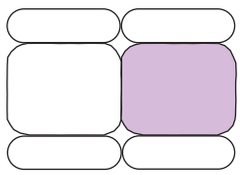
SISTEMI EMBEDDED
CARATTERISTICHE, TECNOLOGIE E MERCATO

Carlo Brandolese
William Fornaciari



Alzarsi la mattina, fare colazione e recarsi con calma in libreria per acquistare un libro. È sabato o domenica, finalmente un giorno un cui non si interagisce con la tecnologia, in cui non si accende il computer. Un tale scenario è ben lontano dalla realtà. Da qualche ora si è probabilmente già entrati in contatto con un centinaio di sistemi basati su microprocessore: decine sull'automobile, due o tre in ogni elettrodomestico o telefono cellulare, almeno uno in ogni apparato alimentato dalla rete elettrica o a pile: condizionatore, apricancello, pace-maker, calzature ginniche evolute, semafori per l'attraversamento pedonale, carte di credito per il pagamento elettronico e così via.

1. INTRODUZIONE

La maggior parte dei sistemi di elaborazione non è costituita da personal computer, bensì da dispositivi in stretta relazione con l'ambiente in cui operano, che in genere hanno una funzione prefissata e per questo motivo non richiedono di caricare programmi né, in molti casi, di avere interfacce tradizionali con tastiera e monitor.
Si parla di sistemi invisibili, ubiqui e pervasivi, per rappresentare la loro trasparente invasività nella vita quotidiana. Questi sistemi, detti dedicati o *embedded*, dominano numericamente il mercato, anche se in modo non conclamato, rispetto ai più noti e visibili sistemi *general purpose* come i personal computer. Si ritiene infatti che attualmente un cittadino del mondo occidentale entri in contatto giornalmente con circa 100 sistemi dedicati. Per convincersi di ciò basti pensare che ciascuno dei seguenti apparati contiene normalmente sistemi elettronici dotati di almeno un microprocessore: telefono cellulare, apricancello, bancomat, testina di una stampante a getto d'inchiostro, lavatrice, lavapiatti, serratura elettronica, navigatore, carta di credito e così via, fino all'automobile ove si contano addirittura decine di microprocessori.
La presenza di tali sistemi in una varietà di prodotti non ha solamente lo scopo di realizzare le funzionalità desiderate ma sempre più spesso diviene il veicolo per introdurre innovazione. La rilevanza dei sistemi *embedded* in alcuni importanti settori applicativi è mostrata nella figura 1 che evidenzia come all'incirca la metà del costo finale di un prodotto è rappresentato dall'elettronica.
L'architettura di un sistema dedicato ricorda quella di un generico sistema di calcolo, con sezioni di elaborazione, comunicazione e memorizzazione. I requisiti operativi di tali sistemi tuttavia li rendono talmente peculiari da richiedere metodologie di progetto, e un approccio alla ricerca di soluzioni tecnologiche, tali da giustificare la nascita di una vera e propria nuova disciplina dell'ingegneria.

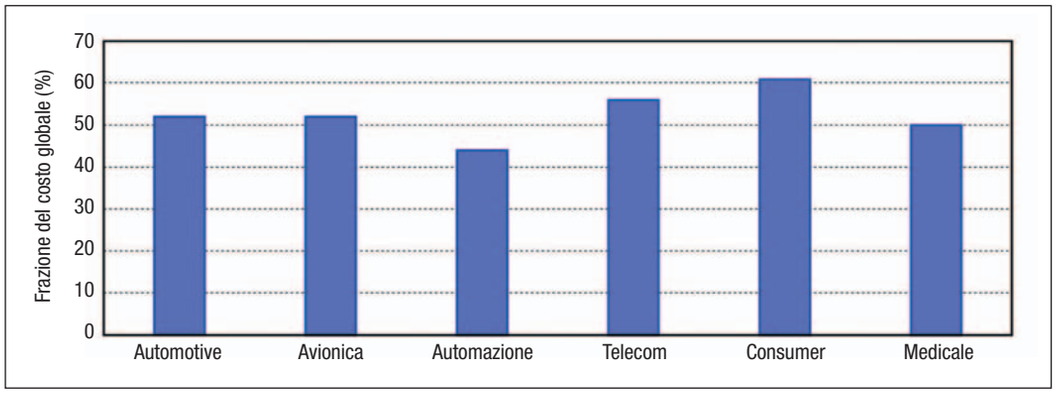


FIGURA 1
Incidenza percentuale dei sistemi embedded nel costo finale dei prodotti

Componente	2004	2009	AAGR[1]
Software	1.6	3.4	16.0%
Circuiti Integrati	40.5	78.7	14.2%
Board	3.7	6.0	10.0%

[1] Average Annual Growth Rate, Tasso medio annuo di crescita.

TABELLA 1
Mercato mondiale dei sistemi embedded per componente (miliardi di dollari)

Settore	2004	2009	AAGR
Computer	18.1	30.5	14.1%
Telecom	11.2	18.0	12.6%
Consumer	12.0	22.5	17.0%
Automotive	3.8	7.8	19.7%
Medicale	3.0	3.7	5.4%
Industriale	3.4	4.5	7.3%

TABELLA 2
Mercato mondiale dei sistemi embedded per macro-settore (miliardi di dollari)

Tra i vincoli di natura non direttamente tecnologica cui sono soggetti i sistemi *embedded* si hanno il *costo* del prodotto (costo unitario e costo non ricorrente) ed il tempo di realizzazione (*time-to-prototype*, *time-to-market e time-to-volume*). Tali sistemi, inoltre, sono fortemente specializzati e ottimizzati per svolgere un ristretto numero di compiti, e sono soggetti a vincoli molto stringenti di consumo di potenza, prestazioni, dimensioni, affidabilità, tempo reale e così via.
Le principali conseguenze di un così elevato numero di requisiti e vincoli sono, da un lato, la scelta di architetture miste hardware/software per trovare il migliore equilibrio tra costo, prestazioni e tempo di sviluppo e, dall'altro, il ricorso a metodologie di progetto ben strutturate e ad una moltitudine di strumenti di *Electronic Design Automation* (EDA).

2. MERCATO

A partire dagli anni '60 abbiamo assistito ad una sorta di seconda "rivoluzione industriale", quella associata allo sviluppo e alla diffusione delle tecnologie digitali. Negli ultimi quarant'anni, infatti, si è passati dai *mainframe* ai *personal computer*, per arrivare oggi ad una larga diffusione di *laptopt*, *PDA* e *smart phone* con un utilizzo prevalentemente legato ad applicazioni di *office automation* e di accesso ad internet.
Nonostante tale sviluppo sia molto evidente, attualmente le statistiche mostrano che il mercato e le applicazioni dominanti – per un fattore pari a circa 100 – sono quelle dei sistemi *embedded*. In un ambiente domestico, per esempio, vi sono al massimo due o tre computer, ma non meno di 40 sistemi *embedded*. Tale diffusione si traduce in una percentuale di mercato dei sistemi embedded rispetto all'intero settore elettronico superiore al 90%. Secondo i dati del *World Trade Statistics*, attualmente il 98% dei dispositivi programmabili sono embedded e altre previsioni di crescita stimano che nel 2010 vi saranno circa 16 miliardi di sistemi *embedded*, circa 3 per ogni abitante della terra, mentre nel 2020 si dovrebbe raggiungere un valore pari a 40 miliardi di unità.
Tali proiezioni, per quanto già estremamente significative, possono essere analizzate con maggiore dettaglio considerando orizzontalmente il mercato globale dell'hardware e del software per applicazioni *embedded* (Tabella 1) e quello più verticale dei vari macro-settori applicativi quali telecomunicazioni, elettronica di consumo, *automotive* e così via (Tabella 2).

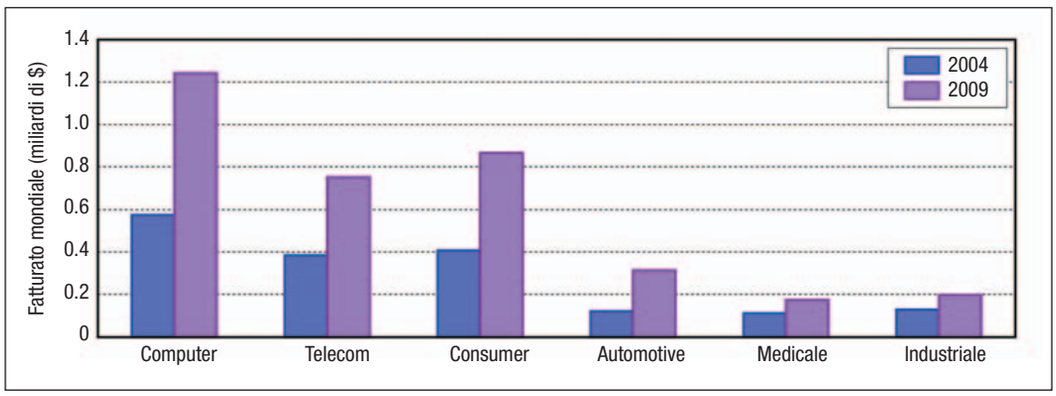


FIGURA 2
Suddivisione per applicazione del fatturato del software embedded

2.1. Software

Analizzare in modo compiuto il mercato del *software embedded* e la sua evoluzione richiede alcune precisazioni. Benché il mercato degli strumenti per il *testing* e quello dei sistemi operativi mostri una notevole crescita (circa il 20% all'anno) il reale impatto del software è ancora superiore poiché la maggior parte delle stime non tiene conto del codice dell'applicazione e dell'eventuale *middleware* realizzato *ad-hoc* dal produttore finale del sistema embedded. Tale frazione non è pertanto compresa nei dati della tabella 1.
Complessivamente il software rappresenta spesso una notevole frazione dell'attività di sviluppo. Circa il 48% del fatturato relativo al software è assorbito dagli USA, mentre il resto del mercato è diviso in parti più o meno uguali fra Asia, Giappone ed Europa. Osservando la suddivisione per tipologia di software (Figura 2), si nota che quello delle telecomunicazioni, dei computer e dell'elettronica di consumo sono i mercati trainanti, grazie anche alla convergenza dei dispositivi palmari (PDA, telefoni cellulari, organizer elettronici, e così via). Si conferma l'effervescenza del settore *automotive*, i cui tassi di crescita battono abbondantemente quelli degli altri settori applicativi, superando il 21%, grazie anche all'aggiunta al semplice controllo del veicolo, di molte applicazioni un tempo peculiari di altri settori come per esempio telefono, la connettività internet e vari sistemi d'intrattenimento.
Si osserva inoltre una crescente presenza dei sistemi operativi in molte applicazioni embedded che testimonia l'interesse dei produttori, oltre che verso i tradizionali microcontrollori, anche nella direzione dei processori di classe superiore (i più diffusi sono

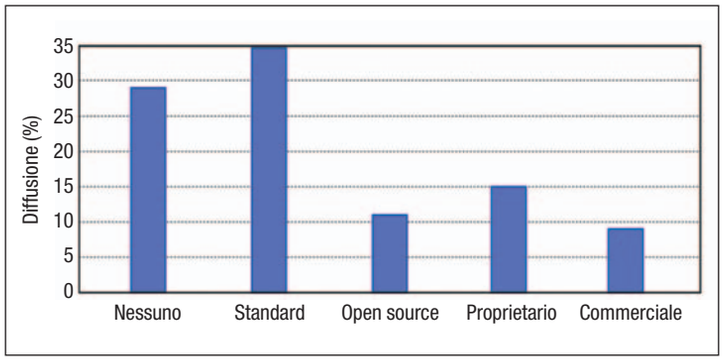


FIGURA 3
Utilizzo recente dei sistemi operativi per sistemi embedded

Freescale, PowerPC e ARM), in grado di ospitare appunto un sistema operativo utile allo sviluppo di applicazioni complesse con interfacce utente evolute. È tuttavia interessante notare che, come emerge dai dati della figura 3, oltre un quarto dei dispositivi embedded non utilizza alcun sistema operativo. Tale fenomeno non dove stupire: in moltissimi apparati, infatti, si utilizzano architetture con memoria integrata sullo stesso chip del processore, di dimensione non sufficiente a ospitare sia un sistema operativo sia l'applicazione. Inoltre, molti sistemi operativi per applicazioni embedded di natura commerciale richiedono licenze d'uso con costi non inferiori a qualche euro (per sistema venduto), valore molte volte superiore al costo dell'intero hardware e, di conseguenza, difficile da proporre a un mercato in cui in molti casi il prezzo è il fattore determinante. La tipologia di sistema operativo è decisamente varia, si spazia dall'*open source* di Linux e le sue varianti, sino ai prodotti commerciali come Windows CE. I dati più dettagliati della figura 4 quantificano la frammentazione del mercato,

sebbene i principali fornitori per i nuovi progetti in ambito real-time siano stati ultimamente Windriver (con il sistema VxWorks) e Linux, mentre Microsoft (Windows CE e Windows XP Embedded) si è rivolta soprattutto al mercato PDA e telefonia.

2.2 Hardware

L'industria dell'hardware per sistemi embedded è sicuramente considerata una delle più promettenti sia come volume, sia per l'elevato tasso di crescita (Tabella 1). Com'è ovvio attendersi, a livello geografico gli USA sono i leader, con un fatturato all'incirca equivalente al resto del mondo. Mentre in Europa le maggiori spinte provengono e proverranno dal mercato telefonico (grazie alla presenza di Nokia, Philips, Siemens ed STM), il Giappone sarà sostenuto principalmente da applicazioni consumer e satellitari (Samsung, LG, Sony e Toshiba) e gli USA sembra si stiano focalizzeranno su internet e dispositivi di comunicazione.
Sotto il profilo applicativo, la distribuzione è quella mostrata nella figura 5, con volumi che si consolidano per computer, telecomunicazioni ed elettronica di consumo e ingenti tassi di crescita per il settore automotive. Come già ricordato per il software, i nuovi veicoli continueranno a incorporare funzioni tradizionali gestite elettronicamente (controllo trazione e motore, freni, sospensioni, sensoristica, e così via), assieme ai nuovi servizi come router GPS, DVD player, dispositivi per accesso a internet o per reti wireless, fino a utilizzare più di un centinaio di processori su una sola auto. La convergenza delle applicazioni verso gli stessi dispositivi (per esempio, macchina fotografica, telefono, comunicazione e media player) costituisce e continuerà a costituire un'importante tendenza a sostegno del mercato consumer.
I componenti più utilizzati nel mercato dell'hardware sono i microprocessori *general purpose* e i microcontrollori, seguiti dai circuiti integrati per applicazioni dedicate (ASIC, PLD ed FPGA), dai DSP (*Digital Signal Processor*), e dalle memorie, il cui mercato si attende abbia i tassi di crescita più elevati (Figura 6).
Il componente più importante per il mercato dei sistemi embedded rimane il microprocessore: le architetture CISC (*Complex Instruction Set Computer*) sono principalmente legate al

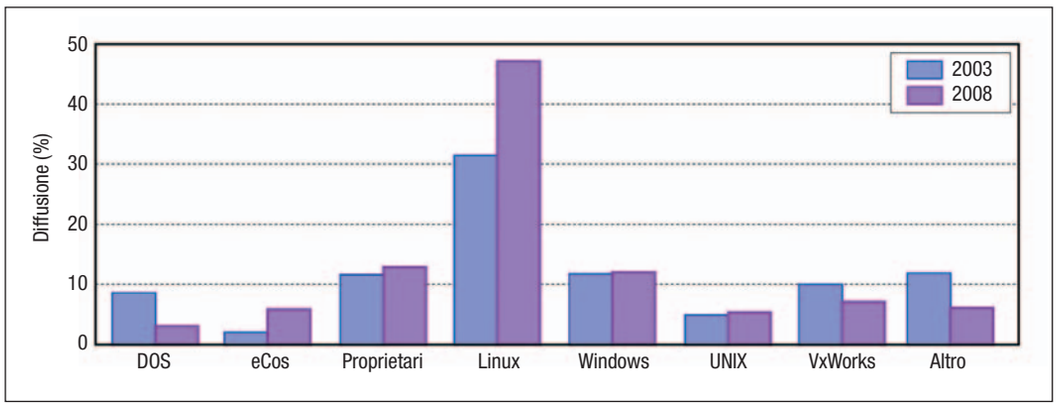


FIGURA 4
Andamento dell'uso dei principali sistemi operativi in ambito embedded

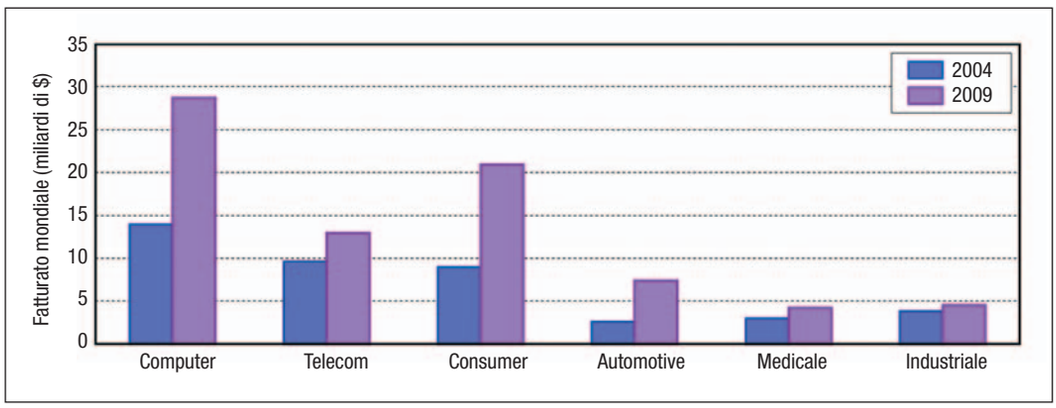


FIGURA 5
Mercato dell'hardware embedded per settore applicativo

settore dei calcolatori, mentre soluzioni di tipo RISC (*Reduced Instruction Set Computer*) sono le preferite per applicazioni *embedded*, in virtù della buona relazione che lega costo, consumo energetico e prestazioni. Benché il fatturato dei microprocessori nel settore dei computer sia ancora rilevante (circa il 20%), non lo si deve confondere i volumi in termini di numero di pezzi. I processori per applicazioni embedded hanno infatti costi industriali che vanno da meno di un dollaro a pochi dollari mentre i processori per calcolatori hanno prezzi fino a due ordini di grandezza superiori. Un'analisi della tipologia dei processori mostra che oltre il 60% di questi sono a 4, 8 o 16 bit e solo il 30% è basato su architetture x86.
In maniera facilmente avvertibile anche dall'acquirente finale, si assiste a una parziale rivoluzione del mercato delle memorie (Figura 7), trainato soprattutto dal tasso di crescita delle tecnologie Flash, utilizzate nei lettori mp3, nelle macchine fotografiche digitali, nei telefoni cellulari, nelle chiavette USB e in molti altri prodotti. In generale la dimensione e il tipo di memoria influenzano le prestazioni e il costo di un sistema *embedded*. Al giorno d'oggi, a una crescita di un fattore dieci della banda di comunicazione di una memoria corrisponde solo un aumento di un fattore cinque della densità. Per questa ed altre ragioni, nonostante le memorie statiche (SRAM) siano più veloci, spesso si ricorre alle memorie dinamiche (DRAM) per via del minore costo e in virtù della loro maggiore densità e capacità. Benché siano in sviluppo nuove tecnologie di memoria (magnetoresistive, ferroelettriche, zero-capacitor) che potrebbero in futuro competere con quelle attuali, il loro impatto in termini di volumi di mercato è ancora poco influente.

3. CARATTERISTICHE DEI SISTEMI EMBEDDED

Gli obiettivi e le conseguenti architetture realizzative di un sistema *embedded* sono duali rispetto a quelli di un normale elaboratore. Un personal computer viene realizzato in modo da essere principalmente versatile, in grado cioè di adattarsi a molteplici applicazioni semplicemente caricando programmi differenti. Come conseguenza l'architettura hardware di supporto contiene in genere risorse sovrab-

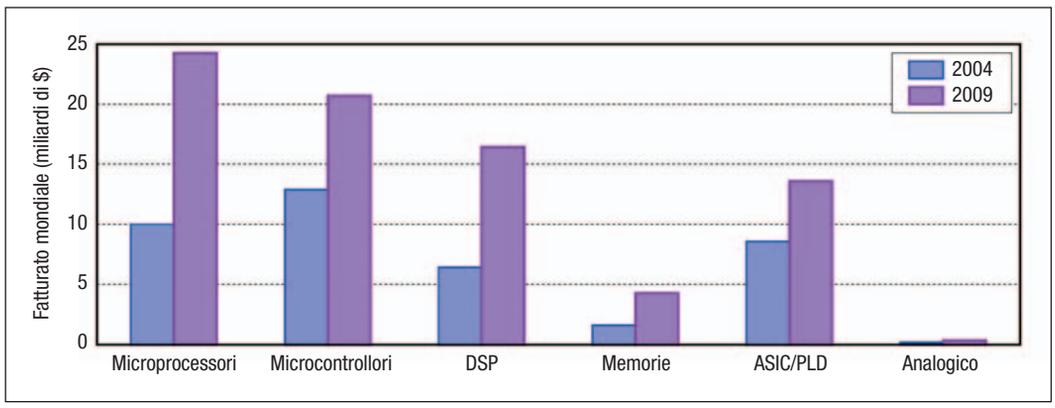


FIGURA 6
Mercato dell'hardware embedded per settore applicativo

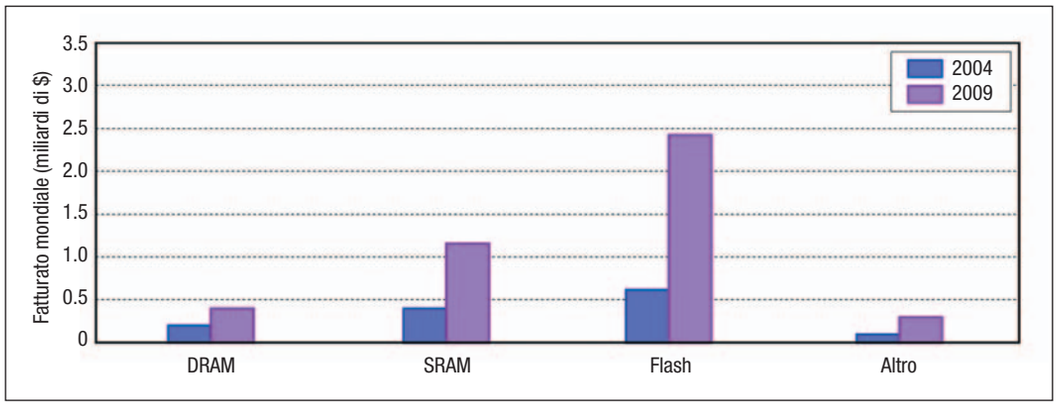


FIGURA 7
Mercato delle memorie suddiviso per tecnologia

bondanti rispetto alle singole applicazioni poiché versatilità ed ottimizzazione sono obiettivi contrastanti. Per esempio, l'utilizzo di un'applicazione di elaborazione di testi difficilmente richiede più di qualche punto percentuale della piena potenza di calcolo dei moderni microprocessori, valore che risulta peraltro spesso insufficiente per la gestione efficace dei requisiti della grafica dei videogiochi, al punto da richiedere la presenza di schede grafiche dedicate.
Un sistema dedicato ad un'applicazione o ad una classe molto specifica di applicazioni può invece essere fortemente *ottimizzato*, venendo meno il requisito di garantire versatilità. In questo caso, sulla base di una conoscenza approfondita dell'applicazione, si può dimensionare in modo corretto ed ottimale la capacità di calcolo scegliendo opportunamente il microprocessore ed eventualmente utilizzando componenti hardware aggiuntivi per ridurre la potenza di elaborazione richiesta al processore stesso.
Non è pertanto affatto facile standardizzare l'architettura di un sistema *embedded*, poiché – anche a parità di requisiti funzionali – i vincoli posti dalla specifica applicazione possono portare a sistemi sensibilmente diversi. Avere per esempio l'obiettivo tassativo di realizzare il sistema in pochi mesi, spesso sbilancia le scelte verso soluzioni principalmente software, mentre le esigenze di raggiungere ingombri limitati o bassi costi unitari per elevati volumi di produzione possono richiedere lo sviluppo di hardware dedicato.
Dato lo spettro amplissimo degli ambiti applicativi e le peculiarità di ognuno di essi il tipo e l'importanza dei vincoli cui un sistema può essere soggetto variano moltissimo.

3.1. Peso e dimensioni

Spesso l'ingombro fisico del sistema è determinante, soprattutto per i dispositivi che non prevedono una collocazione fissa.

3.2. Costo

È un fattore determinante soprattutto per le produzioni in grandi volumi, tipiche dell'elettronica di consumo.

3.3. Consumo energetico

Sempre più sistemi stanno diventando portabili, pertanto un basso livello di consumo energetico consente di avere batterie meno costose, più piccole e leggere e/o di ottenere durate del sistema che ne rendano pratico l'uso. Poiché il miglioramento tecnologico delle batterie è piuttosto lento, si deve affrontare il problema del consumo energetico in sede di progetto, sia per quanto concerne lo sviluppo dell'hardware, sia nella stesura del codice dell'applicazione nonché nella scelta dell'eventuale sistema operativo.

3.4. Dimensione del codice

Nella maggior parte dei casi, i sistemi *embedded* sono completi, ovvero il loro software risiede in un supporto di memoria permanente integrato nello stesso *chip* o sulla stessa *board* del microprocessore. Questo vincolo si riflette, soprattutto per motivi di costo e ingombro, sulla dimensione del codice che deve quindi essere il più possibile contenuta.

3.5. Prestazioni

Le prestazioni non sono un obiettivo generico, come spesso accade nella progettazione di un computer *general purpose*, ma dipendono dall'applicazione. In generale si devono soddisfare due tipi di vincoli: il tempo di reazione a un evento e il tempo di gestione dell'evento stesso eseguendo il codice ad esso associato. Questi vincoli si tradurranno in soluzioni architetturali mirate a soddisfare tali requisiti mantenendo un costo limitato. Bilanciare costo e prestazioni, molto spesso significa ricorrere a soluzioni miste hardware/software, rendendo più complesso il processo di progettazione e sviluppo.

3.6. Tempo-reale

Non necessariamente legato al concetto di prestazione è quello di funzionamento in tempo reale o *real-time*. Nell'accezione più generale possibile si dice che un sistema o una parte di esso è soggetto a vincoli di tempo reale se alcune delle operazioni svolte devono essere iniziate o completate entro un tempo prefissato e ben preciso. Si distinguono vincoli *soft real-time*, la cui violazione porta unicamente ad un degrado delle prestazioni del sistema, e vincoli *hard real-time*, il cui mancato soddisfacimento comporta effetti catastrofici sul sistema, nonché, in alcuni casi, sull'ambiente circostante.

3.7. Affidabilità

In molti casi è necessario garantire il rispetto di rigide norme di certificazione che prevedono un'accurata analisi dei potenziali guasti. In tal caso è necessario disporre di una valutazione della probabilità che il sistema si guasti, cioè appunto della sua affidabilità. Si noti che poter disporre di tale informazione è di per sé un vincolo, che eventualmente diviene più stringente nella misura in cui si fissa una soglia minima di affidabilità.

3.8. Safety

Indica una misura della possibilità che a fronte di un guasto il sistema non provochi conseguenze gravi alle cose o alle persone con cui interagisce. A seconda dell'ambito applicativo – per esempio automotive, biomedicale, nucleare ecc. – si definiscono diversi criteri e diversi livelli di *safety*. In questo senso sono ormai consolidati diversi standard e metodi per la valutazione della *safety* quali ad esempio il *Safety Integrity Level* o SIL.

3.9. Sicurezza

Da non confondere con la *safety* è la sicurezza o security, ovvero la capacità di un sistema di proteggere le informazioni e di verificarne l'autenticità.

3.10. Time-to-market e flessibilità

Le metodologie e le tecnologie scelte per il progetto devono essere scelte in modo da consentire di arrivare al prodotto entro tempi stringenti, in modo da cogliere il massimo delle opportunità di mercato. Questo significa, per esempio, ricorrere ad un approccio alla progettazione che consenta di modificare il progetto in corso d'opera, visto che le fasi iniziali della progettazione avvengono quasi in parallelo alla stesura delle specifiche.

4. UN ESEMPIO TIPICO

Concludiamo questa breve introduzione ai sistemi embedded discutendo un semplice caso di studio relativo ad una funzionalità sempre più diffusa: ovvero la capacità di effettuare fotografie digitali. Tale funzionalità è ormai inclusa sia nelle macchine fotografiche digitali, sia in molti telefoni digitali e PDA, così com'è possibile scattare fotografie a partire da una videocamera digitale o una webcam. Un primo passo verso l'analisi e la progettazione di un sistema per la fotografia digitale consiste nel delineare lo specifico scenario di utilizzo, allo scopo di definire sia i requisiti funzionali sia i vincoli di progetto. In sintesi, questa prima fase ha lo scopo di identificare uno spazio delle soluzioni, entro cui prendere le decisioni che guideranno il successivo processo di sviluppo. I principali scenari d'uso potrebbero realisticamente essere i seguenti.

4.1. Macchina fotografica per consumer electronics

È l'utilizzo più naturale, in cui abbiamo un sistema concepito per fornire tale servizio. L'obiettivo è raggiungere le prestazioni migliori in termini di risoluzione, velocità di scatto, capacità di memorizzazione e di gestione delle foto memorizzate. L'interfaccia utente deve essere semplice e prevedere un piccolo schermo grafico. Il sistema deve offrire la possibilità di connessione sia verso un PC sia verso uno schermo televisivo. Visto che il sistema rimarrà acceso solo saltuariamente, i requisiti di consumo di potenza sono mediamente stringenti. Dal punto di vista meccanico è desiderabile, benché non indispensabile, ridurre le dimensioni. Un aspetto importante è il costo, dato che per un tale prodotto si prevedono volumi piuttosto elevati.

4.2. Funzione fotografica integrata in un cellulare/PDA

Per un telefono, nonostante il processo di convergenza del settore ICT, la capacità fotografica non dovrebbe essere un fattore cruciale. L'utilizzo tipico deve prevedere la presenza di un'interfaccia basata su pochi tasti, possibilmente gli stessi del telefono, con controlli e gestione molto semplificati, senza particolari elaborazioni dell'immagine se non una buona compressione. La risoluzione richiesta è medio-bassa e la velocità di scatto o la capacità di lavorare con bassa luminosità non sono determinanti. Non sono previste interfacce dedicate per le fotografie ma si dovrà fare uso di quelle del telefono, per esempio WiFi o Bluetooth. Il consumo di potenza e le dimensioni devono essere invece molto contenuti. Anche il costo finale della funzionalità deve essere estremamente ridotto.

4.3. Funzione di scatto singolo in videocamere digitali

La funzionalità in questo caso non è in realtà legata allo sviluppo di un sistema *ad-hoc*, ma viene in massima parte *derivata* da quella del sistema di acquisizione video esistente. Ciò significa avere forti vincoli, per esempio, riguardo al tipo e alla risoluzione del sensore ottico, che, essendo quello della videocamera sarà probabilmente inferiore al megapixel. Per contro la disponibilità di risorse di elaborazione, la dimensione delle batterie, la quantità di memoria disponibile, gli ingombri e la varietà di interfacce non sono aspetti critici e possono essere facilmente adattati per supportare la nuova funzionalità.

4.4. Funzionalità di scatto con webcam per PC fisso o portatile

Questo tipo di scenario d'uso ha molti punti in comune con il precedente, poiché anche in questo caso si tratta di una funzione *derivata* da quelle di un diverso sistema ospite. Le maggiori differenze risiedono nel tipo di interfaccia utente, nella memorizzazione e nel collegamento: si utilizzerà infatti un'interfaccia grafica e si ricorrerà ai sistemi di trasferimento dati e di memorizzazione normalmente presenti in un personal computer. Né la velocità di scatto né le dimensioni sono influenti.

Benché sintetica, una simile descrizione degli scenari applicativi di una stessa funzionalità mette comunque in luce buona parte dei vincoli iniziali di un progetto realistico. Tali considerazioni sono il punto di partenza per il raffinamento dell'architettura hardware e software mirato alla realizzazione della funzione in esame.

A questo punto entrano in gioco conoscenze, metodologie e strumenti specifici dei diversi domini: progettazione di sistemi elettronici analogici, di componenti digitali, di *printed circuit board* e di parti meccaniche, sviluppo di software di base e di software applicativo, conoscenze approfondite dei sistemi operativi e dei problemi di networking, progettazione e sviluppo di interfacce uomo-macchina e molti altri ancora. Tutto ciò rende la progettazione dei sistemi embedded alquanto complessa e critica ma ne fa al tempo stesso una delle più interessanti sfide per i progettisti e gli ingegneri di oggi.

Bibliografia

[1] Brandolese C., Fornaciari W.: *Sistemi Embedded*. Pearson Education, Milano, 2007.

[2] Marwedel P.: *Embedded System Design*. Kluwer Academ1ic Publisher, Boston 2003.

[3] Vahid F., Givargis T.: *Embedded System Design - A unified Hardware/Software Introduction*. John Wiley & Sons, 2002.

[4] HiPEAC: European Network of Excellence on High-Performance EmbeddedArchitecture and Compilation. *The HiPEAC Roadmap on Embedded Systems*. Sixth Framework Programme, 2006.

[5] FAST GmbH: *Study of Worldwide Trends and R\&D Programmes in Embedded Systems in View of Maximising the Impact of a Technology Platform in the Area*. Final Report for the European Commission, Bruxelles 2005.

[6] MEDEA+ Office: *Design Automation Roadmap e Applications Technology Roadmap*. http://www.medeaplus.org Parigi, 2007 (Aggiornato annualmente).

[7] ITRS: *International Technology Roadmap for Semiconductors*. Web site: http://www.itrs.net/reports.html 2007 (Aggiornato annualmente).

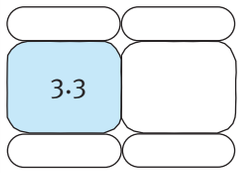
Carlo Brandolese, Ricercatore del Dipartimento di Elettronica e Informazione del Politecnico di Milano dal 2004 si occupa di problemi metodologici legati alla progettazione dei sistemi digitali e più in generale dei sistemi embedded. Negli ultimi anni i suoi interessi si sono rivolti in particolare alle problematiche di stima e ottimizzazione della potenza assorbita dalle componenti software dei sistemi misti. Collabora inoltre da più di 10 anni con il Cefriel occupandosi sia di trasferimento tecnologico e sviluppo hardware/software sia di formazione tecnica rivolta a studenti laureati e a professionisti.
E-mail: carlo.brandolese@polimi.it

William Fornaciari è professore associato presso il Politecnico di Milano, Dipartimento di Elettronica e Informazione. Si occupa da circa 15 anni di trasferimento tecnologico con aziende attive nel settore dei sistemi *embedded* e *consumer electronics*. I suoi interessi di ricerca riguardano le strategie per la riduzione del consumo energetico, sotto il profilo sia hardware sia software, le metodologie di progettazione di sistemi misti hardware-software, i sistemi operativi e le reti di sensori wireless. A partire dal 1990 ha collaborato a 7 progetti di ricerca internazionali su tali tematiche. È autore di oltre 100 articoli pubblicati in conferenze/riviste internazionali ed ha ricevuto 4 International Awards. Nel corso del 2007 ha pubblicato il primo libro italiano sui sistemi embedded.
E-mail: william.fornaciari@polimi.it

MODELLI PER L'ANALISI DI SISTEMI CRITICI

Francesco Flammini
Nicola Mazzocca
Valeria Vittorini

L'impiego di elaboratori in contesti diversi e sempre più delicati richiede di prestare particolare attenzione alle problematiche di sicurezza e affidabilità. In questo articolo, dopo una breve introduzione ai concetti base della *dependability* dei sistemi di elaborazione, verrà presentata una panoramica sulle tecniche modellistiche più diffuse per l'analisi di sistemi critici. Basandosi su vantaggi e limitazioni dei diversi approcci descritti, sarà introdotto il tema della modellazione multi-formalismo, riferendosi ad alcune delle prospettive di ricerca più interessanti in tale ambito.



1. INTRODUZIONE

L'evoluzione dei sistemi di elaborazione negli ultimi decenni ha portato alla nascita di calcolatori sempre più funzionali e performanti, ma non necessariamente più affidabili. In altre parole, all'aumento di complessità è corrisposta una riduzione di affidabilità. L'unica eccezione è costituita dal settore dei sistemi cosiddetti "*dependable*", a cui appartengono, per esempio, i sistemi di controllo in tempo reale impiegati in applicazioni critiche per la sicurezza di persone e beni materiali. La ragione di ciò è da ricercarsi nelle tecniche di sviluppo e verifica di tali sistemi, imposte da rigorosi standard internazionali. Tali tecniche fanno pesantemente uso di modelli al fine di dimostrare formalmente il rispetto dei requisiti qualitativi e quantitativi imposti dalla specifica. Nonostante ciò, anche tali sistemi non sono stati esenti da clamorosi fallimenti, il cui rischio aumenta rapidamente al crescere della complessità [14].
In questo articolo, dopo una breve introduzione ai concetti base della *dependability* dei sistemi di elaborazione, verrà presentata una panoramica sulle tecniche modellistiche più diffuse per l'analisi di sistemi *dependable*. Basandosi su vantaggi e limitazioni dei diversi approcci presentati, sarà introdotto il tema della modellazione multi-formalismo e dello sviluppo di modelli basato su componenti (che consente di governare sotto diversi aspetti la crescente complessità dei sistemi *dependable*), riferendosi anche ad alcune delle prospettive di ricerca più interessanti in tale ambito.

2. CONCETTI BASE E TECNICHE PER LO SVILUPPO DI SISTEMI *DEPENDABLE*

Una particolare classe di sistemi di elaborazione è quella dei sistemi cosiddetti "critici", ovvero tali che un fallimento nell'assolvere una o più delle funzioni ad essi richieste può comportare conseguenze di notevole gravità in termini di perdite economiche e/o di danni a persone o ambiente (Figura 1). Tra questi rientra-

no, in particolare, i sistemi cosiddetti in tempo reale (*real-time*) usati in applicazioni di controllo industriale, i quali devono reagire in tempi prestabiliti a stimoli esterni (riquadro 1).

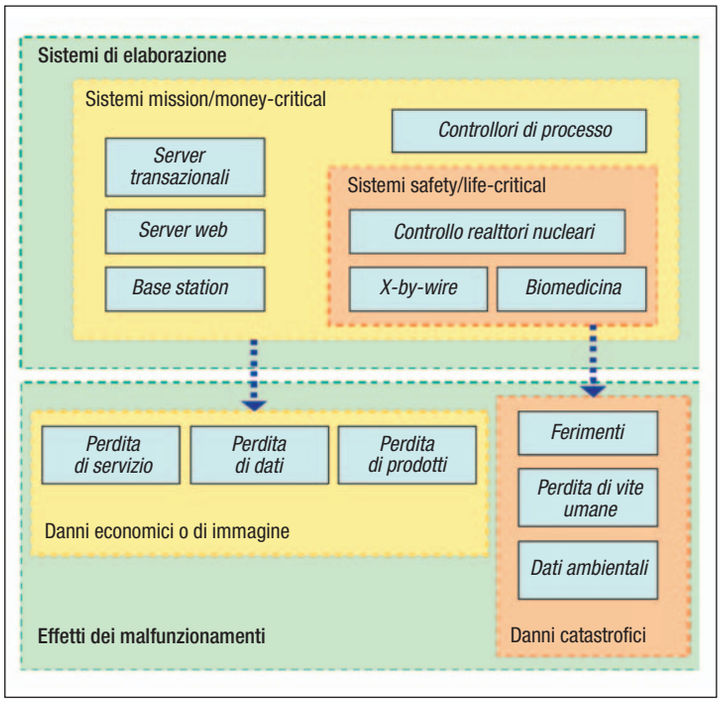


FIGURA 1

Classificazione di alcune tipologie di sistemi critici

Allo scopo di definire una tassonomia di riferimento non ambigua per lo studio di tali sistemi, si è introdotto in tempi relativamente recenti il concetto integrato di *dependability*. Secondo la definizione fornita da Avizienis et al. [1], un sistema può definirsi "*dependable*" se è possibile riporre fiducia nel suo funzionamento in modo giustificato. Chiaramente, questa definizione risulta piuttosto generica, dal momento che è possibile raggiungere diversi gradi di *dependability*, in funzione della particolare applicazione in cui il sistema andrà ad operare.

Un'esposizione sistematica della *dependability* può essere effettuata ricorrendo a tre concetti fondamentali: le minacce, gli attributi e gli strumenti (si faccia riferimento alla Figura 2) [1].

Gli attributi base di un sistema *dependable* sono: affidabilità, disponibilità, manutenibilità, innocuità (*safety*), integrità e riservatezza (*privacy*). A partire da questi, è possibile definire tutta una serie di attributi derivati: resilienza, plasticità, fidatezza (*trustworthiness*), sopravvivenza (*survivability*) ecc.. Tra questi è significativo citare la sicurezza (*security*), composta dagli attributi di disponibilità (per utenti autorizzati), integrità e riservatezza, e la performa-

RIQUADRO 1 - Sistemi *dependable* e sistemi critici

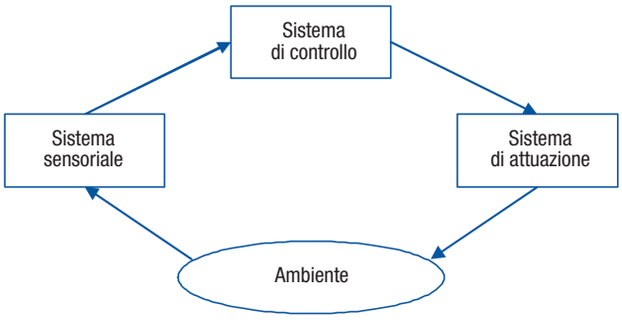
Una particolare classe di sistemi per i quali gli attributi di *dependability* assumono una rilevanza fondamentale è costituita dai sistemi cosiddetti "critici". Esistono diverse tipologie di sistemi critici (si veda la Figura 1):

- Sistemi *mission-critical* o *money-critical*, ovvero tali che un fallimento può comportare perdite notevoli in termini finanziari (*web server*, *base stations* ecc.).
- Sistemi *safety-critical* o *life-critical*, ovvero tali che un fallimento può comportare danni in termini di ferimenti o perdita di vite umane (macchine per dialisi, sistemi di guida automatica *X-by-wire* per autoveicoli, treni o aerei, controllo delle centrali nucleari ecc.).

Tipicamente un sistema *life-critical* è anche *money-critical*, ma ciò non sempre è vero, come nel caso dei sistemi biomedici. È invece vero che un sistema l*ife-critical*, dovendo reagire a stimoli esterni entro tempi rigidamente prestabiliti (Figura **A**), è anche di tipo *hard real-time*. Solo in linea teorica sarebbero classificabili come *life-critical* anche alcuni tipi di sistemi non di tipo *real-time* come quelli usati per il calcolo scientifico o strutturale: infatti, anche se i tempi di calcolo in tal caso non risultano rilevanti, risultati errati potrebbero ugualmente compromettere l'incolumità degli individui.

Per rendere rigorosa una tale classificazione si ricorre al concetto di "costo previsto del fallimento" (CP_M, a volte definito anche "rischio"), definito come segue:

$$CP_F = p_F \cdot C_F$$



A - *Schema teorico di un sistema di controllo*

Dove:

p_F è la probabilità di occorrenza del fallimento (esempio, numero di volte in un anno);

C_F è una stima delle conseguenze del fallimento (esempio, espresso in numero di vittime o danni monetari).

bilità (*performability*), ovvero la capacità di fornire prestazioni adeguate anche in presenza di uno o più malfunzionamenti.
Si noti che i due termini *safety* e *security*, entrambi traducibili in italiano come "sicurezza", esprimono concetti diversi. Il primo è relativo alla sicurezza in termini di incolumità di persone e ambiente, il secondo si riferisce alla sicurezza intesa in senso informatico[1].
Minacce alla *dependability* di un sistema sono guasti, errori e fallimenti. Un errore rappresenta l'alterazione dello stato del sistema causato da un guasto, mentre si ha un fallimento quando l'errore si propaga all'interfaccia del sistema (o del sotto-sistema), causando un'alterazione percettibile di uno o più servizi forniti. La distinzione tra errore e fallimento deriva dalla considerazione che in un sistema complesso l'errore può essere confinato ad uno o più componenti senza raggiungere l'interfaccia di servizio del sistema. Le minacce possono essere classificate in diversi modi, in funzione di origine (naturali, intenzionali), consistenza (riproducibili, elusive), persistenza (transitorie, intermittenti, permanenti), severità (minori, catastrofiche) ecc.. La figura 3 mostra un possibile diagramma di contesto per un sistema di controllo computerizzato, in cui sono evidenziate le entità esterne di diversa natura con cui il sistema interagisce e che possono essere sorgenti di guasti (errori di progettazione, sbalzi di tensione, surriscaldamento, errato utilizzo, manomissione ecc.).
Gli strumenti disponibili per assicurare la *dependability* consistono nel prevedere, preve-

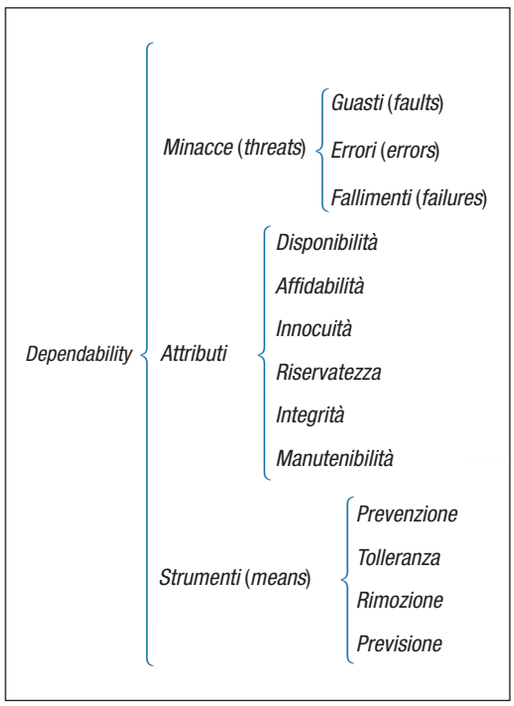


FIGURA 2
L'albero della dependability

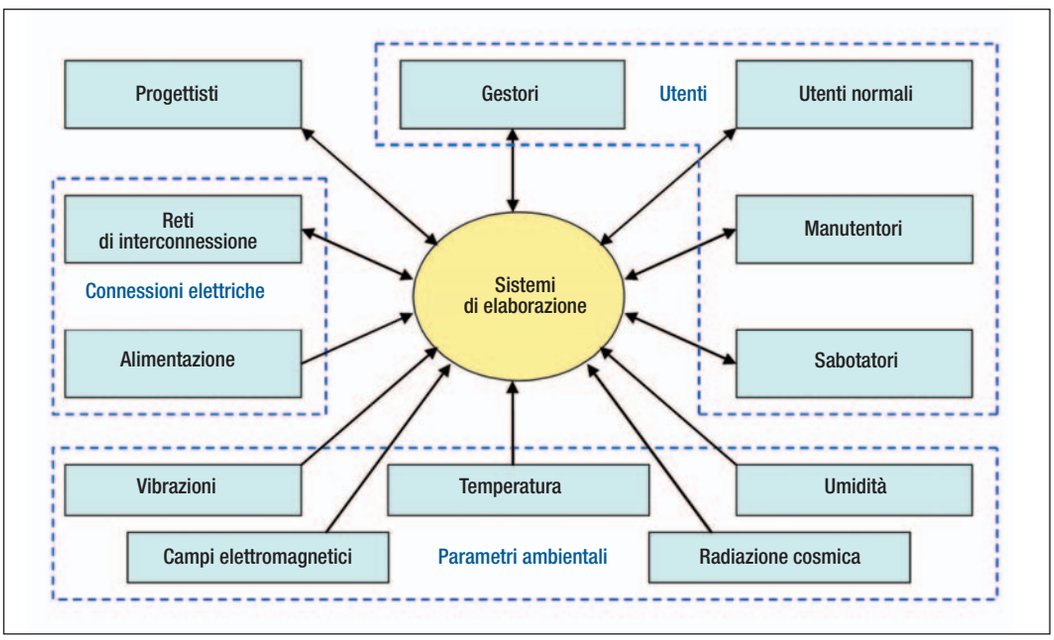


FIGURA 3
Diagramma di contesto di un sistema di controllo riferito alle sorgenti di guasti

[1] Nel contesto delle infrastrutture critiche, la parola *security* può anche essere intesa come protezione nei confronti di minacce esterne al sistema, di origine naturale o intenzionale.

RIQUADRO 2 - Tecniche più diffuse per lo sviluppo e l'analisi dei sistemi dependable

Tra le tecniche più diffuse e di successo in ambito industriale (*best practices*), è opportuno elencare le seguenti (trattasi di approcci complementari applicabili a diversi livelli di astrazione e in fasi distinte del ciclo di vita del sistema):

- Analisi degli azzardi (*Hazard Analysis*, HA) e dei modi di fallimento (*Failure Modes and Effects Analysis*, FMEA), che hanno lo scopo di individuare criticità del sistema ed opportuni meccanismi di protezione.
- Ispezioni tecniche della specifica (*Formal Technical Inspection*, FTI) per garantire non ambiguità, consistenza, completezza e coerenza dei requisiti.
- Iniezione dei guasti a livello hardware (*fault injection*), eseguita sul sistema reale o un suo modello simulato, allo scopo di valutare l'efficacia dei meccanismi di tolleranza ai guasti.
- Verifica del software statica (analisi, ispezione, *walk-trough*) o dinamica (*testing*), finalizzata ad individuare errori sistematici presenti nel codice prodotto.
- Tolleranza ai guasti basata su ridondanza spaziale (replicazione attiva o passiva), tra cui codici a correzione d'errore (*Error Correcting Codes*, ECC), architetture N-modulari (*N-Modular Redundancy*, NMR) con votazione a maggioranza sulle uscite di sezioni indipendenti (isolate e diversamente sviluppate), *watchdog timers*, *middleware* e protocolli robusti basati su cifratura, firma digitale, marcature temporali e numeri di sequenza.
- Ringiovanimento (*rejuvination*) e tolleranza ai guasti a livello software (*software fault tolerance*), per esempio, basata su ridondanza temporale realizzata con blocchi di recupero (*recovery blocks*).
- Programmazione difensiva, che consiste in una disciplina di codifica che contempla e tratta opportunamente condizioni d'errore e anomalie impreviste.
- Analisi dei ritorni dal campo ed elaborazione delle contromisure (*Failure Reporting Analysis & Corrective Action System*, FRACAS).

Nel caso dei sistemi critici, l'adozione di molte delle suddette tecniche è imposta da rigorosi standard internazionali (per esempio, il DO-178B impiegato in campo aeronautico) in funzione del livello di integrità della sicurezza (*Safety Integrity Level*, SIL) richiesto al sistema.

nire, tollerare e rimuovere le minacce ed i relativi effetti. Alcune delle *best practices* industriali utilizzate nelle diverse fasi del ciclo di sviluppo dei sistemi *dependable* sono riportate nel riquadro 2 (per approfondimenti si può far riferimento a [23]).

La tabella 1 riporta le suddette tecniche classificandole in funzione del ciclo di vita, del livello di astrazione (hardware o software) e delle tipologie di minacce che riescono a fronteggiare.

In particolare, in questo contesto ci concentreremo sulle tecniche di analisi basate su modelli formali. Tali modelli, rispetto agli approcci simulativi, consentono di valutare sia in termini qualitativi (proprietà del sistema) che quantitativi (indici numerici) gli attributi della *dependability* del sistema oggetto di analisi, allo scopo di:

- supportare la scelta dei parametri progettuali prevedendone l'effetto fin dalle prime fasi del ciclo di sviluppo;
- dimostrare il soddisfacimento dei requisiti degli standard RAMS[2] internazionali.

La verifica formale di proprietà è possibile ricorrendo a tecniche di *model-checking* e *theorem-proving* [18], mentre la valutazione di parametri quantitativi è in genere ottenuta attraverso l'analisi di modelli stocastici. Alcuni dei formalismi basati su grafi consentono

TABELLA 1
Metodi o strumenti utilizzati nell'ambito dei sistemi dependable

Metodo o strumento	Fase	HW/SW	Classi di minacce
Hazard analysis, FMEA	Specifica	Entrambi	Tutte
Analisi statica, ispezione codice e testing	Verifica	SW	Sistematiche
Fault Injection	Verifica	HW	Sistematiche
Meccanismi di Fault-Tolerance	Operativa	Entrambi	Casuali (tutte in caso di diversity)
Programmazione difensiva	Codifica	SW	HW casuali, SW sistematiche
Metodi formali	Specifica, verifica	Entrambi	Sistematiche
Modelli stocastici predittivi	Specifica, verifica	HW	Casuali
FRACAS, Measurement Based Analysis	Operativa	Entrambi	Tutte
Software Rejuvination	Operativa	SW	Casuali
Formal Technical Inspection	Specifica	Entrambi	Sistematiche

[2] RAMS: *Reliability, Availability, Maintainability, Safety* (Affidabilità, Disponibilità, Manutenibilità, Sicurezza).

entrambe le tipologie di analisi, come sarà meglio descritto nel paragrafo successivo.

3. FORMALISMI PER L'ANALISI DEI SISTEMI DEPENDABLE

I formalismi per lo studio dell'affidabilità dei sistemi *dependable* possono dividersi in due grandi categorie: quelli che seguono un approccio di modellazione top-down (o deduttivo) e quelli di tipo *bottom-up* (induttivi), a seconda che l'analisi parta dal fallimento a livello di sistema o da un guasto dei costituenti elementari. Il formalismo storicamente più diffuso per le analisi top-down è quello degli Alberi dei Guasti (*Fault Trees*, FT), mentre per le analisi *bottom-up* il formalismo di maggior successo è quello dei Diagrammi a Blocchi Affidabilistici (*Reliability Block Diagrams*, RBD). Entrambi questi formalismi si prestano ad analisi di tipo combinatoriale, che ne limita la potenza espressiva. Per esempio, RBD e FT non consentono la modellazione di tipi di guasto di modo comune, cioè guasti simultanei ad unità identiche causati dallo stesso motivo e nello stesso modo, e quindi tali da far cadere l'ipotesi di indipendenza tra le probabilità di guasto dei singoli costituenti.
Per ovviare a tali limitazioni, possono essere impiegati formalismi basati sull'analisi dello spazio di stato, quali catene di Markov a tempo continuo (*Continuous Time Markov Chains*, CTMC) e reti di Petri temporizzate o stocastiche (*Timed/Stochastic Petri Nets*, TPN/SPN). Tali formalismi consentono tra le altre cose una modellazione dettagliata degli aspetti di manutenibilità (diagnostica e recupero dagli errori).
Esistono diversi altri formalismi impiegabili per l'analisi dei sistemi *dependable*, tra cui gli automi temporizzati (*Timed Automata*, TA), i diagrammi di decisione binari (*Binary Decision Diagrams*, BDD) e le algebre dei processi (*Process Algebras*, PA), che possono risultare utili per l'analisi di particolari proprietà o attributi real-time. Inoltre, formalismi nati per scopi diversi sono risultati utili per l'analisi di certi aspetti dei sistemi *dependable*. Tra questi è opportuno citare le reti di code (*Queueing Networks*, QN), utili per modellare aspetti di *performability* in congiunzione con altri formalismi, e le reti bayesiane (*Bayesian Networks*, BN), che consentono di estendere la potenza espressiva degli alberi dei guasti senza incorrere nel problema dell'esplosione dello spazio di stato [3].
Infine, sono stati definiti dei formalismi derivati dai formalismi "base" sopra menzionati, aggiungendo varie estensioni e ricorrendo a metodi di soluzione spesso basati sulla traduzione dei modelli in formalismi con potenza espressiva più elevata. È il caso ad esempio degli alberi dei guasti, di cui sono state definite diverse estensioni (dinamici, parametrici, riparabili ecc.) [10].
Modelli espressi in taluni formalismi (per esempio, SPN) si prestano sia ad analisi di tipo quantitativo (esempio, tasso di fallimenti) che di verifica di proprietà (esempio, raggiungibilità di uno stato non sicuro). Altri (esempio, BN) consentono analisi qualitative sia di tipo "*what if*?", finalizzate a valutare quali sono le conseguenze provocate dal verificarsi di una determinata condizione, che di tipo "*backward*" (a ritroso) o di "*most probable explanation*" (spiegazione più plausibile), che hanno lo scopo di valutare le cause (esempio, guasti) a partire dagli effetti (esempio, fallimenti), e pertanto risultano particolarmente adatti ad analisi di sensitività parametrica o anche, in fase operativa, di tipo diagnostico.
Anche linguaggi per la specifica di sistemi, quali UML (*Unified Modeling Language*) e SysML (*Systems Modeling Language*), sono impiegabili per analisi formali di *dependability* [4].
In conclusione a questo rapido excursus, che non ha le pretese di essere esaustivo, è opportuno evidenziare come non esista un formalismo che sia in assoluto migliore degli altri. La scelta di adottare un linguaggio piuttosto che un altro va formulata in funzione di tutta una serie di fattori, tra cui la facilità d'uso, la verificabilità del modello, la potenza espressiva e, non ultima, l'efficienza degli algoritmi risolutivi (che potrebbe impedire, di fatto, la soluzione dei modelli di una certa complessità, tutt'altro che infrequenti nelle applicazioni reali). Pertanto, la scelta dei modelli va assolutamente calibrata in funzione degli "aspetti" da modellare (tipologia di componente, livello di astrazione ecc.) e degli obiettivi dell'analisi, ovvero di quali attributi è necessario valutare in corrispondenza di quali minacce [20]. Tipica-

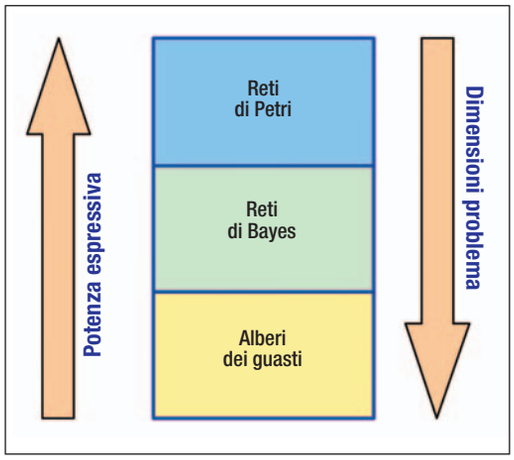


FIGURA 4
Scelta dei formalismi in base a potenza espressiva e dimensioni del problema

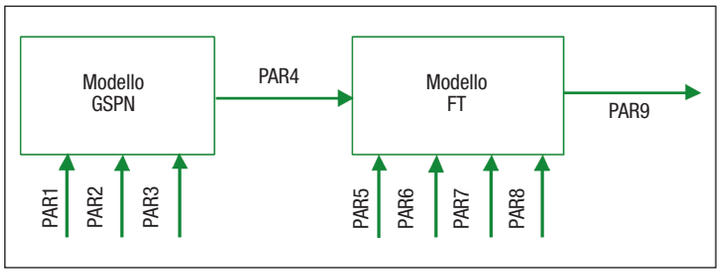


FIGURA 5
Un semplice esempio di composizione

mente, la scelta è dettata da un *trade-off* tra potenza espressiva ed efficienza: il rapporto reciproco tra questi due fattori è rappresentato dalla figura 4, che suggerisce, per esempio, di impiegare alberi dei guasti per sistemi di grandi dimensioni in cui non si abbiano esigenze particolari riguardo gli aspetti da modellare (ovvero si sia disposti a tollerare i limiti modellistici del formalismo).

4. NUOVI APPROCCI ALLA MODELLAZIONE

Si è visto nel precedente paragrafo come la modellazione della *dependability* richieda un compromesso tra diversi fattori, tra cui i più importanti sono la facilità d'uso degli strumenti, la potenza espressiva dei formalismi e l'efficienza dei risolutori. Detto in altre parole, sarebbe impensabile modellare un sistema complesso utilizzando formalismi di elevata potenza espressiva, come quelli basati sulle reti di Petri, perché anche se si riuscisse a gestire la difficoltà modellistica, si otterrebbe un modello impossibile da risolvere su qualsiasi elaboratore commerciale. D'altra parte, è noto come la valutazione degli attributi di *dependability* di un sistema richieda un approccio olistico. Si intuisce subito come l'unica soluzione soddisfacente al problema debba essere ricercata in tecniche modulari o composizionali in cui modelli eterogenei sono integrati in viste coese dell'intero sistema.

Una possibile strada per affrontare il problema consiste nell'adottare tecniche di modellazione **multi-formalismo**[3], il cui obiettivo è di supportare la rappresentazione di sistemi complessi combinando diversi linguaggi e tecnici di analisi. Un modello multi-formalismo è un modello costituito da parti, ciascuna delle quali modella un particolare componente o aspetto del sistema che si intende modellare, utilizzando il formalismo più adatto allo scopo. I diversi sotto-modelli contribuiscono alla creazione del modello complessivo che viene pertanto risolto coinvolgendo diverse tecniche e strumenti di analisi.

Prima di esaminare alcuni concetti fondamentali ed i diversi approcci attualmente utilizzati nell'ambito dello sviluppo di modelli multi-formalismo, può essere utile descrivere un semplice esempio. Supponiamo di volere analizzare un modello costituito da due sotto-modelli, rispettivamente realizzati mediante una Rete di Petri Stocastica Generalizzata (GSPN) e mediante un Albero dei Guasti (FT). Supponiamo che risolvendo il modello complessivo si voglia ottenere una misura della disponibilità del sistema modellato, ottenibile calcolando la probabilità di guasto del sistema data dal modello FT. Supponiamo infine che uno dei parametri necessari all'analisi dell'albero dei guasti (per esempio, la probabilità di occorrenza di un evento) sia fornita dall'analisi del modello GSPN. Si tratta dunque di un semplice caso di composizione sequenziale descritta nella figura 5, dove i due sotto-modelli sono graficamente rappresentati da rettangoli, mentre i dettagli relativi alla loro struttura sono nascosti. In figura sono mostrate solamente le informazioni quantitative che devono essere fornite per popolare il modello e poterlo utilizzare (da PAR1 a PAR3 per il sotto-modello GSPN e

[3] La modellazione multi-formalismo è un aspetto della modellazione multi-paradigma, che comprende anche i concetti di astrazione, meta-modellazione e trasformazione di modelli [19].

da PAR5 a PAR8 per il sotto-modello FT), le informazioni che devono essere prodotte dall'analisi della rete GSPN (PAR4) ed il valore atteso dalla risoluzione del modello complessivo (PAR9). Questo modello può essere risolto utilizzando diversi strumenti di analisi, per esempio Sharpe [21].
Il modello di figura 5 è un modello multi-formalismo, costruito componendo esplicitamente sotto-modelli espressi mediante linguaggi di modellazione differenti. La composizione è realizzata in generale mediante operatori più o meno sofisticati ed è un punto cruciale nella definizione di metodologie di modellazione avanzate. Gli operatori di composizione consentono la condivisione di *stati*, *azioni* o *eventi* tra modelli diversi. Nel caso particolare in cui essi si limitano allo scambio di risultati tra modelli risolti in modo isolato (come nel caso dell'esempio) si parla di "connettori".
Affinché una composizione sia corretta è necessario in generale che:

- il modello composto verifichi determinate proprietà e che garantisca la conservazione di proprietà già verificate dai sotto-modelli (per esempio, proprietà strutturali come l'assenza di situazioni di *deadlock*);
- l'operazione sia semanticamente valida, e cioè che abbia senso l'interazione tra diversi formalismi come definita attraverso la composizione (per esempio, deve aver senso utilizzare la frequenza di scatto di una transizione di una rete di Petri per valutare la probabilità di occorrenza di un evento in un albero dei guasti);
- siano definite opportunamente la sequenza con cui i diversi sotto-modelli devono entrare in gioco nella risoluzione del modello complessivo (per esempio, il modello GSPN deve essere risolto prima del modello FT) e le azioni da effettuare (per esempio, valutare il *throughput* di una transizione del modello GSPN, calcolarne l'inverso, usare questo dato come valore di un parametro del modello FT).

Una volta che i precedenti aspetti sono stati affrontati e risolti (non esiste un metodo generale, queste problematiche vanno affrontate a partire dagli specifici formalismi coinvolti), la composizione è uno strumento molto potente perché consente di introdurre nell'ambito della modellazione di sistemi complessi tecniche simili a quelle utilizzate nel-

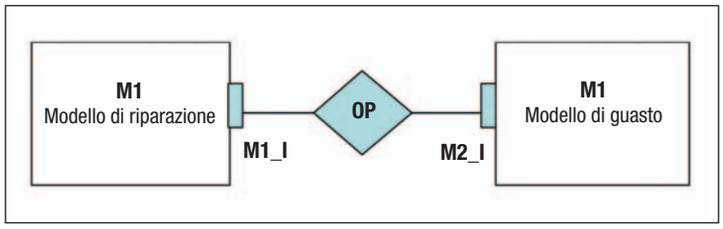


FIGURA 6
Un modello composto

l'ingegneria del software nello sviluppo per componenti.
Nella figura 6 è mostrata una diversa rappresentazione del modello descritto nella figura 5: i due sotto-modelli sono visti come scatole nere (componenti) M1 ed M2, che esportano un'interfaccia (M1_I e M2_I) e forniscono un comportamento noto (che cosa modellano); il connettore/operatore è esplicitato (il rombo etichettato OP); gli archi connettono l'operatore agli "operandi" cui si applicano. L'interfaccia esplicita le informazioni coinvolte ed eventuali vincoli che devono essere rispettati. L'operatore incapsula al suo interno le azioni da compiere e le informazioni circa l'ordine con cui tali azioni devono essere svolte. Nel caso dell'esempio, con riferimento ad un sistema (o ad una sua parte), M1 modella un'azione di riparazione che influenza la probabilità di guasto di uno dei componenti del sistema ed M2 modella le combinazioni di eventi di guasto che possono condurre all'indisponibilità del sistema.
A questo punto è possibile pensare di sostituire, per esempio, il sotto-modello M1 con un altro sotto-modello M espresso mediante un differente formalismo (o un modello espresso mediante lo stesso formalismo ma ad un diverso livello di astrazione), a condizione che quest'ultimo:

a. fornisca un comportamento equivalente (M modella lo stesso aspetto del sistema modellato da M1);
b. rispetti l'interfaccia specificata.

La definizione degli operatori di composizione, delle interfacce ed i problemi relativi alla loro tipizzazione sono un aspetto estremamente delicato e sono un campo aperto di ricerca [17]. L'interesse della comunità scientifica per questo approccio alla modellazione è dettata dai vantaggi che esso comporta: promuove il riuso e la definizione di librerie di

modelli, e quindi lo sviluppo di strumenti che in parte alleviano la difficoltà legata allo sviluppo di modelli complessi; introduce un certo grado di flessibilità e dinamicità nella definizione del modello (si potrebbe definire quale sotto-modello utilizzare per realizzare una parte del modello complessivo sulla base, ad esempio, di valori forniti dalla risoluzione di altre parti); supporta efficacemente metodologie di tipo *divide-et-impera* e la definizione di modelli gerarchici.

A titolo di esempio, la figura 7 mostra un possibile modello multi-formalismo di un complesso sistema di controllo ferroviario. Il sistema preso per esempio deriva dallo standard europeo per le nuove linee ferroviarie interoperabili ad alta velocità. Si tratta di un sistema geograficamente distribuito, in cui si distinguono la parte installata nelle cabine di guida dei treni (*on-board*), quella di terra distribuita lungo linea (*lineside*) e quella di terra concentrata nei centri di controllo (*trackside*). Il modello globale è realizzato attraverso una composizione multi-livello di modelli eterogenei relativi a componenti (o sotto-sistemi) distinti. I modelli al livello di astrazione inferiore rappresentano aspetti hardware, che si modellano tipicamente ricorrendo ad alberi dei guasti (e relative estensioni) e/o reti bayesiane. I modelli al livello di astrazione intermedio rappresentano le diverse modalità operative (marcia a vista, supervisione completa ecc.), che sono legate alla disponibilità dei sottosistemi hardware e a loro volta condizionano le funzionalità del sistema a più alto livello; questi aspetti possono essere modellati agevolmente attraverso automi a stati finiti (o formalismi analoghi). Infine, i livelli più alti rappresentano le procedure operative (es. inizio e fine missione), che richiedono formalismi adatti a modellare aspetti comportamentali, quali per esempio reti di Petri e relative estensioni.

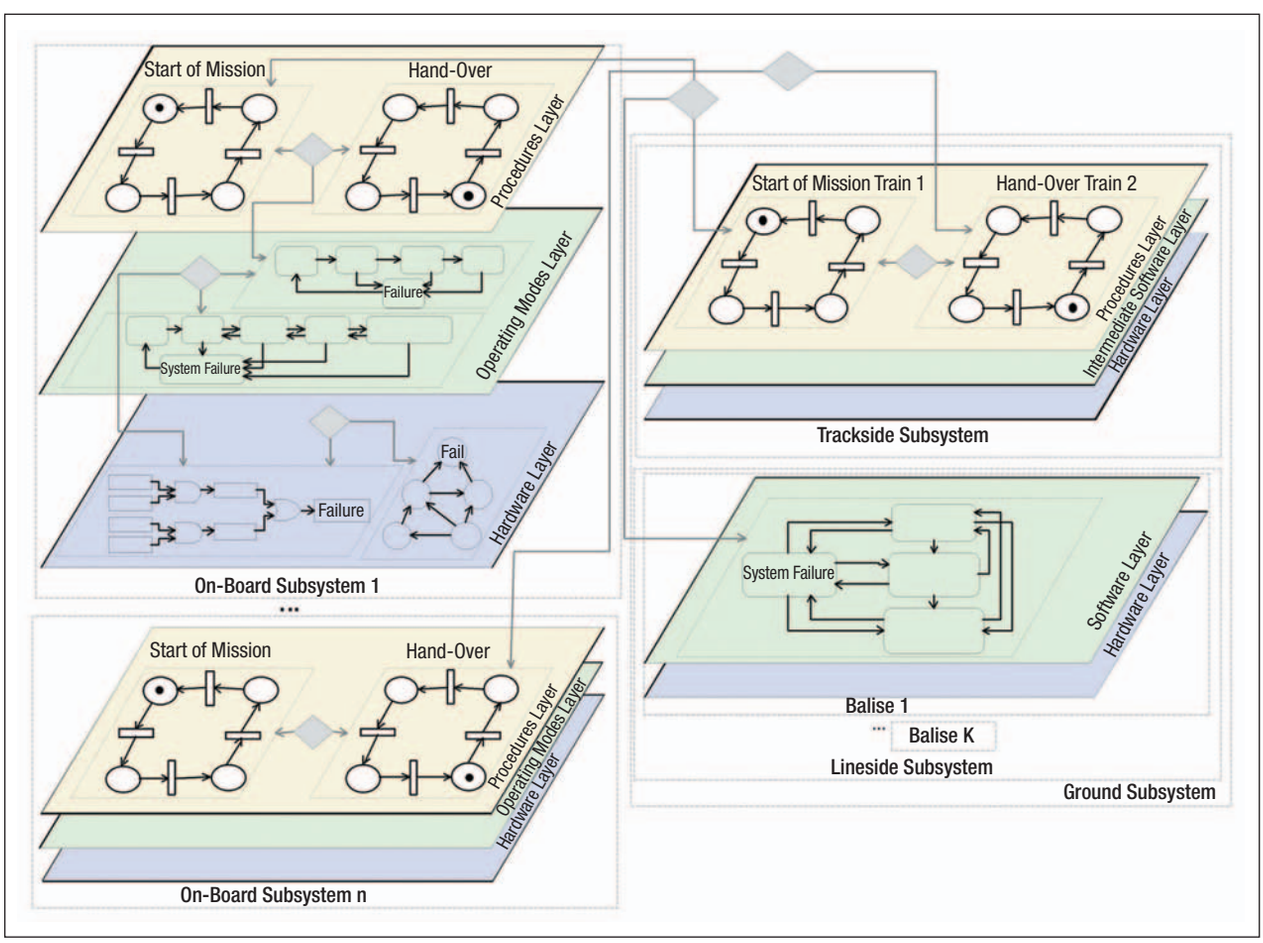


FIGURA 7

Schema per la modellazione multi-formalismo di un complesso sistema di controllo ferroviario

I modelli sono interconnessi attraverso operatori (rappresentati nella Figura 7 dai rombi a sfondo grigio):

- *intra /inter-livello*, a seconda che l'interazione avvenga tra modelli dello stesso livello (*intra*) o di diversi livelli (*inter*);
- *intra /inter-componente*, a seconda che l'interazione avvenga tra modelli dello stesso componente (*intra*) o di componenti diversi (*inter*).

Un siffatto modello consentirebbe una valutazione olistica di attributi di *dependability* (per esempio, disponibilità o sicurezza), relativi a modi di guasto definiti a livello di sistema complessivo, che altrimenti andrebbero valutati attraverso l'analisi di modelli isolati, con la conseguenza di dover effettuare ipotesi poco realistiche o eccessivamente conservative (laddove possibile). Ipotesi poco realistiche possono compromettere la validità dei risultati, perché si trascurano i dettagli di alcune dinamiche di interazione; dal lato opposto, l'essere eccessivamente conservativi costringe ad un dimensionamento non ottimale dei parametri di progetto, con il rischio di comprometterne la fattibilità economica. Inoltre, la valutazione olistica di attributi di *dependability* facilita la produzione delle evidenze formali necessarie ad ottenere eventuali certificazioni.

La definizione di siffatti modelli gerarchici consente di affrontare l'analisi un sistema complesso secondo opportune metodologie. Per esempio è possibile modellare un sistema definendo un livello che descriva il comportamento dei diversi sotto-sistemi, un livello che descriva l'interazione tra i sotto-sistemi mediante una rete e un livello che modelli la rete di comunicazione. Ancora, è possibile sviluppare dei modelli che consentano di considerare l'impatto della manutenzione sull'affidabilità e sicurezza del sistema, sviluppando separatamente e poi componendo un modello di guasto ed uno di riparazione [12].

Modelli multi-formalismo sviluppati secondo un "approccio a componenti" consentono inoltre di gestire la complessità in diversi modi. Per esempio, tra gli approcci finalizzati alla riduzione delle dimensioni dei modelli è possibile individuare quelli:

- modulari, iterativi e gerarchici, ovvero di tipo "*divide et impera*" [7];
- di sostituzione di sottomodelli con modelli equivalenti, ma di dimensioni ridotte (si veda per esempio [15]);
- parametrici, che consentono una compattazione attraverso lo sfruttamento di simmetrie (tecnica nota come "*folding*", letteralmente "ripiegamento") [5].

A più basso livello, si possono impiegare tecniche che ricorrono al parallelismo distribuendo il carico di lavoro necessario alla risoluzione dei sotto-modelli su diverse macchine.

Evidentemente, l'analisi di modelli così complessi implica la disponibilità di opportune tecniche di soluzione e di strumenti automatici che supportino anche lo sviluppo del modello. Se è pensabile di sviluppare e risolvere separatamente utilizzando i *tool* disponibili modelli molto semplici, quale quello di figura 5, realizzando "a mano" l'operatore di composizione, ovvero prelevando il risultato ottenuto dall'analisi del primo modello e immettendolo opportunamente nel secondo), non è possibile procedere in questo modo per modelli quale quello mostrato nella Figura 7, e nei casi in cui gli operatori realizzino funzioni complesse, ad esempio nel caso in cui forniscano informazioni circa l'evoluzione di simulazioni in corso.

Esistono diversi strumenti di supporto allo sviluppo ed alla risoluzione di modelli multi-formalismo abbastanza assestati (uno stato dell'arte è riportato in [22]), tra cui è opportuno citare *SHARPE* [21], *SMART* [6], *DEDS* [2], *AToM*^{3} [16] e *Möbius* [8], che si differenziano tra loro per diversi aspetti. Per esempio, un'importante distinzione è tra i *framework* che ricorrono ad un unico formalismo di più alto livello per la rappresentazione del modello globale e quelli che invece mantengono separati i sottomodelli, che vengono valutati ognuno attraverso un proprio specifico risolutore. Tra questi ultimi è opportuno citare l'approccio innovativo proposto dal *framework OsMoSys* (*Object-baSed Multiformalism mOdeling of SYStems*) [24], che consiste nell'orchestrazione di risolutori esterni attraverso tecniche di *work-flow management* [9].

5. CONCLUSIONI

I modelli di *dependability* risultano di estrema importanza per supportare lo sviluppo e l'analisi dei sistemi *dependable*, in particolare di

quelli dedicati impiegati in applicazioni critiche (controllo di reattori nucleari, *fly/brake/steer-by-wire*, e-medicine, *e-banking* ecc.). Il loro impiego è contemplato da tempo nelle *best-practice* industriali per la valutazione di determinati attributi di affidabilità e sicurezza per sottosistemi limitati. Tre sono i fattori di cui tener conto allo scopo di indirizzare le tematiche di ricerca future in questo ambito:

- la crescente complessità dei sistemi di elaborazione, dovuta all'elevato numero di requisiti funzionali, che impattano su dimensioni del codice, distribuzione ed eterogeneità;
- la crescente criticità dei requisiti non funzionali, che diventano sempre più stringenti dal momento che un fallimento può avere impatti disastrosi in termini economici o di salute;
- la necessità di adottare approcci olistici per assicurare gli obiettivi di progetto, supportando le decisioni in fase di sviluppo e dando evidenza formale dei risultati ottenuti in fase di certificazione (*assessment*).

La gestione attraverso tecniche di modellazione multi-formalismo delle problematiche sopra menzionate richiede il raggiungimento di obiettivi ambiziosi, legati sia ad evoluzioni metodologiche che ad aspetti tecnologici. Tra questi citiamo la riduzione della complessità risolutiva, la ricerca di nuovi operatori per l'interoperabilità tra modelli eterogenei, e la simulazione distribuita. A fronte di uno sforzo sicuramente notevole richiesto in tali ambiti, i nuovi paradigmi di modellazione si prestano ad impieghi che travalicano i diversi settori dell'ingegneria, essendo applicabili a contesti multidisciplinari, quale quello delle infrastrutture critiche [13].

Bibliografia

[1] Avizienis A., Laprie J.-C., Randell B., Landwehr C.: Basic Concepts and Taxonomy of Dependable and Secure Computing. In: *IEEE Transactions on Dependable and Secure Computing*, Vol. 1, 2004, p. 11-33.

[2] Bause F., Buchholz P., Kemper P.: *A toolbox for functional and quantitative analysis of DEDS*. In: Proc. 10-th Int. Conf. on Modelling Techniques and Tools for Computer Performance Evaluation, LNCS 1469, Springer-Verlag, 1998, p. 356-359.

[3] Bobbio A., Bologna S., Ciancamerla E., Franceschinis G., Gaeta R., Minichino M., Portinale L.: *Comparison of Methodologies for the Safety and Dependability Assessment of an Industrial Programmable Logic Controller*. In. Proc. 12-th European Safety and Reliability Conference, 2001.

[4] Bondavalli A., Dal Cin M., Latella D., Majzik I., Pataricza A., Savoia G.: Dependability analysis in the early phases of UML-based system design. In: *Computer Systems Science and Engineering*, Vol. 16, n. 5, 2001, p. 265-275.

[5] Chiola G., Dutheillet C., Franceschinis G., Haddad S.: Stochastic Well-Formed Colored Nets and Symmetric Modeling Applications. In: *IEEE Transactions on Computers*, Vol. 42, 1993, p. 1343-1360.

[6] Ciardo G., Miner A.: SMART: *Simulation and Markovian Analyser for Reliability and Timing*. In Proc. 2-nd International Computer Performance and Dependability Symposium (IPDS'96), IEEE Computer Society Press, 1996, p. 60.

[7] Daly D., Sanders W.H.: *A connection formalism for the solution of large and stiff models*. In: Proc. 34-th Annual Simulation Symposium, 2001, p. 258-265.

[8] Deavours D.D., Clark G., Courtney T., Daly D., Derisavi S., Doyle J.M, Sanders W.H., Webster P.K.: The Möbius Modeling Framework and Its Implementation. *IEEE Trans. Software Engineering*, Vol. 28, n. 10, 2002, p. 956-969.

[9] Di Lorenzo G., Flammini F., Iacono M., Marrone S., Moscato F., Vittorini V.: *The software architecture of the OsMoSys multisolution framework*. In: Proc. 2-nd International Conference on Performance Evaluation Methodologies and Tools, VALUETOOLS'07, Nantes, France, October 23-25, 2007, p. 1-10.

[10] Flammini F., Iacono M., Marrone S., Mazzocca N.: *Using Repairable Fault Trees for the evaluation of design choices for critical repairable systems*. In: Proceedings of the 9-th IEEE Symposium on High Assurance Systems Engineering, HASE'05, Heidelberg, Germany, October 12-14, 2005, p. 163-172.

[11] Flammini F., Marrone S., Mazzocca N., Vittorini V.: *Modelling System Reliability Aspects of ERTMS/ETCS by Fault Trees and Bayesian Networks*. In: Safety and Reliability for Managing Risk: Proceedings of the 15-th European Safety and Reliability Conference, ESREL'06, Estoril, Portugal, September 18-22, 2006, p. 2675-2683.

[12] Flammini F., Marrone S., Mazzocca N., Vittorini V.: *A new modeling approach to the safety evaluation of N-modular redundant computer systems in presence of imperfect maintenance*. In: Reliability Engineering & System Safety, Vol. 94, Issue 9, September 2009, p. 1422-1432.

[13] Flammini F., Mazzocca N., Pragliola C., Vittorini V.: *A Study on Multiformalism Modelling of Critical Infrastructures*. In: Proc. 3-rd International Workshop on Critical Information Infrastructures Security, CRITIS-08, LNCS 5508, 2009, p. 336-343.

[14] Holzmann G.J: Conquering Complexity. In: *IEEE Computer*, Vol. 40, n. 12, 2007, p. 111-113.

[15] Jungnitz H., Desrochers A.A.: *Flow equivalent nets for the performance analysis of Generalized Stochastic Petri Nets*. In: Proceedings of the IEEE International Conference Robotics and Automation, 1991, p. 122-127.

[16] Lara J. d., Vangheluwe H.: AToM3: A *Tool for Multi-formalism and Meta-modelling*. In: Proc. 5-th Intl. Conf. on Fundamental Approaches To Software Engineering, LNCS vol. 2306. Springer-Verlag, London, 2002, p. 174-188.

[17] Laurent Doyen L., Henzinger T.A., Jobstmann B., Petrov T.: *Interface theories with component reuse*. In: Proc. 8-th Annual Conference on Embedded Software (EMSOFT), ACM Press, 2008, p. 79-88.

[18] Merz S.: Model checking: a tutorial overview. In: Modeling and Verification of Parallel Processes. *LNCS*, Vol. 2067, 2001, p. 3-38.

[19] Mosterman P.J., Vangheluwe H.: Computer Automated Multi-Paradigm Modeling: An Introduction. *SIMULATION*, Vol. 80, n. 9, 2004, p. 433-450.

[20] Nicol D.M., Sanders W.H., Trivedi K.S.: Model-based evaluation: from dependability to security. In: *Dependable and Secure Computing, IEEE Transactions* on, Vol.1, Iss.1, 2004, p. 48- 65.

[21] Sahner R.A., Trivedi, K.S., Puliafito, A.: *Performance and Reliability Analysis of Computer Systems. An Example-based Approach Using the SHARPE Software Package*, Kluwer Academic Publishers, 1996.

[22] Sanders W.H.: *Integrated Frameworks for Multi-Level and Multi-Formalism Modeling*. In: Proc. of the 8-th Intl. Workshop on Petri Nets and Performance Models, 1999, p. 2.

[23] Shooman M.L.: *Reliability of Computer Systems and Networks: Fault Tolerance, Analysis, and Design*. John Wiley & Sons Inc., 2002.

[24] Vittorini V., Iacono M., Mazzocca N., Franceschinis G.: The OsMoSys approach to multi-formalism modeling of systems. *Software and Systems Modeling*, Vol. 3, Issue 1, 2004, p. 68–81.

FRANCESCO FLAMMINI ha ottenuto la laurea con lode (2003) ed il dottorato di ricerca in Ingegneria Informatica (2006) presso l'Università di Napoli "Federico II". Dal 2003 lavora in Ansaldo STS come progettista e ricercatore, occupandosi di verifica dei sistemi di controllo e protezione delle infrastrutture. Ha tenuto come professore a contratto corsi di informatica ed ingegneria del software e vari seminari sui sistemi sicuri ed affidabili. È autore di oltre 30 articoli scientifici pubblicati su riviste, libri e atti di congressi internazionali. È membro di IEEE CS, EWICS TC7 ed ERCIM FMICS.
E-mail: frflammi@unina.it

NICOLA MAZZOCCA ha ottenuto la laurea con lode (1987) ed il dottorato di ricerca in Ingegneria Elettronica (1992) presso l'Università di Napoli "Federico II". È professore ordinario di Sistemi di Elaborazione presso l'Università di Napoli "Federico II". Ha tenuto oltre 35 corsi universitari ed è autore di oltre 200 pubblicazioni, riguardanti prevalentemente prestazioni ed affidabilità dei calcolatori. È stato Consigliere del Ministro per l'Innovazione e le Riforme nella Pubblica Amministrazione ed è attualmente Assessore all'Università e alla Ricerca Scientifica per la Regione Campania.
E-mail: nicola.mazzocca@unina.it

VALERIA VITTORINI è professore associato presso l'Università di Napoli "Federico II", dove è docente di Fondamenti di Informatica e di Programmazione. È autrice di numerose pubblicazioni scientifiche sull'utilizzo di tecniche formali di modellazione di sistemi reali in diversi ambiti applicativi. Afferisce al Dipartimento di Informatica e Sistemistica, presso cui svolge la propria attività di ricerca, prevalentemente incentrata sullo studio di sistemi distribuiti e sullo sviluppo di metodologie e strumenti per l'analisi di sistemi critici.
E-mail: valeria.vittorini@unina.it

LA SICUREZZA DELLA FIRMA DIGITALE
STATO E PROSPETTIVE

Armando Leotta

La firma digitale si basa su algoritmi, procedure e schemi ritenuti (computazionalmente) sicuri in un contesto molto diverso da quello in cui operano attualmente. Per una vera conoscenza dello stato dell'arte è opportuno comprendere quali siano le vulnerabilità, i contesti operativi in cui valutare le debolezze degli algoritmi utilizzati valutando i potenziali impatti nel medio termine e i rimedi immediati. Tante le soluzioni proposte a livello internazionale che anticipano soluzioni evolutive innovative come nuove funzioni *hash* e algoritmi basati su curve ellittiche.

3.7

1. INTRODUZIONE

I meccanismi di funzionamento della firma digitale, così come prescritta dal nostro ordinamento, poggiano essenzialmente sugli algoritmi crittografici a chiavi pubbliche detti anche a chiavi asimmetriche poiché utilizzano chiavi diverse per le operazioni di cifratura e decifratura.
Le basi di questo nuovo sistema furono poste nel 1976 da due studiosi, Whitfield Diffie e Martin E. Hellman, *New directions in cryptography*, che elaborarono un protocollo per lo scambio di una chiave segreta sopra un canale pubblico.
Esso segna il primo rivoluzionario progresso nella crittografia moderna.
Gli algoritmi asimmetrici (per esempio Diffie-Hellman, RSA, El Gamal) sono utilizzati in generale per le applicazioni in cui sono richieste la confidenzialità, l'autenticazione e la distribuzione delle chiavi.
In particolare la **firma digitale** fornisce:
- Autenticazione;
- Integrità;
- Non-Ripudio.

2. IMPIEGO DELLE CHIAVI CRITTOGRAFICHE

Le chiavi crittografiche possono essere utilizzate in modi diversi a seconda che lo scopo sia quello di cifrare il messaggio oppure apporvi una firma digitale. La differenza tra le due applicazioni risiede nel ruolo delle chiavi. Qualora si intenda spedire a qualcuno un messaggio cifrato, verrà applicata la chiave pubblica del destinatario all'intero messaggio. In tal modo il testo diviene illeggibile e può essere inviato. Il ricevente sarà in grado di leggere il messaggio solo dopo averlo decifrato apponendovi la propria chiave privata. Nel caso il messaggio arrivasse per errore ad un destinatario diverso, questo non sarebbe in grado di leggerlo perché non in possesso della chiave privata corretta cioè la corrispettiva chiave privata relativa alla chiave pubblica utilizzata nella fase di cifratura.
La firma digitale è pertanto una stringa che associa un messaggio a un originatore. L'algoritmo di generazione di firma digitale è un metodo per produrre una firma digitale. Lo schema di firma digitale consiste di un algo-

ritmo di generazione e di un algoritmo di verifica della firma digitale.
Nella figura 1 è illustrato il modello di funzionamento degli algoritmi asimmetrici.
Ciascun utente genera una coppia di chiavi (chiave pubblica, chiave privata). La chiave pubblica è resa pubblica, quella segreta viene custodita gelosamente. Se Bob desidera inviare un messaggio ad Alice, Bob cifra il messaggio utilizzando la chiave pubblica di Alice. Quando Alice riceve il messaggio, lo decifra utilizzando la sua chiave privata. L'autenticazione con gli algoritmi asimmetrici avviene secondo il modello raffigurato nella figura 2.
Quando si vuole apporre una firma digitale ad un messaggio, il mittente utilizza invece la propria chiave privata, che non deve essere necessariamente applicata a tutto il testo.
L'applicazione di un algoritmo di *encryption* produce [1] un *output* dimensionalmente paragonabile all'*input* trattato. Pertanto, nel caso di documenti corposi la cifratura a chiave pubblica può diventare lenta e particolarmente onerosa da un punto di vista computazionale.

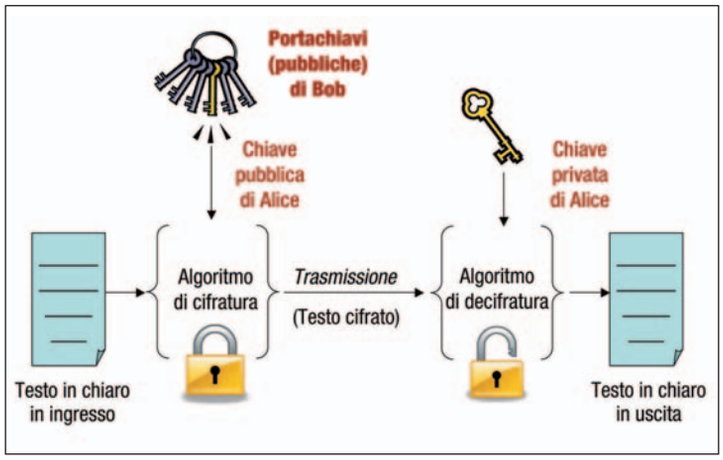


FIGURA 1
Cifratura

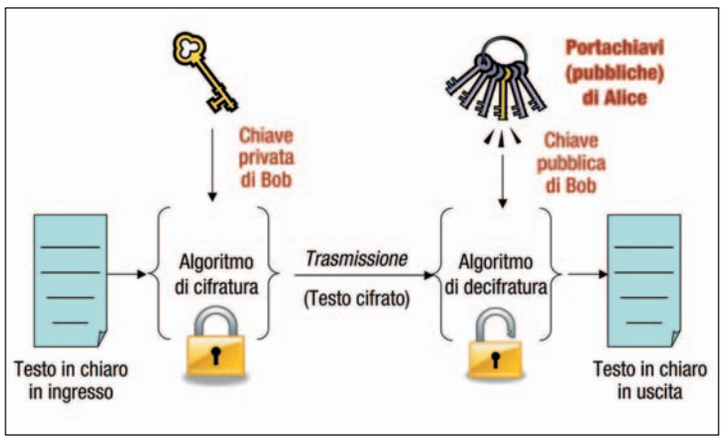


FIGURA 2
Autenticazione/verifica/decifratura

3. RUOLO DELLA FUNZIONE DI *HASH* NEL PROCESSO DI FIRMA DIGITALE

Nel nostro caso l'interesse è unicamente volto all'apposizione di una firma digitale [2] su un documento (e non alla cifratura dello stesso), stiamo pertanto ponendo attenzione agli aspetti di autenticazione e non di riservatezza.
A tal fine viene usata una particolare funzione matematica, chiamata funzione di *hash* che "comprime" il testo in una sorta di "riassunto", che viene anche definito "impronta digitale" (*finger printing* o *message digest*).
Il testo risultante è progettato in modo da minimizzare la probabilità che da testi diversi si possa ottenere il medesimo valore dell'impronta. La dimensione del riassunto è fissa e molto più piccola di quella del messaggio originale, pertanto la generazione della firma risulta molto più rapida.
Quest'ultima consiste quindi semplicemente nella cifratura per mezzo della chiave segreta dell'impronta digitale, generata tramite la funzione di *hash*. La firma è poi aggiunta in una posizione predefinita, normalmente alla fine del testo del documento. Solitamente, insieme con la firma vera e propria, è allegato al documento anche il valore dell'impronta digitale. Nella figura 3 viene schematizzato il processo.
Devono essere inoltre rilevabili attraverso la firma digitale gli elementi identificativi del soggetto titolare della firma, del soggetto che l'ha certificata e del registro sul quale essa è stata pubblicata. La verifica della firma digitale sarà poi fatta dal destinatario applicando la medesima funzione di *hash* usata nella fase di sottoscrizione in modo da ottenere il valore dell'impronta. Tale valore sarà poi confrontato con quello che si ottiene decodificando la firma digitale stessa applicandovi la chiave pubblica del mittente. La disponibilità del valore dell'impronta all'interno del messaggio ha solo la funzione di rendere più veloce la verifica.

Con tale procedura si possono quindi inviare documenti non cifrati e sottoscritti con firma digitale. Questi hanno dunque provenienza certa e colui che li ha inviati non potrà disconoscere di avere inviato il messaggio (non ripudio). È inoltre garantita l'integrità del documento in quanto ogni modifica anche minima apportata al documento firmato in modo digitale è immediatamente riconoscibile in quanto il processo di verifica non si conclude positivamente.
È anche possibile cifrare il contenuto del documento applicando in questo caso la chiave pubblica del destinatario al testo in modo che solo lui potrà decifrarlo utilizzando la corrispondente chiave segreta.

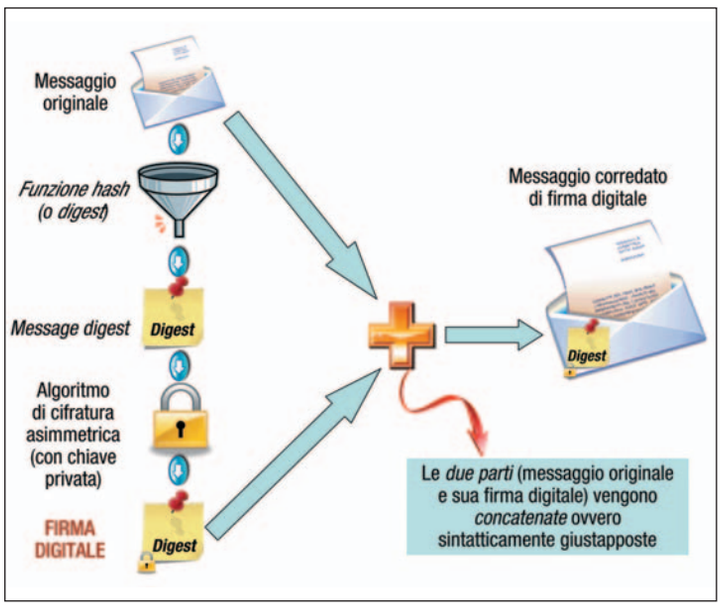


FIGURA 3
Processo di firma

4. HASH

Le funzioni *hash* svolgono un compito fondamentale per l'efficienza degli algoritmi di crittografia e di firma digitale in quanto trasformano una stringa in ingresso di qualsiasi dimensione in un *ouput* di dimensione fissata detto *digest* del messaggio.
Esse sono particolarmente efficienti da un punto di vista computazionale in quanto utilizzano funzioni molto semplici, prevalentemente lo XOR la cui proprietà è quella di "*appiattire*" e "*camuffare*" l'input su cui vengono effettuate le operazioni. Tutte le funzioni *hash* hanno in genere un tempo di esecuzione lineare rispetto all'input, qualsiasi modifica venga fatta al dato in ingresso causerà la produzione di un valore *hash* completamente diverso e sarà sempre un risultato deterministico di lunghezza finita.
Le funzioni *hash* rappresentano una classe di funzioni i cui comportamenti possono essere schematizzate come indicato nella figura 4. La figura rappresenta ad alto livello la struttura di una funzione *hash*.
L'input viene passato ad una funzione di compressione (simbolo a forma di imbuto) che ne riduce la dimensione. Quando la dimensione della computazione parziale è stata ridotta adeguatamente è possibile applicare una trasformazione finale che determinerà l'output finale.
In questo documento, ci concentreremo sulle funzioni *hash* [3] senza chiave.
Tali funzioni possono essere di due tipi:
- funzioni *hash one way* (OWHFs);
- funzioni *hash collision resistant* (CRHFs).

Esse hanno particolari peculiarità matematiche, ovvero sono non invertibili (le OWHF - *pre-image* e 2nd *pre-image resistant*) e sono resistenti alle collisioni (le CRHF). La figura 5 riassume la schematizzazione delle funzioni *hash*.
Affermare che una funzione matematica gode della proprietà di *pre-image resistance*,

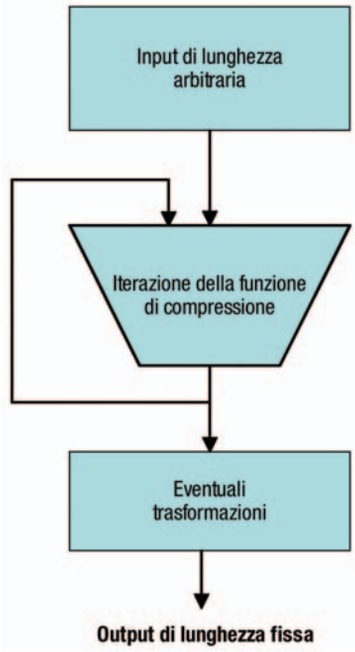


FIGURA 4
Funzione Hash

ovvero la caratteristica di non essere invertibile, significa che dato *y* (= target *hash*) e indicando con *h* la funzione *hash*, è computazionalmente impossibile determinare un valore *x* tale che *h* (*x*) = *y*.
Informalmente, questa caratteristica corrisponde alla difficoltà di trovare un secondo valore input che abbia lo stesso *hash* del primo (non noto).
Un'altra caratteristica importante è la 2nd *pre-image resistant*: dati *x* e *y* = *h* (*x*) e, indicando con *h* la funzione *hash*, è computazionalmente impossibile determinare un valore *x'* tale che *h* (*x'*) = *h* (*x*).
Informalmente, questa caratteristica corrisponde alla difficoltà di trovare un secondo valore input che abbia lo stesso *hash* del primo valore.
Le funzioni *Hash* resistenti alle collisioni (CRHFs) sono particolari funzioni *hash* che oltre alla proprietà di 2nd *pre-image resistance* hanno la proprietà di rendere computazionalmente difficile trovare due input distinti *x* e *x'* i quali generano lo stesso output [*h* (*x*) = *h* (*x'*)].
Essendo la lunghezza del *digest* (output della funzione di *hash*) sensibilmente minore della lunghezza del dato che lo produce (principio della piccionaia), le collisioni sono inevitabili ma le suddette caratteristiche matematiche rendono le collisioni difficili da creare e da gestire e conseguentemente le funzioni *hash* computazionalmente sicure.

5. COLLISIONI

Input diversi possono produrre il medesimo valore di *hash*? Se fossimo in grado di creare due dati in ingresso che generano entrambi lo stesso *digest* (Figura 6) saremmo *potenzialmente* nelle condizioni di cambiare il contenuto di un documento prima che esso venga consegnato alla funzione di *hash* senza che quest'ultima modifichi il *digest* prodotto.
Per mettere a fuoco l'importanza della proprietà *collision resistance* [4] è importante ricordare che nel procedimento di firma digitale quanto viene firmato per motivi di opportunità e di performance non è l'intero messaggio ma il *digest* dello stesso. Tale *digest* viene appunto creato tramite una funzione di *hash*.
Cosa succederebbe se, ipoteticamente, cambiassi un messaggio firmato digitalmente con un altro avente lo stesso *digest*? Nulla, i messaggi pur essendo diversi, avendo lo stesso *digest*, la firma apposta risulterebbe valida e il processo di verifica della stessa sarebbe superato con successo.
Studi matematici e di criptanalisi riportano gli obiettivi progettuali di sicurezza nonché l'onere computazionale necessario a scardinare una determinata proprietà della funzione di *hash* (Tabella 1).
Diremo in generale che un attacco di complessità 2^t richiede approssimativamente 2^t operazioni unitare di riferimento, l'esecuzione di una funzione di compressione, una cifratura ecc..

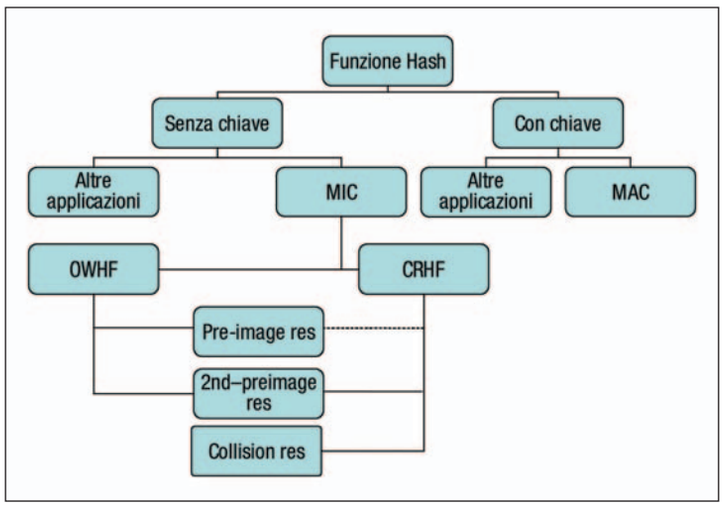


FIGURA 5
Le funzioni hash

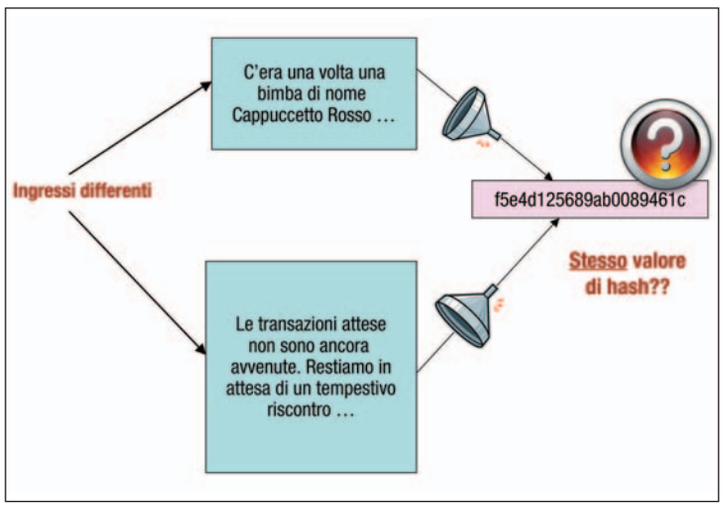


FIGURA 6
Collisioni

Tipo di Funzione Hash	Obiettivo del progettista	Robustezza ideale	Obiettivo dell'avversario
OWHF	Preimage resistance	2^n	Produrre preimage (immagine)
	2^{nd} preimage resistance	2^n	Trovare un secondo input, stessa immagine
CRHF	Collision resistance	$2^{n/2}$	Produrre collisioni

TABELLA 1

Robustezza delle funzioni hash

Per un avversario capace di scegliere i messaggi, un attacco secondo la tecnica del compleanno (*birthday attack*) permette di trovare collisioni in $2^{n/2}$ operazioni.

6. CRITICITÀ E CONTESTO

L'attuale processo di firma digitale prevede l'utilizzo dell'algoritmo di *hash* SHA-1 [5]. Come vedremo, è proprio la debolezza di questo algoritmo a minare la robustezza dell'intero processo di firma digitale.
Esso genera 160 bit di *digest* quindi avremo una collisione al costo di $2^{160/2} = 2^{80}$ operazioni.
Ricordando che, in linea generale, la sicurezza di un algoritmo crittografico è legata al numero di istruzioni necessarie a rompere il sistema, alla capacità computazionale dei sistemi di calcolo, nonché alle risorse disponibili, è determinante porsi un quesito: è ragionevole ritenere ancora sicuro il suddetto limite e quindi computazionalmente irraggiungibile?
Se sì, fino a quando?
Già nel 2005, sul blog di Bruce Schneier [**1, 2**], guru del settore, un team di ricercatori cinesi coordinati dalla Prof.ssa Wang della Shandong University [6, **3, 4**] ha presentato alcuni studi che pubblicavano come riuscire a produrre collisioni nello SHA-1 in 2^{63} operazioni, un fattore 2^{17} inferiore, ben oltre 131000 volte meno oneroso di quanto ritenuto sicuro.
Inoltre, il contesto odierno (*clustering, griding, multicore, parallel processing and distributed computing* ecc.) fa continuamente registrare un aumento sensibile delle performance di calcolo accompagnato da un altrettanto importante abbattimento dei costi sostenuti.
In particolare, nel campo crittografico, occorre segnalare l'impiego [7] delle *Graphical Processing Unit* (GPU), ovvero le unità di elaborazione grafica delle nuove schede grafiche anche nei normali home computer, per elaborazioni crittografiche attraverso il *porting* di opportune applicazioni riscritte con routine grafiche (DirectX, OpenGL).
AMD e Nvidia hanno rilasciato dei tool di sviluppo dedicati (SDK CUDA e AMCL).
Il risultato è un innalzamento consistente delle performance.
Il *Center For Visual Computing and Department of Computer Science* dell'Università Stony Brook di NewYork ha pubblicato uno studio nel quale emergono i benefici di utilizzare un cluster di moderne GPU per impegnativi calcoli scientifici paralleli [8].
I risultati, decisamente interessanti per performance e rapporto GFlops/costo, hanno alimentato ulteriori interessanti studi per accelerare calcoli computazionali di natura non necessariamente grafica.
Poche settimane scorse, la Elcomsoft [**5, 6**], società di software russa con sede a Mosca, ha registrato un brevetto negli Stati Uniti che protegge una tecnica di *password cracking* delle password di accesso ai computer utilizzando le caratteristiche di *parallel processing* delle più recenti GPU in commercio.
Tale tecnica ha fatto registrare un abbattimento di circa 25 volte le risorse temporali necessarie ad una moderna CPU.
Alla luce di questo contesto tecnologico, siamo ancora in piena sicurezza computazionale?
E ancora: se sì, fino a quando?
È possibile isolare la criticità (la funzione di *hash* SHA-1) dal resto del processo di firma digitale? Sarebbe sufficiente o andrebbe affrontato un nuovo schema di firma?

7. INIZIATIVE INTERNAZIONALI

Il *National Institute of Science and Technology* (NIST) americano ha già identificato degli standard per funzioni *hash* di lunghezza e resistenza maggiore: la cosiddetta famiglia SHA-2 (SHA-224, SHA-256, SHA-384 e SHA-512).
L'adozione di un algoritmo della famiglia SHA-2 rappresenta una rapida ed efficace soluzione nel medio e breve termine fino a che i risultati della *call for algorithm* [**7**] non porti i frutti

Utilizzo accettabile fino alla fine del 2009	Uso con certificati qualificati[1]: utilizzo accettabile fino alla fine del 2010	Utilizzo accettabile fino alla fine del 2010	Utilizzo accettabile fino alla fine del 2011
SHA-1	SHA-1	RIPEMD-160	SHA-224, SHA-256, SHA-384, SHA-512

[1] Esclusivamente per ila generazione e la verifica di certificati ma non per la generazione e verifica di altri dati qualificati firmati.

TABELLA 2
Funzioni hash e tempi di "sicurezza"

Fino alla fine del 2007	Fino alla fine del 2008	Fino alla fine del 2009	Fino alla fine del 2010	Fino alla fine del 2011
1024 (minimo)	1280 (minimo)	1563 (minimo)	1728 (minimo)	1976 (minimo)
2048 (raccomandato)	**2048** (raccomandato)	**2048** (raccomandato)	**2048** (raccomandato)	**2048** (raccomandato)

TABELLA 3
Criticità sulla lunghezza chiave RSA

attesi e cioè un nuovo standard condiviso da tutta la comunità internazionale.
Nel marzo del 2006, il governo tedesco ha pubblicato nella gazzetta ufficiale [**8**] quelli che considera tempi di sicurezza per gli algoritmi e le funzioni *hash* utilizzati nel procedimento di firma digitale (Tabella 2).
Oltre alla criticità circa l'utilizzo dello SHA-1 dopo il 2009, viene evidenziata (Tabella 3) la seria preoccupazione del governo tedesco per la lunghezza della chiave RSA [9].
Analoghe preoccupazioni sono sorte in seno al governo austriaco ed in Europa (progetto Nessie [10], [**9**, **10**]).
La *National Security Agency* (NSA) americana presenta la sua raccomandazione, Suite B [**11**] una lista di algoritmi crittografici ritenuti sicuri.
Ad analoghe conclusioni giungono in Europa (progetto NESSIE) e in Giappone (CRYPTREC *Evaluation Committee* [**12**, **13**, 11]).

8. SOLUZIONI IMMEDIATE E TRANSITORIE

"*Aumentare la lunghezza della chiave (per gli algoritmi crittografici, ad esempio RSA) e la dimensione del digest prodotto dalla funzione di hash*".
Queste misure, pur essendo concettualmente e matematicamente valide vanno valutate attentamente in quanto introducono un problema non indifferente di performance spostando solamente di qualche anno il problema.
Per le funzioni *hash* il problema è meno invasivo e l'adozione della famiglia SHA-2 [12, **14**] da 224 a 512 bit risulta un compromesso accettabile ed immediato.
Le medesime considerazioni valgono per l'algoritmo di *hash* WHIRLPOOL [**15**] a 512 bit.
Entrambi gli algoritmi indicati rientrano tra quelli considerati sicuri e adottabili nei diversi studi indipendenti precedentemente citati.

9. NUOVI SCENARI CRITTOGRAFICI

Gli algoritmi basati su curve ellittiche (ECC) [13, 14, 15, 16, 17, **16**] sono legati matematicamente al problema del logaritmo discreto (riquadro a p. 28) e rappresenta un approccio condiviso, robusto, efficace, efficiente e computazionalmente scalabile.
Di seguito, la tabella 4 (fonte NSA [**17**]) che mostra come a "sicurezza equivalente" l'adozione degli algoritmi basati su curve ellittiche riducano sensibilmente la dimensione della chiave rispetto ad algoritmi come RSA e Diffie-Hellman.

10. CONCLUSIONI

È importante chiarire che non vi è una reale emergenza in atto.
Quello che occorre però evitare è ignorare i vari segnali d'allarme e l'elevato livello d'attenzione oramai presente in tutto il settore. Come detto, esistono diverse soluzioni e rimedi temporanei immediati e fortemente consigliati.

Il problema del logaritmo discreto

Il problema del logaritmo discreto è, insieme ad altri, come per esempio quello della fattorizzazione di interi, un problema matematico definito in termini di aritmetica modulare, estremamente importante nell'ambiente crittografico.
Fissato un numero primo p e un intero g tra 0 e $p-1$ si ottiene y dalla relazione:

$y = g^x \pmod p$

Il problema del logaritmo discreto è di determinare l'intero x dati g, y e p.
Per questo problema non è stato trovato nessun algoritmo efficiente in grado di risolverlo.
Taher ElGamal è stato il primo a proporre un sistema crittografico basato su questo problema. In particolare, propose due distinti sistemi: uno schema crittografico e uno schema di firme digitali (DSA [**23**] si basa sul lavoro di ElGamal).
Rompere uno di questi schemi significa risolvere il problema del logaritmo discreto.
Per avere una buona sicurezza il primo p deve essere lungo almeno 230 cifre decimali (760 bit).
Le prestazioni del sistema dipendono dalla velocità di esecuzione dell'esponenziazione modulare.
Il calcolo dominante in ogni trasformazione è:

$g^x \pmod p$

con g compreso tra 0 e $p-1$.
In conclusione la sicurezza di questi sistemi, si basa sul problema del logaritmo discreto modulo p, mentre l'efficienza dipende dalla velocità di esecuzione dell'esponenziazione modulare.

Lunghezza della chiave simmetrica (bit)	Lunghezza della chiave RSA e Diffie-Hellman (bit)	Lunghezza della chiave a curva ellittica (bit)
80	1024	160
112	2048	224
128	3072	256
192	7680	384
256	15360	512

TABELLA 4

Confronto a "sicurezza equivalente"

Sebbene coinvolga anche algoritmi non espressamente affrontati in questo contributo per ovvie ragioni di sintesi e contesto, risulta di estremo interesse la tabella 5, un quadro sinottico che mette a confronto le raccomandazioni frutto di studi indipendenti da parte degli studiosi di tutto il mondo.
Lo SHA-2 in prima istanza per le funzioni di *hash* e l'ECDSA [18] [**18**, **19**, **20**, **21**] (*Elliptic Curve Digital Signature Algorithm*), come schema di firma digitale, rappresentano soluzioni consolidate, sicure, robuste e innovative.
Si pensi al mercato della connettività mobile, dei produttori di chip: adottare ECC significa puntare sul moderno, sullo scalabile, sulle prestazioni con dimensioni decisamente più contenute, dunque risorse hardware inferiori ma con prestazioni superiori.
Nonostante i numerosi vantaggi delle curve ellittiche e la crescente diffusione tra gli utenti, molti nel mondo accademico e industriale ritengono un ostacolo al loro utilizzo alcuni fattori legati alla proprietà intellettuale che sottendono tali curve.
In effetti, vari aspetti della crittografia a curve ellittiche sono stati brevettati sia da individui che da società.
La Certicom Inc. per esempio, è una società canadese che detiene oltre 130 brevetti relativi alle curve ellittiche ed alla crittografia a chiave pubblica in generale.
La NSA ha stabilito un accordo [**22**] con la Certicom circa la proprietà intellettuale e l'utilizzo degli algoritmi brevettati.
Come si evince anche dal suddetto accordo (*agreement*), Certicom ha identificato 26 brevetti che riguardano l'utilizzo. La licenza accordata alla NSA include il diritto di sub-licenza di questi 26 brevetti a produttori incaricati di costruire dei dispositivi nel pieno rispetto della licenza e dei limiti di utilizzo accordati.
Certicom, da parte sua, ha mantenuto il diritto medesimo sia all'interno dell'ambito d'utilizzo che diversamente in base ad una negoziazione diretta con i produttori (*vendor*).
Questi aspetti e la diversa normativa in materia di brevetti su algoritmi (in molte nazioni non è consentito brevettare algoritmi) rappresentano fattori sicuramente non bloccanti ma che non facilitano l'introduzione massiva sul mercato.
Tutti gli esperti del settore sono concordi nel considerare il 2010 una *deadline* per imprimere una svolta ed un cambiamento strutturale nel mondo della crittografia asimmetrica.
È opportuno pertanto considerare seriamente, anche da un punto di vista di mercato produttivo e governativo, l'alternativa offerta dalle curve ellittiche per i vantaggi computazionali, di prestazione e di traffico offerti a parità di sicurezza e robustezza.
Inoltre, essendo le lunghezze delle chiavi sensibilmente minori rispetto alla prima generazione, le soluzioni ECC risultano essere anche più facilmente scalabili nel prossimo futuro.
In Italia, le nuove Regole Tecniche della Firma Digitale sono ad oggi in itinere.
Già nel novembre scorso, le corrispettive del-

		NIST FIPS e SP (considerando un uso fino alla fine del 2030)	Lista di cifrari raccomandati dal CRYPTREC per l'e-government	Cifrari raccomandati (NESSIE)	SO/IEC 18033
Cifrari asimmetrici	Firma	❑ RSA come specificato in ANS X9.31 (2048) ❑ DSA (2048) ❑ ECDSA (224)	❑ DSA (1024) ❑ ECDSA (160) ❑ RSASSA-PKCS1-v1_5 (1024) ❑ RSA-PSS (1024)	❑ RSA-PSS (1536) ❑ ECDSA (160) ❑ SFLASH	(Specificato nell'ISO/IEC 9796-2 e 14888-3)
	Confidenzialità	Nessuna racc.	❑ RSA-OAEP (1024) ❑ RSAES-PKCS1-v1_5 (1024)	❑ PSEC-KEM (160) ❑ RSA-KEM (1536) ❑ ACE-KEM	❑ RSA-KEM ❑ RSA_OAEP ❑ HIME(R) ❑ ACE-KEM ❑ PSEC-KEM ❑ ECIES-KEM
	Scambio chiavi	❑ DH ❑ MKV	❑ DH (1024) ❑ ECDH (160) ❑ PSEC-KEM (160)	Nessuna racc.	(Specificato nell'ISO/IEC 11770-3)
Cifrari simmetrici	Cifrari a blocchi da 64 bit	Triple-Des (a tre chiavi)	❑ CIPHERUNICORN-E ❑ HIEROCRYPT-L1 ❑ MISTY1 ❑ TRIPLE-DES (a tre chiavi)	MISTY1	❑ CAST-128 ❑ MISTY1 ❑ TRIPLE-DES (racc. a tre chiavi)
	Cifrari a blocchi da 128 bit	AES	❑ AES ❑ CAMELLIA ❑ CIPHERUNICORN-A ❑ HIEROCRYPT-3 ❑ SC2000	❑ AES ❑ CAMELLIA	❑ AES ❑ CAMELLIA ❑ SEED
	Cifrari a flusso	Nessuna racc.	❑ MUGI ❑ MULTI-S01 ❑ RC4 (128 bit)	Nessuna racc.	❑ MUGI ❑ SNOW 2.0
Funzioni HASH		❑ SHA-224 ❑ SHA-256 ❑ SHA-384 ❑ SHA-512	❑ RIPEMD-160 ❑ SHA-1 ❑ SHA-256 ❑ SHA-384 ❑ SHA-512	❑ SHA-256 ❑ SHA-384 ❑ SHA-512 ❑ WHIRLPOOL	(Specificato nell'ISO/IEC 10118)

Nota: Le lunghezze delle chiavi in tabella sono espresse in bit ed indicano i valori minimi raccomandati

TABELLA 5

Sintesi e raccomandazioni

la Carta d'Identità Elettronica DM 8/11/2007 nell'allegato B formalizzava l'abbandono dell'RSA a 1024 bit (minimo 2048 o 3072 bit) e dello SHA-1 in favore dello SHA-2.

Inoltre, per la prima volta, viene effettuata la formale apertura verso l'ECDSA con curve raccomandate da 224 e 283 bit.

È stato appena pubblicato sulla *Gazzetta Uf-*

ficiale n. 129 del 6 giugno il DPCM del 30 marzo "Regole tecniche in materia di generazione, apposizione e verifica delle firme digitali e validazione temporale dei documenti informatici". Non appena entrerà in vigore (180 giorni dalla pubblicazione) abrogherà il DPCM del 13 gennaio 2004 recante le "regole tecniche per la formazione, la trasmissione, la conservazione, la duplicazione, la riproduzione e la validazione, anche temporale, dei documenti informatici", pubblicato sulla *Gazzetta Ufficiale* 27 aprile 2004, n. 98.
Estremamente importante l'art. 2 comma 3 che conferma al CNIPA il pieno controllo delle regole tecniche:

"*Gli algoritmi di generazione e verifica delle firme, le caratteristiche delle chiavi utilizzate, le funzioni di hash, i formati e le caratteristiche dei certificati qualificati, le caratteristiche delle firme digitali e delle marche temporali, il formato dell'elenco di cui all'articolo 39 del presente decreto, sono definiti, anche ai fini del riconoscimento e della verifica del documento informatico, con deliberazioni del CNIPA e pubblicati sul sito internet dello stesso Centro nazionale*".

Sarà suo compito pertanto armonizzare le eventuali modifiche del CAD (*Codice dell'Amministrazione Digitale*) e proiettarci in Europa dove la sicurezza (per esempio, il formale impiego di algoritmi basati su curve ellittiche e l'impiego della famiglia SHA-2 per le funzioni di *hash*) e l'interoperabilità dei formati (per esempio, l'uso della sottoscrizione in linguaggio XML, i formati della busta crittografica e di firma come XML e PDF e altri conformi a specifiche pubbliche *Publicly Available Specification* -PAS-) rappresentano aspetti cruciali dell'*e-government* internazionale dei prossimi anni.

Bibliografia

[1] Handbook of Applied Cryptography, Menezes, Ooschot, Vanstone.

[2] Manca G.: Il futuro della firma digitale. *Mondo Digitale*, 2002,

[3] Stinson D.: *Cryptography: theory and practice*. Chapman & Hall.

[4] *Introduction to computer security*. Bishop 2005

[5] *Secure Hash Standard*. SHA-1, FIPS 180-1, NIST.

[6] Wang X.: *Cryptanalysis for Hash Functions and Some Potential Dangers*. RSA Conference, 2006, Cryptographer's Track, 2006 - Xiaoyun Wang, Yiqun Yin, Hongbo Yu, Finding Collisions in the Full SHA-1, Crypto'05 - Collision Search Attacks on SHA1, 2005.

[7] Debra L. Cook, John Ioannidis, Angelos D. Keromytis, Jake Luck: *CryptoGraphics: Secret Key Cryptography Using Graphics Cards*.

[8] *GPU Cluster for High Performance Computing*. ACM/IEEE Supercomputing Conference 2004, November 06-12, Pittsburgh, PA.

[9] Rivest R., Shamir A., Adleman L.: A method for obtaining digital signatures and public key cryptosystems. *Communications of the ACM*, Vol. 21, n. 2, 1978.

[10] *NESSIE è un progetto della Information Society Technologies (IST) Programme della Commissione Europea*. (Key Action II, Action Line II.4.1).

[11] *Year 2010 Issues on Cryptographic Algorithms*. Masashi Une and Masayuki Kanda, March 2007.

[12] NIST: *FIPS Publication 180-2: Secure Hash Standard (SHS-1)*. August 2002.

[13] Miller V.: *Uses of Elliptic Curves in Cryptography*. Advances in Cryptology, CRYPTO '85, Proceedings, Lecture Notes in Computer Science 218, Springer-Verlag, 1986, p. 417-426.

[14] Koblitz N.: Elliptic Curve Cryptography. *Mathematics of Computation*, Vol, 48, 1987, p. 203-209.

[15] Menezes A., Vanstone S.A.: Elliptic curve cryptosystems and their implementation. *Journal of Cryptology*, Vol. 6, 1993, p. 209-224.

[16] Gathani Amit N.: *Implementation of Elliptic Curve Cryptography in Embedded System*. 2001.

[17] *Advances in Elliptic Curve Cryptography, Series*. London Mathematical Society Lecture Note Series (No. 317), Edited by Ian F. Blake, University of Toronto, Gadiel Seroussi, Hewlett-Packard Laboratories, Palo Alto, California, Nigel P. Smart, University of Bristol.

[18] ISO 14888 ANSI X9.62 Public Key Cryptography For The Financial Services Industry: The Elliptic Curve Digital Signature Algorithm (ECDSA), X9.63 Public Key Cryptography for the Financial Services Industry: Key Agreement and Key Transport Using Elliptic Curve Cryptography.

Webgrafia

[1] http://www.schneier.com/essay-074.html

[2] http://www.schneier.com/blog/archives/2005/02/sha1_broken.html

[3] http://news.zdnet.com/2100-1009_22-5598536.html

[4] http://www.infosec.sdu.edu.cn/people/wangxiaoyun.htm

[5] http://www.elcomsoft.com

[6] http://www.elcomsoft.com/EDPR/gpu_en.pdf GPGPU

[7] http://www.nist.gov/public_affairs/techbeat/tb2007_1108.htm#sha

[8] http://www.bundesnetzagentur.de/media/archive/5952.pdf

[9] https://www.cosic.esat.kuleuven.be/nessie/

[10] https://www.cosic.esat.kuleuven.be/nessie/deliverables/decision-final.pdf

[11] http://www.nsa.gov/ia/industry/crypto_suite_b.cfm

[12] http://www.ipa.go.jp/security/enc/CRYPTREC/index-e.html CRYPTREC Giappone

[13] http://www.imes.boj.or.jp/english/publication/mes/2007/me25-1-6.pdf

[14] http://csrc.nist.gov/publications/fips/fips180-2/fips180-2.pdf

[15] http://paginas.terra.com.br/informatica/paulobarreto/WhirlpoolPage.html

[16] The Basics of ECC. http://www.certicom.com

[17] http://www.nsa.gov/ia/industry/crypto_elliptic_curve.cfm

[18] http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=1439

[19] X9 committee http://www.x9.org

[20] CERTICOM Patent letter : http://publicaa.ansi.org/-sites/apdl/Patent%20Letters/PL337.pdf

[21] IEEE P1363 2 http://www.ieee.org/

[22] NSA & CERTICOM Agreement http://www.certicom.com/download/aid-501/FAQThe%20NSA%20ECC%20License%20Agreement.pdf

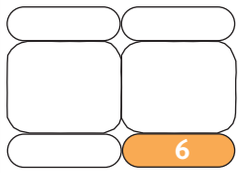
[23] http://www.itl.nist.gov/fipspubs/fip186.htm Digital Signature Standard.

ARMANDO LEOTTA è laureato in Scienze dell'Informazione presso La Sapienza di Roma. Nel 2008 nello stesso Ateneo e Dipartimento ha conseguito il Master di II livello in Gestione della Sicurezza informatica per l'impresa e la Pubblica Amministrazione, concluso con un tirocinio trimestrale presso il CNIPA. Certificato IRCA-RICEC Auditor / Lead Auditor per Sistemi di Gestione per la Sicurezza delle Informazioni (SGSI) a norma ISO/IEC 27001:2005 & BS 7799-2:2002, Certificato OMG UML Professional e OCEB (OMG Certified Expert in BPM), APM Group ITIL v3 e QRP Prince2, dal 2003 si occupa di Gestione della sicurezza e Management dei processi afferenti la sicurezza informativa nella Ragioneria Generale dello Stato. Nel 2009 conclude con AICA un percorso formativo che lo conduce alla certificazione EUCIP Professional/Elective IS Auditor.
E-mail: armando.leotta@gmail.com, info@armandoleotta.it

COMPUTER ETHICS
UN QUADRO CONCETTUALE

Norberto Patrignani

Fin dalla sua nascita, la *Computer Ethics* è stata una materia con diverse direzioni di ricerca: quelle che considerano l'informatica come un campo da studiare con gli strumenti classici della filosofia, quelle che ritengono necessaria una nuova disciplina dedicata ai dilemmi etici specifici legati ai computer, altre che attribuiscono la sua stessa fondazione al padre della cibernetica, Wiener. Questo articolo, dopo una breve introduzione storica, descrive la materia con un approccio da etica applicata per arrivare ad una definizione di una metodologia utile per i corsi universitari di *Computer Ethics*.



1. INTRODUZIONE

Forse il primo scienziato a riflettere sugli aspetti etici e sociali legati ai computer, ad individuare le connessioni tra Etica e *Information Technology* fu Norbert Wiener, professore al MIT negli anni '50 e padre della cibernetica. Egli infatti invitò a riflettere sulle rivoluzionarie conseguenze sociali che questa nuova tecnologia avrebbe prodotto. Nel 1948, nel suo libro sulla cibernetica scrisse: "... mi resi subito conto che eravamo in presenza di una tecnologia che, nel bene o nel male, avrebbe potuto avere potenzialità inaudite da un punto di vista sociale" [40]. Nel suo libro successivo *The Human Use of Human Beings: Cybernetics and Society*, dedicato esclusivamente a riflessioni sulla tecnologia, troviamo già i concetti fondamentali della *Computer Ethics* [41]. Secondo Wiener la diffusione dei computer nella società avrebbe portato a cambiamenti con conseguenze tali da toccare ogni aspetto della nostra vita.

Il primo a raccogliere l'invito a riflettere sulle conseguenze etico-sociali della tecnologia dell'informazione fu, nel 1968, Donn Parker, uno dei più noti ricercatori dello *Stanford Research Institute* (SRI) di Menlo Park. In uno degli articoli che segna un'altra tappa fondamentale nella storia della *Computer Ethics, "Rules of Ethics in Information Processing"*, scrisse: "... sembra che quando le persone entrano nella sala computer lascino l'etica alla porta" [29]. Parker, raccolse e analizzò molti casi di attività legati ai computer e professionalmente poco etici e contribuì in modo sostanziale all'elaborazione del primo Codice Etico di Condotta Professionale della ACM (*Association for Computing Machinery*) la più antica associazione professionale di informatica. Dal 1973 questo codice venne adottato dalla ACM e, da allora, costituisce un punto di riferimento sostanziale.

Sempre negli anni '70 troviamo un altro professore del MIT, Joseph Weizenbaum, che lanciò un forte appello alla comunità scientifica per mettere in guardia dai rischi connessi con le applicazioni militati delle tecnologie dell'informazione [39].

La prima volta che il termine "*Computer Ethics*" venne usato, fu nel 1978, quando Walter Maner,

della Old Dominion University in Virginia, diede la prima definizione: "... campo di indagine che affronta i problemi etici creati, trasformati o aggravati dall'*Information Technology*" [24]. I suoi contributi attivarono il dibattito in molte università statunitensi a seguito del quale furono proposti agli studenti i primi seminari sul tema.
Nel frattempo, i computer di diffondevano nella società e, con la rivoluzione del *personal computing*, con l'apparizione dei primi virus informatici nella nascente rete Internet, con le prime notizie riguardanti invasioni della privacy e disastri causati da malfunzionamenti dei computer, aumentò la consapevolezza delle conseguenze etico-sociali dell'*Information Technology*.
Arriviamo così al 1985, considerato l'anno di fondazione della disciplina *Computer Ethics*. Infatti in quell'anno James Moor, del Dartmouth College, pubblicò il suo famoso articolo *What is Computer Ethics?* [26] e Deborah Johnson, del *Rensselaer Polytechnic Institute*, il suo libro fondamentale *Computer Ethics* [15].
Nel 1983 nasceva a Palo Alto la *Computer Professionals for Social Responsibility* (CPSR), la prima organizzazione internazionale di docenti universitari, ricercatori e *computer professional* con la missione di portare all'attenzione pubblica le scelte critiche riguardanti le applicazioni dell'informatica e le loro conseguenze sulla società, sfatare i miti sull'infallibilità dei sistemi tecnologici e incoraggiare l'uso della tecnologia dell'informazione per migliorare la qualità della vita.
Tutte le energie della CPSR vennero dedicate a contrastare il progetto SDI (*Strategic Defense Iniziative*), annunciato nel 1983 dal Presidente Reagan e più noto come *Star Wars*, dove decisioni molto critiche, come il lancio di missili intercontinentali, rischiavano di essere delegate ai computer.
Le ragioni della CPSR vennero esposte nel famoso libro "*Computers in Battle: Will They Work?*" che rappresenta ancora oggi una pietra miliare nella denuncia dei rischi legati all'utilizzo dei computer in guerra [2].
La prima conferenza internazionale multidisciplinare sulla *Computer Ethics*, con la partecipazione di docenti di Informatica, *computer professional*, filosofi, sociologi, psicologi, giuristi, economisti, giornalisti e rappresentanti della pubblica amministrazione si tenne nel 1991 a New Haven, negli USA. Nello stesso anno la materia *Computer Ethics* venne ufficialmente introdotta nei piani di studi di Informatica (*Computer Science*) delle università statunitensi [6].
Negli anni '90 emerse l'esigenza di avere, da questa disciplina, delle indicazioni anche pratiche, un'*etica applicata* che aiutasse a ridurre la probabilità del verificarsi di effetti imprevisti ed indesiderati derivanti dall'applicazione dell'*Information Technology* e a predisporre delle linee guida, dei codici etici per i progettisti di sistemi complessi che escono dai laboratori di ricerca per diffondersi nella società, offrendo servizi ed applicazioni al pubblico più vasto.
Nasceva così in Europa il *Centre for Computing and Social Responsibility* (CCSR, De Montfort University, UK). Il CCSR, diretto dai professori Simon Rogerson, Terry Ward Bynum e Donald Gotterbarn, oggi è un punto di riferimento mondiale per la ricerca e la formazione sui rischi ed opportunità dell'IT.
Il CCSR è anche l'organizzatore della conferenza internazionale sulla *Computer Ethics* che si tiene regolarmente ogni due anni (l'ultima, ETHICOMP 2008, si è svolta a Mantova, sotto la direzione di Mariacarla Calzarossa e Ivo De Lotto dell'Università di Pavia).
Oggi la *Computer Ethics* sta rapidamente evolvendo verso quella che potremmo definire *Global Information Ethics*: per la prima volta nella storia dell'umanità siamo interconnessi su scala planetaria. Internet ed il World Wide Web ci pongono di fronte ai problemi etici dell'informatica su scala globale: pensiamo per esempio al problema della legislazione nel *cyberspazio* (dove sono le frontiere?) oppure al cyberbusiness (la gestione di transazioni commerciali su scala planetaria), agli effetti sulle culture e tradizioni di una *Global Education*, per non parlare del *Digital Divide* tra Nord e Sud del pianeta.
Dal 1991 i corsi di *Computer Ethics* si sono diffusi in tutto il mondo nelle varie facoltà di *Computer Science*, diventando ormai una parte organica della cultura di base dei professionisti dell'informatica, per citare alcuni corsi: *Ethics and the Law on the Electronic Frontier* al MIT, *Computer and Information Ethics* a Stanford; in Europa, oltre al già citato CCSR in UK, abbiamo *Ethics of Computing* del Prof. Jacques Berleur

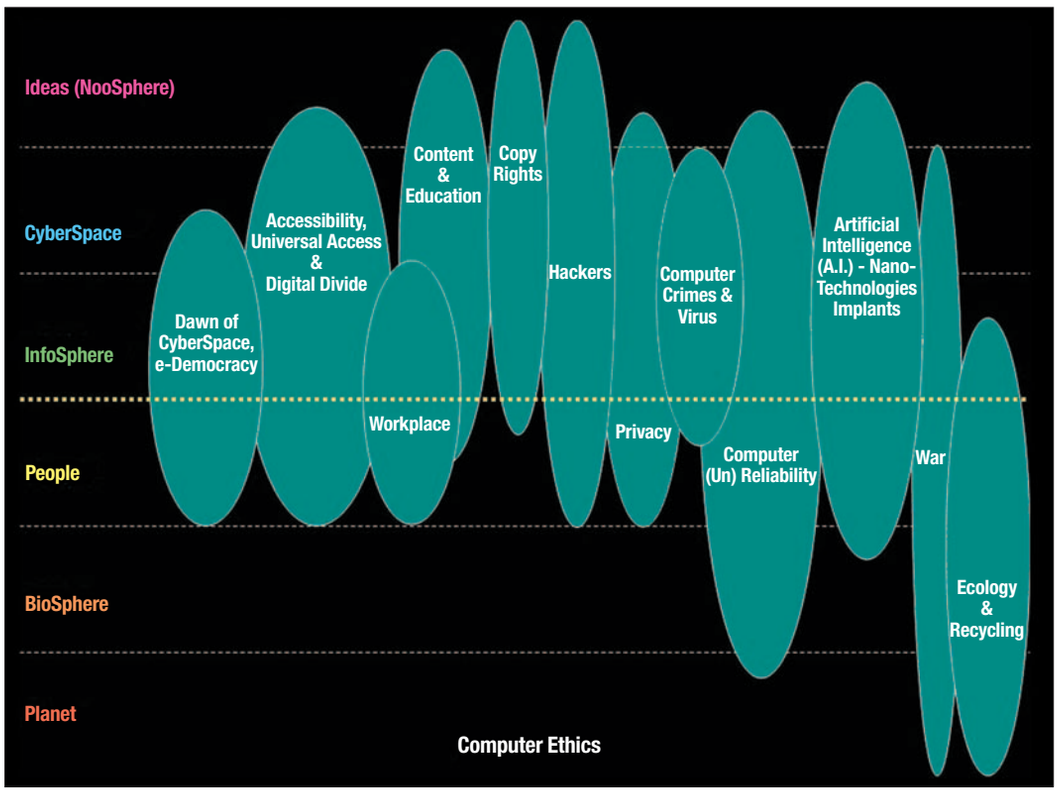


FIGURA 1
Un quadro concettuale per la Computer Ethics

(uno dei pionieri della materia in Europa) all'Università di Namur in Belgio, *Ethics of Information* di Luciano Floridi all'Università di Oxford. In Italia, il primo corso di *Computer Ethics* è stato introdotto dal Politecnico di Torino nel 2008, seguito dall'Università di Trento con *Ethics of Computing*. Al Politecnico di Milano abbiamo "Deontologia ed etica delle tecnologie dell'informazione" di Piercarlo Maggiolini.
Questa materia è, per definizione, interdisciplinare (all'incrocio tra informatica, filosofia, giurisprudenza ecc.) e impone quindi un nuovo approccio che superi le tradizionali barriere tra tecnologi ed umanisti. Un esempio di eccellenza è lo storico dipartimento di *Science, Technology and Society* (STS) dell'Università della Virginia, diretto da Deborah Johnson.
Cosa si intende per etica applicata all'*Information Technology*? Quali sono i principali argomenti e *critical issues* affrontati da questa nuova disciplina? La figura 1 offre una panoramica delle principali questioni emerse negli ultimi decenni. Esse pongono i responsabili dei sistemi informativi, i ricercatori, i *computer professional* e gli utenti finali, di fronte a veri e propri dilemmi etici e mettono in evidenza i punti di maggior vulnerabilità per la società dovuti sia all'uso improprio o scorretto dei computer che al loro malfunzionamento.

2. UN QUADRO CONCETTUALE PER LA COMPUTER ETHICS

Questo quadro (Figura1) può essere semplicemente descritto nelle sue due dimensioni: la dimensione verticale (i "livelli") e la dimensione orizzontale (i "domini"). Ovviamente, come tutti i modelli, esso rappresenta una semplificazione della complessità del mondo reale, una semplice mappa qualitativa.

2.1. I livelli verticali

I livelli verticali rappresentano le diverse aree della realtà potenzialmente influenzate dai computer: dal mondo fisico reale (il pianeta, la biosfera, le persone) al mondo virtuale (l'infosfera, il ciberspazio, le idee). Vediamo una breve descrizione di ciascuno di questi "livelli".

2.1.1. Il pianeta (*Planet*)

Questo è il pianeta Terra, l'unico che abbiamo; nel modello proposto possiamo vedere che alcune aree hanno un impatto su questo livello più basso, come *Computer (Un)Reliability*, *War*

e, ovviamente, *Ecology & Recycling*. Quando analizziamo eticamente le azioni umane e le conseguenze delle tecnologie sviluppate dagli umani, dobbiamo includere anche gli ecosistemi e l'impatto a lungo termine nei confronti delle generazioni future [16].

2.1.2. La biosfera (*Biosphere*)

Questo livello rappresenta l'intero mondo biologico vivente sulla Terra. Tutte le aree che hanno un impatto sul livello *Planet* hanno un impatto anche su questo livello immediatamente superiore; d'altra parte vi è un'area crescente nota come *Artificial Intelligence, Nanotechnologies, Implants* che cambierà il concetto di confine tra entità viventi / non-viventi (pensiamo ai *nano-bots*: macchine autonome a scala nanometrica in grado di muoversi all'interno del nostro corpo per applicazioni mediche come il trasporto di farmaci verso organi-bersaglio; oppure pensiamo all'impianto di dispositivi miniaturizzati all'interno del nostro corpo in grado di interfacciare direttamente il sistema nervoso con l'esterno ecc.).

2.1.3. Le persone (*People*)

Questo livello rappresenta noi stessi, gli esseri umani; noi siamo toccati da tutti i "domini", le aree rappresentate con ellissi nella figura che riassume la *Computer Ethics*. La sfida per noi consiste nel ripensare l'intera disciplina dell'etica includendo nelle nostre analisi anche le speciali "macchine a stati finiti": i computer.

2.1.4. L'infosfera (*Infosphere*)

In questo contesto, con infosfera intendiamo il livello rappresentato dal sottile strato artificiale, costituito dalle tecnologie dell'informazione, che avvolge il pianeta. Esso rappresenta Internet, la Rete: l'insieme di hardware, software, computer, reti, router ecc. che costituisce il prezioso guscio attorno al pianeta che collega tutti noi, qualsiasi cosa, da qualunque punto del pianeta. Con le sue caratteristiche originali quali: standard aperti, architettura *peer-to-peer*, assenza di un punto centrale di controllo (e di guasto), neutralità (solo il mittente e destinatario conoscono i contenuti dei pacchetti di bit, la Rete si limita a smistare con efficienza i dati guardando solo agli indirizzi dei pacchetti), l'infosfera sta diventando talmente importante che, come il pianeta Terra, dovremmo preservarla per le generazioni future.

2.1.5. Il ciberspazio (*Cyberspace*)

Questo nuovo livello, ancora in costruzione, è uno spazio virtuale al di sopra dell'infosfera, dove iniziamo a spendere gran parte della nostra vita [11]. Sono ormai innumerevoli gli esempi di applicazioni in questo spazio dove il confine tra mondo reale e virtuale si sta dissolvendo: social networks con avatar, realtà virtuali, videogame immersivi tridimensionali, giochi online multi-utente ecc..

2.1.6. Le idee "Ideas" (*Noosphere*)

Questo è lo strato al massimo livello di astrazione, probabilmente il più delicato. Esso rappresenta la risorsa più preziosa che abbiamo: la raccolta delle idee che fluiscono nelle menti umane, il pensiero umano nel suo insieme, l'eredità di conoscenza e saggezza dal passato ed il potenziale innovativo verso il futuro, la cosiddetta Noosfera [8]. Il libero flusso delle informazioni e l'approccio alla tecnologia tipico dell'etica hacker possono contribuire fortemente ad alimentare questo spazio (il *public-domain*) che è l'humus di una società davvero innovativa, come la società *knowledge-based* del futuro prossimo. Ecco perché il livello delle idee è toccato dagli ellissi dei domini *Copyrights e Hackers*.

2.2. I domini orizzontali

I domini orizzontali, rappresentati dalle ellissi della figura 1 (con le dimensioni approssimativamente proporzionali alla loro importanza o impatto potenziale), costituiscono una panoramica delle principali aree di criticità o di attenzione della *Computer Ethics*. La posizione verticale di una ellisse individua i livelli interessati da tale dominio. Esse rappresentano anche l'insieme delle questioni più critiche create o aggravate dai computer. Vediamo ora una descrizione dei principali interrogativi relativi a ciascun dominio.

2.2.1. Dawn of cyberspace, e-Democracy

Quali sono i nuovi scenari aperti dall'alba del ciberspazio? Quali sono le nuove metafore di cui abbiamo bisogno per sviluppare un uso appropriato di questo spazio? Come (e dove) usare questi nuovi strumenti di partecipazione alla vita pubblica, per discutere, per contri-

buire a dibattiti su questioni controverse, per confrontarsi con le pubbliche amministrazioni, per prendere decisioni, per votare? [4, 21].
La possibilità di avere identità multiple negli spazi virtuali sta aprendo diverse questioni etiche molto critiche: quale è la responsabilità sociale di un avatar? Quali sono le differenze negli aspetti di partecipazione e democrazia quando passiamo dal mondo fisico al mondo virtuale? Le tecnologie dell'informazione saranno "tecnologie di libertà" o "tecnologie di controllo", utopiche o distopiche? Esse espanderanno la partecipazione informata, la libertà e la democrazia, oppure diventeranno uno strumento per la manipolazione dell'opinione pubblica in una società a "controllo totale"?
In secoli di storia gli umani hanno sviluppato alcuni principi fondamentali che dobbiamo tenere presenti quando progettiamo applicazioni di "democrazia virtuale" o "iper-democrazia":
- la partecipazione diretta (non necessariamente mediata da rappresentanti);
- la discussione di diverse opzioni (non solo il loro conteggio come nei sondaggi);
- l'ascolto, l'apertura alle opinioni delle altre persone, la tolleranza;
- l'opportunità e la capacità di sostenere la propria opinione di fronte agli altri (non solo la comunicazione della propria scelta, spesso ristretta a poche opzioni selezionate "da casa"). Può la partecipazione essere a distanza, anonima, senza contatto fisico? Ha senso una agorà virtuale?

Altre caratteristiche importanti:
- la comunicazione orizzontale (da-molti-a-molti, non solo comunicazione verticale da-uno-a-molti, come in televisione);
- l'assunzione di responsabilità per le conseguenze delle proprie scelte (non solo risposte anonime);
- la pazienza (la democrazia è un processo necessariamente lento, mentre la tecnologia rischia di accelerare drammaticamente i processi decisionali senza rispettare i tempi umani; eppure la democrazia è per-gli-umani, tra-gli-umani);
- la riflessione (la nostra mente richiede tempo, silenzio, se vogliamo avere strumenti critici per poter influenzare la realtà dobbiamo in qualche modo fermare gli input, non possiamo limitarci a ricevere passivamente milioni di messaggi, la nostra mente ha una "banda" limitata che dovremmo rispettare).

Quando progettiamo le applicazioni di *e-participation* dovremmo attentamente considerare tutte queste virtù della democrazia antiche e consolidate.

2.2.2. Accessibility, Universal Access & Digital Divide

Quali sono le nuove barriere (economiche, culturali, sensoriali) che rischiamo di innalzare quando sviluppiamo l'*Information Society*? Come possiamo garantire l'accesso alle risorse di conoscenza usando dei criteri giusti ed inclusivi? Come possiamo evitare di imporre nuove restrizioni alle persone con disabilità, alle persone anziane? [36].
Come possiamo preservare la cosiddetta "neutralità della rete" di fronte alle pressioni ed al massiccio attacco proveniente dai proprietari delle principali tratte della Rete stessa (esempio, le grandi compagnie telefoniche) che vogliono arrivare a segmentare gli utenti offrendo servizi di traffico prioritari di prima classe? Se i pacchetti di prima classe avranno priorità negli snodi della rete, inevitabilmente verranno ritardati i pacchetti degli utenti in seconda classe.
Come possiamo governare e gestire Internet in modo equo ed etico? Come possiamo arrivare a definire una "Costituzione" per questo nuovo spazio virtuale? Come la discutiamo, approviamo (e facciamo applicare)? [33].
La sfida consiste nel trovare la giusta miscela tra i quattro principali strumenti per la *governance* di Internet:
- educazione (codici etici, *soft-laws*);
- leggi (vincoli legali internazionali, *hard-laws*);
- architettura tecnologica (evoluzione dei protocolli di rete, nuovi sviluppi tecnologici);
- mercato (autoregolazione) [19].

2.2.3. Workplace

Quali nuovi problemi e rischi professionali introduce il computer nel luogo di lavoro? Come vengono coinvolti i lavoratori, gli utenti finali nella progettazione dei nuovi sistemi? [7].
Quali sono le principali responsabilità dei *computer professional*, dei progettisti? [12].
Una delle aree che richiedono più ricerca è, ad esempio, quella del cambiamento in qualità (e quantità) dei posti di lavoro quando un processo viene automatizzato, o quando la maggio-

ranza dei processi di un'organizzazione vengono eseguiti da un sistema ibrido dove i compiti sono svolti dalle persone e dai computer.

2.2.4. Content & Education

Come selezionare, raccogliere, organizzare e diffondere i contenuti sulla rete? [18]. Se l'educazione è principalmente un processo di trasmissione di informazioni e conoscenze verso i nostri figli, chi seleziona i contenuti che verranno ereditati dalle future generazioni? Quale è il ruolo etico dei motori di ricerca in queste aree? [42]. E ancora, quale impatto avranno i computer sulle nostre capacità di apprendimento?

2.2.5. CopyRights

Come dovrebbero evolvere i meccanismi di incentivazione per gli artisti, per gli innovatori nel nuovo scenario della società della conoscenza? [20]. Se la tecnologia dell'informazione ha definitivamente cambiato le modalità con le quali possiamo copiare (la copia è indistinguibile dall'originale) e riutilizzare contenuti pre-esistenti, quali sono i nuovi modelli di business in grado di sostenere un'economia della conoscenza? Per quanto riguarda il software, cosa possiamo imparare dall'affermarsi del software *open-source* nei confronti del vecchio software proprietario? Le tecnologie dell'informazione sono forse lo strumento definitivo per, citando Newton [28] "salire sulle spalle dei giganti"?

2.2.6. Hackers

Come dovremmo proteggere i sistemi dalle intrusioni? Come mettere a frutto l'etica *hacker* per il bene collettivo, per migliorare la sicurezza stessa dei sistemi e la loro affidabilità? [22]. Dobbiamo definire molto attentamente il confine tra un *hack* (per esempio, per mettere in evidenza un difetto o trovare un "baco") ed un reato informatico. La società dell'informazione ha bisogno di persone esperte e dotate di speciali capacità tecniche per poter provare e sollecitare i sistemi complessi in condizioni estreme.

2.2.7. Privacy

Come possiamo proteggere i dati sensibili nel ciberspazio? Dove è il limite tra la sicurezza /sorveglianza ed una società Orwelliana? Quali sono le implicazioni relative ai due approcci alternativi "*opt-in*" (un servizio non è sottoscritto se non esplicitamente richiesto dall'utente) e "o*pt-out*" (un servizio è accettato implicitamente dall'utente a meno che non richieda esplicitamente di essere rimosso)? [23].
Le applicazioni di *ambient intelligence* (negli USA viene usato il termine *ubicomp* per definire l'insieme di *ubiquitous computing* e p*ervasive computing*), rappresentano una delle più grandi trasformazioni della storia dell'informatica. L'evoluzione verso il *disappearing computing* dove macchine intelligenti e sensori saranno attorno a noi, metterà l'ambiente stesso in grado di riconoscerci e di reagire alle nostre azioni. Le applicazioni *ubicomp* hanno l'immenso potenziale di migliorare le nostre vite, aumentando le nostre capacità comunicative (per esempio, visualizzando i livelli di attività o di variabili vitali di un familiare a distanza), automatizzando semplici compiti come riordinare medicinali scaduti, assistendo le persone con disabilità, mettendole in grado di partecipare pienamente alla vita sociale (per esempio, con convertitori testo-voce per chi ha problemi di vista), aiutando le persone anziane vulnerabili (ed esempio inviando un allarme quando cadono a terra o quando non prendono medicine vitali). D'altra parte, queste applicazioni hanno anche le potenzialità di condurci verso una società Orwelliana dove ogni azione di una persona viene controllata e registrata [14]: quali linee guida etiche dovrebbero essere usate da questi progetti *ubicomp*?
Dato che le nostre vite possono essere seriamente compromesse da errori nei database, come affronteremo i dilemmi etici (impliciti nella costruzione di database di grandi dimensioni come le banche dati del DNA) derivanti dalla necessità, da una parte, di assicurare il corretto funzionamento di una società democratica controllando i criminali e, dall'altra, di mantenere una società libera dalla sorveglianza? Come possiamo proteggere le nostre libertà civili da organizzazioni molto grandi e potenti, concentrate solo sulle loro missioni ed obiettivi, che usano il loro immenso potenziale tecnologico per controllare enormi quantità di comunicazioni, senza alcuna attenzione verso le conseguenze sociali? [5, 32].

2.2.8. Computer Crimes & Virus

Quale è la definizione di crimine informatico? Come possiamo proteggere i sistemi più importanti dai criminali informatici? [10]. Questo

dominio è separato da quello degli *Hackers* visto in precedenza in quanto gli *hacker* non sono da considerare, a priori, dei criminali[1]. La questione più critica in questo caso è legata al fatto che i reati "senza vittime" (i tipici reati informatici sono "puliti", raramente provocano direttamente dei danni fisici a persone). Questi reati stanno diventando diffusi che dovremmo educare molto attentamente le future generazioni di informatici per affrontare il problema dell'aumento incredibile di reati legati all'abuso dei computer [34].

2.2.9. Computer (Un)Reliability

Come possiamo migliorare l'affidabilità dei computer e proteggerci dai malfunzionamenti dei sistemi più critici per la vita umana? Perché il software è così inaffidabile? Chi e quando dovrebbe essere informato sui difetti più gravi nella sicurezza ed affidabilità delle applicazioni software? [27].
Chi è responsabile dei malfunzionamenti delle applicazioni *life-critical*? (pensiamo ai sistemi di guida automatizzata degli aerei, *fly-by-wire*, o delle auto, *drive-by-wire*, per non parlare delle applicazioni dove il software controlla apparecchiature ospedaliere). Sono responsabili i programmatori? I progettisti dell'hardware? Le aziende fornitrici delle apparecchiature? Gli utilizzatori?
Nell'ottobre 1985 nello storico gruppo di discussione su Internet *comp.risk* dedicato ai rischi ed affidabilità dei computer, apparve una definizione molto efficace del cosiddetto:
"*Effetto Titanic: la gravità delle conseguenze di un malfunzionamento di un sistema è direttamente proporzionale alla convinzione del progettista che ciò non possa mai accadere*".
E del suo corollario: "*La quantità e qualità della ridondanza incorporata in un sistema è direttamente proporzionale al grado di consapevolezza riguardo ai malfunzionamenti*".
Il dilemma etico qui consiste nell'evitare l'effetto Titanic.

[1] In genere la traduzione di *hacker* con "pirata informatico" usata da molti giornali non è propriamente corretta. Infatti letteralmente *hacker* andrebbe tradotto con "persona particolarmente esperta a programmare o a risolvere problemi con i computer". Quando parliamo di criminali informatici il termine inglese più adatto è *cracker*.

2.2.10. Artificial Intelligence

Quali sono le questioni profonde poste all'umanità dallo sviluppo dei robot (sia a scala umana che a scala nanometrica)? Perché delegare a macchine decisioni critiche per le vite umane? Quali questioni farà emergere l'arrivo dei *cyborg*, sistemi ibridi uomo-macchina? [17].
Dobbiamo iniziare a studiare un'etica per i robot come un'etica centrata sugli umani? Un'etica per gli umani coinvolti nello sviluppo dei robot? Come potremo costruire dei robot autonomi dotati di un "codice di condotta" (etica artificiale?) compatibile con lo sviluppo delle persone e della società? [37]. Quali saranno le questioni etiche più critiche quando avremo agenti artificiali a scala nanometrica (*nanobot*)? [1].
Come potremo affrontare aspetti come la nostra privacy (quando vi è qualcosa attorno e dentro di noi che raccoglie informazioni sui di noi), i reati informatici (chi sarà responsabile del controllo di queste nano-macchine?), l'affidabilità dei computer (come verranno provati i *nanobot*? Quali tipi di decisioni sarà in grado di prendere il software di cui sono dotati?), la longevità (saremo in grado di "ritirare" i *nanobot* per evitare l'effetto "vaso di Pandora"?) [31].

2.2.11. War

Quali sono le conseguenze dello sviluppo di armi "intelligenti" dove deleghiamo alle macchine la decisione finale di uccidere? [2]. Possiamo mettere nelle mani delle macchine il destino dell'umanità e del pianeta costruendo sistemi altamente computerizzati di difesa contro attacchi missilistici potenziali? [30].
Non abbiamo ancora idea delle conseguenze dello sviluppo di armi a scala nanometrica, dove avremo armi autonome dotate di sensori che rispondono a stimoli ambientali [3]; in questo caso le questioni etiche attraversano il confine degli umani ed arrivano a toccare anche l'ambiente.

2.2.12. Ecology & Recycling

Come possiamo evitare o minimizzare i rischi ambientali e l'impatto dei cicli di produzione dell'hardware? [9]. Possiamo usare le tecnologie dell'informazione per minimizzare il nostro impatto sul pianeta? Possiamo pensare ad un approccio a ciclo-chiuso per il riciclaggio dei computer? La cosiddetta "legge di Moore", il

raddoppio del numero di transistor in un chip ogni diciotto mesi, è sostenibile da un punto di vista ambientale? [35].

3. UNA METODOLOGIA PER L'ANALISI DI CASI DI STUDIO

Questo quadro concettuale viene attualmente sperimentato nell'analisi di casi di studio nel corso di *Computer Ethics* alla Scuola di Dottorato del Politecnico di Torino (Dottorato di Ricerca in Ingegneria Informatica e dei Sistemi). La metodologia usata parte dall'analisi di un caso reale, preso dalle notizie quotidiane o dalla letteratura, per poi studiarlo in dettaglio attraverso i seguenti quattro stadi.

- **Scenario**: in questa fase viene effettuata una descrizione dettagliata del caso nel modo più completo ed obiettivo possibile.
- **Stakeholders**: tutti i diversi "portatori di interesse" coinvolti vengono identificati con le loro (inter)relazioni.
- **Ethical Issue(s)**: ove vengono individuate le principali aree di attenzione (le ellissi viste prima) coinvolte nel caso in esame.
- **Alternative scenario(s)**: nella situazione descritta, come individuare le scelte giuste o sbagliate? Vi sono degli scenari alternativi possibili? Quale è il ruolo dei computer professional coinvolti nella progettazione delle tecnologie dell'informazione al centro dello scenario analizzato?

In genere, alla fine di questo processo, le questioni etiche più critiche vengono subito identificate. Per riuscire ad evitare (o perlomeno ad affrontare in qualche modo) tali questioni, vengono prodotte delle raccomandazioni a tre livelli:

- ***Education***: quali indicazioni possiamo fornire agli insegnanti, alle università che preparano le prossime generazioni di *computer professional*? Spesso, in classe, dopo discussioni approfondite, vengono identificati molti suggerimenti per gli esperti del settore.
- ***Guideline***: quali indicazioni possiamo fornire alle persone che definiscono i codici di condotta, le linee guida per le organizzazioni, le cosiddette *soft-laws*?
- ***Law***: quali raccomandazioni possiamo fornire a livello politico, alle autorità pubbliche, alle persone responsabili della definizione delle leggi, le *hard-laws*?

4. CONCLUSIONI

Le questioni identificate dalla *Computer Ethics* possono essere affrontare a molti livelli: a livello individuale, professionale e sociale. Un ruolo speciale spetta ai *computer professional*, nella definizione dei loro codici etici di condotta e nel fornire, al pubblico in generale ed ai decisori pubblici, tutte le informazioni sulle potenzialità, sui rischi e limiti delle tecnologie dell'informazione [13].

In Italia diverse università hanno introdotto corsi di *Computer Ethics* e questo contribuirà sicuramente alla formazione di figure complesse, persone che, oltre ad essere esperte, siano anche in grado di dare una giusta valutazione delle implicazioni sociali ed etiche della tecnologia dell'informazione.

Bibliografia

[1] Allhoff F., Lin P., Moor J., Weckert J.: *Nanoethics: The Ethical and Social Implications of Nanotechnology*. Wiley-Interscience, 2007.

[2] Bellin D., Chapman G.: *Computers in Battle: Will They Work*? Harcourt Brace Jovanovich, 1987.

[3] Carrol-Mayer M., Fairweather N., Rogerson S., Carsten-Stahl B.: *Stealing a march on society: The Ethical Implications of the Nanotechnological Weapons Policy*. De Montfort University, Proceedings of Ethicomp2005.

[4] Castells M.: *The Rise of the Network Society*. Blackwell, 1996.

[5] Cauley L.: *NSA has massive database of Americans' phone calls*. USA Today, 10 May 2006

[6] Computing Curricula, Social, Ethical and Professional Issues. *Communications of the ACM*, Vol. 34, n. 6, June 1991.

[7] CPSR: *Computer Professionals for Social Responsibility*. Participatory Design Conference, MIT, 1992.

[8] De Chardin T.: *The Phenomenon of Man*. Harper Perennial, 1976.

[9] EU-WEEE: *EU Directive 2003/108/EC on Waste Electrical and Electronic Equipment*. (WEEE), 2003.

[10] Forester T., Morrison P.: *Computer Ethics, Cautionary Tales and Ethical Dilemmas in Computing*. MIT Press, 1993.

[11] Gibson W.: *Neuromance*. Ace Books, 1984.

[12] Gotterbarn D.: *Informatics and Professional Responsibility, in Computer Ethics and Professional Responsibility*. Ed. by Ward Bynum T., Rogerson S., Blackwell Publishing, 2004.

[13] Gotterbarn D.: *Professionalism Issues*. Software Engineering Ethics Research Institute, Department of Computer and Information Sciences, East Tennessee State University, 2006.

[14] Jean Camp L.: *Design for Privacy: Towards a Methodological Approach to Trustworthy Ubicomp Design*. Proceedings of Ethicomp2005.

[15] Johnson, Deborah G.: *Computer Ethics*. Prentice-Hall; quarta ed. 2009, 1985.

[16] Jonas H.: *Das Prinzip Verantwortung*. Insel Verlag, Frankfurt. English edition (1984): *The Imperative of responsibility. In search for an ethics for the technological age*. University of Chicago Press, 1979.

[17] Joy B.: *Why the future doesn't need us*. Wired 8.04, 2000.

[18] Koyle K.: *Access: not just wires*. CPSR Annual meeting, October 1994.

[19] Lessig L.: *Code and other laws of cyberspace*. Basic Books, New York, 1999.

[20] Lessig L.: *The future of ideas*. Random House, 2003.

[21] Levy P.: *Collective Intelligence: Mankind's emerging World in Cyberspace*. Perseus, 1999.

[22] Levy S.: *Hackers: Heroes of the Computer Revolution*. Anchor Press/Doubleday, 1984.

[23] Lyon D.: *Electronic Eye: The Rise of Surveillance Society*. University of Minnesota Press, 1994.

[24] Maner, Walter: *Starter Kit in Computer Ethics*. Helvetia Press (published in cooperation with the National Information and Resource Center for Teaching Philosophy), 1980. Originally self-published by Maner in 1978

[25] Maner W.: *Unique Ethical Problems in Information Technology*. ETHICOMP95, 1995.

[26] Moor J.H.: *What Is Computer Ethics*? In: Bynum, Terrell Ward, ed. 1985, *Computers and Ethics*. Blackwell, Published as the October 1985 Issue of Metaphilosophy, p. 266-75.

[27] Neumann P.G.: *Computer-Related Risks*. Addison-Wesley, 1994.

[28] Newton I.: *Letter to Robert Hooke*. 15 February 1676.

[29] Parker, D.: Rules of Ethics in Information Processing. *Communications of the ACM*, Vol. 11., 1968, p. 198-201.

[30] Parnas D.: *Why the SDI software system will be untrustworthy*. University of Victoria, 26 June 1995.

[31] Patrignani N.: *NanoBots, the next frontier of Computer Ethics*. Ethics of Human Interaction with Robotic, Bionic, and AI Systems Workshop, supported by the ETHICBOTS European Project, Naples, 17-18 October, 2006.

[32] Risen J.: *Bush Lets U.S. Spy on Callers Without Courts*. New York Times, 16 December 2005.

[33] Rodotà S.: *Internet Bill of Rights*. Internet Governance Forum, Rio de Janeiro, 14 November 2007.

[34] Rogerson S.: IS Security needs ethics. *ETHIcol in the IMIS Journal*, Vol. 12, n. 4, 2002.

[35] Soesterberg: *Trans-Atlantic Network for Clean Production Meeting, Soesterberg*. The Netherlands, May 1999.

[36] UN-WSIS: *Geneva Declaration of Principles*. World Summit on Information Society, 2003.

[37] Veruggio G.: *A Proposal for Roboethics*. First International Symposium on ROBOETHICS, 30-th - 31-st January 2004, Villa Nobel, Sanremo, Italy.

[38] Ward Bynum T.: *The Foundation of Computer Ethics*. AICEC99, 1999.

[39] Weizenbaum, J.: *Computer Power and Human Reason: From Judgment to Calculation*. Freeman, 1976. Trad.it. "*Il potere del computer e la ragione umana. I limiti dell'intelligenza artificiale*", EGA-Edizioni Gruppo Abele - 1987.

[40] Wiener, Norbert: *Cybernetics: or Control and Communication in the Animal and the Machine*. Technology Press., 1948.

[41] Wiener, Norbert: *The Human Use of Human Beings: Cybernetics and Society*. Houghton Mifflin, 1950. (Second Edition Revised, Doubleday Anchor, 1954).

[42] Winograd T.: *Speech at The Ethics and Politics of Search Engines Panel*. Santa Clara University Markkula Center for Applied Ethics, Santa Clara University Center for Science, Technology, and Society, February 27, 2006.

NORBERTO PATRIGNANI, Docente di Computer Ethics alla Scuola di Dottorato del Politecnico di Torino. Laureatosi con lode in Informatica presso l'Università degli Studi di Torino, dopo molti anni nell'Information Technology (1975-1999: ricercatore Olivetti a Ivrea; 1999-2004 analista METAGroup, Stamford, USA) si è concentrato sulle implicazioni sociali ed etiche dei computer, collaborando a progetti di ricerca e innovazione ed insegnando in diversi atenei italiani (Università di Trento, Statale e Cattolica di Milano). Collabora con la EU Commission, DG Science & Society ed è responsabile della sezione Etica dell'IT del Centro Studi Politeia di Milano.
E-mail: norberto.patrignani@unicatt.it

ICT E INNOVAZIONE D'IMPRESA
Casi di successo

Rubrica a cura di
Roberto Bellini, Chiara Francalanci

La rubrica *ICT e Innovazione d'Impresa* vuole promuovere la diffusione di una maggiore sensibilità sul contributo che le tecnologie ICT possono fornire a livello di innovazione di prodotto, di innovazione di processo e di innovazione di management. La rubrica è dedicata all'analisi e all'approfondimento sistematico di singoli casi in cui l'innovazione ICT ha avuto un ruolo critico rispetto al successo nel business, se si tratta di un'impresa, o al miglioramento radicale del livello di servizio e di diffusione di servizi, se si tratta di una organizzazione pubblica.

GREEN IT
Efficienza nei consumi energetici dei data center
Il Caso SeeWeb

Francesco Merlo

1. Introduzione

Il caso riportato descrive le soluzioni IT per il miglioramento dell'efficienza nel consumo di energia elettrica – che rientrano nel più generale contesto del Green IT – adottate da un'azienda operante nel settore della fornitura di servizi IT. La duplice necessità di riduzione dei consumi elettrici e di ottimizzazione dell'utilizzo delle risorse computazionali ha portato SeeWeb ad una continua ricerca di innovazione dal punto di vista dell'infrastruttura del *data center* e dell'architettura logica dei sistemi informativi. L'articolo descrive in dettaglio l'esperienza di SeeWeb analizzandone il core business, il contesto competitivo e le valutazioni e l'impatto che le strategie orientate al Green IT hanno avuto sull'efficienza nel consumo di energia elettrica.

2. SeeWeb

SeeWeb nasce nel maggio del 1998, con la costruzione di una *server farm* innovativa sin dalle fasi di progettazione, per quanto riguarda le caratteristiche tecnologiche e infrastrutturali. L'offerta della società consiste in un insieme di servizi per le imprese comprendenti soluzioni di *hosting* e *housing* per siti e applicazioni web. La società ottiene i primi risultati dall'anno successivo, in quanto già alla fine del 1999 i prodotti SeeWeb la collocavano nelle prime dieci posizioni di fornitori di servizi di registrazione di domini Internet in Italia.
Lo sviluppo della società prosegue nel 2002 con l'acquisizione di Valore, una delle principali aziende italiane nel settore che, con oltre 4000 domini attivi e 3000 clienti, consente a SeeWeb di ampliare l'offerta di prodotti con un brand a basso costo ed entrare nella fascia di mercato *entry level*.
Nel 2005, attraverso la controllata Quadrant, Seeweb apre un nuovo *data center* a Milano, che diventerà il crocevia delle nuove attività di *co-location*, di *clustering* geografico, di *disaster management* e delle attività commerciali per il nord Italia.
Il 2008 si conclude con una posizione di leadership sul mercato italiano, con un fatturato in crescita di oltre il 25%, che sfiora i 4,2 milioni di Euro. Oltre 100.000 domini sono registrati sui server di SeeWeb, alcuni dei quali di grande prestigio e notorietà e un investimento pari al 20% del fatturato destinato all'innovazione tecnologica e di prodotto consente un costante

aggiornamento dell'offerta di servizi. Inoltre, l'azienda è stata certificata ISO9001 e per prima in Italia ISO14001 per l'impatto ambientale controllato delle sue attività.
Attualmente, Seeweb prosegue con una strategia di crescita nell'attività di fornitura di soluzioni tecnologiche per la presenza in rete al fine di consolidarsi come operatore di primo piano nel mercato nazionale e affacciarsi al mercato europeo e internazionale.
Il *core business* di SeeWeb consiste nell'offerta di servizi di hosting, *housing* e *co-location* di alta qualità con management a vari livelli delle risorse offerte, nonché di progettazione e realizzazione di piattaforme telematiche complesse per l'*outsourcing*. Gli obiettivi che guidano l'offerta dei servizi di SeeWeb sono:
❑ sviluppare un'offerta completa e integrata di prodotti in grado di soddisfare i fabbisogni dei clienti sia in termini di flessibilità delle soluzioni, sia in termini di servizi di qualità e professionalità;
❑ fornire prodotti a prezzi competitivi, garantendo la migliore tecnologia, affidabilità e sicurezza disponibile sul mercato;
❑ favorire la crescita del business dei propri clienti attraverso un attento monitoraggio delle loro esigenze e delle loro potenzialità.

3. Il Green IT

L'espressione "Green IT" indica un nuovo campo di ricerca che si rivolge allo studio delle problematiche ambientali ed energetiche legate alle tecnologie dell'informazione1. Più precisamente, il Green IT può considerarsi articolato in tre grandi aree:
❑ il miglioramento dell'efficienza del consumo energetico delle infrastrutture IT;
❑ la gestione eco-compatibile del ciclo di vita delle infrastrutture IT;
❑ l'utilizzo dell'IT come tecnologia abilitante per pratiche di *governance* aziendale orientate all'ecosostenibilità.
La prima area ha l'obiettivo di ottimizzare la progettazione dell'architettura dei sistemi informativi e dei *data center* con lo scopo di massimizzare l'efficienza del consumo di energia elettrica, prendendo in considerazione tutti gli aspetti – dalle specifiche caratteristiche dei modelli dei server alle differenti pratiche di utilizzo di tecniche quali la virtualizzazione – che hanno effetti sul consumo elettrico.
La seconda area di ricerca si rivolge invece alla definizione ed alla promozione di nuove metodologie e nuove tecniche per la gestione eco-compatibile dell'intero ciclo di vita delle tecnologie dell'informazione, prendendo in considerazione aspetti che vanno dalla produzione dell'hardware e all'ottimizzazione degli imballaggi, fino alla gestione delle attività di stoccaggio ed alla rottamazione dell'hardware ormai dismesso.
Infine, la terza area di ricerca ha l'obiettivo di utilizzare le tecnologie dell'informazione come fattore abilitante per misurare e monitorare i parametri relativi all'ecosostenibilità di qualsiasi processo industriale, non per forza limitandosi al contesto dell'IT. Per esempio, per la definizione di sistemi di supporto alle decisioni e cruscotti gestionali che permettano la misurazione ed il controllo di indicatori di performance dei processi come consumo energetico, temperatura, quantità di agenti inquinanti (perd esempio, CO_2 equivalente) emessi durante lo svolgimento delle tipiche attività produttive.
Considerata la vastità delle tematiche trattate dal Green IT e soprattutto la loro rilevanza in termini di impatto ambientale e di possibili riduzioni dei costi operativi, è una conseguenza naturale la nascita di un'estrema attenzione negli ambienti accademici e industriali.
Dal punto di vista dei consumi di energia elettrica è fondamentale osservare come, soprattutto nello sviluppo architetturale dei componenti elettronici destinati alla grande diffusione, i produttori di hardware si siano focalizzati sull'incrementare la frequenza del *clock* di sistema e sulla miniaturizzazione sempre più spinta delle componenti [2]. Se non si considerano le aree di ricerca dedicate specificatamente alla progettazione di componenti a basso consumo (per esempio, i sistemi *embedded*), in generale le tematiche legate all'efficienza dell'impiego dell'energia sono infatti state raramente tenute in considerazione. Il risultato è stato la disponibilità di sistemi molto potenti e veloci, a scapito però di elevati consumi e di scarsa efficienza nell'impiego dell'energia utilizzata. L'impatto degli elevati consumi energetici delle infrastrutture IT può essere analizzato sulla base di tre differenti prospettive.
❑ **Emissioni equivalenti di anidride carbonica (CO_2) e gas serra**
Sulla base di dati riportati da recenti ricerche [1, 3, 4], il settore IT è responsabile di circa il

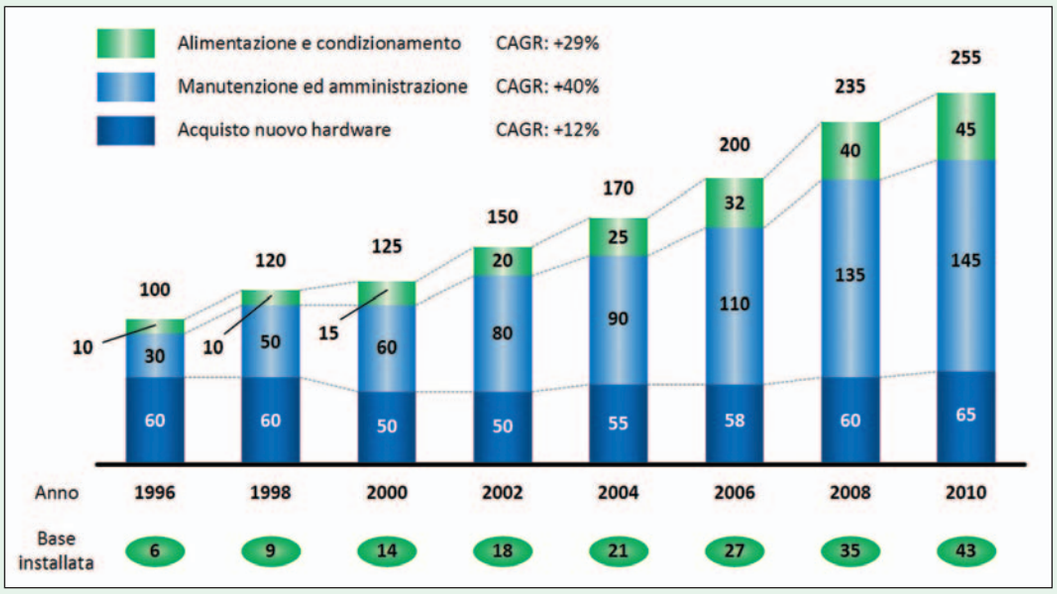


FIGURA 1
Spesa globale nel mercato server (miliardi di Dollari, Fonte: [6])

2% delle emissioni globali di anidride carbonica, una quantità paragonabile a quella dell'industria aeronautica. Il consumo complessivo medio annuale di un PC desktop corrisponde all'emissione di una tonnellata di CO_2, mentre un server ha lo stesso impatto ambientale su base annua di un SUV medio che percorre circa 2,5 km/litro di carburante [5]. Infine, ma certamente non trascurabile, l'aspetto di inquinamento da materiali pesanti, come piombo, cadmio e mercurio, per cui il settore IT contribuisce a livello mondiale al 70% delle emissioni su terreno [3].

❑ Costi operativi

Da un punto di vista prettamente economico, l'impatto degli elevati consumi energetici delle infrastrutture IT ha acquisito importanza sempre più rilevante nel corso degli anni. In particolare, considerando gli ultimi 12 anni (1996 – 2008), il costo legato all'alimentazione ed al condizionamento degli impianti IT è aumentato di ben quattro volte (Figura 1).

Considerando invece il rapporto tra i costi per l'acquisto di nuovo hardware ed i costi per alimentazione e condizionamento, si può notare come attualmente questi ultimi siano pari al 60% dei primi. Ciò significa che per ogni Euro investito nell'acquisto di nuovo hardware, 0,60 € dovranno essere destinati su base annuale per consentirne il funzionamento (Figura 2). Tale rapporto è inoltre destinato a crescere, se non fosse per il continuo aumento del costo dell'energia, che in Italia risulta essere uno dei maggiori a livello mondiale. Dal punto di vista manageriale, queste osservazioni provocano implicazioni di non poco conto, soprattutto considerando il fatto che attualmente pochissime imprese provvedono a rendicontare nel budget IT i relativi consumi energetici, falsando quindi le prospettive.

❑ Scalabilità degli impianti

Infine, l'elevato consumo energetico delle infrastrutture IT rappresenta una limitazione alla scalabilità degli impianti e delle infrastrutture destinate ad ospitarli, come i *data center*. I requisiti energetici delle nuove apparecchiature ad alta densità possono raggiungere soglie estremamente elevate (per esempio, un rack contente 5 *enclosure* per sistemi *blade* può arrivare a richiedere 20 kWatt di potenza), così come la potenza dei più comuni PC desktop continua a crescere ad un ritmo annuale di 8-10%. Tali considerazioni pongono di fronte ad evidenti problematiche nel momento in cui le infrastrutture quali i *data center* siano localizzati in aree densamente abitate, poiché i distributori di energia elettrica potrebbero non essere in grado di convogliarvi la quantità richiesta di energia. Un recente studio [3] ha dimostrato come nei prossimi anni ben il 60% dei *data center* saranno afflitti da problematiche inerenti la scalabilità dei consumi energetici, del condizionamento dei sistemi e dello spazio fisico.

Considerando il caso di un *data center* tipico, esiste un insieme piuttosto articolato di strategie progettuali ed interventi eseguibili a poste-

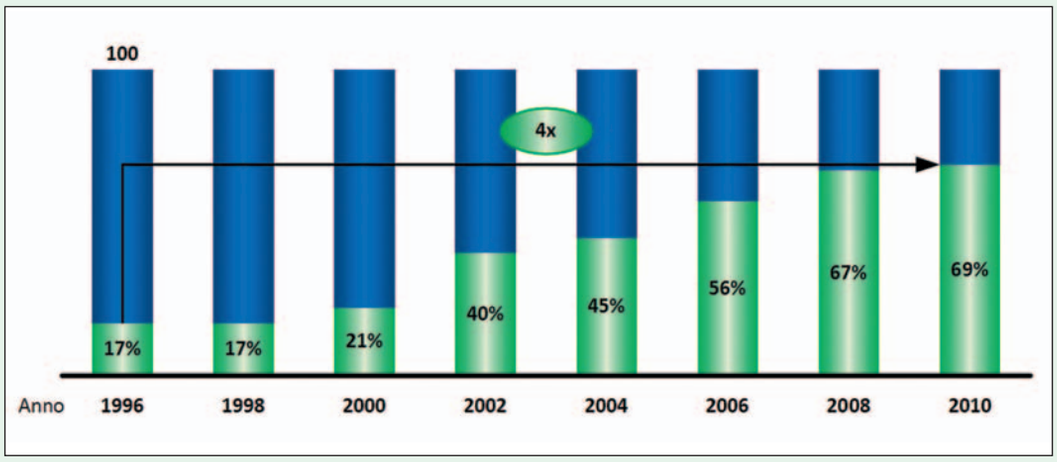


FIGURA 2

Incidenza della spesa per alimentazione e raffreddamento sulla spesa di acquisti del nuovo hardware (Fonte: [6])

riori su impianti già esistenti che consentono di razionalizzare e di ridurre entro certi limiti i consumi delle apparecchiature IT che vi sono contenute. In particolare, è possibile categorizzare tali interventi in tre grandi macrocategorie.

❑ **Interventi sulle infrastrutture**

Una prima categoria di interventi si rivolge alle infrastrutture del *data center* e alla loro disposizione all'interno delle sale. Per esempio, capita spesso nei *data center* che i *rack* non siano completamente occupati da server o altre apparecchiature IT: l'utilizzo di pannelli contenitivi per l'aria negli spazi vuoti permette di gestire al meglio i flussi dell'aria, evitando che l'aria fredda si misceli a quella calda ove non gradito. Tale accorgimento consente di ottenere risparmi potenziali nell'ordine dell'1-2% dei consumi dell'impianto di condizionamento [7]. Inoltre, l'avere un layout dei *rack* efficiente ha un impatto notevole sull'efficienza del sistema di raffreddamento: per esempio, la creazione di corridoi caldi e freddi ed un corretto posizionamento dei componenti dedicati al ricircolo dell'aria all'interno delle sale consente di ottenere risparmi potenziali del 5-12% sui consumi [7].

❑ **Interventi sull'impianto di condizionamento**

La seconda categoria di interventi si rivolge alle configurazioni e alle tecnologie impiegate nell'impianto di condizionamento vero e proprio. Per esempio, il *freecoling* è un sistema di raffreddamento che sfrutta la differenza di temperatura dell'aria all'interno del *data center* con quella dell'ambiente esterno. In particolare, i sistemi che fanno ricorso a tale tecnologia non utilizzano ventole per forzare la circolazione dell'aria o pompe per la movimentazione dell'acqua di raffreddamento, ma si limitano vincolarne la direzione di movimento (per esempio, ascendente/discendente), riducendo quindi i consumi. Le riduzioni legate all'utilizzo di questa tecnologia possono raggiungere il 40% dei consumi di sistemi classici per il raffreddamento dei sistemi [8]. Un secondo esempio di tecnica applicabile agli impianti di condizionamento prevede di isolare fisicamente i corridoi caldi da quelli freddi, consentendo di aumentare l'efficienza delle macchine che raffreddano l'aria (in quanto esse operano con un'efficienza direttamente proporzionale alla differenza di temperatura dei flussi di aria che trattano). In questo caso è possibile raggiungere risparmi nell'ordine del 20-30% [9].

❑ **Interventi sulle apparecchiature IT**

Infine, è possibile ottenere razionalizzazioni dei consumi energetici anche operando direttamente sulle apparecchiature IT ospitate in un *data center*. La maggior parte dei *data center* è attualmente composta da server "a bassa densità", che risalgono mediamente a 3-5 anni fa. I produttori di hardware, come IBM, HP e Dell, hanno introdotto sul mercato nuovi modelli di server progettati e realizzati con particolare attenzione all'efficienza dell'impiego dell'energia elettrica. Il ricorso a tali apparecchiature in luogo di quelle con un'età di almeno 3-5 anni consente di ottenere riduzioni dei consumi energetici nell'ordine del 20% dei loro consumi [7]. È inoltre possibile operare anche sull'architettu-

ra logica del sistema informativo: per esempio, mediante tecniche quali la virtualizzazione si ottimizza l'utilizzo delle risorse computazionali a fronte delle richieste di elaborazione. Mediante l'impiego di tali soluzioni è possibile raggiungere risparmi sul consumo delle infrastrutture IT fino al 65% [7].
La tabella 1 riporta un riepilogo degli esempi di interventi *green* che è possibile applicare in un *data center* tipico, con le relative percentuali di risparmio e i consumi di energia elettrica che è possibile ridurre mediante l'applicazione degli interventi.

4. Il GreenICT in SeeWeb

L'utilizzo delle tecnologie dell'informazione riveste in SeeWeb un ruolo di fondamentale importanza, in quanto esse sono fattore abilitante per la fornitura dei servizi che ne costituiscono il *core business*. Di conseguenza, la corretta progettazione, e soprattutto la corretta gestione delle infrastrutture e delle apparecchiature IT rappresentano per SeeWeb obiettivi strategici fondamentali. Negli ultimi anni l'azienda ha perciò dedicato risorse e investimenti in progetti volti a migliorare costantemente le caratteristiche dei suoi *data center*, anche dal punto di vista delle tematiche del *green* IT. Proprio in questo ambito, due sono le direttive su cui SeeWeb ha concluso o sta svolgendo attività di razionalizzazione e ottimizzazione.
La prima area di attività riguarda gli interventi orientati all'aumento di efficienza dei consumi energetici delle infrastrutture. A partire dal 2007, in entrambi i *data center* di Milano e Frosinone SeeWeb si è concentrata su tre leve per ridurre i consumi:
- tecnologia impiegata nell'impianto di generazione dell'acqua fredda per il raffreddamento, mediante l'introduzione di *chiller* ad alto coefficiente di prestazione (COP > 5.0) impieganti compressori centrifughi a levitazione magnetica;
- strategie per il raffreddamento delle sale dei *data center*, che hanno visto l'impiego di unità di trattamento dell'aria dotate di microprocessori integrati per il coordinamento dinamico delle modalità di funzionamento, che consentono di parzializzare l'utilizzo dell'acqua per il raffreddamento e di adattare dinamicamente la velocità delle ventole per la movimentazione dell'aria;
- tecnologia impiegata nelle infrastrutture per la gestione e la distribuzione dell'energia, mediante l'introduzione di UPS con un rendimento a carico parziale del 95%.

Parallelamente alle attività citate, nel *data center* di Milano è stato avviato anche un progetto pilota che vede l'introduzione di un sistema di raffreddamento *freecooling* aria-aria, che ha permesso l'utilizzo dell'aria dell'ambiente esterno per aumentare l'efficienza del sistema di condizionamento. SeeWeb ha stimato che mediante questa soluzione, è possibile raggiungere una riduzione media del 30% dei costi attuali di climatizzazione.
Complessivamente, i risultati degli interventi di miglioramento dell'efficienza del consumo energetico apportati alle sole infrastrutture porteranno il PUE dei *data center* di SeeWeb da un valore di 2,0 ad un valore di 1,5: se prima degli interventi per ogni Watt di energia erogato alle apparecchiature IT era richiesto un secondo Watt di energia da erogare alle infrastrutture

TABELLA 1
Riepilogo di alcuni interventi green per l'aumento dell'efficienza dei consumi in un data center

Categoria	Intervento	Riduzione potenziale	Risparmi su consumi
Infrastrutture	Pannelli contenitivi per l'aria	1 – 2%	Impianto condizionamento
Infrastrutture	Layout efficiente	5 – 12%	Impianto condizionamento
Condizionamento	Freecoling	fino al 40%	Impianto condizionamento
Condizionamento	Contenimento dell'aria nei corridoi caldo-freddo	20-30%	Impianto condizionamento
Sistema Informativo	Hardware più efficiente nei consumi	a partire dal 20%	Apparecchiature IT
Sistema Informativo	Virtualizzazione	fino al 65%	Apparecchiature IT

del *data center* (alimentazione protetta, distribuzione, climatizzazione, servizi vari), dopo gli interventi, le stime di SeeWeb prevedono di riuscire a ridurre tale quantità a 0,5 W.
La seconda area di attività riguarda invece gli interventi volti al miglioramento dell'efficienza di consumo di energia da parte delle apparecchiature IT. La realizzazione di queste attività si inquadra nella volontà di SeeWeb di entrare nel segmento di mercato dell'*hosting retail low-cost*, fino ad oggi non presidiato. In tale segmento è essenziale, dal punto di vista del vantaggio competitivo rispetto alla concorrenza, presentarsi sul mercato con un prezzo del servizio molto basso, rispettando però i requisiti e i vincoli tecnologici che i potenziali clienti si attendono dal fornitore. Di conseguenza, l'attenzione agli aspetti di efficienza energetica e ottimizzazione dei livelli di utilizzo delle apparecchiature IT riveste un ruolo assolutamente centrale nell'economia complessiva del progetto.
Le infrastrutture e le apparecchiature IT impiegate per la realizzazione del nuovo sistema sono state esplicitamente scelte nell'ottica del "*green IT*", permettendo notevoli risparmi sia dal punto di vista energetico, ma anche, ad esempio, dal punto di vista di spazio fisico occupato all'interno delle sale del *data center*. L'utilizzo di server di tipo *blade*, infatti, consente di ridurre drasticamente lo spazio occupato, consentendo di aumentare di oltre 30 volte l'efficienza di utilizzo dello spazio fisico: l'infrastruttura impiegata da SeeWeb consente di ospitare circa 300.000 siti web per metro quadro occupato, rispetto ad un valore tradizionale tipico di circa 10.000 siti per metro quadro. La progettazione del sistema è stata inoltre focalizzata all'ottimizzazione sia hardware che software: l'impiego di server *blade* e *storage* SAN ha consentito di ottenere altissime densità di potenza per unità di spazio, che unite alle tecnologie di virtualizzazione hanno permesso di ridurre di un fattore 10 i consumi medi di energia per sito ospitato. In un sistema tradizionale, infatti, i consumi IT si attestano tipicamente intorno a 1,0 W/sito, mentre il sistema implementato in SeeWeb riduce tali consumi a meno di 0,1 Watt/sito. Tali riduzioni sono state possibili grazie all'impiego di un sistema software che gestisce lo strato di virtualizzazione, con l'obiettivo di minimizzare il numero di *core* in uso nell'intera piattaforma. In questo modo è possibile adattare dinamicamente l'utilizzo delle risorse computazionali sulla base del carico richiesto, per esempio sulla base della fascia oraria: nelle ore di minor utilizzo delle risorse, interi server possono essere letteralmente spenti, consentendo un notevole risparmio energetico.

5. Conclusioni

Risulta evidente come il ricorso alle nuove tecnologie, sia a livello infrastrutturale che a livello prettamente informatico, possa dare luogo a notevoli miglioramenti nell'impiego dell'energia nel funzionamento di un *data center*.
In particolare, gli interventi messi in atto da SeeWeb dimostrano come la possibilità di migliorare l'efficienza di utilizzo dell'energia elettrica possa tradursi non solo in minori costi di gestione operativa, ma anche in vantaggio competitivo nei confronti dei concorrenti sul mercato. L'Ing. Antonio Baldassarra, CEO di SeeWeb, intervistato per l'occasione ha affermato che sebbene l'azienda stia perseguendo l'obiettivo di un continuo miglioramento nell'efficienza dell'utilizzo dell'energia all'interno dei propri *data center*, rileva ancora forti differenze tra il contesto delle tecnologie dell'informazione e quello di settori industriali più classici. Per esempio, nell'ambito della climatizzazione degli ambienti, i coefficienti di prestazione degli impianti di SeeWeb - sebbene allo stato dell'arte, non riescano ad avvicinarsi a quelli dei climatizzatori industriali impiegati in altri ambiti – sono in grado di ottenere prestazioni pressoché doppie (con valori di COP prossimi a 10).
Ciononostante, risulta indubbio che le tecnologie e le strategie introdotte nel mercato permettano notevoli miglioramenti nell'efficienza di impiego dell'energia elettrica, risultato di assoluta rilevanza nell'ambito di un settore che contribuisce in modo rilevante alle emissioni di gas serra, ma che al contempo presenta notevoli margini di riduzione potenziale dei consumi.

Bibliografia

[1] Murugesan S.: Harnessing Green IT: Principles and Practices. *IT Professional*, Vol. 10, n. 1, 2008, p. 24-33.

[2] Schaller R.R.: Moore's law: past, present and future. *IEEE Spectrum*, Vol. 34, n. 6, 1997, p. 52-59.

[3] Brown E.G., Lee C.: *Topic Overview: Green IT*. Forrester Research report, November 2007.

[4] Kumar R.: *Important Power, Cooling and Green IT Concerns*. Gartner Report, January 2007.

[5] Restorick T.: *An Inefficient Truth*. Global Action Plan Report, 2007.

[6] Josselyin S. L., Dillon B., Nakamura M., Arora R., Lorenz S., Meyer T., Maceska R., Fernandez L.: *Worldwide and regional server 2006-2010 forecast*. IDC Report, 2006.

[7] Rasmussen N.: *Implementing Energy Efficient Data Centers*. APC, White Paper #114.

[8] Club T.I.ER: *Cogenerazione e freecooling: l'esperienza del data center di Acantho*.

[9] Niemann J.: *Hot aisle vs. cold aisle containment*. APC, White Paper #135.

FRANCESCO MERLO, conseguirà il Dottorato di Ricerca in Ingegneria dell'Informazione presso il Politecnico di Milano nel 2010; presso lo stesso ateneo ha conseguito la Laurea Specialistica in Ingegneria Informatica nel 2006 e la Laurea in Ingegneria Informatica nel 2003. Durante gli studi di dottorato è stato *visiting researcher* presso il *College of Management del Georgia Institute of Technology* (Atlanta, USA). Collabora con la Fondazione Politecnico di Milano in qualità di consulente in ambito delle tecnologie dell'informazione. Si occupa di ricerca su tematiche legate al Green ICT, alla qualità ed ai costi di sviluppo e manutenzione del software; su questi temi è autore di pubblicazioni nazionali e internazionali. Dal 2006 tiene presso il Politecnico di Milano parte dei corsi di Sistemi Informativi per l'Impresa e di Gestione della Tecnologia dell'Informazione.
E-mail:merlo@elet.polimi.it

ICT E DIRITTO

Rubrica a cura di

Antonio Piva, David D'Agostini

Scopo di questa rubrica è di illustrare al lettore, in brevi articoli, le tematiche giuridiche più significative del settore ICT: dalla tutela del *domain name* al *copyright* nella rete, dalle licenze software alla *privacy* nell'era digitale. Ogni numero tratterà un argomento, inquadrandolo nel contesto normativo e focalizzandone gli aspetti di informatica giuridica.

Business Continuity: come prevenire i disastri applicando le normative

David D'Agostini, Antonio Piva, Attilio Rampazzo

1. Introduzione

Molti fatti, si pensi per esempio al *black-out* nazionale del 16 settembre 2003 causato da problemi IT e non da mancanza di energia, stanno portando le organizzazioni a porre una particolare attenzione sulla Continuità Operativa conosciuta tra gli addetti ai lavori come *Business Continuity*.

Il D.Lgs 196/2003, meglio noto come Codice della Privacy, recependo i provvedimenti comunitari 96/45/CE e 2002/58/CE ha previsto che il trattamento di dati personali effettuato con strumenti elettronici sia consentito solo se sono adottate procedure per la custodia di copie di sicurezza e il ripristino della disponibilità dei dati ed dei sistemi[1]. La maggior parte dei sistemi IT delle aziende non sono comunque preparati a gestire un blocco delle infrastrutture di business superiore ai sette giorni[2].

Altre normative di settore, emanate sempre negli ultimi anni, prevedono l'adozione di un sistema di continuità operativa. Infatti la *Business Continuity* ormai non risulta solo un'esigenza interna alle organizzazioni, bensì è oggetto di precise richieste settoriali, come per esempio le Linee Guida per la continuità di servizio emanate da Banca d'Italia nel 2004 o le Disposizioni di Vigilanza di Banca d'Italia del 21/3/2007, oppure i quaderni del Centro Nazionale per l'Informatica nella pubblica Amministrazione (CNIPA) sulla continuità operativa nella Pubblica Amministrazione e, infine, la Direttiva Europea per le Infrastrutture Critiche (2008/114/CE dell'8 dicembre 2008).

L'origine della Gestione della Continuità Operativa è spesso ricondotta al *Disaster Recovery* specialmente se si considerano i sistemi IT, ovvero al ripristino dei sistemi IT e delle informazioni dopo un "disastro". Persino i termini "*business continuity*" (continuità operativa) e "*disaster recovery*" vengono spesso confusi: alcune persone li usano per indicare la stessa cosa, per altri significano cose completamente diverse. Altri termini quali: valutazione del rischio, analisi dell'impatto sull'organizzazione (BIA, o *Business Impact Analysis*), gestione delle crisi, gestione degli incidenti, ridondanza, sicurezza, protezione e *governance* sono stati via via aggiunti a questa miscellanea.

Esistono poche norme in tutto il mondo sulla continuità operativa, soprattutto perché è considerata ancora un concetto relativamente nuovo. Tra le norme/leggi/guide che esistono nel mondo, comunque, molte fanno riferimento alla Gestione

[1] Per una completa trattazione si rinvia al numero 2 del giugno 2008, al numero 2 di giugno 2007, al numero 1 di marzo 2004, all'interno della presente rubrica ICT e Diritto di Mondo Digitale.

[2] È questo il riscontro emerso da Gartner Group citato in: "La sicurezza condotta dall'IT è destinata a diventare un ricordo?", News Letter Clusit del 28 febbraio 2008.

della Continuità Operativa (GCO o *Business Continuity Management*, BCM), anche se non utilizzano necessariamente la stessa terminologia. Tra queste norme esistenti si annoverano:

- NFPA 1600 della *National Fire Protection Association* degli USA: questo standard è stato sviluppato dalla lotta agli incendi e vede la continuità operativa da una prospettiva di divieto di accesso;
- HB 221 e HB 292/293: lo standard australiano per la BCM e la guida alla gestione della continuità operativa;
- AS/NZS 4360:2004 una norma in comune tra Australia e Nuova Zelanda, abbinabile al HB 436 per una linea guida alla gestione del rischio;
- NIST *special publication*. 800-34 fornisce le istruzioni, i consigli pratici, le considerazioni per implementare e gestire un *Contingency Planning*[3];
- SPRING TR 19: il riferimento tecnico di Singapore per la gestione della continuità operativa, che si occupa soprattutto degli aspetti tecnici dei sistemi BCM;
- *The King II report of Corporate Governance*: queste linee guida sudafricane sulla gestione del rischio considerano la BCM da un punto di vista di *governance* (governo e gestione dei Sistemi informativi);
- *The Civil Contingencies Act 2004*: questa legge è entrata in vigore nel Regno Unito nel 2004 a seguito dell'approvazione reale (*Royal Assent*) e fornisce una guida alla BCM;
- ISO 27002: una norma per i sistemi di gestione della sicurezza delle informazioni che gestisce e minimizza le minacce alle informazioni[4];
- ISO 22399: una linea guida per la consapevolezza e la preparazione agli incidenti e la gestione della continuità operativa.

In Inghilterra a inizio secolo nasce il *Business Continuity Institute* (BCI), il primo organismo al mondo sull'argomento, che offre accreditamento professionale nel campo della continuità operativa e ha oltre 4.000 membri in più di 85 Paesi: nel 2002 il BCI ha emesso le sue prime Linee Guida di Buona Pratica redatte con la collaborazione di molti esperti del settore. Queste Linee Guida hanno fornito l'ossatura sulla quale si sono modellate le prime attività del BSI (*British Standards Intitution*, ente di normazione britannico) nel campo della gestione della continuità operativa, che hanno portato all'emissione di una specifica sulla *Business Continuity* denominata PAS 56. Nel 2006 la pubblicazione della BS 25999-1 che sostituisce la PAS 56, e successivamente nel 2007 la pubblicazione della BS 25999-2 hanno posto le basi per una nuova famiglia di norme per il mondo anglosassone (anche se già riconosciute a livello mondiale) sulla *Business Continuity*.

La Norma BS 25999 è composta da due parti: la prima elenca i "*consigli pratici*" fornendo la guida e gli obiettivi della norma oltre a spiegare esattamente il campo di applicazione della Gestione della Business Continuità; la seconda parte contiene le specifiche a fronte delle quali le organizzazioni possono essere certificate, dettagliando i requisiti per implementare, documentare e migliorare un Sistema di Gestione della *Business Continuity* (o Sistema di Gestione della Continuità Operativa). In buona sostanza: la parte 1 è il "*dovrebbe*", mentre la parte 2 è il "*deve*".

Il lancio della BS 25999-1, il codice di pratica, ha senza dubbio modificato l'opinione diffusa che la Gestione della *Business Continuity* (BCM) non avesse alcun vantaggio rispetto ad altre discipline di gestione. Una nuova norma raramente viene accolta con un entusiasmo, essendo spesso percepita più come "un aggravio burocratico" che come un'opportunità imprenditoriale. Invece quando la bozza iniziale dei "consigli pratici" è stata diffusa pubblicamente per i commenti, il documento è stato scaricato 5.000 volte in tutto il mondo; prima di allora, nella storia ultracentenaria del BSI non si erano mai superate le poche centinaia. Ne è seguita un'enorme quantità di commenti da recepire prima che la norma potesse essere pubblicata. Molti dei commenti erano positivi, anche se una parte era ferocemente contraria ad una qualsiasi forma di norma in questo campo. Analogamente, la BS 25999-2, la norma che dettaglia i requisiti a fronte dei quali un'organizzazione può essere oggetto di *audit* e certificata, ha suscitato un incredibile interesse, così grande che il BSI ha dovuto organizzare un tour mondiale delle città più importanti, impazienti di saperne di più sulla nuova norma.

---

[3] Inteso come il processo attraverso il quale vengno pianificate le azioni da intraprendere in caso di incidenti, in modo che i servizi IT possano continuare a funzionare, o a riprendere a funzionare dopo un interruzione.

[4] Per una completa trattazione si rinvia al numero 3 di settembre 2008, all'interno della presente rubrica ICT e Diritto di Mondo Digitale.

Attualmente si sta lavorando in tutto il mondo per ottenere un'accettazione chiara e standardizzata della gestione della *Business Continuity*: è un compito arduo perché ciò che risulta applicabile in un settore o in uno Stato potrebbe non essere accettabile in un altro.
Negli USA, la *National Fire Protection Association* (NFPA) ha emesso un regolamento per la gestione delle emergenze e la risposta ai disastri mentre il *National Institute of Standards and Tecnologies* (NIST) ha emesso una linea guida sulla pianificazione degli eventi nel settore IT.
Non si tratta, tuttavia, di norme e quindi le organizzazioni non possono ottenere la certificazione che tanto desiderano: ne consegue che al momento la BS 25999-2 è l'unica norma che possa certificare un Sistema di Gestione della *Business Continuity*.

2. BS 25999-1:2006 Code of Practice for Business Continuity Management

Come già citato la Norma BS 25999-1 è una linea guida che stabilisce i principi, la terminologia e il processo per la gestione della continuità operativa e del business; la norma stabilisce le attività che risultano necessarie nella progettazione del processo di gestione della continuità operativa e i documenti da produrre e, inoltre, fornisce i passi necessari per mettere in atto il processo della continuità del business.
Questa norma è orientata a tutte le organizzazioni indipendentemente dalla dimensione, settore commerciale o tipologia e fornisce assistenza ai responsabili di gestione del programma della continuità del business. L'approccio si basa sull'analisi del contesto dell'organizzazione dove l'obiettivo è quello di capire quali sono i prodotti/servizi critici e a quali esigenze dei portatori di interessi (degli *stakeholder*) principali a questi sono collegabili; tale prima analisi formale porta a individuare quello che in termini di prodotti e servizi sarà il campo di applicazione del modello.
Il sistema proposto parte dall'analisi del rischio (identificazione, analisi e valutazione), per proseguire con l'individuazione degli aspetti di rischio che superano il livello di soglia valutato accettabile e che hanno influenza sulla *Business Continuity*, ossia la capacità di mantenere i livelli di servizio/prodotto collegati all'evento/crisi ed emergenza in atto. Se i rischi non sono collegabili ai suddetti livelli di servizio non rientrano nel modello di *Business Continuity* e nel relativo approccio sistemico: ciò, ovviamente, deve essere da un lato coerente rispetto al rapporto tra costo e beneficio e dall'altro rispettoso dei requisiti contrattuali, di legge e normativi.
Si devono definire per ogni rischio collegabile alla Gestione della *Business Continuity* tre documenti:
- piano di gestione incidenti (per minimizzare l'evento);
- piano per la *business continuity*;
- programma per il ripristino della situazione normale.

Il contenuto della Norma BS25999-1 è il seguente:
- *Scope and applicabilità*;
- *Terms and definitions*;
- *Overview of business continuity management (BCM)*;
- *The Business Continuity Management policy*;
- *BCM Programme Management*;
- *Understanding the organization*;
- *Determining business continuity strategy*;
- *Developing and implementing a BCM response*;
- *Exercising, maintaining and reviewing BCM arrangements*;
- *Embedding BCM in the organization's culture*.

La metodologia proposta dalla norma è suddivisa in fasi (capitoli): ogni fase, concepita per essere eseguita in sequenza, è costituita da specifiche attività che implicano l'utilizzo di strumenti/tecniche. L'insieme completo delle fasi costituisce il processo di *Business Continuity Management*, contraddistinto dall'essere un processo continuo che evolve nel tempo e recepisce i mutamenti di business, organizzativi e tecnologici delle realtà in cui è applicato.
Il ciclo di vita della Gestione della *Business Continuity* (o Continuità Operativa), rappresentato nella figura 1, comprende sei elementi:
- programma della Gestione della *Business Continuity* (cap. 5);
- comprendere l'attività / analizzare l'organizzazione (cap. 6);
- determinare la strategia della Gestione della *Business Continuity* (cap. 7);
- sviluppare e implementare la Gestione della *Business Continuity* (cap. 8);
- mettere in pratica, mantenere e revisionare la Gestione della *Business Continuity* (cap. 9);
- incorporare la Gestione della *Business Conti-*

FIGURA 1

Il ciclo di vita della Continuità Operativa (BS 25999-1:2006)

nuity nella cultura dell'organizzazione (cap.10). Tutte le fasi (esposte nei capitoli della norma) sono molto importanti nel ciclo di vita, tuttavia si evidenzia che il capitolo 5 rappresenta il cuore, come dice la norma, di tutto il processo di Gestione della *Business Continuity*: il programma deve definire tutti gli aspetti organizzativi necessari ad assicurare che le attività richieste per la realizzazione della *Business Continuity* siano realizzate nei modi e nei tempi richiesti dall'Organizzazione.

Altro capitolo di rilievo è il 10 nel quale viene evidenziato che, per avere successo, la *Business Continuity* deve diventare parte del modo in cui l'organizzazione viene gestita, indipendentemente dalla dimensione/settore. Infatti nel disegno del ciclo di vita del BCM risulta evidente come questa fase avvolga tutta l'attività di Gestione della *Business Continuity*.

Per gli addetti ai lavori, specialmente dell'ambiente IT, abituati a valutare il *Disaster Recovery* o la *Business Continuity* di un sistema IT tramite i due indici *Recovery Time Objective* (RTO) e *Recovery Point Objective* (RPO), nella nuova norma hanno trovato un nuovo indice: *Maximum Tolerable Period of Disruption* (MTPOD), mentre non è fatto cenno dell'RPO.

L'MTPOD è la durata massima dopo la quale un'organizzazione sarà irrevocabilmente minacciata se la sua operatività non può essere ripresa, mentre l'RPO è il massimo tempo fissato per l'indisponibilità di un prodotto, del servizio o di un'attività, ovvero il tempo entro il quale il prodotto, servizio, attività da proteggere deve essere ripristinato dopo un incidente, con l'attenzione che l'RTO deve essere sempre inferiore all'MTPOD.

3. BS 25999-2:2007 Requisiti Sistema di Gestione della Business Continuity

La BS 25999-2 fornisce le indicazioni e i requisiti per la messa in esercizio di un Sistema di Gestione della *Business Continuity* (o Gestione Operativa) certificabile.

I contenuti della norma sono riassunti nell'indice seguente:

- *Scope and applicability*;
- *Terms and definitions*;
- *Planning the business continuity management system (cap. 3)*;
  - *General*;
  - *Establishing and managing the BCMS*;
  - *Embedding BCM in the organization's culture*;
  - *BCMS documentation and records*;
- *Implementing and operating the BCMS (cap. 4)*;
  - *Understanding the organization*;
  - *Determining business continuity strategy*;
  - *Developing and implementing a BCM response*;
  - *Exercising, maintaining and reviewing BCM arrangements*;
- *Monitoring and reviewing the BCMS (cap. 5)*;
  - *Internal audit*;
  - *Management review of the BCMS*;
- *Maintaining and improving the BCMS (cap. 6)*;
  - *Preventive and corrective actions*;
  - *Continual improvement*.

La BS 25999-2:2007 è strutturata macroscopicamente in 3 blocchi di requisiti:

- dal capitolo 0 al capitolo 2: requisiti introduttivi e di spiegazione della norma (come quasi tutte le norme);
- dal capitolo 3 al capitolo 6: requisiti applicativi del BCMS;
- allegati.

All'interno dei capitoli (detti anche *requisiti* o *clausole*) dal 3 al 6 sono distribuiti i veri e propri requisiti applicativi per il Sistema di Gestione della *Business Continuity* (BCMS).

Lo schema di figura 2, sebbene non esaustivo, fornisce un'indicazione di massima dei capitoli

FIGURA 2
Capitoli dei requisisti della norma con i relativi collegamenti al ciclo PDCA

dei requisisti della norma con i relativi collegamenti al ciclo *Plan-Do-Check-Act* (PDCA – ciclo di DEMING). In buona sostanza il ciclo PDCA è il motore dello standard BS 25999-2:2007 come tutte le norme che hanno preceduto questo standard (ISO 9001:2000, ISO 14001:2004, ISO 20000-1:2005, ISO 27001:2005).
L'infrastruttura di Gestione della *Business Continuity* si basa quindi sul ciclo di continuo di miglioramento *Plan-Do-Check-Act* (Figura 3) applicato in modo continuativo su più livelli dell'organizzazione e sull'approccio sistemico allo scopo di identificare, capire e gestire i processi tra loro correlati contribuendo all'efficacia ed all'efficienza dell'organizzazione nel conseguire i propri obiettivi.
Il capitolo 3 relativo alla fase *PLAN* prevede come stabilire la politica di *Business Continuity*, gli obiettivi, i controlli, i processi e le procedure pertinenti per gestire i rischi e migliorare la *Business Continuity* al fine di produrre risultati conformi alle politiche e agli obiettivi generali dell'organizzazione.
Il capitolo 4 relativo alla fase *DO* prevede di attuare e rendere operativa la politica della *Business Continuity*, i controlli, i processi e le procedure.
Il capitolo 5 relativo alla fase *CHECH* prevede di valutare e, ove applicabile, misurare le prestazioni a fronte della politica della *Business Continuity*, degli obiettivi e delle esperienze pratiche, quindi riportare i risultati alla direzione ai fini del riesame e del miglioramento.
Il capitolo 6 relativo alla fase *ACT* prevede di mantenere attivo, aggiornato e migliorato il BCMS intraprendendo azioni correttive e preventive, basate sui risultati delle verifiche/riesame e sulla rivalutazione del campo di applicazione, della politica e degli obiettivi del Sistema di Gestione della *Business Continuity*.
Ne consegue che un Sistema di Gestione della *Business Continuity* (BCMS) conforme alla BS 25999-2, deve soddisfare i seguenti capisaldi (cfr. requisito 3.1 della norma):
❑ definizione di campo di applicazione, politica ed obiettivi;

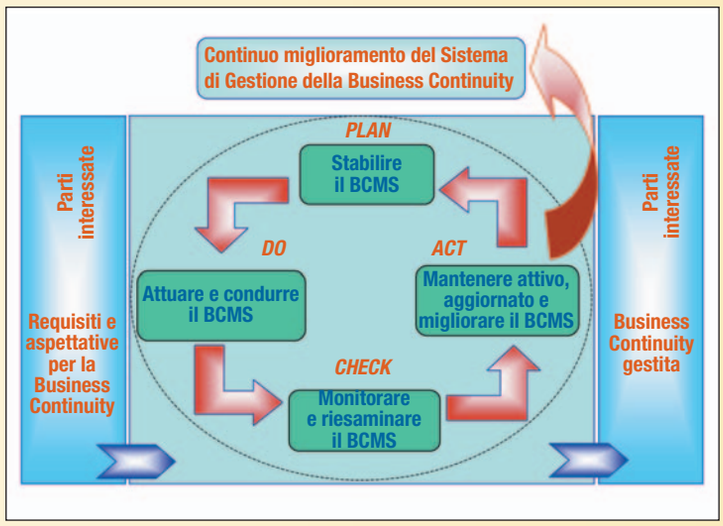


FIGURA 3
Ciclo PDCA come modello di riferimento per la descrizione dei processi e dei requisiti dello standard

❑ allocazione di risorse;
❑ attribuzione di ruoli e responsabilità;
❑ consistenza del sistema documentale e coerenza col campo di applicazione;
❑ accuratezza della *Business Impact Analysis* e del *Risk Assessment*;
❑ preparazione di adeguati piani di *Business Continuity* e di *Incident Management*;
❑ evidenza dell'implementazione del BCMS, sia nelle sue componenti operative (policy, procedure, istruzioni tecniche), che organizzative (*job description, audit* interni, *management review*).
Si rilevano sia l'efficacia del BCMS nel raggiungere gli obiettivi di continuità operativa stabiliti in Politica, sia l'efficienza del medesimo nel migliorare la gestione dei processi operativi nei tre possibili "stati" definiti dalla BS 25999:
• risposta agli incidenti;
• gestione della continuità;
• ripristino della normalità.
In questo ambito deve essere prestata molta attenzione all'esercizio (prova) dei piani di *Bu-*

siness Continuity (requisito 4.4.2 della Norma), il quale consente di essere certi che quanto pianificato trovi un'efficace attuazione in caso di incidente; per essere attendibili e verosimili, i test dei piani devono riprodurre quanto più fedelmente possibile la situazione ipotizzata.

4. Conclusioni

Lo standard BS 25999 risulta ancora giovane e dovrà maturare ulteriormente per la sua accettazione a livello mondiale, per essere con tutta probabilità adottato dall'ISO.
A meno di due anni dalla sua applicazione si contano già numerose aziende, che in tutto il mondo, lo hanno adottato e hanno conseguito la certificazione di BSI tramite accreditamento UKAS[5]: si tratta di circa una cinquantina di soggetti (nella Tabella 1 sono esposte il numero di aziende per nazione) e tra queste anche un'organizzazione italiana[6].

Nazione	n° aziende	n° siti
Brasil	1	1
India	6	61
Italy	1	1
Japan	3	19
Mexico	1	1
Netherlands	1	1
Poland	2	11
South Korea	3	3
Spain	1	1
Sweden	1	1
Taiwan	3	4
UAE	1	24
United Kingdom	21	43
U.S.A.	4	11
Totali	49	182

TABELLA 1
Le certificazioni BS 25999- 2 emesse da BSI sotto accreditamento UKAS, Fonte: BSI Group - United Kingdom (agg. al 20 agosto 2009)

Da ultimo l'ente normatore anglosassone a fine settembre 2008 ha pubblicato la versione draft di una nuova norma con i "consigli pratici per la continuità delle tecnologie dell'informazione e della comunicazione" per organizzazioni sia pubbliche che private. Tale proposta nasce dalla necessità di nuove specifiche di continuità per le organizzazioni con dipendenza dalle tecnologie dell'informazione e della comunicazione (ICT).
La nuova BS 25777, pubblicata recentemente, mira a colmare questa lacuna individuata da BSI, sostituirà la PAS 77 pubblicata nel 2006 per fornire un orientamento sulla continuità IT: per i suoi contenuti andrà a fiancheggiare la BS 25999-1 e integrerà il ciclo di vita del BCM con il nuovo ciclo di vita del BCM in ambiente ICT.
Anche l'ISO, ente internazionale, non è rimasto indietro: acquisendo le esperienze degli enti nazionali (Australia, Israele, Giappone, Inghilterra e Stati Uniti) ha aperto più progetti per nuove norme sull'argomento tra le quali:
- ISO/IEC CD 27031 ICT *readiness for Business Continuity*;
- ISO/PAS 22399:2007 *Societal security - Guideline for incident preparedness and operational continuity management*;
- ISO/CD 22301 *Societal security - Preparedness and continuity management systems – Requirements*.

Bibliografia

[1] BSI BS 25999-1:2006 *Business continuity management. Code of practice*.

[2] BSI BS 25999-2:2007 *Business continuity management. Specification*.

[3] BSI BS 25777:2008 I*nformation and communications technology continuity management. Code of practice*.

[4] BSI PAS 56:2003 - *Guide to Business Continuity Management*.

[5] Ente britannico di accreditamento organizzazioni di certificazione: www.ukas.com

[6] È l'ICCREA, istituto Centrale delle Banche di Credito Cooperativo che ha voluto mettere a frutto il lavoro svolto per aderire alle disposizioni della Banca d'Italia e nel giugno 2008 ha conseguito la certificazione del suo Sistema di Gestione della *Business Continuity*.
Si noti che altri organismi indipendenti di certificazione stanno proponendo la certificazione; infatti sempre in Italia un'altra organizzazione del mondo finanziario la SIA-SSB ha conseguito la certificazione tramite DNV (*Det Norske Veritas*).

[5] BSI PAS 77:2006 - *IT Service Continuity Management. Code of Practice*.

[6] Business Continuity Institute – *GPG 2008 Good Practice Guidelines A framework for Business Continuity mgmt*.

[7] ISO/IEC 27001:2005 - *Security techniques - Information security management systems - Requirements*.

[8] ISO/IEC 27002:2005 – *Information tecnology – Security tecniques – Code of practice for information security management*.

[9] CNIPA - Quaderno n. 28 del 6/2006: *Linee guida per la continuità operativa nella PA*.

[10] CNIPA - Quaderno n. 35 del 8/5/2008: *La continuità operativa nella PA - Casi di studio*.

[11] ISCOM - *Gestione delle emergenze locali*.

[12] Banca d'Italia: *Gestione della continuità operativa*. Bollettino di Vigilanza luglio 2004.

[13] ABI Lab: *Metodologia per la realizzazione del piano di continuità operativa*.

[14] ENISA: *IT & Business Continuity Overview and implementation principles*.

[15] ENISA: *How to raise information security awareness*.

[16] Quaderno n. 28 AICQ – Comitato per la Qualità del Software e dei Servizi IT: *La Gestione della Continuità Operativa: dal Disaster recovery, alla Business Continuity ad un Sistema di Gestione della Business Continuity*.

DAVID D'AGOSTINI avvocato, master in informatica giuridica e diritto delle nuove tecnologie, collabora all'attività di ricerca scientifica dell'Università degli studi di Udine e ha fondato l'associazione "*Centro Innovazione & Diritto*". È componente della Commissione Informatica dei Consigli dell'Ordine del Triveneto, responsabile dell'area "*Diritto & informatica*" della rivista "*Il foro friulano*", membro dell'organo di Audit Interno di Autovie Venete SpA.
E-mail: studio@avvocatodagostini.it

ANTONIO PIVA, laureato in Scienze dell'Informazione, *Vice Presidente dell'ALSI* (Associazione Nazionale Laureati in Scienze dell'Informazione ed Informatica) e Presidente della commissione di informatica giuridica. Docente a contratto di *diritto dell'ICT e qualità* all'Università di Udine. Consulente sistemi informatici e Governo Elettronico nella PA locale, valutatore di sistemi di qualità ISO9000 ed ispettore AICA.
E-mail: antonio@piva.mobi

ATTILIO RAMPAZZO, consulente di Sistemi Informativi e Sicurezza delle Informazioni in AlmavivA Finance SpA. Ha maturato un'esperienza più che trentennale nello sviluppo e conduzione di progetti informatici in ambito bancario e finanziario, nei quali la qualità e la sicurezza hanno ricoperto un ruolo determinante. È Vice Presidente del Comitato AICQ "Qualità del Software e dei Servizi IT", valutatore Sistemi di Sicurezza delle Informazioni R.G.V.I. (AICQ_SICEV certificato n.3), socio AIPSI-Associazione Italiana Professionisti Sicurezza Informatica.
E-mail: attilio@rampazzo.it

PROFESSIONE ICT
Competenze e professionalità per l'innovazione digitale

Rubrica a cura di

Roberto Bellini, Federico Butera, Alfonso Fuggetta

Il tema dell'innovazione e della competitività del sistema Italia è all'ordine del giorno della discussione economica e di quella sulle politiche industriali; sono promosse iniziative istituzionali a supporto dell'innovazione e si auspica un maggiore contributo della ricerca a livello universitario e privato. Anche l'Unione Europea spinge sul tema dell'innovazione, in particolare sul ruolo che le tecnologie ICT possono svolgere sia nei sistemi industriali che nei sistemi di governo e sull'importanza che può avere la definizione di un *framework* comune delle competenze ICT, compatibile con quanto previsto dall'EQF - *European Qualification Framework* - recentemente approvato dall'Unione Europea (2006).

Mondo Digitale vuole sostenere la diffusione di una maggiore sensibilità sul contributo che le competenze e le professionalità relative alle tecnologie digitali possono fornire in termini di innovazione dei servizi e del business dell'Impresa e di servizi per la cittadinanza erogati dagli enti della Pubblica Amministrazione. Questa nuova rubrica è dedicata appunto all'approfondimento sistematico di tutti gli aspetti che riguardano i progetti di analisi e di miglioramento delle competenze per l'innovazione digitale, il monitoraggio dei bisogni di competenza richiesti dal mercato e la valutazione delle offerte di qualificazione e aggiornamento delle competenze proposte dalle istituzioni educative di base e dagli operatori della formazione professionale e permanente.

La rubrica analizzerà l'andamento del mercato del lavoro delle professionalità ICT, i casi di successo nella crescita di competenze del personale dei fornitori di tecnologie e servizi e degli specialisti ICT, sia delle imprese manifatturiere e di servizio che degli enti della Pubblica Amministrazione, nonché l'andamento delle retribuzioni a livello nazionale e internazionale, usando come riferimento i profili e le competenze dello Standard EUCIP che AICA promuove in Italia.

La certificazione professionale in IBM come fattore competitivo

Walter Pistarini, Raffaele Stifani

1. Introduzione: le professioni in IBM

La certificazione professionale, in IBM, gode di grande considerazione e ha una lunga storia alle spalle. Infatti, fu introdotta più di quindici anni fa, insieme al modello "per professioni", quando IBM, fino ad allora focalizzata su hardware e software, decise che anche i servizi professionali dovevano entrare a far parte del *core business*. Perchè potesse funzionare ed avere un alto livello qualitativo erano necessari alcuni cambiamenti fondamentali. L'obiettivo era sicuramente quello di creare le "condizioni ambientali" affinchè i talenti emergessero e si potessero identificare come professionisti altamente qualificati.

IBM è sempre stata all'avanguardia nella gestione e nella certificazione dei professionisti dell'informatica

Sino agli inizi degli anni '90, il modo più comune per fare carriera era quello di abbandonare il contesto professionale e inserirsi in una carriera manageriale. Quel modello non poteva andare bene in un'azienda di servizi che invece

doveva (e deve) poter offrire percorsi di crescita anche ai suoi migliori "tecnici" (considerando con il termine "tecnici" tutti coloro che ricoprono ruoli non manageriali). Si stabilì quindi che le prospettive di carriera, inclusa la dirigenza fino ai massimi livelli, dovevano essere "aperte" anche alle figure non manageriali, ovviamente con qualche adattamento. Mentre infatti per un manager esiste un indicatore relativamente semplice delle sue capacità, basato sulla gestione del suo team ed i relativi risultati di business, per una persona tecnica è decisamente più difficile misurare il suo contributo ai risultati aziendali.
Quando si decise di avviare un *framework* di professioni ci si basò su quanto esisteva già nei laboratori di sviluppo e nei centri di ricerca IBM. Le professioni così nate furono subito intese come comunità di persone con competenze e interessi omogenei e distribuite in tutte le linee di business e in tutte le aree geografiche. Da subito erano (e lo sono tuttora) indipendenti dalle strutture di business e dalle organizzazioni territoriali e tendono a raggruppare persone con comuni obiettivi di conoscenza.
Il processo di certificazione fu il passo identificato per effettuare una selezione accurata per i livelli più alti della carriera "tecnica". Ovviamente questo passaggio doveva associarsi comunque con l'impatto sul business. Le professioni oggetto di questo approccio furono quelle che avevano un rapporto più stretto con i clienti ove l'attenzione e la sensibilità doveva essere maggiore: rappresentanti commerciali, specialisti, architetti, capi progetto e consulenti.
Con la certificazione professionale si volevano indirizzare tre obiettivi complementari tra loro. Il valore per il cliente: avere la certezza che le persone con cui il cliente interagisce avessero i giusti livelli di conoscenza ed esperienza. Il valore per IBM: permettere l'identificazione delle persone con i giusti livelli di competenza e più idonee ad indirizzare le richieste del mercato; in questo modo veniva anche stabilita una base per l'identificazione dei potenziali talenti e per la gestione delle risorse. Il valore per l'individuo: il professionista poteva determinare un suo preciso percorso di crescita e individuare le aree da sviluppare.
Questi processi e criteri di certificazione sono stati ampiamente rivisti dapprima internamente con l'organizzazione delle risorse umane e successivamente con autorità esterne ed indipendenti come *British Computer Society* (BCS) e *Institution of Engineering and Technology* (IET). Sono in corso, per l'Italia, le procedure di riconoscimento di alcune professioni IBM a cominciare da quella dell'*IT Architect*, rispetto allo Standard EUCIP, promosso da AICA che rappresenta il CEPIS, proprietario di EUCIP ed ECDL.
IBM crede fortemente infatti che il valore delle sue professioni viene arricchito e consolidato dal riconoscimento delle principali istituzioni esterne indipendenti; IBM opera quindi per sviluppare la propria posizione di *leadership*, ma favorisce anche lo sviluppo di un professionalismo sempre maggiore che potrà essere un vantaggio sia per i propri clienti che per tutti gli specialisti di tecnologie informatiche che operano in generale presso le imprese e le organizzazioni che effettuano domanda di ICT.

2. Il processo di certificazione

Oggi in IBM, con alcune variazioni a seconda della professione, esistono tre livelli di certificazione e sono collegati a diversi livelli di carriera. I tre livelli sono conosciuti come accreditamento, certificazione e certificazione senior, nella maggioranza dei casi. Tuttavia comincia ad affacciarsi una nuova terminologia proveniente dall'organizzazione "*The Open Group*" e che identifica i tre livelli come certificazione di livello 1, di livello 2 e di livello 3. In particolare, per quanto riguarda la professione di "*IT Architect*", sulla quale questo articolo pone una focalizzazione maggiore, l'organizzazione "*The Open Group*" definisce così i tre livelli di certificazione:

- livello 1: il candidato è in grado di operare con assssistenza e supervisione e possiede comunque un ampio spettro di conoscenze nell'area delle architetture informatiche;
- livello 2: indicato come *Master Certified IT Architect*: il candidate è in gracdo di operare in modo autonomo e di prendere in carico la responsabilità come lead architect per la realizzazione di sistemi e soluzioni;
- livello 3: noto come *Distinguished Certified IT Architect*: il candidato possiede un'ampia esperienza ed una profonda conoscenza per realizzare architetture e soluzioni che hanno impatto sul business aziendale.

Ovviamente i requisiti da soddisfare variano da livello a livello. In questa breve relazione ci si focalizza sulla certificazione di livello 2, quella nata per prima tanti anni fa. Inoltre, per poter fare esempi concreti, si prende come modello di riferimento la professione di *IT Architect* (riquadro).
La figura 1, che concettualizza quanto detto finora, mette in relazione i vari livelli di carriera (che per semplicità espositiva sono stati raggruppati seguendo una terminologia abbastanza diffusa) con i livelli di certificazione menzionati. Si osserva che in questo contesto la certificazione, di qualsiasi livello si tratti, è un presupposto indispensabile per un avanzamento di carriera, ma da sola non è sufficente in quanto devono esserci anche le condizioni di mercato.
La caratteristica fondamentale della certificazione di livello 2 è che ha una approccio detto di "*peer review*": i membri della commissione esaminatrice sono colleghi più esperti della stessa professione e già certificati (quindi gli esaminatori non sono i manager nella linea gerarchica del candidato).
I passi fondamentali del processo sono sintetizzati nella figura 2 e sono i seguenti:

- valutazione dei prerequisiti;
- preparazione del "*package*";
- esame con la commissione.

2.1. Valutazione dei prerequisiti

Prima di procedere con la compilazione del "*package*" (il documento da sottoporre alla commissione che contiene anche le guide di compilazione) il candidato deve verificare il possesso dei requisiti fondamentali, che variano, ovviamente, da professione a professione. In questa fase è fondamentale fare un attento esame con il proprio manager in quanto il *package* richiede almeno due livelli manageriali di approvazione. In questa fase è importante il ruolo del *mentor* (una persona che svolge la stessa professione, già certificata) che aiuterà il candidato a fare una valutazione attenta dei requisiti e a colmare eventuali lacune sia di competenza che di esperienza.

2.2. Preparazione del package

Il "*nomination package*" è il documento che il candidato deve compilare e sottoporre alla commissione esaminatrice. Il *package* è diverso per ogni professione, pur avendo molto concetti similari. Nel caso della professione *IT Architect*, il *package* è strutturato in sei sezioni in cui il candidato deve dimostrare la conformità con i requisiti di:

1. competenza (*compliance with skill requirements*);
2. esperienza (*compliance with experience requirements*);

La professione di IT Architect

"Il ruolo dell'*IT Architect* è quello di definire e progettare soluzioni ai problemi di *business* dei clienti attraverso l'uso esperto e ragionato delle tecnologie informatiche. Queste soluzioni si estrinsecano mediante "architetture" e possono includere componenti di sistemi, applicazioni e processi. Possono includere l'utilizzo e l'integrazione di un'ampia gamma di prodotti, tecnologie e servizi, varie architetture di sistemi e applicazioni, come pure componenti hardware e software".
Per poter svolgere questo tipo di attività in modo proficuo è evidente che ci deve essere, alla base, una certa esperienza sull'IT in generale e una conoscenza profonda relativamente ad alcune aree. Pertanto un *IT Architect* lavorerà per acquisire competenze e maturare esperienza soprattutto nella progettazione e nell'integrazione di componenti differenti.

FIGURA 1
I livelli di certificazione

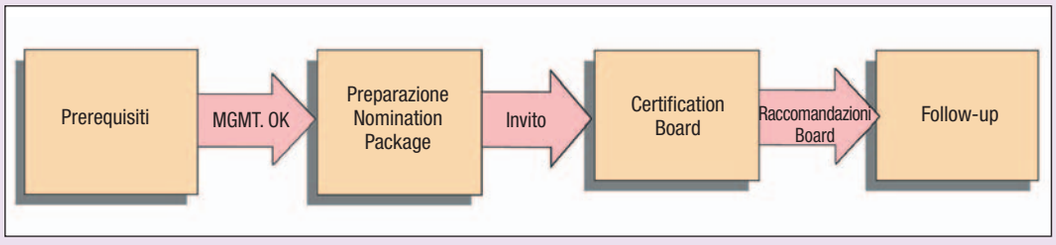


FIGURA 2
Il processo di certificazione

3. formazione (*compliance with education requirements*);
4. contributo alla comunità professionale (*contributions to the profession*);
5. progetti realizzati e ruolo svolto (*projects profiles*);
6. storia professionale (*professional history*).

Si comincia con la sezione delle competenze. Ogni *IT Architect* dovrà avere una certa dose di competenze "di base", tipiche della professione, e un insieme di competenze relative alla disciplina in cui opera e per la quale intende ottenere la certificazione. La professione di *IT Architect* definisce le seguenti discipline: *Enterprise Architecture, Information Architecture, Application Architecture, Infrastructure Architecture, Integration Architecture, Operations Architecture, Systems Engineering and Architecture*. Le discipline tengono conto dell'area ove il candidato mostra una più elevata competenza ed ha accumulato una maggiore esperienza.

Le competenze fondamentali contemplano competenze di comunicazione, leadership e capacità negoziali, ma anche competenze più legate alle tecnologie come sviluppare architetture, usare tecniche di modellazione, utilizzare metodologie note e condivise (per evitare che ognuno inventi la sua metodologia), conoscere e applicare gli standard, conoscere ed usare tecniche consulenziali, conoscere e attuare il piano di qualità ecc..

Approfondendo ulteriormente l'aspetto delle competenze, vanno poi identificate quelle tipiche della disciplina in cui si opera. Per esempio un *IT Architect* "applicativo" dovrà essere in grado di saper selezionare con metodo eventuali pacchetti applicativi, dovrà avere la capacità di determinare i requisiti funzionali e non-funzionali (compresi, per esempio, tutti gli aspetti di sicurezza, di alta affidabilità, di prestazioni ecc.), dovrà essere in grado di guidare il gruppo di lavoro nell'ambito di un progetto di sviluppo applicativo ecc.. Per converso un *IT Architect* in ambito infrastrutturale dovrà aver maturato competenze, per esempio, nel fare *assessment* di infrastruture IT, dovrà saper sviluppare un modello concettuale ed un modello fisico dell'architettura infrastrutturale, dovrà saper selezionare i prodotti per una gestione efficiente dei sistemi.

Dopo la sezione delle competenze si passa alla sezione, molto più importante, relativa alle esperienze maturate. Sin dalle origini è stato evidente che una certificazione professionale, per poter offrire e garantire il massimo del risultato, doveva dare un peso rilevante all'esperienza del candidato. I corsi, i diplomi, gli attestati sono indispensabili per avere un buon punto di partenza, ma è solo tramite l'esperienza che si acquisisce quella sensibilità necessaria a garantire un alto livello qualitativo ed il successo dei progetti.

Le esperienze richieste variano a seconda della professione in cui si è inseriti. Per la professione degli *IT Architect*, che abbiamo deciso di prendere ad esempio in questo articolo , e per la certificazione di secondo livello che stiamo esaminando, vengono richieste molte specifiche esperienze, a partire dalla definizione di architetture da almeno quattro anni. Inoltre è necessario dimostrare l'aver operato in progetti che prevedevano l'integrazione di differenti prodotti e tecnologie, l'aver progettato soluzioni informatiche con sistemi differenti, con architetture differenti, con piattaforme hardware e software diversi. Va da sè che il tutto (o buona parte di esso) va fatto seguendo una metodologia riconosciuta (dall'IBM o dal mercato). Inoltre un *IT Architect*, per meritarsi la certificazione, deve dimostrare di aver coordinato un gruppo di lavoro su alcuni progetti significativi e di successo; deve essere stato coinvolto su tutte le fasi di un progetto, deve possedere conoscenze almeno in un settore di industria e altro ancora. Tutte queste esperienze devono essere suffragate da riferimenti concreti ai progetti in cui sono maturate e devono essere verificabili da parte dei membri della commissione esaminatrice.

I requisiti di formazione che il candidato deve dimostrare riguardano non solo quelli relativi alla tecnologia ma anche quelli relativi all'approccio consulenziale, alle nozioni che regolano la gestione dei progetti e alla capacità di gestire end-to-end gli aspetti della soluzione. È normale che una persona, che sta già svolgendo il ruolo di *IT Architect* da alcuni anni, abbia naturalmente frequentato corsi sui modelli architetturali e sulle tecnologie. L'attenzione che qui si pone è più su temi metodologici e formativi che tecnici: la gestione dei progetti, le tecniche consulenziali, le meto-

dologie ed un corso particolare, sviluppato all'interno della comunità degli *IT Architect* IBM e chiamato "*Architectural Thinking*". Tale corso, di grande valore, viene tenuto da architetti molto esperti e tratta tutti i più importanti aspetti da tenere in conto nella definizione di una architettura informatica; inoltre permette uno scambio di esperienze e la discussione su importanti problematiche di questa professione.
Un'altra sezione del *package* di certificazione riguarda i contributi alla professione (*givebacks*). Un *IT Architect* certificato deve rappresentare un "*role model*" per la comunità e per essere tale non basta essere tecnicamente molto bravi. È necessario dare un qualcosa in più, all'azienda e ai colleghi. Bisogna "restituire" (*giveback*) un pò degli investimenti fatti sulla persona. Lo si può fare in diversi modi. Nel nostro processo vengono elencate sei tipologie di contributi e viene richiesto che almeno due siano soddisfatte. Può trattarsi della disponibilità a fare mentoring verso altri colleghi, oppure possedere certificazioni su prodotti, oppure tenere corsi ai colleghi od ai clienti, oppure accrescere il parco di capitale intellettuale IBM (con brevetti, creazione di programmi ed asset, documenti o altro che sia di valore e che possa essere riutilizzato dai colleghi). Ovviamente sono apprezzate anche le attività fatte per aiutare e sostenere la professione, come per esempio far parte della commissione esaminatrice delle certificazioni oppure aver effettuato uno studio commissionato dai membri della direzione generale. Va sottolineato che molto spesso questi contributi non fanno parte della normale attività lavorativa ma rappresentano veramente un contributo che il candidato produce volontariamente in quanto lo ritiene utile per la comunità.
La parte più cospicua del package è quella dedicata ai profili dei progetti, dove il candidato descrive tre progetti di successo, in cui ha operato con ruolo di "*lead IT Architect*". Ogni profilo, in poche pagine, dovrà contenere la tematica di business affrontata, le sfide architetturali affrontate, le metodologie utilizzate, gli elaborati prodotti e le "*lesson learned*". Queste esperienze progettuali spesso sono utilizzate dai membri della commissione esaminatrice per capire quanto il candidato è pronto ed attento ai livelli qualitativi richiesti per una certificazione di questo livello.
In ultimo, viene presentata la storia professionale del candidato mediante un dettagliato curriculum vitae.

2.3. La commissione esaminatrice (il *certification board*)

La commissione esaminatrice ha il compito di verificare e validare il *package*, intervistare il candidato (mediante tre interviste separate condotte da tre membri della commissione), discutere il caso in commissione e raggiungere una posizione comune. La commissione esaminatrice è presieduta dal leader nazionale della professione (il "*profession leader*", uno per ogni professione) che ha il compito di coordinare tutte le attività e accertarsi che le regole stabilite a livello globale siano rispettate (segnalando eventuali miglioramenti). Il processo è soggetto ad audit ad ulteriore garanzia del rispetto formale delle regole.
La commissione è composta da professionisti già certificati che svolgono il loro lavoro presso i clienti e che danno la loro disponibilità a far parte della commissione per un numero limitato di volte all'anno. Per i membri della commissione questa attività diventa un contributo alla professione. Il fatto che si tratti di persone che operano in progetti concreti presso i clienti, permette di mantenere quel legame con la realtà e quella vitalità che è ritenuta indispensabile per ottenere la certificazione.
Il *package*, una volta completato e approvato dalla linea manageriale, viene rivisto da tre membri della commissione (i revisori) che verificano se quanto contenuto sia in linea con i criteri richiesti per il rilascio della certificazione. In particolare, si fa in modo di avere un revisore appartenente alla stessa disciplina del candidato (per capire a fondo l'esperienza maturata in quella disciplina) e gli altri due revisori appartenenti a discipline differenti (per valutare se la persona è un *IT Architect* completo, capace di ragionare e muoversi anche al di fuori dalla sua specifica disciplina).
Alla fine si tiene quello che viene chiamato "*consensus meeting*", ovvero la riunione al completo della commissione esaminatrice ove i revisori espongono la loro posizione. Ognuno dei membri della commissione può dare il suo contributo in relazione alla conoscenza e alle

esperienze maturate insieme con il candidato. Ovviamente per ogni candidato avranno maggiore titolo le considerazioni dei tre revisori che hanno esaminato il package e condotto le interviste.
La valutazione della commissione verrà comunicata alla linea manageriale ed alla persona, oltre ad essere pubblicata in un bollettino della intranet aziendale consultabile da tutti i dipendenti. Nel caso di parere negativo, al candidato verranno indicati i punti di debolezza e verranno date alcune indicazioni precise per poterlo indirizzare sui requisiti non soddisfatti.
In ultimo un accenno alla *Senior Certification,* quella di livello 3. In questo caso la commissione esaminatrice è di tipo internazionale (a livello europeo), le interviste sono condotte in lingua inglese e il candidato deve dimostrare di avere le competenze tecnologiche, di esperienza e di leadership in grado di svolgere con autonomia la sua attività in ogni nazione. Si precisa che per il livello 2 la commissione è di tipo nazionale ma comunque si richiede che il package sia prodotto in lingua inglese.

3. Aspetti motivazionali

Come è stato già accennato all'inizio di questo articolo, l'intero processo di certificazione ha diversi obiettivi:

- aumentare il valore del servizio ai clienti, garantendo un livello qualitativo di alto profilo e con standard definiti a livello internazionale; in questo modo il mondo IBM può presentarsi in modo coeso e facilitare la comunicazione tra i team;
- dare valore alle persone, aumentandone la visibilità, fornendo loro informazioni dettagliate su come poter concorrere allo sviluppo professionale (tutte le informazioni sono disponibili a tutti nella intranet aziendale), dando l'opportunità di avere un mentor, supportando la normale e giusta aspirazione di carriera, chiarendo quali sono gli step da seguire e dando il controllo al singolo individuo (ovviamente ci vorrà poi anche una condizione di mercato favorevole per completare il quadro);
- portare valore a IBM, in quando aiuterà sicuramente ad aumentare la soddisfazione dei clienti e la motivazione delle persone; al contempo permetterà di avere un aiuto oggettivo nell'identificazione dei possibili talenti in azienda, aumentando le opportunità di utilizzarli al meglio e farli crescere più rapidamente. Sempre dal punto di vista aziendale non meno importante è la possibilità di avere un alto standard interno a livello globale, con valori similari e condivisi in tutto il globo.

4. Benefici e svantaggi

L'ultra decennale esperienza maturata permette di fare delle verifiche regolari sui pro e sui contro di questo approccio organizzativo e di gestione delle risorse professionali a livello aziendale. In generale, in base ad alcuni studi approfonditi effettuati in azienda, si può dire che le persone certificate sono mediamente più richieste dal mercato, anche senza la conoscenza del loro status di certificati e quindi solamente in base alle competenze dimostrate. Inoltre è stato verificato che le persone certificate sono mediamente più motivate, hanno valutazioni migliori e sono più legate all'azienda. Sono anche le persone che sottopongono più richieste di brevetti. Queste stime sono state riscontrate per tutte le professioni prese in esame.

5. Le certificazioni in Italia e nel mondo

Il concetto di certificazione, in uso da molto tempo nel mondo anglosassone, sta entrando nell'uso comune anche in Italia, anche se è ancora un processo in divenire.
Nel mondo dell'informatica sono dapprima comparse le certificazioni di prodotto: quasi ogni produttore di hardware o di software offre percorsi di certificazione per i tecnici che operano con i sui suoi prodotti. Sia IBM che altri fornitori di tecnologie hanno procedure ben definite per conferire certificazioni di prodotto. Ed è una cosa saggia; si tratta di una forma di garanzia che chi è chiamato ad operare su un certo bene o servizio sappia cosa stia facendo.
Un passo "culturale" in avanti si è fatto con l'ECDL (*European Computer Driving Licence*), la patente europea del computer, introdotta dal CEPIS (*Council of European Professional Informatics Societies*) dieci anni fa e che ha conosciuto ottima diffusione anche in Italia.
In fatto di certificazioni professionali approvate

da enti "*super partes*", negli USA, da qualche anno, l'organizzazione "*The Open Group*" (TOG, www.opengroup.org), oltre ad essere un ente di certificazione per alcuni standard aperti come CORBA®, LDAP, POSIX®, UNIX® e altri, si occupa anche della certificazione professionale di *IT Architect* e *IT Specialist*. Il processo utilizzato da "*The Open Group*" è similare a quello adottato da tempo da IBM che è stata una delle prime aziende ad essere "accreditata" a livello globale come ente certificatore per queste professioni. Agli IT Architect ed agli IT Specialist certificati internamente ad IBM viene offerta la possibilità di avere anche la certificazione TOG (*The Open Group*).

La *British Computer Society* (BCS), che opera da tempo principalmente in Gran Bretagna, offre diversi livelli di certificazione professionale riconosciuti nel mondo anglosassone.

In Italia, l'Associazione Italiana di Informatica e Calcolo Automatico (AICA, www.aicanet.net) rappresenta il *Council of European Professionals Informatics Societies* (CEPIS) e fortemente sostiene la cultura della certificazione professionale (a partire dal fortunato ECDL); AICA attualmente sta effettuando un lavoro encomiabile con la certificazione europea EUCIP (*EUropean Certification of Informatics Professionals*), ormai riconosciuta sia dalle principali Istituzioni come il CNEL, il CINI per quanto riguarda l'Università e il MIUR per quanto riguarda la Scuola, sia dalle principali Associazioni di Imprese come Confindustria-ASSINFORM, Confcommercio, ASSINTEL, Confapi per le Piccole e medie imprese e le varie Associazioni di Imprese Artigiane.

I consulenti hanno diversi enti certificatori fra cui, ad esempio, l'Associazione Professionale Italiana dei Consulenti di Direzione e Organizzazione (APCO, http://www.apcoitalia.it/) con la certificazione CMC (*Certified Management Consultant*).

In Italia infine è molto conosciuto il *Project Management Institute* (PMI) per la professione dei capi progetto.

In questo contesto è rimarcabile lo sforzo del CNIPA, che nelle sue linee guide per la Pubblica Amministrazione pone un'enfasi particolare al controllo di qualità anche nel settore dei sistemi informativi e ha recentemente pubblicato un aggiornamento su EUCIP nelle proprie Linee Guida.

6. Conclusioni

IBM è stata sempre all'avanguardia nel favorire l'adozione di standard professionali nell'industria delle tecnologie informatiche. Il perseguimento di un alto livello qualitativo da parte dei professionisti IBM avviene mediante diversi processi interni ed una costante validazione delle competenze e delle esperienze maturate e rappresenta il riconoscimento di una carriera di successo. In questo modo i professionisti IBM sono in grado di concentrarsi sulle attività a più valore aggiunto durante la definizione e la realizzazione del progetto con il cliente.

Tutto questo permette di incrementare il numero di progetti completati con successo, di ridurne il rischio e di aumentare la soddisfazione dei clienti. IBM ha sempre creduto che la professionalità sia uno degli elementi critici fondamentali per il successo sul mercato, soprattutto per un'azienda che fornisce ai clienti non solo prodotti ma anche servizi (ben il 60% del fatturato 2008 è stato costituito da servizi).

Un processo di certificazione apporta notevoli benefici alla qualità dei servizi offerti. Se si ricerca un alto standard qualitativo nell'organizzazione del proprio lavoro così come nella ricerca dei fornitori migliori, un buon punto di partenza è quello di ricercare professionisti con certificazione basata un processo solido. Ci sarà un costo iniziale, ma sarà ampiamente ripagato durante l'attività progettuale con minori problemi e migliori risultati.

I professionisti IBM sono spinti a sviluppare sempre di più le loro capacità mantenendone traccia negli archivi aziendali e questo permette all'azienda di poter facilmente selezionare, per i progetti con i clienti, quei professionisti che hanno le competenze e le esperienze migliori per quel tipo di progetto.

Il mondo dell'IT non solo non fa eccezione, ma al contrario, essendo una realtà in cui ben poco è definito e molto va capito, creato e ottimizzato, è estremamente sensibile a tutto questo.

L'augurio, ovviamente, è che l'apprezzamento delle certificazioni professionali, oltre a quelle di prodotto già presenti in Italia, sia al più presto riconosciuto, ricercato e diffuso presso tutti i clienti, come requisito minimo per ogni fornitura.

Bibliografia

[1] L'intranet della IBM Italia, accessibile a tutti i dipendenti, è molto ricca di tutto quanto serve

per approfondire i temi della propria professione. Esternamente si possono invece accedere.

[2] Le certificazioni EUCIP: www.eucip.com e www.eucip.it

[3] Il programma di certificazione IT Architect proposto da The Open Group: http://www.opengroup.org/itac/

[4] SFIA foundation: http://www.sfia.org.uk

[5] CNIPA: Dizionario dei profili di competenza per le professioni ICT: http://www.cnipa.gov.it/site/it-IT/Attivit%c3%ao/Qualit%c3%ao_delle_forniture_ICT/Manuali/Dizionario_dei_profili_di_competenza_per_le_professioni_ICT/

Walter Pistarini è leader della professione IT Architect in IBM Italia. Assunto in IBM nel 1976, ha ricoperto molti incarichi: Tecnico di manutenzione HW e poi SW, Istruttore, people manager, consulente, principal. Ha creato e guidato per un certo periodo in IBM Italia l'Object Technology Practice. Si è certificato a livello principal in quel contesto. Attualmente si occupa di Professional Development in GBS Italia ed è il Profession Leader per la comunità degli I/T Architects della IBM Italia. Nel 1990 ha pubblicato il libro "Introduzione al Project Management" per Franco Muzzio Editore.
E-mail: pistarini@it.ibm.com

Raffaele Stifani, laureato con lode in Ingegneria, è Executive IT Architect in IBM. Il suo compito è aiutare le aziende ad utilizzare le tecnologie informatiche, creando innovazione, per rispondere velocemente alle mutevoli esigenze del mercato. È riconosciuto, sia in IBM che presso i clienti dove opera, per la sua esperienza nella integrazione di sistemi complessi, nella definizione di architetture d'impresa e nella realizzazione di progetti. È docente di "Impianti di elaborazione" presso la Facoltà di Ingegneria dell'Università di Pavia. È membro del Consiglio Direttivo Centrale di AICA. È autore di articoli e pubblicazioni su riviste specializzate. Ha conseguito il titolo di "Master Certified IT Architect", rilasciato dall'organizzazione "The Open Group".
E-mail: raffaele_stifani@it.ibm.com

GREEN SOFTWARE

Giovanna Sissa

Una riflessione approfondita sulla sostenibilità ambientale dell'ICT richiede attenzione anche alle caratteristiche del software, in quanto corresponsabile delle emissioni di CO_2: i consumi energetici durante il funzionamento di un elaboratore non dipendono solo dalle caratteristiche hardware, ma anche dalla configurazione software e dalla sua efficienza. Alcuni dati offrono spunti di riflessione su come interventi sul software riducano i consumi di energia e su come sia possibile usare più a lungo i PC rispettando l'ambiente, risparmiando e nel contempo proiettandosi nella dimensione del *cloud computing*.

3.2

1. L'IMPATTO DEL SETTORE ICT SULL'AMBIENTE

Il settore ICT produce il 2% delle emissioni globali di CO_2 [1] e contribuisce così al riscaldamento climatico. Nel 2007, l'impronta complessiva del settore ICT è stata di 830 Mt di emissioni - equivalenti di CO_2 (Figura 1), un contributo analogo a quello del settore aereonautico civile.
Computer, dispositivi elettronici e infrastrutture ICT incidono sull'ambiente in ogni fase del loro ciclo di vita: consumando risorse non rinnovabili e inquinando durante i processi di produzione e trasporto, producendo a fine vita rifiuti difficili da trattare e pericolosi (*e-waste*) [2, 3] e consumando molta energia durante il loro uso. Questa tendenza è in forte crescita, almeno nello scenario attuale (BAU - *Business As Usual*) [1].
Gli effetti dell'ICT si possono classificare come di primo, secondo e terzo ordine [3, 4]. I primi sono quelli diretti, derivanti dalla mera esistenza degli elaboratori e comprendono produzione, uso e smaltimento a fine vita. Gli effetti di secondo ordine discendono dalle applicazioni dell'ICT e includono l'ottimizzazione dei processi indotta in altri settori (per esempio, sul traffico), gli effetti di sostituzione (per esempio, la teleconferenza che elimina gli spostamenti) e gli effetti indotti, quando l'ICT crea più domanda in altri settori.
Gli effetti del terzo ordine sono relativi ai cambiamenti sociali che l'ICT induce e sono i più difficili da prevedere e quantificare. Comprendono i cambiamenti nei comportamenti derivanti dalla de-materializzazione, gli effetti di rimbalzo e la crescente dipendenza dall'infrastruttura critica ICT.

1.1. Un effetto di rimbalzo

Ogni PC in uso produce all'incirca una tonnellata di CO_2 l'anno [5]. La richiesta di energia per alimentare server, computer, monitor, apparecchiature di comunicazione, sistemi di raffreddamento per i *data center* è in crescita.
L'efficienza energetica (in termini di prestazioni per watt) dei processori – grazie a tecniche innovative di progettazione, che vanno da aspetti tecnologici a soluzioni che ottimizzano la gestione dinamica dell'architettura

durante l'esecuzione – è migliorata in modo notevole [6], anche se la densità di potenza è a sua volta aumentata.
D'altro canto la legge di Moore - per cui il numero di componenti per microchip raddoppia ogni 18-24 mesi – non corrisponde alla legge economica del settore ICT: a fronte di una crescita esponenziale di prestazioni per chip abbiamo una crescita doppia delle prestazioni rispetto al costo [3, 6]. Sembra un paradosso, ma il costo dell'hardware decresce più in fretta di quanto aumenti la sua miniaturizzazione. Questo fa aumentare la domanda di servizi ICT e il risultato è che gli enormi miglioramenti nell'efficienza energetica non tengono testa all'aumento della richiesta di uso di computer, internet e cellulari; dunque il saldo netto dei consumi energetici del settore ICT è negativo: la domanda totale di energia dell'hardware installato è in crescita [5].

2. GREEN ICT E GREEN SOFTWARE

Il *green computing* prevede che, la progettazione, la costruzione, l'uso e la dismissione di computer, server (e relativi sottosistemi) siano fatte nel rispetto dell'ambiente, tenendo conto dunque dell'intero ciclo di vita del prodotto [7], incluso lo smaltimento e il riciclo. I benefici per l'ambiente discendono dall'uso di materiali ecocompatibili, dall'eliminazione di sostanze pericolose, dalla progettazione orientata all'*upgrade* e alla manutenzione (prevedendo sin dalla fase di design il riuso e riciclo) e, soprattutto, dall'aumento dell'efficienza energetica.
Centrale, ai fini della riduzione delle emissioni di gas serra, è la diminuzione dei consumi di energia.
Il *low power software* - usato da tempo in alcuni settori quali, per esempio, i sistemi *embedded* - si occupa specificamente di progettazione orientata al risparmio energetico, a partire da aspetti di compilazione e programmazione a basso livello (quali per esempio la strutturazione dei dati) per arrivare ai sistemi operativi [27, 28, 29]. La pervasività dell'ICT, presente e futuro, determina un'attenzione alle questioni ambientali in tutti i settori, contrariamente a quanto avveniva prima,

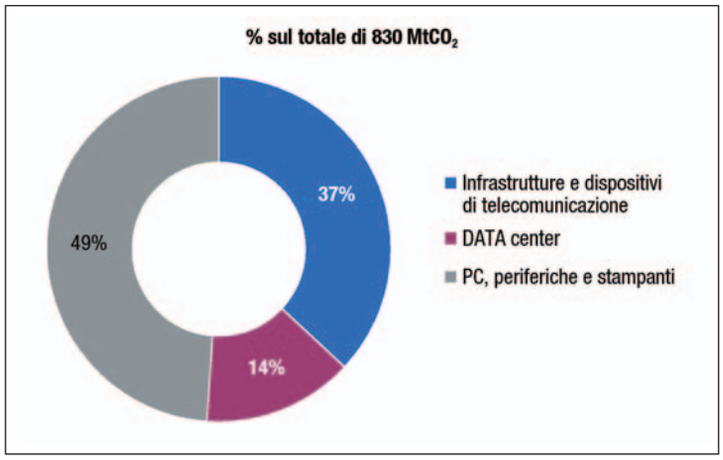


FIGURA 1

Impronta ambientale per settore ICT, 2007 (Fonte: GeSI/The Climate Group: SMART 2020: Enabling the low carbon economy in the information age)

quando solo specifici settori si preoccupavano dei consumi energetici.
L'incremento della richiesta di Internet e applicazioni web ha, per esempio, determinato la rapida crescita dei *data center*. Il numero di server è aumentato di sei volte in dieci anni e ogni server consuma più elettricità dei precedenti modelli [5, 8], poiché molto più potente.
Una ICT davvero "*green*" non può essere ottenuta intervenendo solo sull'hardware, come potrebbe sembrare invece guardando all'offerta di mercato del *green computing*. Il software infatti gioca un ruolo molto importante sull'intero ciclo di vita: sia nella fase d'uso, come vedremo in un'intervista che quantificherà i risparmi ottenibili con interventi sull'efficienza, ma anche nella fine vita dei computer. Il ruolo del software nel *green ICT* non è solo monitoraggio e ottimizzazione di risorse, ma può andare oltre, consentendo ad esempio di soddisfare nuove esigenze con le infrastrutture esistenti. Al termine *green software* mi piace dare qui una connotazione più ampia di quella che si trova talvolta in letteratura: diciamo che è "green" quel software che consente all'ICT di essere più attento all'ambiente. Vediamo come, iniziando dai luoghi in cui si concentrano dati e applicazioni: i *data center*.

2.1. Come rendere più verdi i data center

In un *data center* l'energia complessiva è usata per il 55% dai sistemi di alimentazione e raffreddamento e per il 45% dal carico IT. Nei server il 70% è consumato dai sistemi di

alimentazione, memorie, ventilatori, *planary drive* e solo il 20% del carico dei server è usato per la computazione effettiva, poiché l'80% sono processi *idle* [9].
Per ogni Watt usato effettivamente per il calcolo, il processore consuma 5 W, il server 16 W e il *data center* nel suo insieme 27 W [10].
I metodi per migliorare l'efficienza dei *data center* consistono nell'uso di apparecchiature efficienti sul piano energetico, migliorando l'areazione per ridurre le esigenze di raffreddamento, con una progettazione attenta all'ambiente e investendo in software di gestione energetica.
Il rapporto fra il consumo elettrico globale del *data center* e quello dei server è indicato come PUE (*Power Usage Effectiveness*) [17]. Se il rapporto del consumo del raffreddamento rispetto al consumo effettivo per il server aumenta, il PUE cresce. Il PUE medio di un *data center* è 2.5 ed è considerato buono se arriva ad essere minore di due. L'ottimo sarebbe 1. Alla base dell'efficienza energetica vi sono soluzioni software: virtualizzazione, efficienza, consolidamento, gestione della potenza dei server.

2.2. Virtualizzazione

La virtualizzazione è la prima tecnologia cha ha consentito un approccio al *Green ICT* centrato sul software. L'idea è semplice: un *software host* (un programma di controllo) crea un ambiente simulato (la v*irtual machine*) per il suo software ospite. Il software ospite, che può essere a sua volta un sistema operativo completo, gira come se fosse installato su una piattaforma hardware *stand-alone*. In questo modo ogni server fisico ospita server virtuali multipli e ciò consente ai *data center* di consolidare la loro infrastruttura, usando così meno elettricità e semplificando i *data center* stessi. La virtualizzazione dei server prevede un'applicazione software che divide un server fisico in più ambienti virtuali. Un miglior utilizzo dei server riduce le necessità di spazio fisico dove allocarli e riduce la richiesta di energia di un *data center* [5].

2.3. Il software e l'efficienza energetica

Sono varie le misure che possono migliorare l'efficienza energetica di un *data center*. La catena causa-effetto inizia con le applicazioni software e continua passando per i dispositivi hardware e l'alimentazione, fino alla progettazione dell'edificio e al sistema di raffreddamento. La considerazione chiave è che tali misure sono tanto più efficaci quanto più prossime all'inizio della catena causa-effetto.
Sembra banale, ma è utile ricordare che se un'applicazione non è più necessaria e il server su cui risiede viene spento, viene usata meno potenza, le perdite nel sistema UPS diminuiscono e di conseguenza anche il carico del sistema di raffreddamento.
Il punto di partenza da affrontare dunque per avere un *data center* efficiente sono i dati e gli applicativi [11]. Accade spesso che, nonostante un terzo delle applicazioni non siano più necessarie, esse continuano ad essere attive su un server [11]. Non è trascurabile quantitativamente l'uso di risorse hardware da parte di applicazioni identiche. Nella guida *Energy Efficiency in Data Centres*, BIT KOM [12] si ipotizza che in molte aziende una parte considerevole di dati consistano in foto, video e file MP3 dei dipendenti – ovvero al di fuori dell'attività dell'azienda. Questo significa che applicazioni e dati devono essere tenuti sotto controllo e quando non più necessari, devono essere eliminati.

2.4. De-duplicazione dei Dati

Nel 2006, sono state creati e copiati 161 exabytes (161×10^{12} bytes) di informazioni digitali. L'equivalente di tre milioni di volte le informazioni contenute in tutti i libri mai scritti. L'equivalente di 12 pile di libri, ognuna alta 93 milioni di miglia - la distanza dalla terra al sole. Nel 2010 tale quantità sarà sei volte più grande [13].
L'idea, ampliamente diffusa negli ultimi decenni, che le risorse di storage dei dati abbiano costi trascurabili, ha creato l'abitudine di duplicare e duplicare i dati, senza considerare le ridondanze inutili.
Si indica come *data de-duplication* l'eliminazione dei dati inutilmente ridondanti. Nel processo di de-duplicazione i dati duplicati più volte sono eliminati in modo da lasciare solo una copia dei dati da memorizzare; l'operazione può arrivare a ridurre lo spazio disco per il *back-up* notevolmente. I livelli di ripristino di servizio sono più alti, gli errori da gestire diminuiscono in media e si rendono disponibili più punti di *recovery* sui media per il *recovery* veloce.
La de-duplicazione dei dati riduce inoltre anche la mole di dati da inviare su WAN per *back-up* remoti, replicazione dati e *disaster recovery*.

3. IT MANAGEMENT SOFTWARE

I produttori di hardware continuano a puntare ad una maggior efficienza nei consumi elettrici e nel calore totale prodotto e aumentano la capacità delle tecnologie per server, reti e device. Anche qui il software può giocare un ruolo importante per l'IT management. Nel prendere in considerazione soluzioni "verdi", infatti i responsabili IT si stanno accorgendo che la sostituzione delle attuali risorse hardware con tecnologie più efficienti dal punto di vista energetico comporta spesso costi proibitivi, mentre l'implementazione di software per l'IT management può generare efficienza e risparmi significativi.
Nello studio del 2007 dell'EPA (*Environmental Protection Agency*) [14] sui *data center* si osserva che "le tecnologie esistenti e le strategie di progettazione hanno dimostrato come sia possibile ridurre del 25% il consumo energetico di un server". Anche con gli attuali dispositivi IT, realizzare pratiche di miglior *energy management* e il consolidamento da molti server su uno solo può ridurre i consumi di un *data center* del 20%. Vari studi e iniziative sono volte ad introdurre indicazioni, parametri e metriche di valutazione [16, 17].
Un *ICT green* dunque non richiede necessariamente investimenti sostanziali in nuove tecnologie hardware. Per ottimizzare con la virtualizzazione e il consolidamento, le organizzazioni devono essere consapevoli del corrente stato di utilizzazione e delle dipendenze associate.
La britannica *National Energy Foundation* [18] stima che l'implementazione di power management software abbia la potenzialità di ridurre le emissioni di CO_2 in UK di 140,000 t l'anno mettendo, da remoto, gli apparati in *sleep mode* e attivandoli quando necessario.
L'*IT management software* può fare dunque un miglior uso delle risorse disponibili e non rispondere alla necessità di porre attenzione all'ambiente con la sola acquisizione di nuovo hardware. Il ruolo del software, più in generale, non si limita a questo.

3.1. I risparmi energetici ottenibili con miglioramenti sul software

Nell'intervista presentata nel riquadro a p. 50, l'Ing. Castelli fornisce esempi concreti di casi in cui interventi software sulle architetture dei dati o sulle caratteristiche dei database hanno consentito riduzioni estremamente consistenti nei consumi energetici. Evidenzia anche, in accordo con quanto detto nei paragrafi precedenti, che gli interventi sono tanto più efficaci quanto più si interviene alla base della catena causa-effetto.
La produzione di CO_2 nelle grandi organizzazioni discende anche dalla quantità di semplici operazioni, come le interrogazioni ai database: migliorando l'accesso a una vista dello statino di un dipendente in una organizzazione che ne ha decine di migliaia, la potenza dissipata si riduce di oltre 300 volte. I risparmi ottenibili intervenendo proprio sull'efficienza software sono impressionanti, come i dati forniti sempre dall'Ing. Castelli e relativi all'esperienza di ENI (Tabella 1) dimostrano.

4. RIDUZIONE DEL CONSUMO DEI PC

Abbiamo visto come interventi sul software nei *data center* permettano consistenti risparmi energetici e riduzioni di emissioni. Certamente però la porzione più rilevante degli 830 Mt di CO_2 prodotti complessivamente dal settore ICT (Figura 1) dipende da PC e periferiche. A questo settore dunque va dedicata l'attenzione.
Il numero di PC in uso è così elevato che qualunque intervento atto a ridurre i consumi energetici induce risparmi enormi; come noto il numero di computer venduti nel mondo ha da tempo superato il miliardo e questo importo sta per raddoppiare. Si possono realizzare grandi risparmi con piccoli accorgimenti, come abilitare le funzioni di *power management*. Senza sacrificare alcuna funzionalità è possibile programmare i PC in modo tale che quando non sono usati siano in stato *energy-saving*.
La *US Environmental Protection Agency* (EPA) stima che lo *sleep mode* possa consentire risparmi nei consumi energetici del 60–70% [14, 19].
Questo per quanto riguarda l'aspetto quantitativo dei consumi dell'informatica individuale tradizionale. Ma in prospettiva va tenuto presente anche come sta cambiando il modo di usare il PC da parte sia di utenti *simple user* che utenti *power user*. Sempre di più l'uso di servizi web rende il *browser* lo strumento di lavoro, che funge da interfaccia unica per accedere a servizi, applicazioni o dati

INTERVISTA A GIANLUIGI CASTELLI
V. President - Direzione Information & Communication Technology, ENI

In un recente evento lei ha parlato dei consistenti risparmi energetici realizzati da ENI nei suoi data center a fronte di interventi sul software. Può descriverci come avete identificato le applicazioni interne e quali erano?
L'attenzione si è rivolta alle applicazioni in corso di realizzazione o prossime a essere messe in esercizio, poiché chiaramente è molto meno oneroso intervenire in fase di progettazione/affinamento che non su applicazioni ormai consolidate. Poiché abbiamo recentemente introdotto una fase di performance test per tutte le nuove applicazioni abbiamo potuto raccogliere dati molto interessanti nella fase di pre-produzione. È qui che sono emerse le aree di miglioramento, principalmente legate a cattive architetture di dati o a cattivo uso delle caratteristiche dei DB. Stupefacente è il risultato ottenuto intervenendo solo su queste aree: in molti casi con miglioramento prestazionale di 3 o anche 4 ordini di grandezza. Ciò non toglie che molto di più si potrebbe fare se si riflettesse di più sull'inefficienza intrinseca di certi ambienti di sviluppo o linguaggi che adottano intensivamente l'interpretazione. Ma vorrebbe dire porsi al di fuori del mainstream delle soluzioni commerciali e degli ambienti di programmazione.

Avete quantificato i risparmi nei consumi energetici ottenuti a fronte degli interventi sul software?
È molto difficile quantificarli nella realtà poiché molto dipende dall'intensità di utilizzo che a sua volta dipende da abitudini, errori umani, cicli e ricicli determinati dai processi adottati. In assenza di strumenti di monitoraggio, di cui oggi non disponiamo, anche la determinazione di un "consumo specifico" del software non è possibile. Inoltre spesso il dimensionamento dell'hardware è stato fatto accettando le prestazioni esistenti del software, dunque è spesso sovrabbondante, ma la potenza in eccesso non è immediatamente riutilizzabile. Però, che il software possa rappresentare il vero elemento di efficienza, quindi di riduzione dei consumi, è indubbio. Il software sta all'hardware come il pittore sta al pennello. Ma, almeno per il momento, è più facile, quindi anche più economico, comprare dei pennelli nuovi, piuttosto che dei pittori migliori. Ma le cose sono destinate a cambiare: le soluzioni di *cloud computing* implementate da Google sono l'esempio di come il software debba essere il vero elemento di efficienza. Credo che si potrà diffondere un'attenzione diversa all'efficienza del software. Certo siamo condizionati da molti anni di consuetudine nel sopperire con potenza hardware a software non performante. Ciò ha creato una cultura negativa che sarà difficile sradicare.

Gli interventi sono stati molto onerosi per l'azienda?
No, per le ragioni descritte in precedenza, ovvero la fase progettuale in cui siamo intervenuti. Sono però fermamente convinto che sia necessario investire molto per migliorare la cultura e la competenza tecnica che si è persa nell'ultimo decennio.

Secondo lei quali sono gli aspetti del software su cui è più produttivo intervenire per ridurre i consumi energetici ed in genere l'impatto ambientale?
Non c'è una ricetta che vada bene per tutto. Nel software applicativo di una grande azienda sono gli approcci architetturali e strumentali dove si potrebbe fare molto: multiplexing, multi threading, linguaggi efficienti, DB orizzontali. Ma servirebbe avere un numero sufficiente di pittori. Nel medio lungo periodo il *cloud computing* potrà essere il drive dell'efficienza.

Lei pensa che l'esigenza di riduzione dei consumi energetici nei data center continuerà, anche oltre l'attuale crisi economica?
Penso che quello che è avvenuto e sta avvenendo intorno ai *Data Center*, non possa essere catalogato come "riduzione dei consumi". Si trattava di spreco: i consumi accessori dei Data Center si sono ridotti a circa un decimo nel giro di pochi anni. Nei prossimi anni l'attenzione dovrà spostarsi all'interno del computer (efficienza dei trasformatori, sistemi integrati di raffreddamento, ventilazione forzata ecc. ecc.) mentre per i Data Center la vera sfida sarà quella di passare dal "risparmio" al "recupero" energetico, per la precisione del recupero del calore prodotto dai computer stessi. In altri termini la sfida sarà di arrivare ad un PUE inferiore ad 1.

Pensa che si stiano definendo delle professionalità informatiche orientate a questi aspetti? Quali sono? Sono fra quelle già esistenti o se ne definiranno di nuove?
I progettisti di Data Center hanno sempre pensato che degli stupidissimi computer venissero inseriti in complesse opere di ingegneria. Ovviamente è vero il contrario. Un Data Center ha, al massimo, qualche kW al m^3. Le CPU hanno densità di energia di GW al m^3. Un Data Center è fatto da decine di elementi il software da milioni di istruzioni. Insomma la complessità sta nell'informatica e gli informatici diventeranno i gestori dell'infrastruttura integrata rappresentata dal connubio Software, Hardware, Data Center. Da qui lo sviluppo di qualche nuova professionalità.

remoti. Uno per tutti vale l'esempio della *webmail*. Stiamo passando dall'era in cui era indispensabile installare un applicativo in locale per poterlo utilizzare (o quantomeno installare il *software client* per applicazioni su Internet) ad un modo di lavorare in cui per accedere a servizi web di ogni tipo è sufficiente disporre di un account personale. Sul fronte dell'uso dell'ICT individuale assistiamo a due fenomeni principali. Uno è la diffusione dei

dispositivi mobili, con le relative esigenze di disporre di *device* sempre più performanti. L'altro consiste nella criticità dell'infrastruttura di comunicazione, che nessuna organizzazione può permettersi di eludere. Senza la disponibilità di accesso a banda larga si aprono infatti rischi di divario digitale.
Tenere conto di quest'ultima considerazione, ovvero che siamo e saremo *always on*, può dare utili indicazioni. E un vecchio concetto dell'informatica, quello di *thin client*, può ritornare alla ribalta nel delineare scenari futuri che siano più ecocompatibili.

4.1. Thin-client

"*Thin*" è un aggettivo della lingua inglese che significa magro, sottile. Il "*thin client*" realizza il disaccoppiamento tra l'interfaccia utente e le risorse di computazione; separando le funzioni di Input/Output dal calcolo vero e proprio e "remotizzandole" si rende il sistema a disposizione dell'utente estremamente leggero, essenziale, indipendente dalla piattaforma di calcolo e senza necessità di manutenzione.
Il calcolo e le risorse necessarie ad esso invece sono posizionate sul *terminal server*, quindi saranno centralizzate e facilmente manutenibili dall'amministratore di sistema [2].
Il *thin client* è un computer che dipende per tutte le attività di elaborazione da un *terminal server*. Come *thin client* si possono utilizzare dispositivi *ad hoc* (progettati dunque per fungere da terminali), oppure dei PC con hardware poco potente. I benefici ambientali derivanti dalla sostituzione di PC desktop con *thin client* sono molteplici. I dispositivi hardware *thin client* consentono un risparmio di energia elettrica almeno del 50%. Un dispositivo [20] di nuova generazione ha una potenza di 5-30 W, contro i 150 W di un *desktop tower*.
Il *thin client* consente all'utente di effettuare le operazioni di Input/Output con il server (e ridotte elaborazioni locali). Inoltre consente all'utente di mantenere il necessario e voluto controllo sul suo "*virtual desktop*" e sul suo accesso personale. Rappresentano l'alternativa per compagnie, organizzazioni o istituzioni che abbiano un consistente parco installato di desktop. Non è, come evidente, valida per le organizzazioni che prevedono uso di *laptop* o altri dispositivi mobili, ma l'universo di situazioni in cui si opera su postazioni fissa resta enorme.
Oltre ai vantaggi ambientali, presentano notevoli vantaggi organizzativi, per esempio nelle scuole, come quello di alleggerire enormemente la gestione delle aule informatiche e del parco *desktop* poiché la manutenzione avviene solo sul *terminal server* [2, 20].
Il *Borderstep Institute for Innovation and Sustainability* ha calcolato che nelle scuole tedesche, in uno scenario in cui si preveda di aumentare la consistenza del parco macchine del 6.7% rispetto all'attuale, se invece di usare solo *desktop* ci fossero il 22% di *notebook*, il 12% di *thin client* e il 7% di computer appliance, il risparmio energetico sarebbe di 178 GWh [20].
Molti *thin client* servono solo per far girare un *browser* o un software di *remote desktop*; tutta l'elaborazione avviene sul server. Quello di *thin client* è un concetto software, per questo si possono usare anche dei PC obsoleti come

Statini paga - Valori per dipendente (circa 35.000)

10,53 kWh di energia l'anno
6 kg di CO_2 l'anno
1,2 € costo dell'energia l'anno

Ottimizzazione Software	Prima	Dopo
Accesso a una vista (tempo)	3.5 s	0,01 s
Accesso a una vista (#buffer)	≈223.000	163

Rapporto di potenza < 1/350

Arricchimento dati cliente (1.800.000 clienti)
≈930.000.000 linee di codice da gestire

Ottimizzazione Software	Prima	Dopo
Elapsed time	200 h	55 m

Rapporto di Potenza < 1/218

Ricerca dati cliente

Ottimizzazione Software	Prima	Dopo
Ricerca	6,62 s	0,01 s
Ricerca per codice	14,49 s	0,01 s
Ricerca per nome della ditta	18,26 s	0,05 s
Ricerca per nome utente	2,38 s	0,01 s

1/1449 < Rapporto di Potenza < 1/238

TABELLA 1
Riduzione dei consumi energetici ottenuta mediante interventi sul software

thin client, come vedremo. Ma prima di affrontare il tema del riuso, vediamo se esiste una relazione fra obsolescenza dei computer e software.

4.2. E-waste e SIHO

La pervasività dell'ICT è in aumento e il ciclo di vita di molti dispositivi è sempre più breve. L'UNEP (*United Nations Environment Program*) stima che ci siano fra i 20 e i 50 milioni di tonnellate l'anno di rifiuti elettronici [21] e che la tendenza sia in crescita.
Il ciclo di vita effettivo di un computer non coincide con il ciclo di vita potenziale dell'hardware, che è ben più lungo. Va considerato infatti anche il software: Il ciclo di vita del software non è regolato dalle stesse leggi che regolano il TCO (*Total Cost of Ownership*) dell'hardware, quali rotture di pezzi e sostituzioni di parti [22]. Dipende piuttosto da un numero di fattori correlati fra loro, principalmente la disponibilità del prodotto, del supporto, le funzionalità e le specifiche dell'hardware. Nel caso di software proprietario il ciclo di vita si identifica con il periodo durante il quale il produttore vende e supporta il suo prodotto. In genere i prodotti proprietari sono tolti dai canali di vendita prima che il supporto sia disattivato.
Se il software contribuisse a far raggiungere all'hardware il suo punto ideale di richiesta di potenza in proporzione al carico ci sarebbe un'ottimizzazione nella fase di uso.
Se una nuova versione di software richiede hardware più capace, questo aumenta la richiesta riducendo la durata della fase d'uso dell'hardware. Tale effetto è definito "obsolescenza indotta" (*Software Induced Hardware Obsolescence* - SIHO) [23]. Questo è un altro effetto di rimbalzo [2, 23]. Se il software proprietario è funzionale ad una logica di mercato push per l'hardware, il lato pull della domanda può favorire logiche di riuso.
Si può pensare a come usare più a lungo i computer, ma non secondo la logica tradizionale, propria del *refurbishing* che è mutuata dal settore dell'automobile - dove si attribuisce valore residui ai componenti, visti come potenziali pezzi di ricambio. E neanche vedendoli come dispositivi di seconda mano, destinati a sostituire equivalenti nuovi prodotti presso utenti di categorie deboli – tipicamente nel terzo mondo o paesi in via di sviluppo. Sistemi di seconda mano, che possono essere solo una fotocopia sfuocata di quelli di nuova generazione.
Si tratta invece di ripensare le modalità di riuso, in modo che sia sostenibile anche sul piano sociale, non fornendo dunque agli utenti finali l'impressione di usare computer di seconda mano. Si può fare basandosi proprio sul paradigma più nuovo, il *cloud computing*. Ancora una volta si tratta di software, di conoscenza e skill professionali.

4.3. Il Green Data Cloud

Il *cloud computing* è una delle stelle nascenti del firmamento ICT: la "nuvola" è infatti una metafora di Internet. È uno stile di computazione in cui il software è fornito come un servizio, consentendo all'utente di accedervi senza necessità di specifico *know-how* e soprattutto senza la necessità di avere un controllo diretto sulle infrastrutture di supporto.
Come è accaduto per la virtualizzazione anni orsono, il *cloud computing* sta manifestando delle intrinseche potenzialità *green*.
Avere allocato su server remoti e virtualizzati i servizi, diminuisce la necessità di gestire infrastrutture fisiche e dunque anche l'impatto ambientale delle aziende o organizzazioni che scelgono il modello SaaS (*Software as a Service*), piuttosto che PaaS (*Platform as a Service*) o IaaS (*Infrastructure as a Service*) [30].
Analogamente l'uso di *thin client* nelle organizzazioni riduce la necessità di *desktop* di *front end* fornendo agli utenti la possibilità, a basso impatto ambientale, di accedere al proprio *virtual desktop* [25]. Un esempio di tale soluzione è stato sperimentato in alcune scuole [2, 26] italiane dove, a differenza di quanto indicato nel modello proposto per le scuole tedesche [25] i *thin client* non erano dispositivi specifici di nuova generazione, ma PC obsoleti riusati come *thin client*. Il riuso di PC obsoleti come thin client non conduce certo ai risparmi energetici che gli specifici dispositivi di ultima generazione consentono [18, 11], però questa soluzione permette di prolungare anche del 100% il ciclo di vita di PC, riducendo dunque la quantità di *e-waste* da smaltire [2, 26].
Possiamo dire che il *cloud computing* contri-

buisce a ridurre in vari modi le emissioni di CO_2. Questo ci riporta al concetto base che sono le scelte software soggiacenti a orientare in un'ottica green i sistemi ICT.
Va detto, per completezza che la remotizzazione diminuisce, ma non azzera le emissioni di CO_2. Evidentemente è necessario calcolare anche l'impatto, ancorché di entità minima, degli accessi a Internet. Secondo la stessa google, una query produce 0.2 g di CO_2, ma l'uso di un *laptop* per un'ora ne produce 20 g, l'uso di un PC con monitor per un'ora 75 g . Una copia fisica di quotidiano invece ne produce 173 g, mentre andare da Parigi a Ginevra in TGV 13 kg e 56 kg in aereo (Economy Class) [24].

5. CONCLUSIONI

Ridurre l'impatto dell'ICT richiede skill nuovi, che coniughino competenze ICT e competenze ambientali. Non solo, l'uso di applicazioni ICT richiede anche la consapevolezza dell'impatto ambientale dei comportamenti dei singoli. Questo necessita di accrescere la conoscenza diffusa relativa all'impatto che l'ICT ha sull'ambiente e di supportare lo sviluppo di nuove professionalità, di una nuova cultura tecnologico - informatica attenta all'ambiente e di una generale consapevolezza.
Se gli informatici lavorano in un'ottica verde, operano alla base della catena di causa effetto e, come l'esperienza di Eni indica, possono svolgere un ruolo fondamentale nell'abbattimento delle emissioni.
Gli sviluppatori software devono essere consapevoli che basarsi sulla disponibilità di hardware sempre più economico da parte degli utenti finali è ambientalmente costoso, perché accresce la richiesta permanente di hardware. Andrebbe dunque evitata l'obsolescenza indotta dal software [2, 3, 23, 26].
Sembra pertinente ricordare una indicazione base dell'informatica: cercare di far fronte alla varietà delle esigenze degli utenti con chiarezza concettuale, basandosi su un insieme finito di concetti e principi ben-definiti. Se le funzionalità software fossero presentate con uno stile "assiomatico" invece che con uno stile additivo, ovvero con attenzione alla coerenza complessiva dei sistemi in cui si introducono, la loro complessità sarebbe meglio gestibile. Concludo ricordando una regola aurea dell'informatica: *Avoid unmastered complexity* (E. W. Dijkstra).

Bibliografia

[1] The Climate Group and GeSi (Global e-sustainability Initiative): *Smart 2020: Enabling the low carbon economy in the information age*.

[2] Sissa G.: *Il computer sostenibile*. Ed. Franco Angeli, 2008.

[3] Hilty Lorenz M.: *Environmental impact of ICT- A conceptual framework and some strategic recommendations*. OECD workshop on ICT and environmental challenges, Copenhagen , 20-22 May 2008. http://www.oecd.org/dataoecd/42/13/40833380.pdf

[4] Hilty L.M., et al.: *The precautionary principle in the information society – effects of pervasive computing on health and environment*. Second Revised Edition. Edited by the Swiss Center for Technology Assessment (TA-SWISS), Bern, Switzerland (TA46e/2005) and the Scientific Technology Options Assessment at the European Parliament (STOA 125 EN), 2005.

[5] Murugesan S.: Harnessing Green IT: Principles and Practices. *IT Professional*, Vol. 10, n. 1, 2008, p. 24-33.

[6] Mattern F.: *Die technische Basis für das Internet der Dinge*. In: Fleisch E., Mattern F. (Hrsg.): Das Internet der Dinge. Springer, 2005, p. 39-66.

[7] Williams J., Curtis L.: Green: The New Computing Coat of Arms. *IT Professional*, Vol. 10, n. 1, 2008, p. 12-16.

[8] Restorick T.: *An Inefficient Truth*. Global Action Plan Report, 2007. www.globalactionplan.org.uk/research.aspx

[9] Ebbers M., Galea A. Schaefer M., Tu Duy Khiem M.: *The Green Data Center: Steps for the Journey*. IBM Redpaper ibm.com/redbooks

[10] Renzi F.: *L'innovazione che fa la differenza- la strategia IBM e la tecnologia a supporto della flessibilità d'impresa e dei risparmi energetici*. ForumPA 2008.

[11] Fichter Klaus, Clausen Jens, Eimertenbrink Maik: Borderstep Institute for Innovation and Sustainability, *Energy-Efficient Data Centres Best-Practice Examples from Europe, the USA and Asia*, Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU)

[12] BITKOM: *Energy efficiency in the data centre. A guide for the planning, the modernisation and the operation of data centres*. 2008 www.bitkom.org.

[13] Gantzet al John F.: *IDC – EMC2 White paper, The Expanding Digital Universe*. Marzo 2007.

[14] Report to Congress on Server and Data Center Energy Efficiency - Public Law 109-431 U.S. Environmental Protection Agency, August 2, 2007.

[15] Scheerer Benjamin, Somashekar Sam: *The Role of IT Management Software in the Greening of IT*. Jan. 2009.

[16] European Commission: *Code of Conduct on Data Centres*. 2008. http://re.jrc.ec.europa.eu/energyefficiency/html/standby _initiative_data centres.htm

[17] The Green Grid: *The Green Grid Metrics: Data Centre Infrastructure Efficiency (DCIE)*. Detailed Analysis, 2008, available under: www.thegreengrid.org

[18] http://www.nef.org.uk/news-events/documents/1E_report.pdf

[19] UCSU Environmental Center: *Green computer guide*. http://ecenter.colorado.edu/energy/projects/green_computing.html

[20] Clausen J., Fichter K., Hintemann R.: *How to make computers green? Ressource-efficient innovationsin schools*. Einblicke n. 49/Fruejahr 2009 - Carl Von Ossietzky Universitaet, Oldenburg, p. 48-51.

[21] www.unep.org/Documents.Multilingual/Default.asp?DocumentID=485& ArticleID=5431&l=en

[22] *Paying the price? A Total Cost of Ownership comparison between new and refurbished PCs in the small business, NGO and school in Africa*. By Open Research www.openresearch.co.za Dec. 2004, p. 51.

[23] Hilty L.M.: *Information Technology and Sustainability*. Essays on the Relationship between ICT and Sustainability, 2008.

[24] Kott Benjamin: *Head Green Business Operations Europe, Middle East & Africa, Google: Tackling energy use at Google's data centres*. OECD Conference ICTs, the environment and climate change, Helsingor, 2009.

[25] Fichter Klaus: *ICT Innovation for Behavioural Adaptation – The Case of Thin Client &Server-Based Computing*. High-level OECD Conference ICTs, the Environment and Climate Change, Helsingor 27-28 May 2009.

[26] Sissa G.: *How Public Administration can contribute to a sustainable ICT*. OECD Workshop on ICT and environmental challenges, Copenhagen, 20-22 May 2008. http://www.oecd.org/dataoecd/42/24/40833678.pdf

[27] De Micheli G., Benini L.: *Networks on Chips: Technology and Tools*. San Francisco, Morgan Kaufmann, 2006, p. 408 (Systems on Silicon).

[28] Bogliolo A., Benini L., Lattanzi E., De Micheli G.: *Specification and analysis of power-managed systems*. Proceedings of the IEEE, Vol, 92, 2004, p. 1308-1346.

[29] Bertozzi D., Benini L., De Micheli G.: *Energy-efficient Network-on-chip design - Low power NoC design techniques*. In: E. MACII, Ultra Low Power Electronics and Design, London, Kluwer Academic Publishers, 2004, p. 14-42.

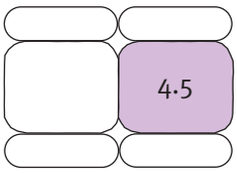
[30] Sun Microsystem: *Take your business at a higher level – A cloud computing primer*. 2009.

GIOVANNA SISSA, laureata in Fisica a Genova nel 1978, si è occupata dapprima di ricerca sulle energie rinnovabili per poi dedicarsi alla società dell'informazione ed all'innovazione tecnologica. È stata consulente di grosse aziende e come esperta indipendente collabora con varie istituzioni nazionali ed internazionali. Ha ideato e diretto l'Osservatorio Tecnologico per le scuole ed è stata membro della Commissione Open Source del Ministro della Funzione Pubblica. Ha scritto libri ed articoli sui temi emergenti dell'ICT. Sulla sostenibilità ambientale dell'ICT ha pubblicato il volume "Il computer sostenibile".
E-mail: giosissa@tin.it

TECNOLOGIE DELLA PRESENZA
CONCETTI E APPLICAZIONI

Giuseppe Riva
Francesco Vatalaro
Gianluca Zaffiro

Condotte per vari decenni, le ricerche sulla "presenza" approdano oggi sul mercato con i primi prodotti e le prime applicazioni. Ma il settore, caratterizzato da numerose sfide dovute anche alla sua natura fortemente interdisciplinare, lascia prevedere altre dirompenti scoperte tecnologiche nel prossimo futuro. L'articolo esamina le prospettive della ricerca sulla presenza e illustra i risultati conseguiti in alcuni settori fra i più vicini alle applicazioni di mercato.



1. INTRODUZIONE

Il poderoso processo di convergenza in corso tra i settori delle tecnologie dell'informazione determinerà sempre più la compenetrazione di informatica, comunicazioni e media. Nei Paesi più avanzati – dall'Estremo Oriente, agli Usa, all'Europa – da questa fusione nasce un settore unificato, sempre più integrato e interdipendente, quello delle ICMT (*Information Communication and Media Technologies*). Questo settore al contempo traina ed è trainato da una decisa progressione verso l'Ultra Larga Banda (ULB), intesa come disponibilità a casa del cliente di collegamenti fissi in fibra ottica, possibilmente in tecnologia FTTH (*Fibre To The Home*), con velocità di trasmissione elevatissime, qualità di servizio (QoS) garantita e collegamenti pienamente simmetrici per consentire non solo il classico *download* da rete internet, ma anche di poter fruire e generare ovunque numerosi servizi di fonia, video e dati, alcuni dei quali oggi ancora non definiti e persino non noti.
Con la futura ULB sono attesi, infatti, livelli di capacità di traffico molto maggiori dei valori odierni (non meno di 50-100 Mbit/s) e resi disponibili sempre e comunque sia in luoghi residenziali che in piena libertà di movimento. Con la penetrazione della fibra ottica nelle zone urbane più popolose si potranno anche creare molte isole *wireless* (grazie al *radio-over-fiber* o a simili tecnologie) di raggio tipicamente minore di 200 m e capacità di picco garantita assai maggiore di quella oggi offerta dai sistemi 3G/HSPA[1] (qualche Mbit/s).
Con l'avvento della ULB è prevedibile anche un balzo in avanti nei servizi, passando dall'odierno Web 2.0 al Web 3.0 degli anni 2010-20 che sarà caratterizzato dall'uso delle tecnologie del Web semantico sia nelle attività umane che in quelle delle macchine. Si avrà anche la definitiva affermazione delle tecnologie *peer-to-peer* (P2P), già oggi diffuse tra i giovani, anche nella platea generale degli utenti, inclusa l'utenza affari; i servizi video bidirezionali saranno molto impiegati sulla base del cosiddetto paradigma del *prosu-*

[1] HSPA: *High Speed Packet Access*.

mer, in cui ogni utente potrà essere al contempo sia consumatore che produttore di contenuti multimediali e la rete, in virtù delle tecnologie di *cloud computing*, offrirà servizi di memorizzazione ed elaborazione in tempo reale tanto efficienti e a basso costo da potersi avere terminali portatili semplici ed economici pur ottenendo un livello di servizio oggi inimmaginabile.
Prevedendo che entro 5 – 10 anni i fenomeni della convergenza dei servizi e la disponibilità di ultra larga banda producano i benefici attesi sull'utenza sia residenziale che *business* in Europa, specialmente nelle aree urbane, potranno allora diffondersi le nuove tecnologie fameliche di banda (*bandwidth hungry*): tutto ciò trascinerà le tecnologie della "*presenza*" che faranno finalmente il salto dai laboratori di ricerca al mondo delle applicazioni sul mercato. Questo articolo ha lo scopo di descrivere lo stato dell'arte del settore tecnologico della presenza, esaminando le prospettive che apre in tutte le attività umane verso forme sempre più avanzate di interazione digitalmente mediata.
Il seguito dell'articolo è così organizzato. Nel paragrafo 2 si fornisce una panoramica delle tecnologie ICMT previste per gli anni 2010, con particolare riferimento a quelle della presenza e di realtà aumentata, inquadrandole nella cornice dei più generali trend tecnologici. Nel paragrafo 3 viene esaminato in dettaglio il concetto di presenza esaminato secondo i tre punti di vista più significativi: la dimensione tecnologica, quella cognitiva e, infine, quella comunicativa. Segue il paragrafo 4 che prima prende in rassegna le discipline alla base dello sviluppo delle tecnologie della presenza e poi descrive le aree applicative e di mercato, esaminando l'attuale livello di sviluppo. Infine, prima di riassumere le principali conclusioni, nel paragrafo 5 si riportano alcuni casi di successo di applicazione delle tecnologie della presenza negli ambiti della medicina e delle telecomunicazioni, mettendo anche in luce alcune delle sfide tecnologiche in corso.

2. LE TECNOLOGIE ICMT PER GLI ANNI 2010

È consuetudine condurre esercizi di predizione tecnologica (*forecasting*) per individuare le direzioni in cui è più promettente indirizzare gli investimenti. Annualmente lo stato dello sviluppo delle tecnologie ICMT è rappresentato tramite la famosa curva di Gartner delle tecnologie (*Gartner's Hype Cycle*) che è una rappresentazione grafica del livello di maturità delle tecnologie che, di anno in anno, ne consente di seguire lo sviluppo fino all'adozione sul mercato. Osserviamo per esempio l'ultima "fotografia" fatta dal Gartner Group nel 2008 (Figura 1) che mostra chiaramente come alcune delle tecnologie necessarie a sostenere lo sviluppo sul mercato delle tecnologie della presenza stiano rapidamente avanzando.
Affinché si possa creare l'ecosistema tecnologico che trascina anche le tecnologie della presenza occorre che maturino in parallelo alcune tecnologie oggi previste per il decennio 2010-2020; fra queste, sulla base sia del rapporto Gartner 2008 che di altri esercizi di predizione, richiamiamo le seguenti:
- *Context Delivery Architecture* (CoDA): un'architettura, prevista matura intorno al 2018, in grado di fornire funzionalità di calcolo consapevoli del contesto (*context-aware*) e servizi arricchiti dalle informazioni di contesto dell'utente (luogo, preferenze, identità ecc.) erogati in modalità naturale e in ogni situazione operativa.
- *Augmented Reality* (AR): classe di tecnologie, previste intorno al 2020, con lo scopo di sovrapporre alla realtà percepita dal soggetto immagini ed etichette digitali allo scopo di amplificare e arricchire la percezione dell'utente da oggetti virtuali che possono fornire informazioni aggiuntive sull'ambiente reale.
- *Tecnologie per le interfacce naturali*: queste tecnologie, previste intorno al 2015, hanno lo scopo di abbattere le classiche barriere che si incontrano nella fruizione dei servizi ICMT, assicurando l'interazione con la rete attraverso gesti, espressioni, movimenti, proprio come nell'interazione quotidiana con il mondo reale.
- *Green computing e virtualizzazione*: questo nuovo paradigma informatico degli anni 2010-2013 con finalità di natura sistemica intende sostituire per quanto possibile le tecnologie classiche con le tecnologie ICMT per aumentare la soddisfazione degli utenti e la qualità della vita anche attraverso la virtua-

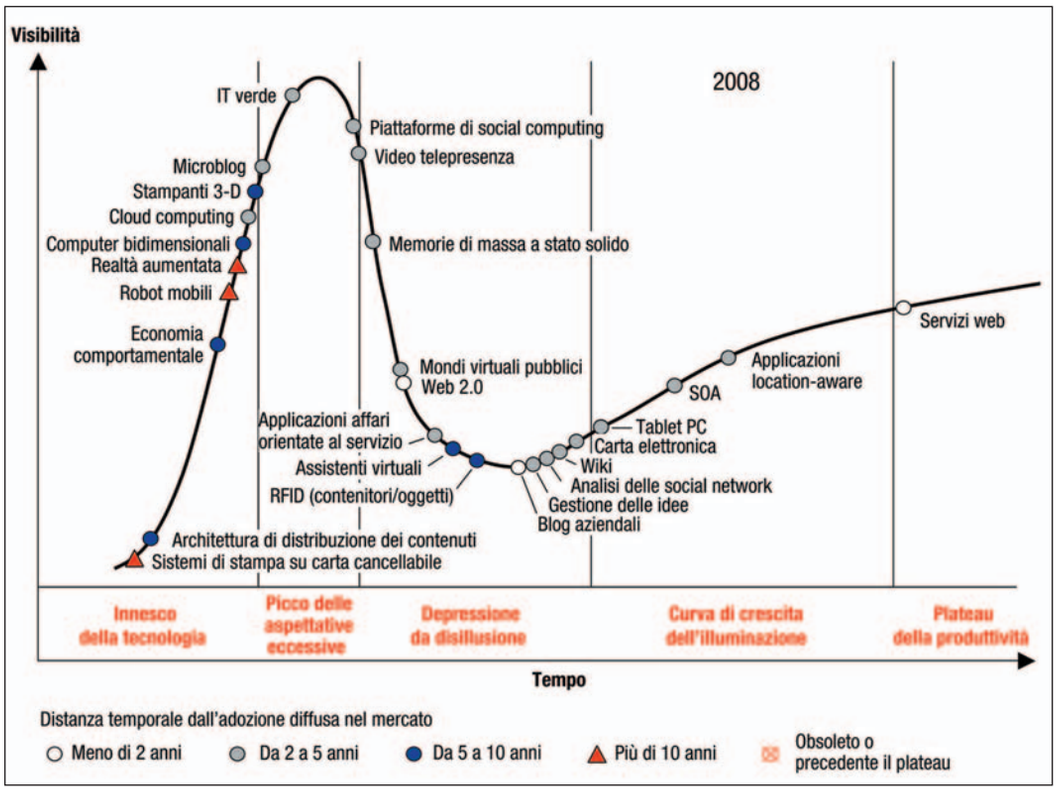


FIGURA 1
La curva di Gartner delle tecnologie (Fonte: Gartner Group, 2008)

lizzazione delle risorse informatiche e il telelavoro.

❑ *Cloud computing*: intorno al 2010-2015 potrà svolgere il ruolo di tecnologia abilitante di numerosi servizi futuri con la distribuzione di risorse informatiche via internet, consentendo di remotizzare i servizi di calcolo e di svolgere localmente solo poche semplici operazioni, pur ottenendo risultati di elaborazioni che richiederebbero la disponibilità locale di un supercomputer.

3. CONCETTO DI PRESENZA: DEFINIZIONE E DISCIPLINE COINVOLTE

Anche senza invocare la fantascienza, fin dagli albori il cinema ha sempre sfidato la nostra immaginazione intorno al concetto di presenza, nel mondo reale, in quelli virtuali e nella percezione del trasporto del "Sé" avanti e indietro tra questi mondi abbattendo le barriere spaziali e temporali (Figura 2).
In generale, la *Scienza della presenza* studia come la mente umana costruisce un modello della realtà e di sé analizzando e integrando dati sensoriali e di interazione. Questa Scienza appartiene a una classe più ampia di campi di ricerca che studiano come i sistemi cognitivi costruiscano modelli del proprio ambiente attraverso l'interazione con esso. Secondo un'interpretazione condivisa, la mente è un sistema atto alla continua generazione di modelli e la presenza è lo strumento che le consente di creare un modello della realtà attraverso l'interazione e la manipolazione sensoriale. Se si verifica che il "Sé" ha la sensazione di essere immerso in una situazione (*being there*) significa che la mente ha attivato quel particolare meccanismo cognitivo-sensoriale detto, appunto, presenza.

3.1. Il concetto di presenza

Il termine "*presenza*" si diffonde nella comunità scientifica a partire dal 1992, quando gli studiosi americani Sheridan e Furness lo utilizzano come titolo di una nuova rivista scientifica dedicata allo studio dei sistemi di realtà virtuale e di teleoperazione: "*Presence, Teleoperators and Virtual Environments*". Nel primo numero Sheridan, descrive la presenza partendo dalle esperienze sperimentate dagli utilizzatori dei teleoperatori. Un teleoperatore non è un robot in senso stretto. Pur condivi-

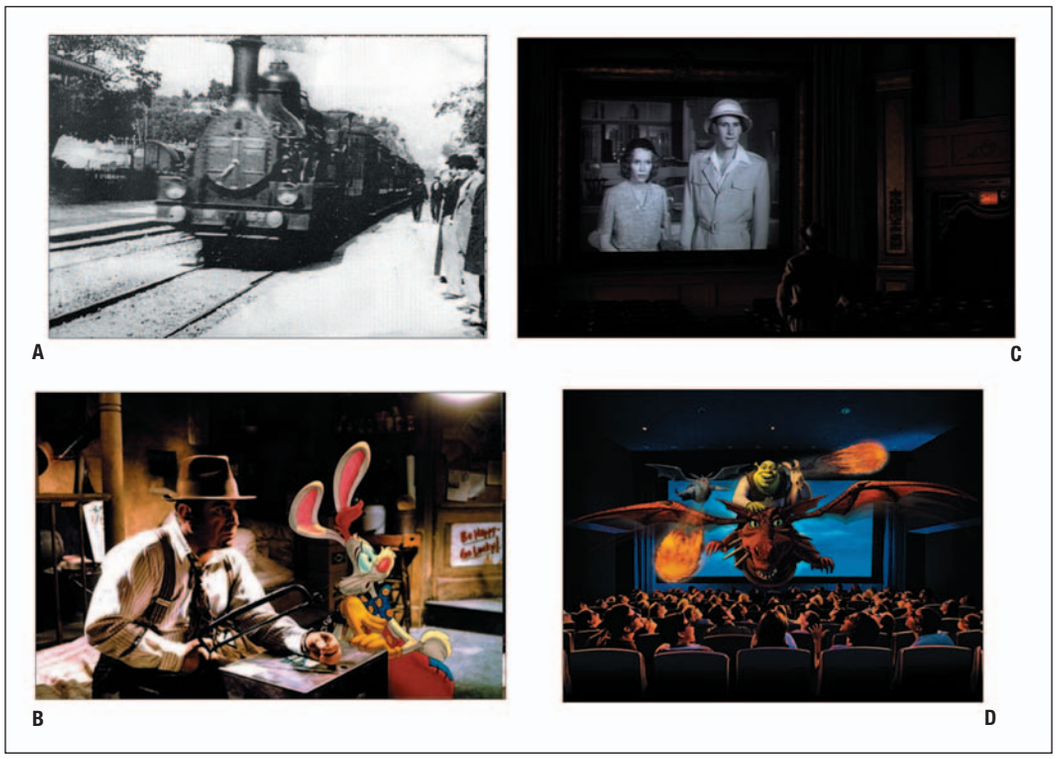


FIGURA 2

***Il potere evocativo ed anticipatorio del cinema*. A - Fratelli Lumière**, *L'Arrivée d'un train à La Ciotat (1895).* L'abbattimento della barriera spaziale nel mondo reale, alias "il Reale (contemporaneo) irrompe nel Reale"*;* **B - R. Zemeckis**, *Who framed Roger Rabbit (1988).* L'abbattimento della barriera tra i mondi, alias "il Reale irrompe nel Virtuale"; **C - W. Allen**, *The Purple Rose of Cairo (1985).* L'abbattimento della barriera spazio-temporale nel mondo reale, alias "il Reale (non contemporaneo) irrompe nel Reale"*;* **D - S.J. Smith**, *Shrek 3-D, IMAX edition (2003).* L'abbattimento della barriera tra i mondi, alias "il Virtuale irrompe nel Reale"

dendone gli equipaggiamenti *hardware* – bracci meccanici, motore, cingoli ecc. – si differenzia nelle modalità di funzionamento. Infatti, le sue operazioni non sono dirette dal processore di un computer ma dalle capacità percettive e motorie di un operatore umano, che lo controlla a distanza [1]. Secondo Sheridan [2] a caratterizzare un teleoperatore sono i seguenti elementi: un sistema di sensori di posizione collegati all'operatore umano, un braccio e una mano meccanica controllati a distanza dall'operatore umano, un canale di comunicazione tra l'operatore e i diversi strumenti tecnologici. Più in generale il termine teleoperazione si riferisce al controllo umano, diretto e continuo, sulle operazioni svolte dal teleoperatore. Con il termine *tele-presenza* si intende invece la sensazione dell'operatore umano di sentirsi presente nel luogo remoto in cui si trova il teleoperatore.

A caratterizzare l'analisi di Sheridan è proprio questo punto: gli utenti dei teleoperatori, dopo qualche minuto di utilizzo, smettevano di sentirsi nella loro postazione di lavoro ma invece si sentivano presenti nello spazio remoto del teleoperatore. Gli utenti della realtà virtuale sperimentano una simile sensazione di *presenza*: bastano pochi minuti di interazione nell'ambiente virtuale per sentirsi immersi nell'ambiente tridimensionale generato dal computer.

Come spiegare la sensazione di presenza sperimentata durante l'utilizzo dei teleoperatori? In relazione alle finalità scientifiche che ci si prefigge e secondo il punto di vista espresso da ciascuna disciplina che la studia, esistono tre principali definizioni operative di presenza, ciascuna caratterizzata da una specifica focalizzazione: tecnologica, cognitiva o comunicativa. Cerchiamo perciò di analizzare meglio queste tre definizioni e i principali aspetti che le caratterizzano.

3.2. Presenza e tecnologia: l'illusione di non mediazione

L'informatica o, più propriamente, le discipline dell'informazione, danno una definizione pragmatica di *presenza* attraverso lo sviluppo di tecnologie idonee a generare e valorizzare le esperienze cognitive dell' "esserci", senza che ciò sia effettivamente "vero" in senso fisico. Inoltre si occupano dello studio dei problemi di misurabilità dell'*effetto presenza* attraverso l'analisi della risposta umana a una situazione potenzialmente reale in termini fisiologici, comportamentali, emotivi, cognitivi e soggettivi.

Più precisamente, a caratterizzare la definizione tecnologica di presenza è il modo in cui è concepito il rapporto tra "realtà" e "virtualità", tra spazio oggettivo e spazio simulato: alla realtà, intesa come un insieme di oggetti collocati fuori dalla mente e dotati di caratteristiche spaziali ben definite, si contrappone l'ambiente virtuale, inteso come simulazione degli oggetti reali.

In questo contesto, il concetto di presenza ha un ruolo funzionale: maggiore è il *livello di presenza*, maggiore è la facilità con cui il soggetto riesce ad operare nell'ambiente virtuale. L'idea di base è quindi ridurre la complessità della tecnologia, consentendo all'operatore umano di interagire con l'ambiente remoto come se fosse presente in esso.

A partire da questo punto di vista Lombard e Ditton [3] hanno elaborato la definizione di presenza che oggi è maggiormente condivisa all'interno del mondo delle tecnologie: l'illusione di non sperimentare una mediazione della tecnologia. In pratica, secondo questo approccio, essere presenti all'interno di un medium significa provare un'esperienza in cui il soggetto non è consapevole del ruolo della tecnologia nel supportarne l'azione e/o la comunicazione. Ciò significa che, meno il soggetto è consapevole della mediazione operata dal medium durante la propria azione, maggiore è il livello di presenza.

La principale conseguenza di questa impostazione è la seguente: un ambiente virtuale produce più o meno presenza a seconda della fedeltà, in particolare grafica, con cui riesce a ricreare gli oggetti "reali".

L'esperienza degli sviluppatori di ambienti virtuali sembra però indicare una direzione differente: non è il realismo grafico ma la capacità d'interazione l'elemento chiave per indurre più *presenza*. Più precisamente, la sensazione di presenza dipende in modo significativo dal livello di interattività che gli attori sperimentano nell'ambiente, sia "reale" che simulato. In particolare, un ambiente virtuale offre un elevato livello di presenza quando l'utente è in grado di navigare, scegliere, spostare e muovere gli oggetti intuitivamente. In questo senso l'attenzione del progettista si sposta dalla qualità delle immagini alla libertà di movimento, dalla perfezione grafica del sistema all'azione dell'attore nell'ambiente.

Questo dato ha spinto numerosi studiosi a cercare di arrivare a una diversa formulazione teorica del concetto di presenza. In particolare ad una definizione di presenza che non fosse legata al concetto di medium: noi non siamo presenti solo in un ambiente virtuale. Questo è uno degli obiettivi del programma di ricerca della Comunità Europea *Future and Emerging Technologies* che ha attivato a partire dal 2002 una linea di ricerca specifica su questi temi, attualmente intitolata *Human-Computer Confluence*.

3.3. Presenza e cognizione: la capacità di mettere in atto le proprie intenzioni

Le scienze cognitive hanno recentemente iniziato a descrivere la presenza come un meccanismo selettivo e adattativo, che permette al Sé di definire il confine dell'azione mediante la distinzione tra "interno" ed "esterno" in un flusso sensoriale. In pratica la presenza può essere considerata la sensazione di "essere" all'interno di un ambiente, reale o virtuale, risultato della capacità di mettere in atto nell'ambiente le proprie intenzioni attraverso le opportunità (*affordance*) che questo offre [1].

Cerchiamo di spiegare meglio questa definizione con un semplice esempio. Supponiamo di volere inviare un SMS usando un nuovo cellulare ancora mai usato. Se l'interfaccia del nuovo cellulare corrisponde a quella del precedente – ha un tasto con la busta che consente l'accesso diretto al menù d'invio degli SMS (*affordance*) – ci si può concentrare sull'intenzione – che cosa scrivere – senza fare attenzione alle azioni da compiere e si è "presenti" nell'interfaccia del cellulare perché in grado di sfruttarne le *affordance*. Quando ciò

avviene l'interfaccia diventa trasparente, ossia non richiede attenzione per realizzare con successo l'azione. In pratica, un'interfaccia trasparente è un'interfaccia efficace, cioè in grado di incanalare con successo l'azione del soggetto verso la sua intenzione senza alcuna coercizione. Il punto d'arrivo è l'idea che, attraverso l'apprendimento dell'uso di un medium, il soggetto allarghi in maniera trasparente i confini del Sé, diventando "presente" nel medium che sta usando.
Se, però, l'interfaccia è diversa e non prevede un tasto specifico per gli SMS, prima di scrivere ci si dovrà soffermare (*break in presence*) a capire dove trovare il menù giusto, cioè a cercare altre *affordance*. In altri termini, ogni interruzione nel *senso di presenza* segnala che l'interfaccia è opaca: richiede attenzione per realizzare con successo l'intenzione. Quando questo avviene, è possibile interagire con l'interfaccia solo dedicando risorse cognitive in modo consapevole.
In quest'ottica, uno degli obiettivi principali dell'usabilità è l'identificazione dei processi e degli strumenti per attivare il massimo livello di presenza nell'interazione con un artefatto, in modo da favorire il raggiungimento delle intenzioni dell'utente.

3.4. Presenza e comunicazione: esserci con un altro

Come abbiamo appena visto, la visione cognitiva della presenza ha come fuoco l'intenzione del soggetto e la sua capacità di metterla in atto nell'azione, compresa quella comunicativa. Tuttavia l'efficacia della comunicazione richiede anche la "*presenza sociale*", la sensazione di "essere con altri Sé" all'interno di un ambiente reale o virtuale, risultato della capacità di riconoscere nell'ambiente le intenzioni degli altri [1]. In pratica, indipendentemente dalla *presenza* fisica o virtuale dell'interlocutore, l'"altro Sé" è presente nell'atto comunicativo solo nel momento in cui si è in grado di riconoscerne le intenzioni.
Il filosofo Paul Grice ha sottolineato come nella regolazione del processo comunicativo abbia un ruolo centrale il "principio di cooperazione". In base a questo principio tacito, i soggetti comunicanti si sentono, per così dire, obbligati a dare un loro contributo affinché l'atto comunicativo in cui sono inseriti possa continuare. Tuttavia, se io non sono in grado di comprendere l'intenzione dell'altro, non sarò in grado di cooperare con lui nel proseguire l'interazione.
Come indicato dalle scienze cognitive, la capacità di comprendere le intenzioni dell'altro è parte di una capacità cognitiva e affettiva complessa, nota come "Teoria della mente": la capacità di comprendere la mente propria e quella altrui in relazione al comportamento manifesto. È grazie ad essa che è possibile la comunicazione: per comunicare devo essere in grado di cogliere la *presenza* dell'altro, comprendere la sua volontà di iniziare un atto comunicativo e identificare le intenzioni che esprime nel processo comunicativo.
In quest'ottica, uno degli obiettivi principali nella creazione di un medium è l'identificazione dei processi e degli strumenti per attivare il massimo livello di *presenza sociale* nell'interazione attraverso di esso, in modo da favorire la comprensione delle intenzioni reciproche dei soggetti interagenti. Un esempio di questo approccio viene offerto dal progetto europeo PASION [4] che ha cercato di identificare strumenti, come i biosensori, e indici, come i parametri di centralità di un utente, in grado di aumentare il senso di presenza sociale sperimentato nei nuovi media.

4. TECNOLOGIE E MERCATI DELLA PRESENZA

4.1. Le tecnologie della presenza

Le tecnologie della presenza hanno anche la capacità di consentire alle persone di agire liberamente e di interagire in ambienti virtuali altamente realistici con gli occhi, le orecchie e le mani. L'espressione chiave qui è "interazione multimodale": i sensi umani sono integrati in un'unica esperienza che consente l'immersione completa.
La maggior parte degli attuali sistemi accoglie l'utente come un semplice osservatore passivo. Ogni volta che l'interazione con il mondo virtuale risulta inevitabile, come nel caso dei giochi via computer, l'azione umana è limitata da dispositivi di base compromettendo in maniera significativa la sensazione di "essere lì".
La *Scienza dell'immersività* si occupa di cambiare radicalmente questa condizione molto restrittiva. Gli utenti immersi negli ambienti

virtuali sono in grado di manipolare oggetti di varie forme, dimensioni e tessuti, nonché di interagire con altri utenti per mezzo del contatto fisico, anche eseguendo operazioni congiunte su oggetti virtuali.
Al fine di realizzare questo nuovo livello di *presenza immersiva*, un importante obiettivo della Scienza dell'immersività consiste nell'indagine della dimensione tattile, al fine di recuperare il ritardo con i notevoli progressi compiuti nel campo dei dispositivi sia visuali che uditivi. Lo sviluppo di nuove tecniche di trattamento del segnale dovrà finalmente portare alla sintesi di tutte le modalità sensoriali in una singola percezione, tale da permettere il pieno riscontro multimodale nei previsti scenari virtuali.
Le tecnologie che abilitano il senso di presenza coprono una vasta gamma di dispositivi e sistemi, tra cui la parte più importante è l'interfaccia uomo-macchina. Gli agenti sia umani che sintetici hanno attuatori e sensori e sono consentite le interfacce dirette tra loro.
Nel riquadro sono elencate alcune tra le più significative tecnologie abilitanti della presenza.

Esempi di tecnologie della presenza

- **Riconoscimento vocale**: sistemi di interpretazione della voce e traduzione in testo o comandi (dettatura, controllo di interfaccia, comandi di navigazione).
- **Riconoscimento della scrittura**: sistemi basati su analisi delle forme per convertire in tempo reale la scrittura a mano nel corrispondente testo alfanumerico o in comandi.
- **Tessuti sensorizzati**: materiali speciali e tessuti dotati di sensori che trasformano l'abbigliamento in interfacce di input.
- **Tracciamento dello sguardo**: sistemi per determinare il punto di attenzione visivo e il suo spostamento (in genere attraverso l'uso di telecamere).
- **Riconoscimento dello stato emotivo**: analisi di segnali (vocali ecc.) atti a riconoscere lo stato emotivo di un soggetto.
- **Riconoscimento dei gesti**: sistemi per determinare la posizione e lo spostamento di mano, braccio, testa o corpo attraverso l'uso di una telecamera o di un dispositivo indossabile dotato di sensori.
- **Interfacce mente-calcolatore**: tecnologia già disponibile commercialmente in grado di interpretare le onde cerebrali generate volontariamente da un utente come, per esempio, comandi per un calcolatore.
- **Microproiettori**: piccoli dispositivi per la proiezione spesso basati sulla diffrazione della luce (laser) e associati a dispositivi portatili.
- **Schermi LCD tridimensionali**: tecnologie che permettono la visualizzazione di immagini tridimensionali (3-D), senza dover indossare lenti speciali.
- **Visori indossabili**: occhiali o lenti su casco con tecnologie integrate per proiezione di immagini.
- **Interfacce tattili e cinestetiche**: interfacce, come per esempio guanti o *joystick*, atte a produrre un *feedback* al tocco o alla pressione, per fornire la sensazione di tensione o movimento in muscoli, tendini e articolazioni.
- **Telepresenza video**: sistemi di videocomunicazione immersiva studiati per creare la sensazione che tutti i partecipanti alla conferenza si trovino nella stessa stanza.
- **Teleoperazione**: controllo a distanza di robot o di altri dispositivi mobili, in cui l'operatore assume il punto di vista del dispositivo e attraverso un'interfaccia cinestetica percepisce i *feedback* legati alla manipolazione.
- **Realtà aumentata o mista**: tecniche atte ad arricchire la visione del mondo reale con informazioni supplementari e virtuali, di solito per mezzo di un *display* indossabile che sovrappone testo o grafica alla visione naturale.
- **Realtà virtuale**: tecniche atte a fornire un ambiente sintetico 3-D generato da calcolatore in cui l'utente si immerge e l'ambiente interagisce in modo naturale al movimento dell'utente attraverso il tracciamento della testa (si utilizzano visori indossabili o stanze stereoscopiche dette CAVE - *Computer Automatic Virtual Environment*).
- **Ambienti e mondi virtuali**: piattaforme *online* in cui i partecipanti, attraverso la rappresentazione di Sé tramite *avatar*, si proiettano in spazi virtuali 3-D generati nel calcolatore.
- **Ricerca in linguaggio naturale**: tecnologia che analizza le domande poste nel corso di normali conversazioni umane e fornisce risposte in base alle parole di senso comune, in luogo di singole parole chiave.
- **Agenti virtuali e chatbot**: entità spesso antropomorfe generate dal calcolatore, dotate di capacità di interazione sociale con gli utenti, in genere in supporto all'esecuzione di compiti.
- **Traduzione automatica**: tecnologia per tradurre automaticamente il testo di una conversazione o un discorso da una lingua ad un'altra.
- **Intelligenza artificiale (IA)**: disciplina per lo studio e la progettazione di agenti intelligenti, ovvero entità in grado di relazionarsi autonomamente con l'ambiente e agire per massimizzare le probabilità di successo di un dato compito.
- **Intelligenza artificiale sociale**: applicazione della IA agli aspetti di interazione sociale, alla codifica e decodifica di segnali non-verbali e alla gestione del dialogo per gli agenti antropomorfi.

4.2. I settori applicativi delle tecnologie della presenza

I settori tecnologici della presenza possono essere classificati in otto categorie e, per ciascuna di esse, le principali applicazioni vengono discusse nel seguito.

4.2.1. Architettura e costruzioni edili

Sono disponibili applicazioni (*architectural walkthrough*) che consentono, attraverso l'uso del calcolatore e di un sistema interattivo, di simulare un percorso all'interno di uno spazio architettonico 3-D per aiutare a verificare l'impatto di un progetto urbano proposto. Un'altra applicazione è il modello di produzione 4-D, che consente di verificare su schermo tridimensionale come un progetto edilizio evolve sul piano temporale, permettendo di rivedere la pianificazione o di verificarne lo stato di avanzamento.

4.2.2. Didattica e formazione

In questo ambito le tecnologie della presenza consentono di sviluppare strumenti di apprendimento e addestramento altamente immersivi (per esempio, attraverso interfacce di simulazione e di realtà virtuale, e giochi di ruolo) che consentono ai discenti di provare ciò che in un'esperienza reale sarebbe difficile da gestire e generare. D'altronde è scientificamente provato che le reazioni psicofisiche e comportamentali indotte dall'esperienza virtuale nel soggetto sono confrontabili con quelle che si avrebbero nell'ambiente reale corrispondente, nonostante la consapevolezza delle condizioni artificiali: pertanto, applicazioni come le esercitazioni in caso di incendio o in altre situazioni di pericolo assumono notevole valore. Infine, le tecnologie della presenza si rivelano utili nell'educazione scolastica per visualizzare in 3-D elementi non visibili perché inaccessibili, troppo piccoli o troppo grandi.

4.2.3. Intrattenimento

L'intrattenimento trae giovamento dalle tecnologie della presenza rivoluzionando il modo di fare e vedere un film aggiungendo interattività e immersività e offrendo l'esperienza di incontro con i personaggi di fantasia in un ambiente simulato. Un altro vantaggio delle tecnologie della presenza in questo ambito è abilitare le persone ad interagire con il contesto di gioco e a comportarsi come se fossero realmente al suo interno. Inoltre va registrata la grande diffusione dell'intrattenimento con ambienti simulati al calcolatore, i cosiddetti mondi virtuali, costruiti per essere abitati e vissuti in modo interattivo dagli utenti attraverso i loro avatar. Esempi sono Second Life, There.com, o Club Penguin. Infine, sono oggi applicate a videogiochi e giocattoli le nuove interfacce mente-calcolatore, che consentono la comunicazione diretta attraverso la lettura e l'interpretazione dell'elettroencefalogramma.

4.2.4. Manifatturiero e design

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 migliorare la sicurezza e l'efficienza di un posto di lavoro usando strumenti di ergonomia virtuale. La visualizzazione di dati complessi, come ad esempio giacimenti petroliferi profondi, è anche un altro modo efficace di contenere i costi industriali.

4.2.5. Marketing

Con le tecnologie della presenza un prodotto, anche se non ancora realizzato, può essere visto e manipolato con un'animazione 3-D a 360°. È anche possibile indossare abiti in camerini virtuali, per farsi un'idea del prodotto prima di acquistarlo e decidere come personalizzarlo sulla base delle proprie preferenze, realizzando modifiche sul modello in *real-time*. I pubblicitari possono adattare la pubblicità a uno specifico contesto usando la virtualizzazione di immagini. Infine nei mondi virtuali gli avatar di persone reali possono vedere e acquistare prodotti reali (o virtuali).

4.2.6. Medicina e psicologia

La medicina e la psicologia traggono beneficio dalle tecnologie della presenza in diversi modi: dalla visualizzazione "aumentata" che aiuta il chirurgo durante un'operazione, alle nuove opportunità che derivano dall'impiego di ambienti immersivi nel trattamento di fobie e altre malattie mentali e nella formazione con l'uso di simulatori di parti anatomiche e simulatori di strumenti medicali.

4.2.7. Militare

Le tecnologie della presenza offrono in questo ambito i vantaggi competitivi derivabili dall'aumento delle capacità umane in azioni di guerra. Adottare visori con sistemi a realtà aumentata aiuta a fornire informazioni aggiuntive ai soldati. Inoltre, le forze militari possono effettuare addestramenti simulati di impiego di armi e veicoli militari in condizioni controllate e di sicurezza, e impiegare ambienti virtuali e la tecnologia dei videogiochi per l'addestramento tattico e per il reclutamento.

4.2.8. Telecomunicazioni

La telepresenza 2-D o 3-D è un esempio già oggi evidente di come le tecnologie della presenza aiutino a ricreare l'illusione di comunicare come se ci si trovasse riuniti nello stesso luogo. Queste tecnologie ricreano e potenziano l'esperienza sociale di comunicazione e collaborazione abbattendo la barriera della distanza: 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 presenza sociale (co-presenza) [5, 6], aggiungendo alle comunicazioni mediate tradizionali gli elementi informativi non-verbali [7], gli aspetti emozionali e le informazioni di contesto. Infine è possibile dotare gli agenti virtuali, 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.

4.3. Analisi dei mercati della presenza

Nell'ambito dell'azione coordinata "Peach" del 6° programma quadro [8] è stato svolto uno studio delle aree di applicazione delle tecnologie della presenza che già oggi possono essere individuate nel mercato. L'analisi svolta da Telecom Italia ha consentito di selezionare 152 aziende su un totale di 211 esaminate per mezzo di diversi strumenti di indagine [9]. Sulla base dell'analisi le aziende sono state classificate sia per collocazione geografica che per area di intervento. La figura 3 mostra la distribuzione delle aziende per settori tecnologici di attività.

La prevalenza di imprese nei settori della realtà virtuale e della computer grafica è il risultato della visione tradizionale di queste tecnologie, ma in parte anche del fatto che queste aziende sono quelle più note. Basandosi sulla distribuzione delle aziende per area di competenza, le stesse possono essere collocate sui tre pilastri disciplinari della presenza [10]: interazione uomo-calcolatore, scienze cognitive umane (la comprensione dei processi cognitivi umani), scienze cognitive della macchina (la riproduzione dei processi cognitivi umani su sistemi artificiali). Nella figura 4 il numero contenuto nei cerchi bianchi mostra la quantità di aziende che appartengono a quella determinata area.

In virtù dell'elevato numero di aziende con

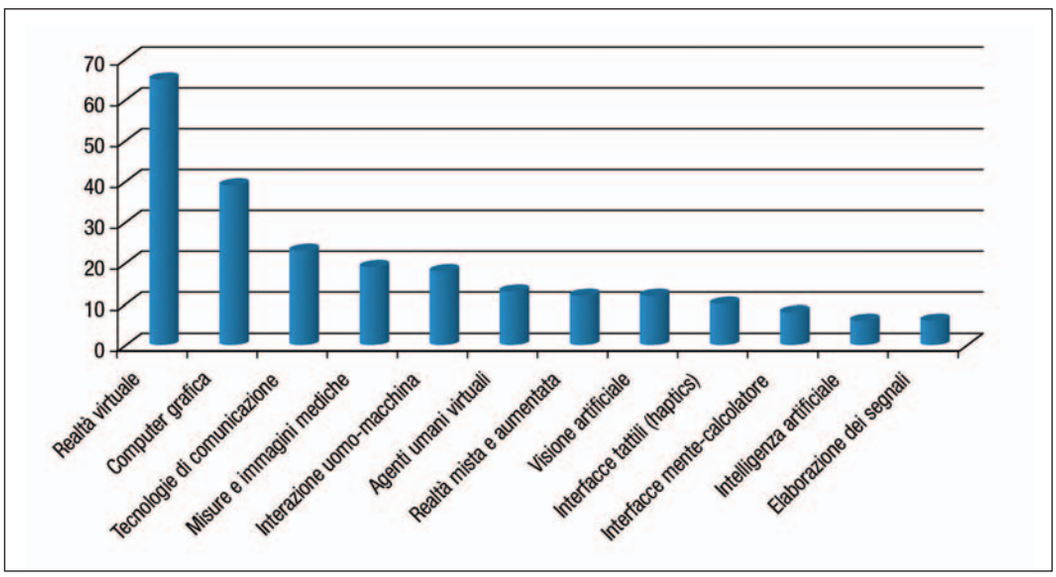


FIGURA 3
Distribuzione delle aziende attive nei settori tecnologici della presenza

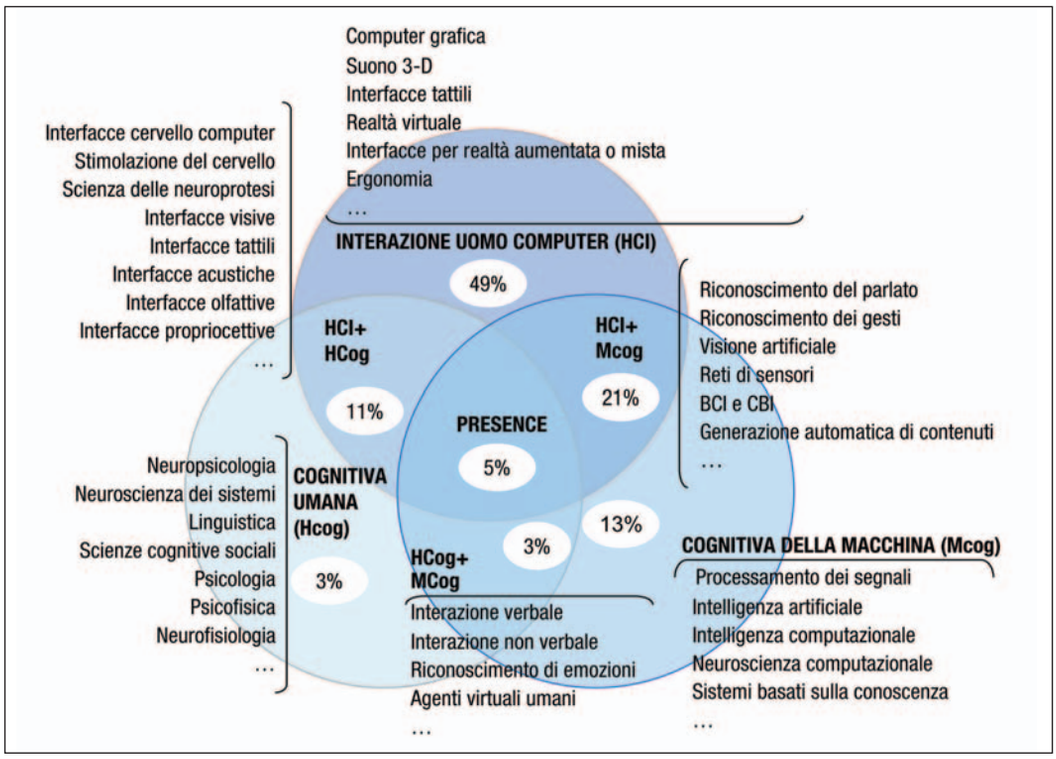


FIGURA 4
Distribuzione delle aziende nei settori disciplinari della presenza

competenza nella grafica computerizzata e nella realtà virtuale, il campo dell'interazione uomo-calcolatore è oggi il più popolato. Si osserva ancora una certa debolezza di attività industriale nel campo delle scienze cognitive umane.
La distribuzione geografica delle aziende che operano nei settori tecnologici della presenza è riportata nella figura 5. Il numero totale di aziende europee è circa pari a quello presente nel continente americano; nelle restanti aree del mondo sono state identificate aziende in Israele, Cina, Sud Corea e Giappone.

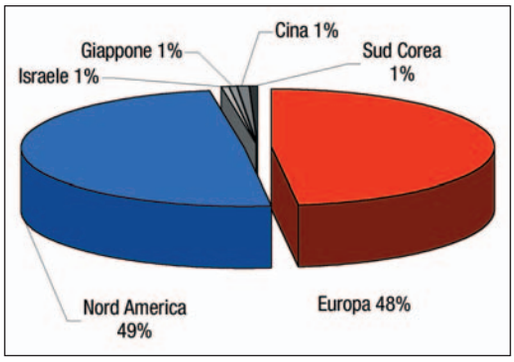


FIGURA 5
Distribuzione geografica delle aziende che operano nei settori tecnologici della presenza

5. ALCUNE APPLICAZIONI DELLA PRESENZA

Tra i settori della presenza oggi più dinamici in termini di applicazioni già sul mercato e prevedibili nel breve termine si possono, a fini esemplificativi, esaminare quelli della medicina e delle telecomunicazioni.

5.1. Applicazioni nel campo della medicina

Le caratteristiche delle tecnologie della presenza si prestano a un impiego con successo in ambito medico in numerosi ambiti che spaziano dalla formazione alla terapia.

Nell'ambito della formazione la principale opportunità offerta dall'odierna realtà virtuale è la possibilità di partecipare attivamente alla creazione e allo sviluppo della propria conoscenza: l'apprendimento è legato allo "scoprire" e al "fare" in prima persona. Come nelle ormai diffuse tecniche di simulazione chirurgica (Figura 6) esiste la possibilità di utilizzare la Presenza per far sperimentare all'utente situazioni ed esperienze complesse o problematiche come se ci fosse davvero dentro.
Per esempio, negli Stati Uniti la DARPA (*Defense Advanced Research Projects Agency*)

ha commissionato lo sviluppo di un sistema di realtà virtuale per addestrare i propri operatori sanitari utilizzati per il primo intervento sul campo di battaglia (Figura 7).
Rispetto ad una tradizionale lezione in aula i vantaggi offerti da queste tecnologie sono due. In primo luogo si permette all'operatore di interagire con i pazienti e con i diversi strumenti richiesti come se questi fossero davvero insieme all'utente. Ciò consente di imparare attraverso l'esperienza diretta e in tempo reale dai risultati delle proprie azioni: se l'intervento è sbagliato il paziente virtuale simula il dolore o il pianto. In secondo luogo è possibile riprodurre le caratteristiche ambientali della situazione che producono nell'operatore una risposta emotiva simile a quella che proverebbe in battaglia. Questo permette non solo di imparare una tecnica ma anche di sperimentare emozioni (panico e tensione) e di imparare a controllarle grazie alla supervisione di uno psicologo.

FIGURA 6
Simulatore chirurgico

FIGURA 7
Addestramento di operatori per primo intervento ai feriti

Questa componente esperienziale viene utilizzata anche nel trattamento di diversi disturbi psicologici. In particolare è stata utilizzata all'interno del progetto americano "Virtual Iraq" [11] per aiutare i veterani a superare i traumi della guerra e dal progetto di ricerca europeo "VEPSY UPDATED" [12] per combattere le fobie.
Le fobie sono paure irragionevoli di un determinato oggetto, luogo o situazione (la paura di volare, dell'altezza, dei ragni ecc.) e sono incontrollabili sebbene spesso i soggetti che ne soffrono siano consapevoli del loro stato. I pazienti coinvolti nelle sedute terapeutiche non immaginano le diverse situazioni ma le "vivono" interagendo all'interno degli ambienti virtuali.
Un ulteriore vantaggio è la flessibilità delle tecnologie della presenza. L'ambiente artificiale può essere adattato in tempo reale alle necessità di ogni paziente e/o alle strategie del terapeuta: per esempio l'ascensore o il vagone della metropolitana possono essere vuoti o pieni di persone, a seconda della fase della terapia e delle risposte del paziente [13]. Se il paziente ha difficoltà ad entrare nell'ascensore affollato il terapeuta può decidere di iniziare la seduta da un ascensore vuoto e progressivamente aggiungere nuove persone appena il paziente si adatta alla situazione. Se invece il paziente non ha problemi ad affrontare la situazione il terapeuta può passare ad uno scenario più complesso, come la metropolitana.
Tuttavia, il modo migliore per combattere queste paure è l'esposizione controllata del soggetto agli stimoli scatenanti e la generazione di risposte antagoniste rispetto a quelle disadattive. In pratica, è possibile desensibilizzare un soggetto con la paura di volare abituandolo

progressivamente alle emozioni che prova salendo sull'aereo e durante la fase di decollo.
Quando, come in questo caso, l'esposizione diretta (in vivo) non è possibile, l'efficacia della terapia dipende dalla capacità del paziente di produrre le immagini mentali che riproducano le situazioni ansiogene. Ma non tutti i soggetti riescono ad immaginare ciò di cui hanno paura. Nel progetto VEPSY UPDATED, il gruppo di ricerca della Facoltà di Psicologia dell'Università Cattolica di Milano, in collaborazione con la California School of Professional Psychology e l'Università di Quebec, ha realizzato una serie di contesti e situazioni simili a quelle reali da utilizzare per la terapia. Per esempio, i pazienti con disturbi di panico sperimentano prima l'entrata in un ascensore, per poi salire su una metropolitana.
Recentemente queste tecnologie hanno trovato applicazione anche nell'ambito della riabilitazione [14]. Dato che la riabilitazione deve consentire al paziente di riappropriarsi delle capacità di programmare, eseguire e controllare sequenze di azioni e comportamenti complessi nello svolgersi della vita quotidiana, le tecnologie della presenza consentono di riprodurre in contesti controllati compiti e situazioni della vita reale. In particolare esse permettono di costruire scenari spaziali e temporali realistici che possono essere usati per ampliare la sensibilità diagnostica dei test programmati.

5.2. Applicazioni alle telecomunicazioni

Oggi sono già disponibili sul mercato sistemi per telepresenza 2-D che si caratterizzano per un'alta qualità audio-video e per la possibilità di consentire la visione su grandi schermi delle figure di dimensioni naturali (Figura 8) [15]. Questi sistemi si prestano a svolgere teleconferenze *multi-party* per uso professionale. Il grado di immersività è già apprezzabile pur in assenza di visione 3-D.
Cominciano peraltro ad essere disponibili sistemi moderni con *display* 3-D che potrebbero rappresentare la tecnologia di base per le telecomunicazioni immersive del prossimo futuro. Come è noto i sistemi di *display* 3-D riproducono l'effetto parallasse bioculare fornendo informazione visiva differente all'occhio sinistro e all'occhio destro. In particolare hanno interesse le moderne tecniche dei *display* autostereoscopici [16] e dei *display* quasi-olografici [17]. Essi potranno rappresentare il vero salto di qualità rispetto alle attuali tecnologie invasive che richiedono l'uso di occhiali speciali o caschi.
Il principio di funzionamento si basa sulla trasmissione di una coppia di immagini riprese simultaneamente con un piccolo angolo di se-

FIGURA 8
Concetto di telepresenza 2-D (soluzione Halo di HP)

parazione (stereoscopia) si induce nel sistema visivo-percettivo umano l'effetto di parallasse che consente di apprezzare la terza dimensione (profondità). Affinché le due immagini vengano lette separatamente si impiegano come schermi due strutture filtranti in grado di separare le immagini ricevute. Nella tradizione cinematografica 3-D per la separazione dei due "canali" ci si basa sulla trasmissione di una coppia di polarizzazioni ortogonali e lenti polarizzate, sebbene esistano altri metodi come la trasmissione ortogonale alternata nel tempo, con lenti che vengono aperte e chiuse elettronicamente in modo sincrono.
Per evitare il problema dell'invasività della soluzione che produce fastidio e stanchezza nell'osservatore, negli ultimi anni si sono affacciati nuovi tipi di schermi che tendono essi stessi a produrre l'effetto di profondità, operando direttamente sul *display* del PC senza bisogno di indossare occhiali. Con questa tecnica ognuna delle due immagini viene suddivisa in piccole strisce verticali e le due immagini generate dai proiettori di sinistra (S) e di destra (D) vengono alternate sul *display*. Questi *display* detti "autostereoscopici" sono tipicamente fabbricati usando i filtri ottici collocati a contatto con il convenzionale *display* LCD[2] o OLED[3]; si sono affermati due tipi di filtri, ad ottica lenticolare e a barriera di parallasse. In entrambi i casi, adattando i parametri geometrici, è possibile fare in modo che l'occhio sinistro veda solo le immagini S e quello destro solo le immagini D. Se le due immagini sono una coppia stereoscopica si percepisce perciò una sola immagine 3-D.
Esistono anche soluzioni che utilizzano opportune tecniche di elaborazione di segnale per consentire sia la visione di immagini 3-D "native" che di immagini 2-D sul medesimo schermo auto stereoscopico.
Conviene osservare che gli schermi autostereoscopici sono affetti da vari tipi di difetti (detti "artefatti") per correggere i quali si devono applicare tecniche di *post-processing* atte a assicurare una migliore qualità dell'immagine e a consentire di percepire la stessa nitidezza dell'immagine che si ha con i moderni schermi monoscopici. Ogni tecnica autostereoscopica proposta presenta da questi punti di vista vantaggi e svantaggi.
Allo stato dell'arte si dispone anche di *display* quasi-olografici in grado di ricostruire un campo di luce quasi perfetto consentendo all'osservatore una visione di un oggetto a 360°, ruotando intorno all'immagine 3-D riprodotta. Questi riproduttori quasi-olografici consentono anche la proiezione di immagini 3-D in movimento.

[2] LCD: *Liquid Crystal Display*.

[3] OLED: *Organic Light Emitting Diode*.

6. CONCLUSIONI

Le tecnologie della presenza, il cui sviluppo si è svolto per decenni nei laboratori universitari e industriali, sono oggi fra le tecnologie più promettenti per determinare i progressi attesi nel settore delle ICMT nel decennio entrante.
Dopo avere illustrato alcuni degli elementi della *road-map* tecnologica prevista per il prossimo futuro, questo articolo ha prima discusso il concetto di presenza e si è poi soffermato sulle tecnologie di base e sulle applicazioni che più di tutte stanno determinando il successo delle tecnologie della presenza in numerosi segmenti di mercato.

Bibliografia

[1] Riva, G.: *Psicologia dei Nuovi Media*. Bologna: Il Mulino, 2008.

[2] Sheridan, T.B.: Telerobotics. *Automatica*, Vol. 25, n. 4, 1989, p. 487-507.

[3] Lombard M., Ditton T.: At the heart of it all: The concept of presence. *Journal of Computer Mediated-Communication*, Vol. 3, n. 2, 1997.

[4] PASION FP6 Integrated Project No 27654. May 2009 URL http://www.ist-pasion.com

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

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

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

[8] PEACH FP6 Coordination Action No 33909. May 2009 URL: peachbit.org

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

[10] Ruffini G., et al.: *Deliverable D11 Visions, Roadmaps, the ERA [Issue 3].* Peach FP6 Coordination Action No 33909. April 2009.

[11] Online Defence Magazine, Virtual Iraq. May 2009 URL: defense-update.com/products/v/VR-PTSD.htm

[12] VEPSY UPDATED European Research Project No. 2220. May 2009 URL: www.cybertheraphy.info

[13] Riva G., Gaggioli A.: Virtual clinical therapy. *Lecture Notes in Computer Sciences*, 2008, 4650, p. 90-107.

[14] Riva G., Castelnuovo G., Mantovani F.: Transformation of flow in rehabilitation: the role of advanced communication technologies. *Behavior Research Methods*, Vol. 38, n. 2, 2006 p. 237-44.

[15] URL: www.francetelecom.com/sirius/rd/en/galerie/mur_telepresence/flash_sonorise.php

[16] Dodgson N.A.: Autostereoscopic 3D Displays. *IEEE Computer*, Vol.38, n.8, Aug. 2005, p. 31-36.

[17] Jones A., McDowall I., Yamada H., Bolas M., Debevec P.: *Rendering for an Interactive 360° Light Field Display*. ACM SIGGRAPH 2007.

Giuseppe Riva, Professore Associato di Psicologia e Nuove Tecnologie della Comunicazione presso l'Università Cattolica di Milano, dove dirige il Laboratorio di studio dell'Interazione Comunicativa e delle Nuove Tecnologie. Autore di oltre 100 articoli scientifici legati agli aspetti psicologici ed ergonomici delle nuove tecnologie è responsabile editoriale delle riviste scientifiche internazionali "CyberPsychology & Behavior" e "Annual Review of CyberTherapy and Telemedicine".
E-mail: giuseppe.riva@unicatt.it

Francesco Vatalaro, Professore ordinario di Telecomunicazioni presso l'Università di Roma "Tor Vergata", ha oltre 25 anni di esperienza nell'industria e nell'università. È membro del comitato direttivo della IEEE Italy Section, oltre che numerosi comitati scientifici e di redazione. È autore di oltre 150 pubblicazioni e i suoi principali interessi scientifici sono nelle comunicazioni e reti wireless.
E-mail: vatalaro@uniroma2.it

Gianluca Zaffiro, Ricercatore presso Telecom Italia. 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 Presence. Si occupa di elaborare scenari innovativi di medio/lungo periodo di interesse per le telecomunicazioni.
E-mail: gianluca.zaffiro@telecomitalia.it