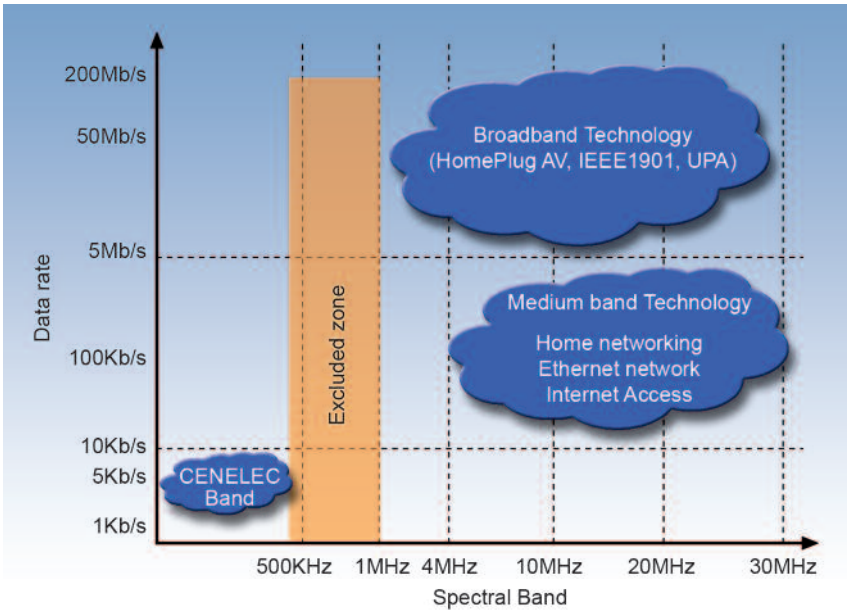


Figura 1 – Applicazioni della tecnologia PLT

2.1 *Il canale trasmissivo utilizzato per le PLT*

Le linee elettriche rappresentano un canale trasmissivo piuttosto severo per i segnali a frequenze maggiori di 50-60 Hz. Infatti, la funzione di trasferimento tra due prese di un'abitazione può essere equiparata ad una linea di trasmissione con molte diramazioni (stub) aventi ognuna un carico variabile. La risposta in ampiezza e fase di tale rete non è costante, ma cambia in modo significativo sia con le frequenze, sia nel tempo, a causa della variabilità dell'impedenza dei dispositivi connessi alla linea elettrica. La variabilità del canale trasmissivo richiede un uso efficiente del mezzo mediante approcci adattativi molto spinti.

Anche la presenza di disturbi sulla linea elettrica, sia impulsivi che selettivi in frequenza, possono rappresentare un problema per un utilizzo efficiente della rete elettrica. Esempi di sorgenti di rumore possono essere: alimentatori di tipo switching, lampade a basso consumo energetico, motori elettrici, regolatori di luminosità, segnali a radiofrequenza indotti sulle linee da trasmettitori radio. L'impatto di queste sorgenti di rumore si ripercuote sul numero di errori nella trasmissione PLT. In generale, quindi, i dispositivi PLT devono disporre di sistemi di correzione degli errori estremamente efficienti del tipo Forward Error Correction (FEC), interleaving, error detection e Automatic Repeat Request (ARQ), in modo da garantire un'adeguata affidabilità del canale.

Inoltre, riveste una certa importanza anche la topologia della rete. In particolare, esiste la possibilità che i segnali presenti in un'abitazione si propaghino in un'altra, con conseguenti possibili problemi di coesistenza, che tendono ad aumentare man mano che aumenta la diffusione dei dispositivi PLT.

2.2 *La tecnologia delle PLT*

Le tecniche di trasmissione utilizzate dalle PLT sono fortemente condizionate da due fattori:

- rendere i segnali immuni ai disturbi ed in generale alle caratteristiche del canale trasmissivo;
- limitare i livelli dei segnali trasmessi per minimizzare le interferenze con i sistemi radio operanti nelle stesse bande di frequenza.

Queste due limitazioni impongono, per avere un bit rate elevato, la distribuzione del segnale su una banda in frequenza molto estesa, tipica-

mente da 1 a 30 MHz. In questo senso, la mancanza di un particolare piano di assegnazione delle frequenze costituisce una limitazione per la coesistenza delle PLT.

La tecnica trasmissiva utilizzata dai sistemi PLT maggiormente diffusi, HomePlug/Intellon e UPA/DS2, è l'Orthogonal Frequency Division Multiplexing (OFDM), che risulta molto robusta specialmente per canali che presentano riflessioni come nel caso delle reti elettriche.

Con la tecnica OFDM, il flusso dati ad alta velocità viene suddiviso in tanti flussi paralleli di velocità molto più bassa. Ogni flusso va a modulare una precisa sottoportante di una numerosa serie che viene trasmessa simultaneamente alle altre, occupando singolarmente una ristretta parte della banda complessiva: la risposta in frequenza del canale si può così considerare piatta su ogni singola frequenza e quindi semplice da equalizzare. Tale meccanismo consente una migliore protezione del "fading" selettivo, in quanto consente di lavorare dinamicamente nelle zone di frequenza dove l'attenuazione ed il rumore sono minori.

Le singole sottoportanti sono modulate con vari schemi, tra cui BPSK, QPSK, QAM, con costellazioni più o meno estese a seconda delle caratteristiche del canale. Con queste modulazioni si può raggiungere un'efficienza spettrale piuttosto elevata, fino a 7,5 bit/s/Hz.

Uno degli svantaggi della tecnologia OFDM è però l'elevato rapporto picco/valor medio dei segnali e ciò può creare potenzialmente problemi di interferenza verso altri sistemi.

La tecnologia OFDM, anche se complessa e ben conosciuta da molto tempo, ha avuto un forte sviluppo con l'evoluzione dei sistemi DSP a basso costo, e oltre all'applicazione nei sistemi PLT, ha avuto già successo con altre tecnologie trasmissive, quali DVB (Digital Video Broadcasting), DAB (Digital Audio Broadcasting) e xDSL.

2.3 *Caratteristiche tecnologiche delle PLT di ultima generazione*

Intellon ha recentemente messo sul mercato un chipset denominato INT 6400 conforme alla specifica HomePlug AV. Questo dispositivo è stato ottimizzato per applicazioni multimediali e ha la capacità di fornire, sulla rete elettrica, uno streaming dati a livello fisico fino ad un massimo di 200 Mbps, con un throughput di picco di 60 Mbps a livello TCP e di 120 Mbps a livello UDP.

All'interno del chipset INT 6400 è stato integrato un microcontrollore della serie ARM 9 con cui vengono gestite tutte le operazioni del sistema. Ad esempio, attraverso la specializzazione del firmware del microcontrollore, possono essere eseguite all'accensione alcune verifiche sul mezzo trasmissivo per fornire indicazioni all'utente sullo stato del sistema attraverso dei LED.

Le caratteristiche principali del chipset INT 6400 possono essere riassunte come segue:

- conformità alla specifica HomePlug AV per la gestione dei livelli MAC e fisico;
- accesso al canale con protocollo MAC basato su TDMA e con priorità di accesso basata su CSMA/CA;
- interfaccia Ethernet indipendente dal mezzo (media indipendent interface - MII);
- supporto funzionalità IGMP e sessioni multicast;
- Windowed OFDM con più di 1000 sottoportanti ortogonali con adattamento individuale al canale in modo dinamico con la possibilità di essere spente singolarmente (notch fino a -30dB);
- Advanced Turbo Code Forward Error Correction (su brevetto di France Telecom);
- supporto di schemi di modulazione a 1024/256/64/16/8-QAM, QPSK, BPSK e la modulazione robusta ROBO (per gestione traffico multicast);
- 128-bit Advanced Encryption System (AES) con gestione programmabile delle chiavi;
- filtri notch programmabili per ridurre il problema dell'interferenza elettromagnetica.

Oltre alla tecnologia HomePlug AV, è disponibile sul mercato anche una tecnologia alternativa che si basa su una specifica preliminare in fase di definizione da parte dell'Universal Powerline Association (UPA). A tale proposito, l'azienda

spagnola DS2 ha messo in commercio un chip-set denominato DSS9010, con le seguenti caratteristiche principali:

- ottimizzazione per la trasmissione di segnali video, VoIP e dati, per comunicazione fino ad una velocità di 200 Mbps a livello fisico;
- gestione della qualità di servizio (QoS);
- gestione trasmissione multicast;
- protezione dei dati con una codifica 3DES a 168 bit;
- filtri notch programmabili, anche da remoto, per ridurre il problema dell'interferenza elettromagnetica.

2.4 *Aspetti normativi*

Lo sviluppo applicativo delle PLT su ampia scala è condizionato dallo sviluppo di standard per:

- i diversi apparati (bridge Ethernet, USB, modem, ecc.) che consenta la connessione alla rete a 230 Vac;
- definire i limiti di emissione al fine di garantire la coesistenza con i servizi radio operanti nella stessa banda di frequenza (1 - 30 MHz e in prospettiva fino a 100 MHz);
- definire meccanismi relativi alla coesistenza di diverse tecnologie PLT nella stessa rete elettrica o della stessa tecnologia di diversi costruttori.

Per quanto riguarda gli aspetti di emissione, occorre precisare che la Commissione Europea ha pubblicato nel 2001 un mandato verso l'ETSI ed il CENELEC per lo sviluppo di una norma che permetta di gestire le problematiche EMC per tutte le reti di telecomunicazione, incluse le linee elettriche con dispositivi PLT. Gli aspetti normativi concernenti le emissioni degli apparati PLT sono attualmente demandati all'IEC/CISPR, che sta ormai da tempo lavorando alla modifica dell'attuale norma CISPR 22 (equivalente in Europa alla norma CENELEC EN 55022) [5], applicabile a tutte le apparecchiature della tecnologia dell'informazione e TLC, per includere i requisiti di emissione dei disturbi anche delle PLT.

2.4.1 Aspetti normativi per la compatibilità elettromagnetica

Come visto in precedenza, gli apparati PLT utilizzano frequenze fino a 30 MHz (ed in prospettiva si raggiungeranno i 100 MHz) su un mezzo condiviso non bilanciato e non schermato, con conseguenti impatti relativi alle emissioni nello spettro comunemente utilizzato dai servizi radio.

Inoltre, l'impiego di tali frequenze, necessario per trasmettere a bit rate elevato, aumenta la tendenza all'irradiazione di campi elettromagnetici.

A livello normativo internazionale, il fatto che non siano stati definiti i limiti di irradiazione costituisce un punto di attenzione da non trascurare in merito all'impiego di sistemi PLT. La contrapposizione di interessi tra gli utilizzatori dello spettro radio ed i costruttori di apparati PLT ha frenato lo sviluppo di una normativa di riferimento con conseguente adozione, in Europa, di approcci diversi per la marcatura CE e l'immissione sul mercato dei dispositivi PLT.

In particolare, i sistemi radio operanti nella stessa banda delle PLT che possono essere soggetti ad interferenza sono:

- **servizi di radio diffusione**: sono essenzialmente di due tipi: onde medie (MF) da 0,5265 a 1,6065 MHz e onde corte da 3,9 a 26,1 MHz;
- **servizi radiomatoriali**: generalmente sfruttano antenne installate sui tetti delle abitazioni per ricevere dei segnali anche piuttosto deboli appena sopra il rumore di fondo, e per questo sono facilmente disturbabili;
- **servizi mobili**: in genere sono servizi marittimi, aeronautici, terrestri, civili o militari: la maggior parte di questi utilizza le onde corte per trasmettere a lunga distanza e vengono utilizzati quando a causa della distanza non è possibile utilizzare le trasmissioni VHF. Considerando l'importanza di questi servizi, la normativa USA FCC proibisce l'uso delle PLT nelle zone in prossimità di porti, aeroporti e basi militari;
- **servizi fissi**: si tratta di collegamenti punto-punto utilizzati per servizi meteorologici o per

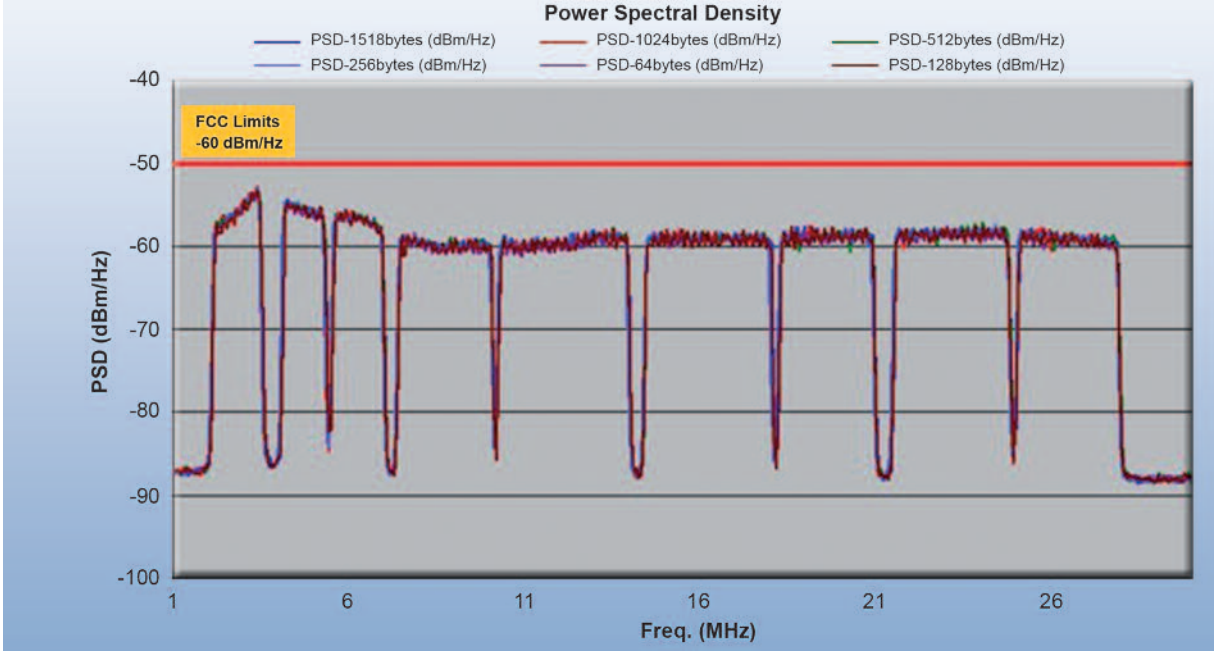
trasmissioni di dati aeronautici in genere curate da amministrazioni governative. Valgono le stesse raccomandazioni per i servizi mobili;
- **ricerca e soccorso**: sono allocate alcune frequenze specifiche per l'emergenza marittima. Questi servizi richiedono ovviamente la massima protezione dalle interferenze;
- **altri servizi**: ricerca spaziale e radioastronomia, servizi di radiolocalizzazione.

Per lo sviluppo di una norma di prodotto, al fine di proteggere i servizi elencati, l'IEC/CISPR, già all'inizio del 2005 ha attivato un New Work Item Proposal (NWIP), con l'obiettivo di emendare l'esistente norma CISPR 22 per includere i requisiti relativi all'emissione delle apparecchiature PLT.

L'attività di questo task group, tuttavia, non ha portato, nel corso di questi anni, ad una condivisione dei requisiti. Tuttora i punti di maggiore contrasto riguardano i limiti e le metodologie di misura dei segnali irradiati dagli apparati PLT in ambiente residenziale. Le nuove ipotesi di lavoro prevedono, oltre alla definizione di limiti di emissione un po' più blandi, anche la possibilità di imporre ai costruttori di PLT delle tecniche di mitigazione basate sull'uso di filtri notch statici e/o programmabili, anche da remoto, e di meccanismi automatici di controllo della potenza iniettata sulle linee elettriche. Attualmente, infatti, tutti i dispositivi trasmettono una potenza spettrale di -50dBm/Hz indipendentemente dall'attenuazione della linea (*figura 2*).

Parallelamente allo sviluppo di una norma di prodotto per le PLT, in ambito europeo, è in fase di redazione da parte di un comitato congiunto ETSI/CENELEC una normativa relativa alle emissioni dalle reti di telecomunicazioni incluse le reti elettriche con apparati PLT. Dopo cinque anni di attività, recentemente è stato raggiunto un primo accordo con la preparazione di tre bozze di norma attualmente (maggio 2009) in fase di inchiesta pubblica:
- pr EN 50521-1 EMC network standards Part 1: Wire-line telecommunications networks using telephone wires;
- pr EN 50521-2 EMC network standards Part 2: Wire-line telecommunications networks using coaxial cables;
- pr EN 50521-3 EMC network standards Part 3: Wire-line telecommunications networks using power lines.

Figura 2 – Esempio di PSD misurata in laboratorio su apparati PLT



In aggiunta agli aspetti normativi accennati in precedenza, attualmente la regolamentazione per la protezione dei servizi radio si basa sulla seguente legislazione:

- a livello europeo: Direttiva 2004/108/EC ai fini dell'immissione sul mercato dei prodotti, e varie Raccomandazioni CEPT relative alla pianificazione dello spettro radio;
- a livello nazionale: Codice delle Comunicazioni (D.lgs.vo 1 agosto 2003, n. 259) ai fini della gestione delle interferenze dei servizi radio e Legge 24 novembre 1981, n. 689 per le relative violazioni amministrative.

La Conferenza Europea delle Amministrazioni Postali (CEPT), attraverso il suo comitato ECC, ha pubblicato nel mese di maggio del 2003 un report (ECC report 24) con uno studio sulla compatibilità tra sistemi di comunicazione in cavo e i servizi radio. Sulla base di questo report, l'ECC ha pubblicato, nel mese di giugno 2005, la raccomandazione ECC 05/04, con la definizione dei criteri di valutazione delle interferenze prodotte dagli impianti verso i servizi radio. L'applicazione dei principi definiti in ambito CEPT è demandata ai singoli stati membri.

Un metodo per limitare le problematiche di interferenza verso servizi radio è l'inserzione di filtri, a determinate frequenze, sulla banda usata dagli apparati PLT. In particolare, alcuni costruttori implementano di default i filtri per le frequenze dei radioamatori (IARU): si veda *figura 2*. La norma ETSI TS 102 578 ("Coexistence between PLT Modems and Short Wave Radio broadcasting Services") prevede invece l'inserimento di filtri dinamici. Il sistema PLT rivela eventuali portanti radio presenti nelle vicinanze e, in modo automatico, inserisce il filtro per quella frequenza. Benché esistano prototipi, apparati PLT che implementino questa funzionalità non sono ancora commercialmente disponibili.

2.4.2 Evoluzione della normativa sulla coesistenza

La coesistenza per gli apparati PLT riguarda i seguenti aspetti:

- **coesistenza tra gli apparati di costruttori diversi**; la coesistenza tra apparati di differenti costruttori rappresenta al momento uno degli aspetti di maggior attenzione quando si devono realizzare LAN basate su powerline. I tre principali competitor in questa tecnologia (HomePlug AV-Intellon, UPA-DS2, CEPCA-Panasonic) basano le specifiche dei propri prodotti su "standard" proprietari e non compatibili tra loro. L'utilizzo quindi di tecnologie differenti sulla stessa rete elettrica porta alla conseguenza di un malfunzionamento complessivo della LAN su powerline;
- **coesistenza tra gli apparati dello stesso costruttore**; la coesistenza tra apparati dello stesso costruttore ha come effetto principale la riduzione delle prestazioni in presenza di più di una coppia di apparati connessi alla stessa rete elettrica. Pur essendo questo un fenomeno difficilmente eliminabile, il recente standard ITU-T G.9960 (ex G.hn) [6] ha preso in considerazione la possibilità di unificare le tecnologie residenziali su rame (powerline, doppino telefonico e coassiale), cercando di minimizzare, mediante opportune tecniche, questo specifico problema di coesistenza. A tale standard hanno contribuito anche i produttori sopra menzionati;
- **coesistenza tra gli apparati PLT ed altre tecnologie trasmissive (per esempio VDSL2)**; tra le varie problematiche di coesistenza dei sistemi powerline, esiste quella con le tecnologie di accesso su doppino telefonico VDSL2 operanti anch'esse a frequenze fino a 30 MHz. È infatti possibile un'induzione tra il cavo elettrico dove sono attestate le PLT per connettere AG e STB nel servizio IPTV e il doppino telefonico su cui è presente il segnale VDSL2. Benché non esistano ancora normative specifiche per questa tematica, ETSI (su proposta ed esperienza di Telecom Italia) ha programmato per maggio 2009 un PlugTest proprio sulla coesistenza tra queste due tecnologie, con l'obiettivo di valutazione degli effetti di questo accoppiamento sulle prestazioni del VDSL2. Da questi risultati potranno poi essere implementate funzionalità sui nuovi ap-

parati VDSL2 per limitare anche in questo caso la coesistenza delle tecnologie.

2.5 *Scenari applicativi delle PLT*

Di seguito si riportano brevemente alcuni scenari applicativi legati all'impiego delle PLT.

2.5.1 Servizio IPTV

In ambito residenziale, la tecnologia PLT è utilizzata come "Cable Replacement" per il collegamento punto-punto tra l'AG e il STB, specialmente nei casi in cui il tradizionale cavo UTP per Ethernet non sia applicabile, per esempio quando i due apparati da connettere sono ubicati in stanze diverse.

I bridge Ethernet/PLT sono stati ottimizzati per avere una bassa latenza, un'alta affidabilità proprio per applicazioni di streaming video.

In prospettiva è ipotizzabile l'impiego delle PLT in scenari più complessi, quali il servizio IPTV in configurazione multi-room, oppure per la realizzazione dell'intera LAN residenziale, come mostrato in *figura 3*.

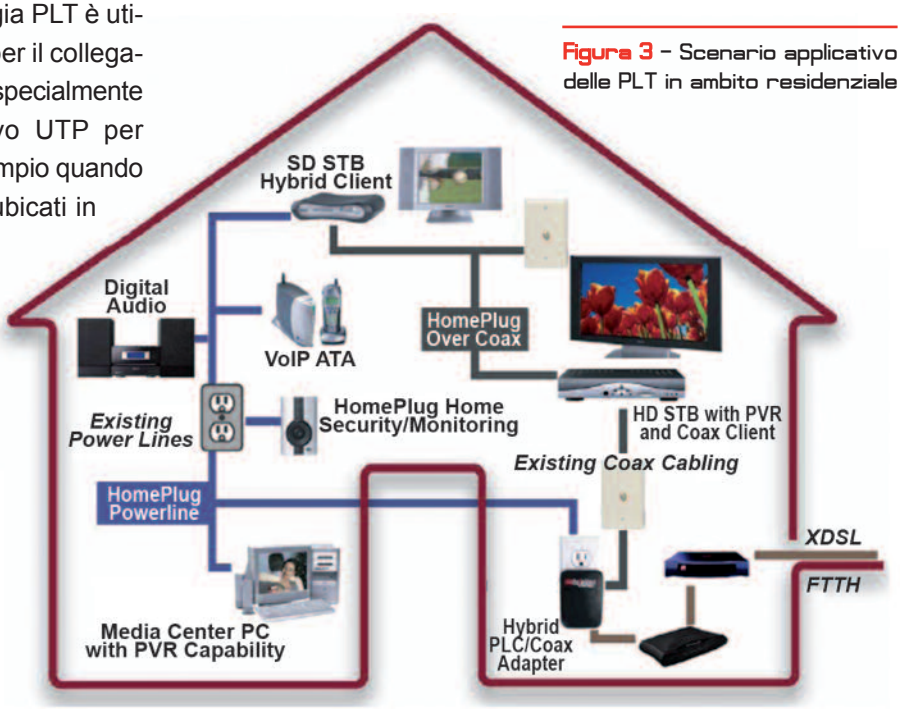


Figura 3 – Scenario applicativo delle PLT in ambito residenziale

2.5.2 Domotica

Tra le applicazioni di domotica che possono essere realizzate sfruttando le potenzialità della tecnologia delle PLT, si citano, oltre alle classiche HVAC (Heating, Ventilation and Air Conditioning), l'Energy Saving, la sorveglianza e il controllo di accesso (es. sistemi anti-intrusione), la telemedicina e il telesoccorso.

Nel corso degli anni, sono state sviluppate diverse soluzioni tecnologiche per le modalità di comunicazione su rete elettrica, tra cui:

- KONNEX: standard europeo CENELEC EN 50090;
- LonWorks: standard sviluppato da Echelon e recepito in ambito EIA (Electronic Industries Alliance) con la norma EIA-709.

Tutte le soluzioni prevedono l'impiego delle PLT a bassa velocità, che risultano adeguate per le applicazione tipiche di domotica.

La *figura 4* mostra un esempio di utilizzo delle PLT per il collegamento di un elettrodomestico con l'AG (Access Gateway), al fine di consentirne il controllo da remoto.

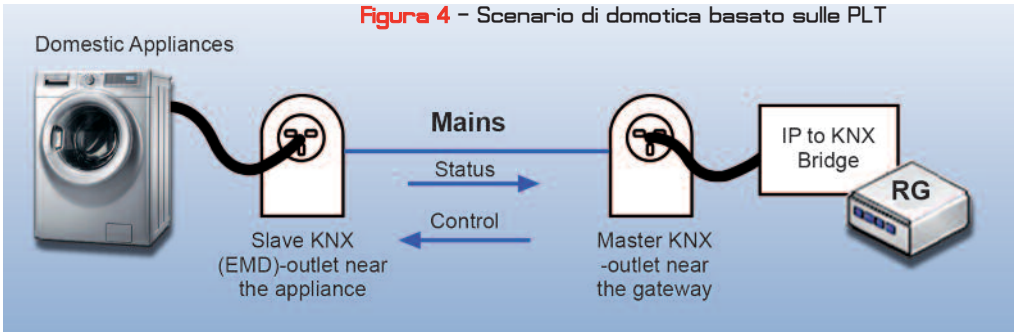


Figura 4 – Scenario di domotica basato sulle PLT

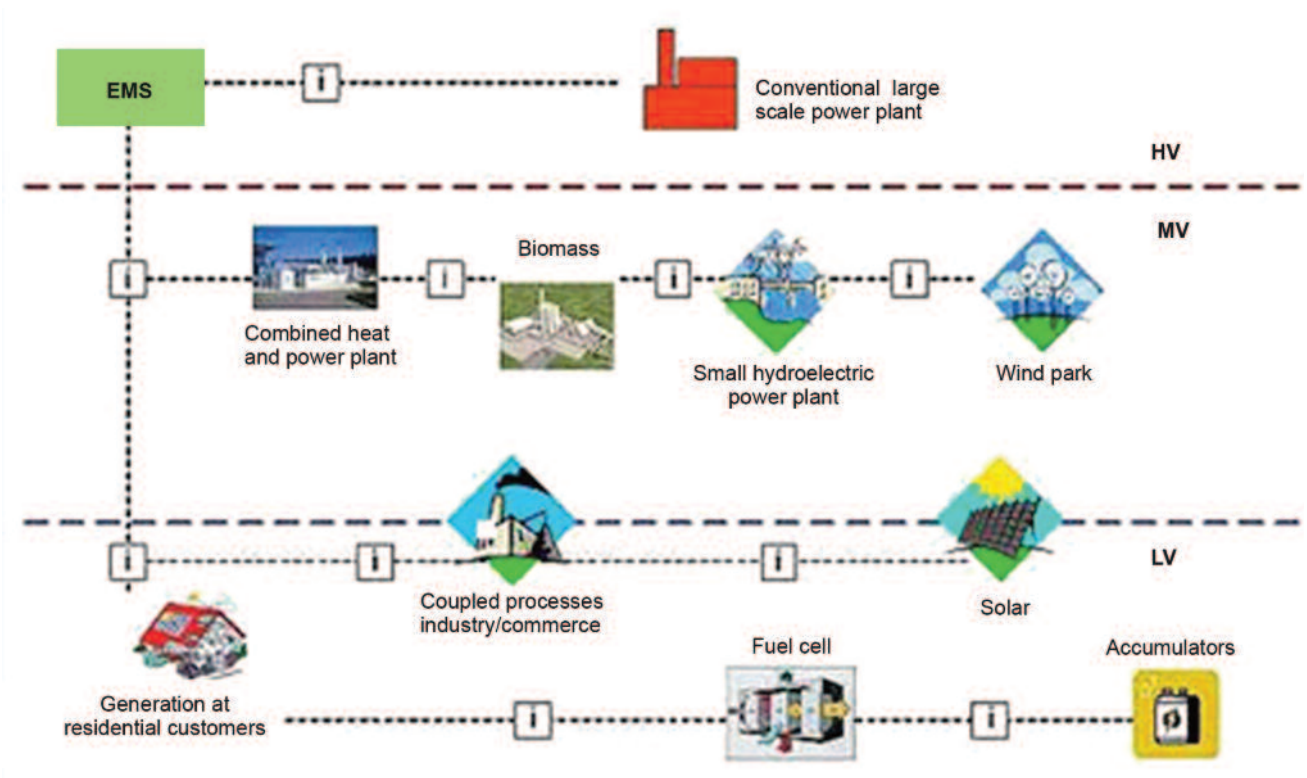
2.5.3 Smart Grid

L'interconnessione delle sorgenti distribuite di energia, incluse le varie forme di energie rinnovabili, richiedono una complessa infrastruttura di sensori ed attuatori. Questi dispositivi devono operare in modo integrato attraverso un sistema di controllo con scambi continui di informazioni che garantiscano una certa sicurezza ed affidabilità. In *figura 5* è riportato uno scenario tipico di generazione distribuita di energia.

In un simile contesto, l'uso delle PLT costituisce uno strumento fondamentale per la trasmissione delle informazioni tra i vari sistemi, proprio per la loro peculiarità di non richiedere nuove infrastrutture trasmissive, sfruttando la rete elettrica pre-esistente con vantaggi sia economici sia ambientali. Anche la Commissione Europea promuove attività di Energy Saving & Energy Control basate anche su tecnologie PLT [7].

Per quanto riguarda la standardizzazione, in ambito IEEE, il progetto P2030 focalizzato su Smart Grid sta prendendo in considerazione le PLT come mezzo principale per la trasmissione dei dati.

Figura 5 – Esempio di generazione distribuita di energia



3 Le fibre ottiche in plastica

Le fibre ottiche in plastica (Plastic Optical Fibre-POF) sono composte, come quelle convenzionali in vetro, da un nucleo (core) rivestito da un mantello (cladding), avente indice di rifrazione più basso. Il nucleo è normalmente costituito di Polimetilmetacrilato (PMMA), rivestito da un sottile strato di polimero fluorurato. Il profilo di indice di rifrazione adottato è quello a gradino (SI-Step Index).

La luce si propaga all'interno del nucleo della fibra plastica grazie al mantello, che agisce come uno specchio, riflettendola e guidandola lungo il cammino descritto dalla fibra (vedi *figura 6*).

La fibra plastica funziona sostanzialmente come una comune fibra ottica in vetro, pur differenziandosi per alcune caratteristiche. In partico-

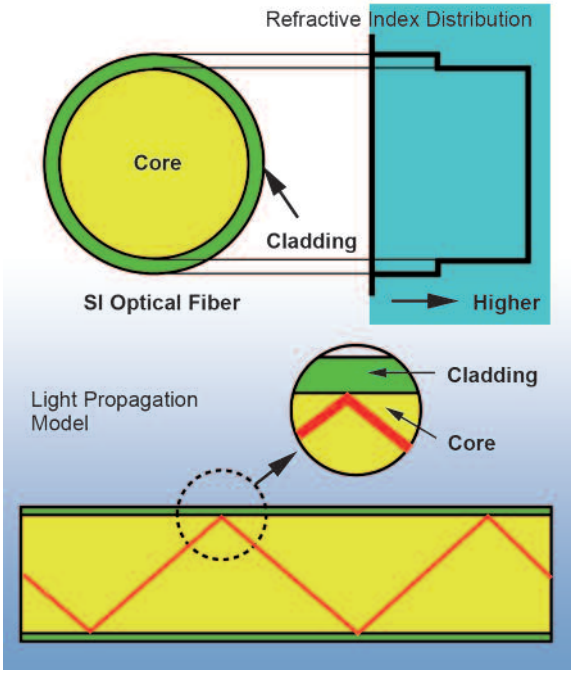


Figura 6 – Struttura di una fibra ottica in plastica

lare, il diametro del nucleo della fibra plastica è di 980 μm, quindi molto maggiore rispetto a quello di una fibra ottica convenzionale che ha un diametro del nucleo compreso tra 8 e 10 μm per quanto riguarda le fibre monomodali, ed un diametro di 50 o 62,5 μm se riferito a fibre multimodali.

Le dimensioni generose del nucleo consentono il funzionamento anche in caso di allineamento non perfetto tra i componenti elettro-ottici e quindi, non sono necessarie tecniche sofisticate per l'installazione che altrimenti richiederebbero personale specializzato.

La fibra plastica è trasparente a lunghezze d'onda diverse da quelle della fibra ottica in vetro. Mentre le fibre ottiche in vetro hanno un massimo di trasparenza nella regione vicino all'infrarosso (800-1600 nm), le fibre ottiche plastiche hanno un massimo di trasparenza, o detto in altri termini un minimo di attenuazione, nella zona corrispondente alla luce visibile.

Come si vede in *figura 7*, vi sono tre finestre di trasmissione corrispondenti ai minimi di attenuazione: 525 nm (luce verde), 575 nm (luce ambra), 650 nm (luce rossa). L'utilizzo di luce visibile permette di verificare immediatamente il funzionamento del cavo: basta controllare se "esce" la luce. Si evidenza che con la luce visibile, inoltre, non si rischiano inconsapevoli esposizioni alla luce, come invece nel caso degli infrarossi (tradizionali fibre in vetro).

I componenti ottici (LED e fotodiodi) più utilizzati per la trasmissione sono quelli a luce rossa, perchè più comuni ed economici, anche se sono stati realizzati sistemi di trasmissione a luce verde.

La massima distanza raggiungibile con questa tecnologia è minore di quella ottenibile con le fibre in vetro, a causa sia del fatto che l'attenuazione delle fibre plastiche è circa 1000 volte superiore a quelle delle fibre in vetro, sia dal fatto

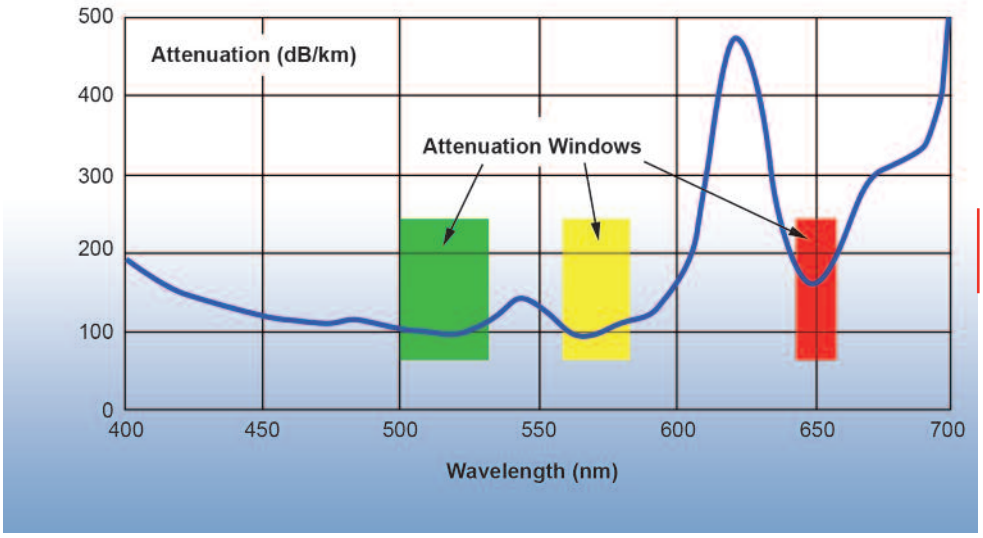


Figura 7 – Attenuazione spettrale delle fibre in plastica

che, essendo fibre multimodali, hanno il problema di una maggiore dispersione modale e dispersione cromatica [8]. Le distanze sono comunque adeguate per gli scenari applicativi in ambito indoor.

3.1 *Possibili applicazioni delle POF*

Le fibre ottiche plastiche hanno trovato originariamente il loro impiego soprattutto nel campo dell'automotive, mentre attualmente le applicazioni riguardano anche i settori della sensoristica e delle telecomunicazioni.

3.1.1 Il servizio IPTV

Nel campo delle telecomunicazioni, l'interesse dell'impiego di fibre è principalmente rivolto alle reti a larga banda e all'offerta di servizi video di IPTV all'interno della casa (home networking): in particolare, l'utilizzo delle fibre plastiche (vedi *figura 8*) riguarda il collegamento tra il modem/router broadband (Access Gateway - AG) connesso ad una delle prese dell'impianto telefonico e il Set Top Box (STB) posizionato vicino la televisione. Le attuali tecnologie permettono facilmente di avere un bit rate pari a quello della tecnologia Fast Ethernet (100 Mbit/s).

Ad oggi, per garantire la comunicazione bi-direzionale tra AG e STB, necessaria per il servizio IPTV, è necessario utilizzare una coppia di fibre plastiche. In futuro, in base alla roadmap presentata da alcune aziende del settore, potrebbe essere sufficiente utilizzare una singola fibra.

Il collegamento punto-punto tra AG e STB può essere realizzato mediante due adattatori elettro-ottici (Ethernet/POF), entrambi alimentati da corrente. Alcuni produttori stanno integrando l'interfaccia per fibra plastica su modem/router e STB, eliminando la necessità degli alimentatori.

3.1.2 Altri scenari applicativi in ambito residenziale

In merito all'evoluzione previste per i servizi offerti, è possibile ipotizzare i seguenti scenari:

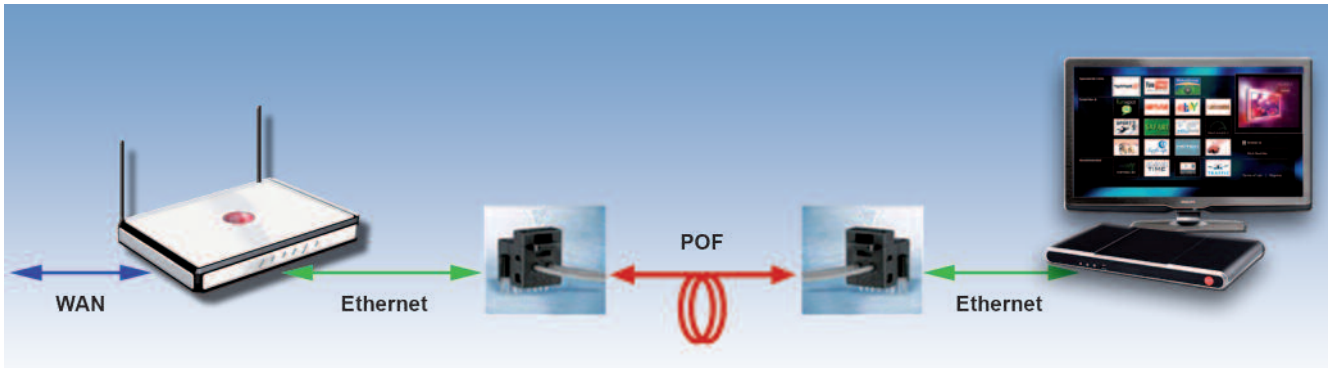
- IPTV multi-room;
- estensione della copertura radio (Wi-Fi).

Le figure della pagina seguente si riferiscono all'ambito Fibre To The Home (FTTH) "Brownfield", dove si prevede l'impiego di una terminazione di rete (ONT) separata dall'AG.

3.2 *Panoramica dei prodotti commerciali*

Di seguito sono presentate le tipologie e le caratteristiche principali dei prodotti richiesti per l'impiego delle POF come soluzione di connettività indoor. Tutte le soluzioni proposte sono basate sulla conversione del segnale elettrico-ottico e sulla tecnologia Fast Ethernet.

Figura 8 – Utilizzo delle fibre plastiche in ambito home networking



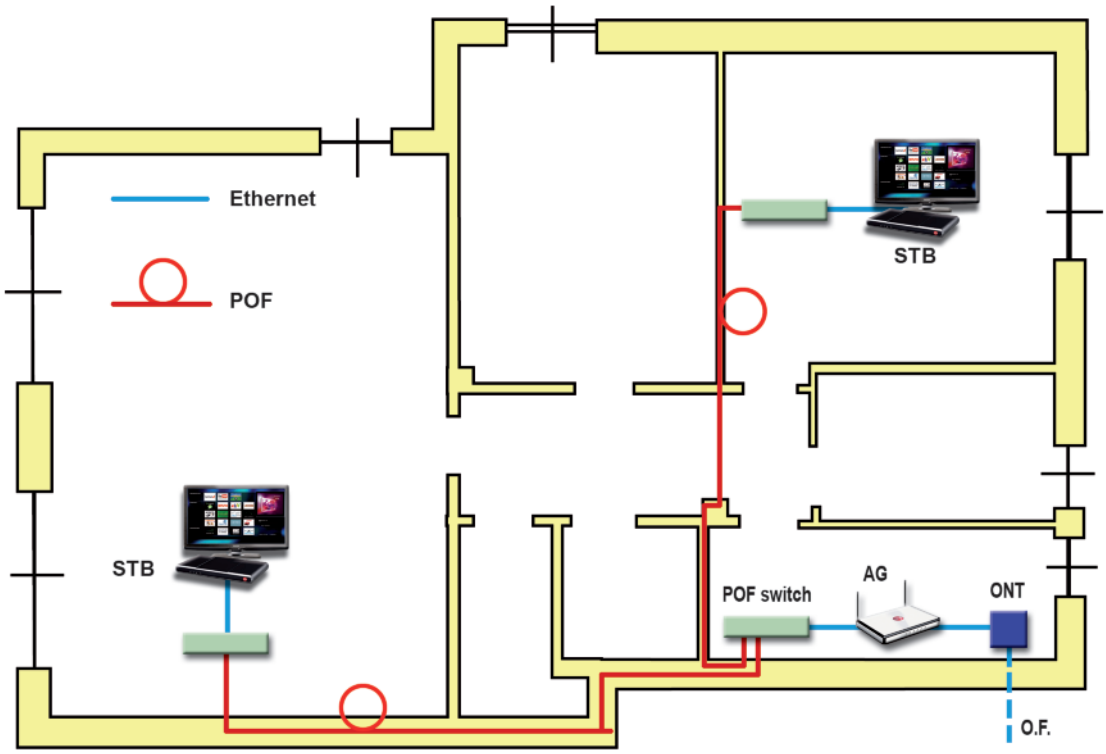


Figura 9 – IPTV multi-room

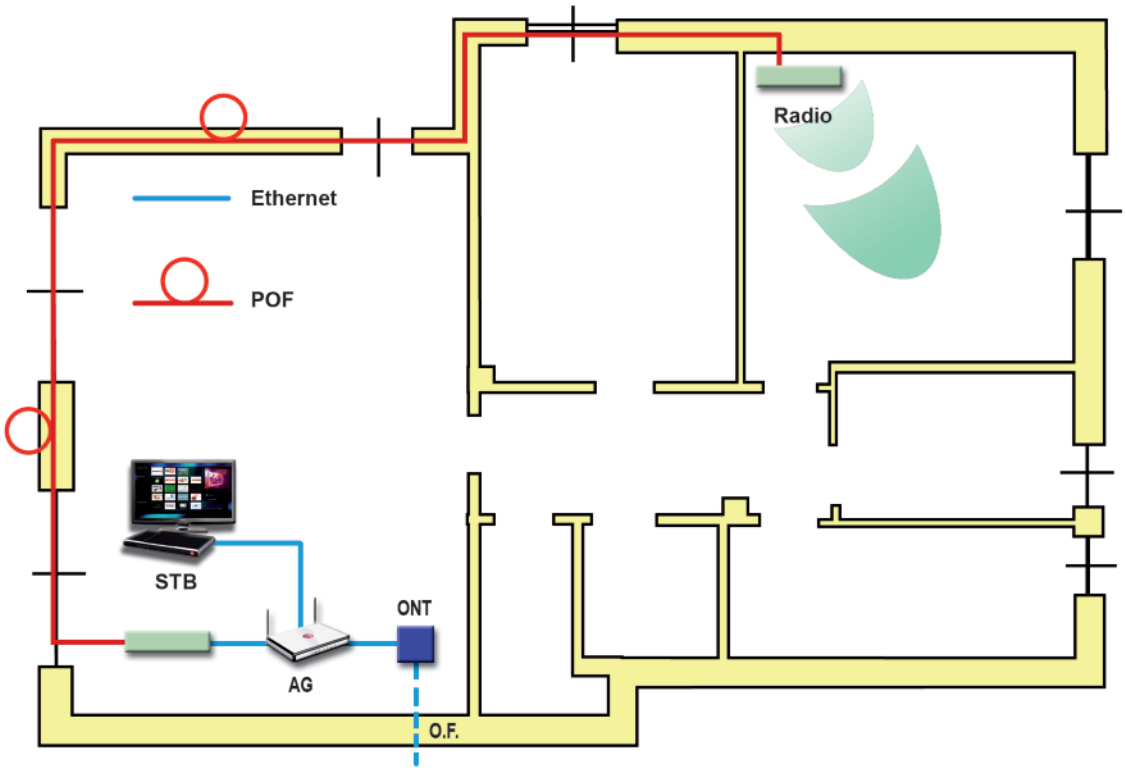


Figura 10 – Estensione della copertura radio

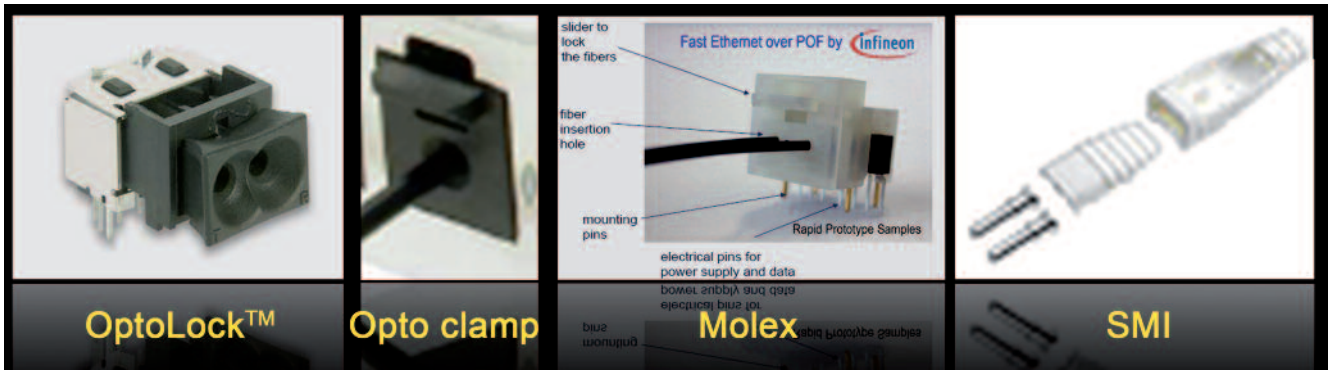


Figura 11 – Principali connettori per POF

3.2.1 Adattatori Fast Ethernet - POF

Si tratta di apparati attivi, ossia che richiedono di essere alimentati. L'alimentatore può essere integrato nello stesso adattatore, oppure esterno (tipicamente di tipo switching). L'adattatore è caratterizzato dal connettore per POF (*figura 11*), le cui tipologie principali sono:

- OptoLock™ (marchio registrato da Firecomms);
- Opto clamp;
- Molex;
- SMI (Small Multimedia Interface).

3.2.2 Fibra ottica plastica

Si ricorda che per un collegamento punto-punto, per garantire una comunicazione bi-direzionale, è necessario utilizzare una coppia di POF. Le caratteristiche delle fibre possono variare, in particolare in relazione a:

- diametro esterno, tipicamente pari a 1,5 o 2,2 mm per singola fibra;
- apertura numerica (NA), tipicamente di valore 0,3 o 0,5, da cui dipendono le prestazioni, per esempio in termini di massima distanza raggiungibile per una connessione punto-punto per garantire un bit-rate fissato (es. 100 Mbit/s). Minore è l'apertura numerica e migliori sono le prestazioni (a scapito di un costo maggiore).

3.2.3 Switch ottici

Sono apparati attivi, dotati di alimentatore esterno, che consentono la realizzazione di scenari di tipo punto-multipunto. Uno switch ottico, tipicamente, dispone di una porta Fast Ethernet e più interfacce POF, anche se esistono apparati solo con interfacce POF.

3.2.4 POF kit

Sono disponibili in commercio i cosiddetti POF kit, soluzioni adatte per connessioni punto-punto e nate per l'ambito residenziale, in particolare proprio per la connessione tra modem/router e STB necessaria per il servizio IPTV.

Un POF kit (*figura 12*) comprende:

- 2 adattatori elettro-ottici (con alimentatore esterno o integrato);
- 30 m di POF (cavo con una coppia di fibre);
- 1 cutter;
- 2 cavi UTP (es. di lunghezza pari a 1 m).

3.2.5 Altri apparati

Sono inoltre disponibili in commercio altre tipologie di prodotti, quali per esempio:

- adattatore elettro-ottico Fast Ethernet - POF con alimentatore esterno e con possibilità di fissaggio alla parete mediante viti;

Figura 12 – Esempio di POF kit

- adattatore elettro-ottico USB - POF alimentato via USB;
- adattatore elettro-ottico Fast Ethernet - POF alimentato via USB;
- adattatore elettro-ottico PCI - POF per desktop.

3.2.6 Evoluzione tecnologica

Sono disponibili a livello prototipale "prese a muro" con adattatori elettro-ottici integrati, al momento non ancora presenti sul mercato italiano. Esternamente sono dotate di 2 connettori RJ45 (per cavi UTP/Ethernet), mentre internamente sono disponibili 2 connettori ottici per POF, che possono essere quindi inseriti nelle canaline pre-esistenti dell'impianto elettrico (*figura 13*).

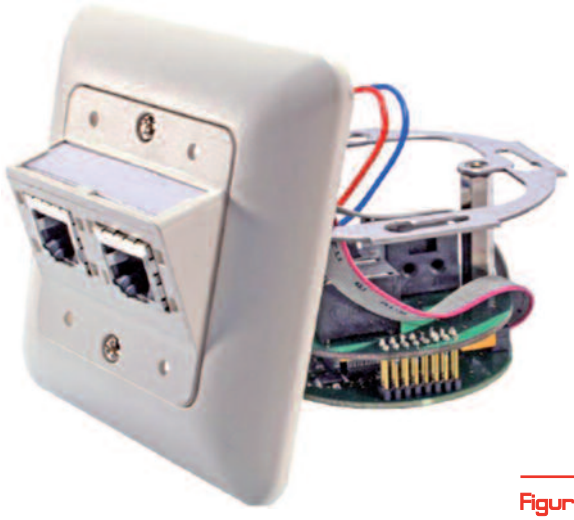


Figura 13 – Prototipo di presa a muro che abilita l'impiego di POF

I prototipi sono in fase di valutazione mediante prove di laboratorio e in ambiente reale per comprendere l'effettiva possibilità di impiegare tali soluzioni innovative negli impianti elettrici pre-esistenti.

3.3 *Installazione*

In base al particolare ambiente, la posa delle POF potrà avvenire sfruttando la controsoffittatura e le canaline pre-esistenti esterne o interne alla parete. Si fa presente che grazie alle dimensioni ridotte delle fibre (1,5 o 2,2 mm di diametro) e alle ottime caratteristiche meccaniche (possono essere piegate e tirate senza essere danneggiate), le POF possono essere inserite anche nelle canaline pre-esistenti, tipicamente utilizzate all'interno delle pareti per gli impianti elettrici, non essendoci problemi di interferenza, né di coesistenza di alcun tipo.

Alcune prove preliminari effettuate in laboratorio hanno mostrato che per tiraggi fino a circa 65 N applicati ad una singola fibra da 1,5 mm, non si percepiscono degradi prestazionali. Tale valore è coerente con i dati di targa forniti dai costruttori.

A livello normativo, in ambito CEI (Comitato Elettrotecnico Italiano), grazie anche al contributo di Telecom Italia, è stata legalizzata proprio l'operazione di impiego di fibre ottiche in coesistenza con altri mezzi di trasporto (norma CEI 64-8): in questo modo, le operazioni di cablaggio risultano notevolmente agevolate.

La procedura di installazione delle POF è estremamente semplice e non richiede la presenza di tecnici specializzati per le operazioni di connettorizzazione per la terminazione delle fibre. Infatti, le dimensioni generose del nucleo (circa 1 mm di diametro) consentono il funziona-

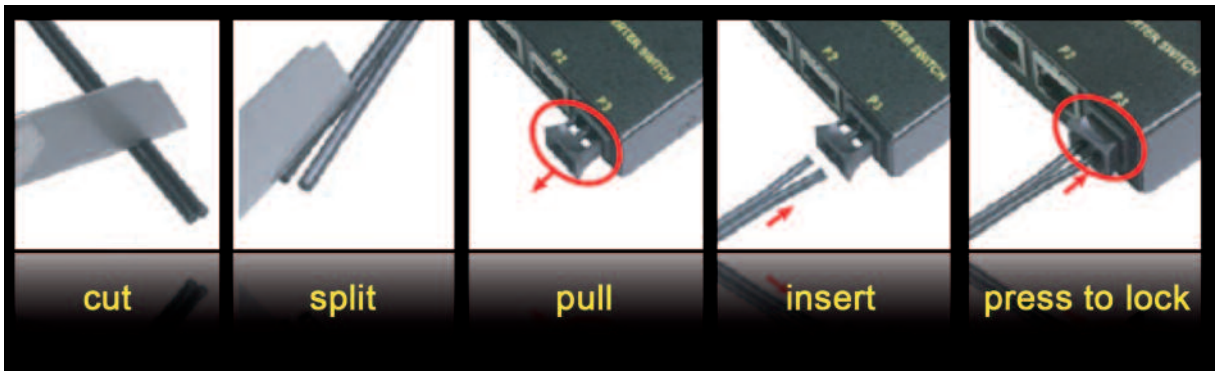


Figura 14 – Operazioni per l'installazione delle POF

mento anche in caso di allineamento non perfetto, rendendo quindi le POF affidabili anche senza interventi di installazione particolarmente sofisticati.

A titolo di esempio, in *figura 14* sono riportate in maniera semplificata le operazioni necessarie per realizzare l'inserimento delle POF nel connettore OptoLock™. In pratica, la procedura prevede:

1) tagliare la coppia di POF mediante l'apposito cutter oppure utilizzando un normale taglierino o delle forbici;
2) separare le 2 POF per pochi cm;
3) estrarre verso l'esterno il "cassettino" del connettore OptoLock™;
4) inserire le 2 POF nel connettore;
5) spingere verso l'interno il "cassettino" del connettore OptoLock™.

L'unica avvertenza è di connettere correttamente la coppia di POF in modo tale che, per un collegamento punto-punto, il trasmettitore di un apparato sia collegato al ricevitore dell'altro. Anche questo aspetto non risulta critico, poiché le POF trasportano informazioni mediante luce visibile, che consente di individuare il trasmettitore in maniera semplice.

Esempio applicativo: Telecom Italia per il Sociale - Progetto "Smart Inclusion"

A cura di Filippo Tempia

Ponte tecnologico tra Scuola e Ospedale: Soluzioni e servizi di teledidattica per bambini ospedalizzati

"*È compito della Repubblica rimuovere gli ostacoli di ordine economico e sociale, che, limitando di fatto la libertà e l'eguaglianza dei cittadini, impediscono il pieno sviluppo della persona umana e l'effettiva partecipazione di tutti i lavoratori all'organizzazione politica, economica e sociale del Paese*". È l'art. 3 della Costituzione Italiana, che, tra gli altri, sancisce il diritto all'educazione e all'istruzione anche per i soggetti emarginati o potenzialmente tali.

È su questi principi che nasce il progetto Smart Inclusion, la prima iniziativa in Italia che integra su un'unica piattaforma tecnologica, sviluppata da Telecom Italia, con la supervisione scientifica del CNR - ISOF di Bologna, servizi di teledidattica, intrattenimento e gestione dei dati clinici, consentendo ai bambini lungodegenti di partecipare alla vita sociale, collegandosi con il mondo >

esterno in maniera semplice ed immediata e al contempo ai medici di disporre di strumenti evoluti a supporto dei processi di cura del paziente.
Inaugurato nel febbraio 2009, il progetto è stato realizzato in meno di 4 mesi presso il Policlinico Sant'Orsola-Malpighi e la Scuola Media Irnerio di Bologna, e si rivolge principalmente ai bambini ospedalizzati nei reparti di Oncoematologia Pediatrica e Cardiologia e Cardiochirurgia Pediatrica. L'obiettivo è quello di ridurre la distanza umana, sociale e culturale tra i piccoli pazienti lungodegenti in ospedale e il mondo esterno, la famiglia e la scuola, e favorire al tempo stesso l'assistenza ospedaliera. Attraverso il terminale innovativo "SmartCare Terminal", all-in-one, con monitor LCD a 17" di tipo touchscreen, dotato di webcam e microfono orientabili a bordo letto, il bambino può:

- accedere al mondo esterno attraverso una finestra virtuale;
- entrare nella sezione ***Intrattenimento***, videocomunicando con i propri famigliari e amici a casa, oppure allietarsi con video divertenti on-demand o canali cartoons tematici;
- partecipare in diretta ad una lezione in classe equivalente presso la scuola collegata, entrando nella sezione ***Scuola***, dialogando con l'insegnante e i compagni ed interagire con una Lavagna Interattiva Multimediale collocata presso le 5 classi della scuola Irnerio coinvolte.

Il tutto attraverso il semplice utilizzo di un dito, grazie ad un'interfaccia di fruizione molto semplice che rende accessibile ogni funzionalità in modo facilitato e diretto, senza l'uso di una tastiera o di un mouse. Smart Inclusion rappresenta anche un canale di accesso alla piattaforma "Innova Scuola", sviluppata e diffusa in Italia dal ministero dell'Innovazione, per la condivisione di lezioni inter-scuola e la messa a disposizione di insegnanti e studenti di lezioni registrate e contenuti e strumenti a supporto della didattica.
Smart Inclusion, attraverso i propri terminali, rappresenta anche il punto informatizzato a bordo letto a disposizione del personale medico, il canale d'accesso ai sistemi informativi ospedalieri, in modo immediato ed efficace. Il personale, infatti, fruendo delle funzionalità esposte dalla sezione Ospedale, è già in grado di visualizzare la cartella clinica del piccolo paziente, trovandosi di fronte ad uno strumento tecnologico, che permette una migliore gestione dell'assistenza e della cura ospedaliera del paziente.
Per gli aspetti di connettività indoor, il progetto si avvale di soluzioni basate sulle tecnologie POF e PLT, come spiegato nel seguito.
Il progetto rappresenterà alla fine del 2010 uno standard per tutti i reparti di oncoematologia pediatrica di strutture ospedaliere pubbliche in Italia, con realtà analoghe al reparto relativo al Policlinico Sant'Orsola-Malpighi. L'obiettivo è quello di creare un substrato tecnologico sperimentale, in grado di evolvere con servizi ICT a supporto dell'attività ospedaliera. Su indicazione del Ministro per la Pubblica Amministrazione e l'Innovazione, infatti, il progetto è entrato a far parte del piano **e-Government 2012**, approvato dalla Presidenza del Consiglio, e prevede l'estensione su 18 ospedali pediatrici italiani entro il 2010. La prossima realizzazione coinvolgerà l'Ospedale Pediatrico Bambino Gesù a Roma, per poi proseguire nel 2009 a Firenze (Meyer), a Torino (Regina Margherita), a Padova (Azienda Ospedaliera), a Genova (Istituto I. Gaslini) e a Pavia (Policlinico S. Matteo), e su ulteriori 12 realtà ospedaliere nel 2010.

> L'impiego delle powerline

Per il progetto pilota di Bologna, presso il reparto di Oncoematologia dell'ospedale Sant'Orsola-Malpighi, è stata impiegata la tecnologia delle PLT per il collegamento dei singoli terminali (SmartCare) all'unico punto LAN disponibile in reparto. Sono stati utilizzati adattatori Ethernet - PLT come "add-on" esterno per ogni terminale.
La stessa soluzione è stata adottata presso la scuola Irnerio di Bologna, dove gli adattatori Ethernet - PLT sono stati integrati nei terminali presenti in ogni aula per l'attività di teledidattica.

L'impiego delle POF

Presso il reparto di Cardiologia dell'ospedale Sant'Orsola-Malpighi, è stata realizzata un'architettura a stella per il collegamento dei singoli terminali (SmartCare) all'unico punto LAN disponibile in reparto.
Per ciascuno dei 7 terminali è stato utilizzato un adattatore elettro-ottico (con alimentatore integrato), mentre la connessione tra i singoli adattatori e il punto LAN ha richiesto l'impiego di 2 switch ottici con 5 interfacce POF ciascuno (anch'essi con relativo alimentatore).
In altre realtà ospedaliere, si potrà valutare la possibilità di definire architetture diverse da quella a stella, oppure architetture "miste" Ethernet + POF mediante l'impiego di switch Ethernet (oltre a quelli ottici).
La posa delle POF è avvenuta prevalentemente sfruttando la controsoffittatura e le canaline preesistenti esterne alle pareti.

filippo.tempia@telecomitalia.it

5 Conclusioni

Il presente articolo ha illustrato le caratteristiche principali delle tecnologie powerline e POF, evidenziandone punti di forza e potenziali criticità. È stato posto l'accento sugli scenari applicativi, non necessariamente legati all'ambito residenziale. In particolare, è stato presentato l'esempio del progetto Smart Inclusion, che ha dimostrato la possibilità di impiegare le tecnologie negli ambiti ospedalieri e scolastici.

ACRONIMI

AES Advanced Encryption System
AG Access Gateway
ARM Advanced Risc Machine
ARQ Automatic Repeat request
BPSK Binary Phase-Shift Keying
CEI Comitato Elettrotecnico Italiano
CENELEC Comité Européen de Normalisation ELECtrotechnique
CEPCA Consumer Electronics Powerline Communication Alliance
CEPT Comité Européen des Postes et Telecommunications
CISPR Comité International Spécial des Perturbations Radioélectriques
CNR-ISOF Consiglio Nazionale delle Ricerche - Istituto per la Sintesi Organica e la Fotoreattività
CSMA/CA Carrier Sense Multiple Access / Collision Avoidance
DAB Digital Audio Broadcasting
DES Digital Encryption Standard
DSP Digital Signal Processing
DVB Digital Video Broadcasting

ECC Electronic Communications Committee
EIA Electronic Industries Alliance
EMC Electro-Magnetic Compatibility
ETSI European Telcoomunication Standard Institute
FCC Federal Communication Commission
FEC Forward Error Correction
FTTH Fibre To The Home
HVAC Heating Ventilation and Air Conditioning
IARU International Amateur Radio Union
ICT Information and Communication Technology
IEC International Electrotechnical Commission
IGMP Internet Group Management Protocol
IP Internet Protocol
IPTV Internet Protocol TV
LAN Local Area Network
LCD Liquid Cristal Display
LED Light Emitter Diode
MAC Medium Access Control
MF Medium Frequency
MII Media Indipendent Interface
MIMO Multiple Input Multiple Output
NA Numerical Aperture
NWIP New Work Item Proposal
OFDM Orthogonal Frequency Division Multiplexing
ONT Optical Network Termination
PCI Peripheral Component Interconnect
PLT Power Line Telecommunication
POF Plastic Optical Fibre
PMMA Polimetilmetacrilato
QAM Quadrature Ampliture Modulation
QoS Quality of Service
QPSK Quadrature Phase-Shift Keying
RJ Registered Jack
ROBO ROBust OFDM
SI Step Index
SMI Small Multimedia Interface
STB Set Top Box
TCP Transmission Control Protocol
TDMA Time Division Multiple Access
UDP User Datagram Protocol
UPA Universal Powerline Association
USB Universal Serial Bus
UTP Unshielded Twisted Pair
VDSL Very high speed Digital Subscriber Line
VHF Very High Frequency
VoIP Voice over IP
WAN Wide Area Network
Wi-Fi Wireless-Fidelity
xDSL x - Digital Subscriber Line

BIBLIOGRAFIA

[1] Guida CEI 306-2, "Guida per il cablaggio per telecomunicazioni e distribuzione multime-diale negli edifici residenziali"
[2] Progetto OMEGA (Home Gigabit Access), http://www.ict-omega.eu/
[3] IEEE International Symposium on Power Line Communications and its Applications, http://www.ieee-isplc.org/2009/
[4] CENELEC EN 50065-1, "Signalling on Low-Voltage Electrical Installations in the Frequency Range 3 kHz to 148,5 kHz Part 1: General Requirements, Frequency Bands and Electromagnetic Disturbances"
[5] CISPR 22, "Information technology equipment - Radio disturbance characteristics - Limits and methods of measurement"
[6] ITU-T G.9960, "Next generation wireline based home networking transceivers"
[7] http://www.smartgrids.eu/
[8] "Le Plastic Optical Fiber e le TLC", Notiziario Tecnico Telecom Italia n. 3 - Dicembre 2007, pagg. 97-100

andrea.bergaglio@telecomitalia.it
mariano.giunta@telecomitalia.it
angelantonio.gnazzo@telecomitalia.it

AUTORI

Andrea Bergaglio

laureato in Ingegneria Elettronica con indirizzo Telecomunicazioni, nel 1996 è entrato in Azienda, dove, fino al 2003, si è occupato della progettazione e dello sviluppo software di applicazioni per la gestione di apparati di rete (NT1Plus per ISDN) e per l'esercizio e la manutenzione della stessa rete. Successivamente ha lavorato nel campo dei terminali di rete fissa.
Dal 2006, si occupa di tematiche di home networking, con l'obiettivo di analizzare e sperimentare le tecnologie di connettività di interesse per il contesto di servizio multi-play di Telecom Italia, in relazione alle evoluzioni previste nello scenario NGN2, sia per l'ambito residenziale che business. Attualmente è responsabile del progetto di "Home/indoor networking technology - Innovation" ■

Mariano Giunta

laureato in Ingegneria Elettronica, nel 1989 entra in Azienda, dove si occupa delle attività di sperimentazione, standardizzazione e qualificazione per gli aspetti di Compatibilità Elettromagnetica (EMC) relativi al settore delle telecomunicazioni e della tecnologia dell'informazione. Attualmente è responsabile del Laboratorio di Compatibilità Elettromagnetica ed è coinvolto oltre che nell'attività di normativa internazionale (ETSI, CENELEC, IEC), in progetti relativi alle problematiche EMC delle reti di telecomunicazione a larga banda e dei sistemi di Power Line Telecommunication. Ha tenuto corsi presso il Politecnico di Torino ed è autore di diverse pubblicazioni presentate in congressi internazionali (IEEE EMC Symposium, Zurich EMC Symposium, ecc.).
Dal 2002 al 2005 è stato membro dello Steering Committee del progetto europeo FOR-EMC finanziato dalla Commissione Europea ■

Angelantonio Gnazzo

laureato in Fisica, nel 1988 è entrato in Azienda, dove, fino al 1996, ha lavorato nel campo delle tecnologie per le fibre ottiche e ottica integrata. Ha contribuito al progetto e alla realizzazione di fibre ottiche speciali e di dispositivi quali i diramatori di potenza, gli amplificatori integrati e i dispositivi selettivi in lunghezza d'onda.
Dal 1996 e fino al 2000, la sua attività ha riguardato gli aspetti di misura su portanti fisici e sugli impianti di telecomunicazione.
Dal 2000 sta lavorando su tematiche di home networking, con particolare attenzione alle attività riguardanti lo studio e l'integrazione delle reti e dei terminali nell'ambito di scenari di servizio multi-play. Ha partecipato a diversi progetti nazionali ed europei, nonché seguito gruppi di normativa. Attualmente è responsabile del laboratorio di Home Networking di Telecom Italia Lab ■