

# IL TEST DI TURING

## STORIA E SIGNIFICATO

Appare molto difficile individuare un criterio oggettivo che ci consenta di distinguere il pensiero da una sua imitazione meccanica. Il punto critico è la definizione di pensiero. In un articolo del 1950 Turing trasformò la domanda se una macchina possa pensare in un problema indiziario basato su una prova comportamentistica di tipo dialogico: se un giudice non riesce a distinguere il comportamento verbale di un computer da quello di un essere umano, che per definizione pensa, allora il computer pensa. Il *test di Turing* ha suscitato negli anni, e ancora suscita, molte discussioni, sia per la vaghezza della formulazione originale sia in relazione ai progressi via via compiuti dall'intelligenza artificiale.

### 1. INTRODUZIONE

“Il metodo della scienza è lo stesso che serve per tutta la conoscenza delle cose di questo mondo: consiste nel limite, o addirittura nell'ignoranza. Posti di fronte alla grande, illimitata unità della natura, noi non possiamo far altro che manipolare col pensiero, scegliendo certi dettagli e (deliberatamente o inconsciamente) isolandoli da tutto il resto”

Edward Carpenter,  
*Civilisation: its Cause and Cure*

“Soltanto il nostro cervello umano può prestare un significato alla cieca capacità dei calcolatori di produrre verità”

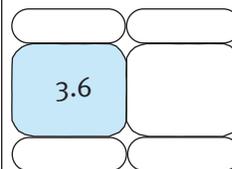
Karl Popper

Nell'articolo *Computing Machinery and Intelligence*, pubblicato sulla rivista *Mind* nel 1950, il grande matematico e logico inglese Alan Mathison Turing (Londra, 1912 - Manchester 1954) descrisse un *Gedankenexperiment*, cioè un esperimento concettuale, per stabilire se una macchina (un computer) sia o no in grado

di pensare. L'esperimento, oggi noto come *test di Turing*, prende lo spunto dal “gioco dell'imitazione”, nel quale un esaminatore cerca di distinguere un uomo da una donna basandosi soltanto sulle risposte a una serie di domande. Turing immagina di sostituire alla donna una macchina, assegnando all'esaminatore il compito di distinguere l'uomo dal computer. La versione modificata del gioco consente a Turing di affrontare il problema: *le macchine possono pensare?* evitando le difficoltà definitorie legate ai termini “macchina” e “pensare”, il cui significato dovrebbe essere precisato in via preliminare. L'utilità di questo esperimento concettuale sta non tanto nella risposta che esso consente di fornire al problema di partenza (non è scontato che dia una risposta), quanto nella possibilità che esso offre di analizzare concetti come *mente, pensiero, intelligenza*. Soprattutto il concetto di intelligenza, che in quegli anni stava diventando il terreno d'incontro di discipline quali la logica, la matematica, la psicologia e la filosofia, che a vario titolo si occupavano dell'uomo, o meglio delle sue capacità cognitive, ritenute le facoltà umane



Giuseppe O. Longo



più caratteristiche. Era importante possedere un criterio se non per *definire* l'intelligenza, almeno per riconoscerne la *presenza* sia negli organismi biologici sia negli artefatti. Non dimentichiamo che erano gli anni della *cibernetica*, una disciplina che si proponeva, secondo il programma del suo fondatore Norbert Wiener, di studiare in modo unificato "la comunicazione e il controllo nell'animale e nella macchina". Di lì a poco, nel 1956, per iniziativa del ma-



**FIGURA 1**  
Alan Turing

tematico John McCarthy si tenne a Dartmouth, nel New Hampshire, un seminario in cui fu scelta la locuzione "intelligenza artificiale" (IA) per indicare un insieme di ricerche piuttosto eterogenee (anche per le discipline di afferenza dei vari partecipanti) che avevano tuttavia l'obiettivo comune di produrre simulazioni o emulazioni di comportamenti intelligenti da parte dei calcolatori elettronici.

Nei paragrafi che seguono esporremo il criterio di Turing e lo sottoporremo a un'analisi critica, per metterne in luce i pregi e i limiti. Pur essendo la trattazione autonoma, faremo costante riferimento all'articolo originale di Turing, la cui lettura è ancora di grande interesse e al quale rimandiamo il lettore (Figura 1).

## 2. IL GIOCO DELL'IMITAZIONE

Il gioco dell'imitazione consiste in uno scambio linguistico cui partecipano tre persone: un uomo A, una donna B e un terzo individuo C, il giudice o esaminatore, il cui sesso è irrilevante. C, che è isolato dagli altri due, può porre ad A e B domande di qualunque tipo e natura e, basandosi soltanto sulle risposte ricevute, deve stabilire se A è l'uomo e B la donna o viceversa. A e B si sforzano entrambi di essere identificati nella donna, quindi B cerca di aiutare C e A cerca di ingannarlo, fornendo le risposte che a suo parere darebbe la donna. C può avere a disposizione soltanto le risposte e non

### Turing: cenni biografici

Nato a Londra nel 1912 da una famiglia della classe media, Alan Mathison Turing manifestò fin da piccolo un carattere solitario e indipendente. Con l'aiuto di un manuale, imparò a leggere da sé in tre settimane. A scuola non brillava nelle materie umanistiche e preferiva dedicarsi a letture scientifiche. Ammesso nel 1931 al King's College dell'Università di Cambridge, vi studiò meccanica quantistica, logica e calcolo delle probabilità, laureandosi nel 1934 col massimo dei voti. Nel 1936, trasferitosi alla Princeton University per specializzarsi, pubblicò il fondamentale articolo *On Computable Numbers, with an Application to the Entscheidungsproblem*, in cui descrisse la cosiddetta "macchina di Turing". Durante la seconda guerra mondiale, ebbe una parte di primo piano nella forzatura del sistema *Enigma*, impiegato dai tedeschi per cifrare i messaggi, e fu vicino al gruppo di crittanalisti che progettò Colossus, il primo computer elettronico digitale programmabile, pur senza prender parte direttamente alla costruzione. Verso la fine della guerra, dopo aver lavorato a dispositivi per la crittografia vocale, Turing ricevette un'alta onorificenza, l'OBE (Ordine dell'Impero Britannico), per i suoi servizi a favore della Gran Bretagna, ma i suoi contributi rimasero segreti per molti anni. Dal 1945 al 1947 si dedicò al progetto dell'ACE (*Automatic Computing Engine*), un computer davvero pionieristico, dotato di un'architettura molto diversa da quella, quasi contemporanea, proposta da von Neumann. Nel 1951 fu eletto membro della Royal Society. Oltre all'articolo di cui si parla in queste pagine, Turing scrisse, sullo stesso argomento, altre due importanti relazioni: *Intelligent Machinery* (1948) e *Digital Computers Applied to Games* (1953), divenendo così un antesignano dell'intelligenza artificiale, anche grazie allo studio delle relazioni tra le macchine e la neurologia e fisiologia degli esseri viventi. Dal 1952 alla morte, nel 1954, si dedicò infatti alla formalizzazione matematica della biologia, in particolare alla morfogenesi, studiando, tra l'altro, i collegamenti tra le forme vegetali e i numeri di Fibonacci. Arrestato nel 1952 per omosessualità, fu condannato a una pena severissima, la castrazione chimica, che lo rese impotente. Fu probabilmente questo il motivo che nel 1954 indusse Turing al suicidio, che mise in atto mangiando una mela attossicata col cianuro di potassio. Benché contestata dalla madre, l'ipotesi del suicidio appare quasi certa. Nel libro *Zeros and Ones*, la Plant suggerisce che la Apple Inc. abbia scelto a proprio simbolo la mela morsicata in omaggio ad Alan Turing, ma l'azienda non ha mai né confermato né smentito questa congettura.

altri indizi (l'aspetto, la voce, la grafia) che renderebbero facile l'identificazione dei soggetti, quindi può comunicare con A e B solo per tele-scrittore o con un mezzo altrettanto neutro. Turing immaginò di modificare il gioco dell'imitazione, sostituendo ad A una macchina, e si chiese se in queste condizioni l'esaminatore C avrebbe sbagliato con la stessa frequenza con cui sbagliava nel gioco dell'imitazione classico. C sbaglia quando ritiene che l'essere umano sia A e non B; in tal caso la macchina vince, e viene dichiarata capace di pensare. Vogliamo ora osservare da vicino il gioco del-

l'imitazione e la variante che ne propose Turing. Chiediamoci intanto: quando esce dai salotti per diventare *Gedankenexperiment*, che cosa dimostra o vorrebbe dimostrare il gioco dell'imitazione? (riquadro 1).

Già proporre il gioco significa ritenere che le risposte di un uomo siano distinguibili da quelle di una donna, cioè che il pensiero maschile differisca in misura apprezzabile da quello femminile. Ma significa anche che, pur essendo forse l'intelligenza della donna diversa da quella dell'uomo, questi può essere tuttavia capace di simularla.

### RIQUADRO 1 - Alcune varianti possibili del gioco dell'imitazione

Chiamiamo A, B e C i partecipanti al gioco dell'imitazione. Ecco alcune delle varianti possibili:

1. A uomo, B donna, C giudice umano (gioco originale)
2. A macchina, B umano, C giudice umano (variante di Turing)
3. A macchina, C giudice umano (prima variante a due)
4. A umano, C giudice umano (seconda variante a due)
5. A uomo, B donna, C giudice macchina
6. A umano, C giudice macchina (terza variante a due)
7. A macchina, C giudice macchina (quarta variante a due)

In tutte le varianti il giudice deve stabilire la natura di A. Nella prima variante il giudice deve sapere che ha di fronte, senza vederli, un uomo e una donna. In tutte le altre varianti, il giudice può essere al corrente oppure no delle identità possibili di A (e di B, se c'è). Turing stesso non precisa se nella variante 2 il giudice sappia che uno dei due contendenti è una macchina, anche se è abbastanza ovvio ritenerlo. Se per esempio nella variante 3 il giudice non sa che A può essere una macchina, è possibile che concluda sempre che si tratta di un umano, magari stupido, pazzo o burlone. A maggior ragione ciò accade nella variante 4, in cui gli errori commessi dall'umano per confondere il giudice avrebbero la qualità umana della pertinenza (riquadro 2). Quindi, per rendere più interessante il gioco, il giudice deve essere al corrente delle alternative possibili. Nei casi 5, 6 e 7 fornire al giudice macchina l'informazione sulla possibile natura degli interlocutori è problematico, com'è problematico immaginare quali domande possa fare C per condurre l'esame (e poi c'è da chiedersi se C abbia o no superato a sua volta il test di Turing, vedi Figura). In definitiva sono tre varianti poco significative, ma confermano forse che il test riguarda il giudice quanto e più di A e B: almeno dal nostro punto di vista, solo un'intelligenza umana può giudicare un'intelligenza umana (o d'altro tipo). Ciò sembra confermare che l'intelligenza umana è, da noi, ritenuta superiore a quella della macchina, così come negli esami l'intelligenza (o la sapienza) dell'esaminatore supera (o dovrebbe superare) quella dell'esaminato. Non so se questo antropocentrismo sia inevitabile come mi appare in questo momento. La variante 7 fa tuttavia intravedere una prospettiva nuova: gli umani giocheranno con gli umani e con le macchine fino a quando le macchine non sapranno giocare come gli umani e anche meglio; a quel punto è plausibile che gli umani giochino solo con gli umani e le macchine solo con le macchine (questo comincia ad accadere con i programmi scacchisti e già da un pezzo accade nella corsa: cavalli con cavalli, automobili con automobili e - aggiungiamo - uomini con uomini, donne con donne!). Ma quando le macchine avranno superato gli umani, forse i giudici più qualificati in tutte le varianti saranno macchine, cioè dovremo adottare un punto di vista macchina-centrico.

L'idea di fondo è che, almeno finora, non è possibile formalizzare il concetto generale di intelligenza. Si può convenire di porre il confronto tra la macchina e l'essere umano all'interno di un perimetro ben preciso, una sorta di gioco, appunto, regolato da parametri quantificabili, per esempio - nel caso della conversazione - la ricchezza del vocabolario, le varianti costruttive delle frasi, la concatenazione logica, la frequenza delle metafore e ovviamente la durata della partita. Ma sorge subito un problema: quale dovrebbe essere l'ambito della conversazione? Oppure la conversazione dovrebbe spaziare su qualsiasi tema? S'intuisce che nel secondo caso nessun programma di conversazione può oggi superare il test se il confronto dura anche poco (riquadro 5), mentre nel caso di ambiti molto limitati il superamento del test non dimostra nulla: la conversazione è un processo interattivo, costruttivo, autopoietico, arborescente, in parte aleatorio e in parte determinato, che riflette da vicino gli eventi della vita, con la loro imprevedibilità e la loro capricciosità, ma anche con la loro logica complessa aperta sul mondo. Come dice Licata, "il processo della cognizione umana è ancorato al confine complesso e imprevedibile tra un sistema e l'ambiente che lo circonda".

TURING TEST EXTRA CREDIT:  
CONVINCE THE EXAMINER  
THAT HE'S A COMPUTER.

YOU KNOW, YOU MAKE  
SOME REALLY GOOD POINTS.  
I'M ... NOT EVEN SURE  
WHO I AM ANYMORE.



Una bizzarra variante del gioco dell'imitazione. Il giudice dice al computer: "hai certe idee davvero interessanti, sai" e soggiunge: "non so nemmeno più chi sono..." la dicitura in alto recita: "avrete un punto in più nel test di Turing se convincerete il giudice che lui è un computer"

Non si dimentichi tuttavia che il gioco è a tre: l'identificazione del sesso di A e B è demandata a un giudice. Se il giudice è in gamba può prendere, aiutato da B, una decisione corretta nonostante gli sforzi di A per sviarlo. Quindi sorge la domanda: si tratta di stabilire se A è più in gamba di B oppure se C è abbastanza in gamba? Forse il gioco è un test per C: cambiando giudice, ma non A e B, l'esito del gioco potrebbe cambiare. Si può anche immaginare di ripetere il gioco con lo stesso giudice, facendo molte "mani" e fornendo una valutazione statistica dei risultati. Oppure si può immaginare che vi siano molti giudici in parallelo che, decidendo tutti in base alle risposte fornite alla stessa serie di domande, possono arrivare a decisioni diverse.

Nella versione originale si tenta di distinguere un uomo da una donna. Questa è la dicotomia più classica, ma potremmo immaginare altre coppie di concorrenti: un bianco e un nero, un italiano e un cinese, un poeta e un matematico (tutti di sesso arbitrario). E la cosa si potrebbe raffinare, mettendo a confronto un matematico cinese e un poeta italiano. Con questi esempi si capisce forse meglio che C, il quale conosce le categorie cui appartengono A e B, è portato a decidere sì in base alle risposte, ma anche in base alle sue credenze, ai suoi pregiudizi, alla sua conoscenza del mondo e così via.

In altre parole, anche se ci si serve di una telescrivente, nelle risposte resta un contenuto che non è puramente sintattico, interno alla lingua, ma anche e fortemente semantico, nel senso che è legato alla natura, non solo cognitiva, di A e di B. Naturalmente la presenza e l'entità affiorante di questo contenuto dipende dalle risposte e quindi dipende dalle domande e dalle aspettative del giudice quanto alle risposte.

Se fosse possibile scegliere domande di natura squisitamente sintattica (di tipo logico, matematico) che contemplassero solo risposte di natura sintattica, forse distinguere un uomo da una donna sarebbe impossibile (ma le recenti scoperte dei neurofisiologi sembrano indicare differenze cerebrali notevoli tra i due sessi, quindi una possibile differenza di pensiero). Naturalmente qui si parla di uomo e di donna generici, cioè scelti a caso, e sconosciuti all'esaminatore. Ne seguirebbe che la distinzione tra sesso maschile e femminile si potreb-

be operare solo sulla base di fatti *non* riconducibili a successioni di simboli. O meglio, a fatti extralinguistici che tuttavia possono lasciare la loro impronta sulle successioni di simboli.

Esiste cioè la possibilità che le differenze (di esperienza, di ruolo sociale, di caratteristiche biologiche e via dicendo) tra uomo e donna introducano differenze apprezzabili, e non colmabili mediante imitazione, nelle risposte di A e B a *certe domande*. Ciò indicherebbe che il pensiero non si può mai isolare o astrarre del tutto dal resto dell'essere umano, per esempio dal suo corpo e dai suoi condizionamenti, quindi dal mondo. Insomma, se le differenze nelle risposte di A e B derivassero dalle differenze *tout court* tra A e B, un giudice abbastanza bravo darebbe sempre la risposta corretta perché B non potrebbe mai imitare A: per imitare A dovrebbe *essere* A.

### 3. LA MENTE E IL CORPO

Turing non la pensa in questo modo e ritiene che il pensiero si possa separare dal suo supporto. Infatti, perché il gioco abbia senso, C deve essere isolato da A e B, altrimenti distinguerebbe d'acchito la donna dall'uomo oppure, nella variante, la macchina dall'essere umano in base all'aspetto esteriore. Turing estende questa condizione al divieto di considerare caratteristiche o capacità corporee (la bellezza, l'abilità nel ballo e così via), che sarebbero discriminanti nei confronti della macchina e che in realtà, per lui, non aggiungono nulla all'intelligenza di chi le possiede. Traspare da questa clausola il pregiudizio di Turing, ma non solo suo, secondo cui l'intelligenza è un fatto esprimibile e riconoscibile soltanto mediante sequenze ben costruite di simboli. Il corpo non avrebbe nulla a che fare con l'intelligenza. Questo riduzionismo mentale o informazionale di Turing è confermato dal fatto che le attività umane che egli riteneva adatte all'automazione erano quelle che non comportano contatti con il mondo esterno: il gioco degli scacchi, la matematica, la crittografia, forse anche la traduzione da una lingua all'altra. Ma gli aspetti corporei erano esclusi. In una trasmissione radiofonica dichiarò:

*"Credo di sicuro, e spero, che non si faranno molti sforzi per costruire macchine dotate di*

caratteristiche (a parte quelle intellettuali) specificamente umane, come per esempio la forma del corpo. Mi sembra futile fare tentativi del genere, i cui risultati avrebbero all'incirca la stessa sgradevole qualità dei fiori artificiali. Fare una macchina pensante mi sembra un lavoro che rientra in una categoria diversa”.

Tuttavia nell'articolo del 1950 Turing contraddice in parte questa sua posizione: estendendo l'attività delle macchine intelligenti alla conversazione, anzi facendo della capacità linguistica addirittura il segno distintivo del pensiero, forse senza rendersene conto apre un grande varco e stabilisce un contatto tra la macchina e il mondo esterno. Infatti, l'esaminatore può porre domande arbitrarie, quindi anche domande sui fatti del mondo (cioè del mondo abitato da noi umani). Il legame tra parola e mondo, che si manifesta quando la parola cessa di essere autoreferenziale, cioè puro simbolo sintattico all'interno di un formalismo alla Hilbert, per caricarsi di semantica e alludere alle cose esterne alla lingua, questo legame, per cui la parola crea, modifica e struttura il mondo e ne è condizionata, ora coinvolge anche la macchina (Figura 2).

C'è dunque un certo grado di confusione. Da una parte Turing riconosce che l'intelligenza dell'uomo deriva dalla sua interazione con il mondo e quindi dal suo equipaggiamento corporeo, che è il tramite che consente quell'interazione. Dall'altra emerge la convinzione che la macchina a stati discreti, la cosiddetta “macchina di Turing” (riquadro 2 a p. 16, 17), possa usare le parole in modo umano, al punto da ingannare l'esaminatore, senza essere dotata di organi di senso e di organi effettori che le consentano di interagire con la realtà esterna.

Il gioco dell'imitazione si basa sull'interazione linguistica, dunque su un segno esteriore del pensiero, che è inaccessibile. In questo senso è un criterio comportamentistico. Le sequenze verbali sono lì a manifestare la propria esistenza, mentre il pensiero, da cui (forse) scaturiscono, ha uno statuto esistenziale molto più dubitoso e, a quei tempi, nella prima metà del Novecento, era bandito dalle riflessioni di psicologi e filosofi. Turing accetta dunque in pieno la posizione comportamentista allora imperante. Ma nell'accettarla egli non nega l'esistenza del pensiero, anzi con-



**FIGURA 2**  
David Hilbert  
(1862-1943)

tribuisce alla formulazione di un problema tuttora vivo nelle scienze cognitive, il rapporto tra i vari livelli di funzionamento e di descrizione. Infatti, tra le obiezioni che egli considera<sup>1</sup>, merita attenzione la seguente:

*“Non possono forse le macchine comportarsi in una maniera che deve essere descritta come pensiero, ma che è molto differente da ciò che fa un uomo?”*

<sup>1</sup> Nell'articolo Turing anticipa le obiezioni dei critici, elencando e discutendo nove opinioni contrarie: 1. *Obiezione teologica*: pensare è una funzione dell'anima immortale, che gli uomini posseggono e le macchine no; 2. *Obiezione dello struzzo*: se le macchine pensassero, le conseguenze sarebbero terribili; speriamo e crediamo che esse non pensino; 3. *Obiezione matematica*: ci sono domande alle quali la macchina non può rispondere in base ai teoremi di Gödel; 4. *Argomento della coscienza*: si veda nel testo a proposito delle dichiarazioni di Jefferson; 5. *Argomento fondato su incapacità varie*: le capacità di tipo fisico elencate impedirebbero alla macchina di pensare (si vedano nel testo le considerazioni a proposito delle fragole con la panna); 6. *Obiezione di Lady Lovelace*: la macchina non è in grado di creare alcunché, può solo eseguire i nostri ordini; 7. *Argomento della continuità del sistema nervoso*: siccome il sistema nervoso non è una macchina a stati finiti, non si può sperare di imitarne il funzionamento con una macchina di quel genere; 8. *Argomento del comportamento senza regole rigide*: siccome l'uomo non si comporta in base a un insieme di regole rigide, ma adotta una condotta flessibile, l'uomo non è una macchina; 9. *Argomento fondato sulla percezione extrasensoriale o ESP* (argomento che Turing definisce molto forte): poiché ci sono prove schiaccianti a favore dell'ESP nell'uomo, uomo e macchina si troverebbero in posizioni diverse quanto alla risposta e l'identificazione sarebbe facilitata.

## RIQUADRO 2 - La macchina, Hilbert e Gödel

Come si è accennato, per Turing la “macchina” è una macchina digitale a stati finiti, cioè un calcolatore particolare, prototipo astratto dei calcolatori concreti, chiamato appunto “macchina di Turing” (MT). Si tratta di un dispositivo ideale, munito di un nastro di carta illimitato diviso in caselle, capace di eseguire queste operazioni: far scorrere il nastro avanti e indietro, scrivere un simbolo su una casella vuota e cancellare il contenuto di una casella scritta. Questa macchina può eseguire qualsiasi operazione descrivibile mediante un algoritmo (macchina universale). Turing dimostrò tuttavia che la MT non era in grado di fornire risposta a tutti i problemi matematici e collegò questa scoperta ai teoremi di Gödel (Foto).

Secondo Hodges la MT, che comunica con il mondo solo per il tramite esiguo del programma, rappresenta anche - potenza delle esperienze personali nel foggia- re le concezioni scientifiche! - l'ideale di vita di Turing: “una vita in cui lo si fosse lasciato da solo in una stanza tutta sua, a trattare col mondo esterno esclusivamente per mezzo di argomenti razionali”. Contro questa interpretazione si può tuttavia citare il seguente passo dalla tesi di dottorato di Turing a Princeton.

“Così facendo non teniamo conto di quella importantissima facoltà grazie alla quale gli argomenti di qualche interesse vengono distinti da tutti gli altri; al contrario consideriamo sola funzione del matematico quella di determinare la verità o falsità delle proposizioni”.

La facoltà di cui parla Turing dipende dal contatto con la realtà e non può quindi essere posseduta dalla macchina. Anche il passo seguente indica l'importanza che Turing attribuiva al contatto con il mondo.

“È opinione corrente che gli scienziati procedano inesorabilmente da un fatto ben stabilito a un altro fatto ben stabilito, senza lasciarsi mai influenzare da una congettura non dimostrata. Questa opinione è del tutto sbagliata. Purché venga messo chiaramente in evidenza quali siano i fatti provati e quali siano le congetture, non può aversi alcun danno. Le congetture sono di grande importanza, giacché suggeriscono utili linee di ricerca”.

Egli pare dunque oscillare, senza decidersi, tra una vagheggiata autoreferenzialità ideale del mondo linguistico, in particolare logico-matematico, e un'apertura comunicativa e attiva verso il mondo reale.

Un esempio cospicuo di autoreferenzialità in ambito matematico ci è offerto dal cosiddetto “programma di Hilbert”. All'inizio del Novecento il matematico David Hilbert aveva sostenuto la possibilità di costruire tutta la matematica con il metodo assiomatico, usando un numero finito di proposizioni di partenza (gli assiomi) e un numero finito di regole di inferenza. Gli enti matematici sarebbero stati definiti in modo implicito e astratto, senza legami semantici con il mondo esterno. In altre parole Hilbert evitava deliberatamente e programmaticamente di “definire” gli enti matematici, che sarebbero stati “nominati” e poi collegati tra loro e caratterizzati dalle reciproche relazioni. Poi, usando le regole, dagli assiomi si sarebbero potuti ricavare i teoremi con un procedimento meccanico.

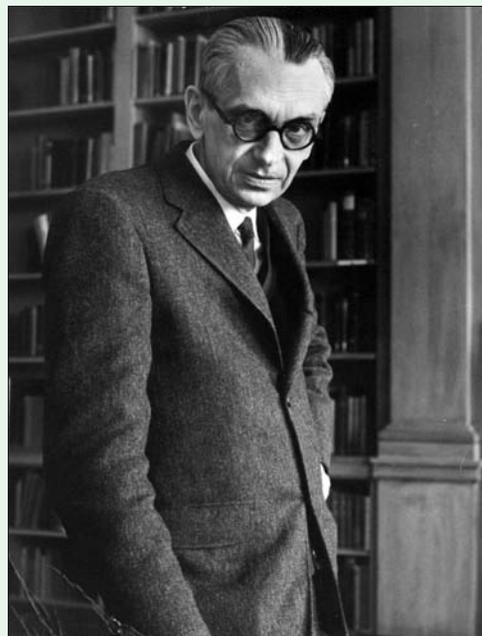
Ma nel 1931 il logico Kurt Gödel fissò limiti molto precisi al programma di Hilbert dimostrando due teoremi. Il primo dimostra che in un sistema formale abbastanza potente da consentire almeno una formalizzazione dell'aritmetica (vale a dire una definizione della struttura dei numeri naturali con le operazioni di somma e prodotto) si può ricavare dagli assiomi una proposizione indecidibile, cioè di cui non si può dimostrare né la verità né la falsità. Il secondo teorema dimostra che all'interno del sistema formale è impossibile dimostrare la sua coerenza. Questi teoremi non implicano tanto uno scacco alla matematica o alla ragione umana, quanto una sconfitta della speranza di Hilbert di separare la matematica da ogni riferimento esterno: un matematico non si limita a manipolare simboli, ma fa riferimento al contesto metamatematico, all'importanza pratica dei problemi, compie scelte e individua legami tra problemi e risultati in apparenza lontani. Di fronte a una proposizione indecidibile, per esempio, ricorre alla semantica per ampliare il sistema formale, creandone un altro, il quale conterrà a sua volta una proposizione indecidibile e quindi richiederà un ulteriore ampliamento, e così via. Solo i sistemi formali molto poveri, che non riescono a formalizzare neppure l'aritmetica, sono coerenti e completi, quelli appena più potenti presentano un'apertura semantica indefinita sul mondo. Ciò contraddice la visione della matematica come tautologia finita, cioè come esplicitazione meccanica di verità teorematice implicite negli assiomi.

Nel 1936 Alan Turing pubblicò *On Computable Numbers*, un articolo in cui illustrava la natura atomica dei passi computazionali mediante l'introduzione della macchina di Turing (MT), protagonista di una proposizione, la tesi di Church-Turing, secondo la quale (ne dò qui una versione selvaggiamente semplificata; per una discussione e altre versioni si veda Hofstadter) *tutti i processi cerebrali derivano da un substrato calcolabile*. Quindi i processi di pensiero dell'uomo sarebbero riproducibili con un automa finito come la MT. Questa proposizione sta alla base della cosiddetta versione forte dell'IA e trova la sua giustificazione nel fatto che la memoria umana è necessariamente limitata.

Nel 1969, tuttavia, Gödel trovò nel ragionamento di Turing una falla che comprometteva la sua visione finitistica della computazione umana. La mente umana, argomentava Gödel, non è statica, ma si sviluppa di continuo durante il suo funzionamento. Benché ad ogni stadio del suo sviluppo il numero dei suoi possibili stati distinti sia finito, non c'è motivo per cui questo numero non possa divergere nel corso di questo sviluppo. La mente è, all'inizio, basata su un certo numero finito di assiomi (che potrebbero corrispondere alle facoltà iniziali determinate dall'informazione genetica), ma poi grazie all'interazione perturbativa con l'ambiente, cioè all'informazione fornita dall'esperienza, il “sistema mente” si complessifica e sfugge alla nostra analisi formale, così come un sistema formale abbastanza potente, nel senso precisato sopra, sfugge ad ogni riduzionismo di tipo assiomatico.

Ciò che caratterizza la mente umana è la capacità di produrre novità, di rompere le simmetrie, di apprendere, di compiere insomma tutte quelle operazioni creative ben descritte da Gregory Bateson nella sua trattazione delle categorie logiche dell'apprendimento e della

*segue*



Kurt Gödel (1906-1978)



comunicazione. L'interazione mente-ambiente è generativa e dinamica e non si può racchiudere in una descrizione *definitiva* della mente mediante una o più MT. Mentre le strutture dell'IA presentano una logica chiusa e hanno successo nei micromondi specializzati o per i problemi giocattolo, la mente ha una struttura logica aperta, cambia di continuo le regole del gioco e si adegua alla mutevolezza dell'interazione con il (macro)mondo.

Del resto anche la fisica classica, basata, almeno in apparenza, su una chiusura che si esprime nell'inesorabile procedere dei fenomeni secondo ferree leggi eterne, è incapace di spiegare la novità, la nascita dell'informazione e lo sviluppo della storia. Soltanto con le conquiste rappresentate dalla complessità, dall'emergenza e dagli aspetti informativi, si è potuto fornire un quadro evolutivo, storico e aperto della realtà fisica filtrata dalla nostra cognizione. In ciò si può ravvisare quell'intreccio tra la fisica e la mente che in forma filosofica e qualitativa era stato indicato da Bateson con la sua "ecologia della mente".

Non bisogna tuttavia credere che Turing fosse così ingenuo da ritenere che una MT o una famiglia di MT potesse essere un modello adeguato della mente umana. L'organizzazione e il funzionamento della macchina, sostiene Turing sia nell'articolo del 1936 sia in un rapporto del 1948, si devono poter modificare. La prima modifica funzionale è costituita dagli *oracoli*, dispositivi capaci di condizionare dall'esterno la computazione di una MT. C'è un collegamento tra oracoli e intelligenza naturale: il programma inscritto nel genoma è via via modificato dall'esperienza. La disciplina, dice Turing con riferimento all'organizzazione di partenza e alla capacità di obbedire ai programmi, da sola non basta: per l'intelligenza ci vuole anche l'*iniziativa*. Quindi Turing si rende conto che la MT è un concetto troppo povero per descrivere i processi cognitivi: paradossalmente, la macchina deve essere capace di sbagliare, se è infallibile non può essere intelligente. Del resto, un matematico è sottoposto a un lungo apprendistato, durante il quale in genere commette parecchi errori. Perché non dovrebbe essere la stessa cosa per una macchina?

Questa apertura all'errore consente di evitare ciò che di ripetitivo e, appunto, macchinico si ravvisa in certi programmi che vorrebbero imitare le capacità dialogiche dell'uomo. Mentre gli errori delle macchine nel rispondere alle domande del test di Turing possono essere arbitrariamente madornali, negli errori commessi dagli umani si ravvisa sempre una certa dose di *pertinenza*, nozione più ampia e sfumata della correttezza. La pertinenza insomma comprende anche certi tipi di errore: l'importante è che le risposte sbagliate non diano l'impressione di arbitrarietà o di follia. La nozione di pertinenza può aiutarci a cogliere l'idea (vaga) di comprensione più e meglio della nozione (assoluta) di esattezza (riquadro 5).

Insomma, anche se l'attività superficiale, affiorante, cioè l'espressione linguistica, è la stessa, può darsi che i meccanismi sottostanti siano molto diversi nella macchina e nell'uomo. Ma, conclude Turing, riadeguandosi al comportamentismo, se una macchina gioca bene al gioco dell'imitazione, la differenza nei meccanismi sottostanti può essere trascurata.

semplicemente, un *gioco d'intelligenza*, un passatempo, al massimo uno spunto di riflessione? Per avere un'indicazione in questo senso, poniamoci domande analoghe a proposito di un essere umano, al quale in genere concediamo senza troppe esitazioni la qualifica di intelligente.

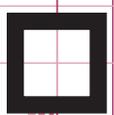
#### 4. LA VARIANTE DI TURING

Consideriamo ora la variante proposta da Turing, in cui A è una macchina e B un essere umano. Anche in questo caso non è molto chiaro quale sia il significato del gioco. Si può immaginare che, ripetendo la prova più volte, cioè facendo molte mani, certe volte C sbaglia l'attribuzione e altre volte no, cioè che la macchina a volte superi e a volte fallisca la prova. È intuitivo che, se le domande possono essere arbitrarie, basta porne un numero sufficiente per far cadere la macchina. È sufficiente uscire da quello che è il suo orizzonte di competenza, ponendo domande sul mondo in grande. Ma ciò vale anche per gli umani. Forse, come per tutti i giochi, bisogna porre dei limiti e delle condizioni. Ammesso allora di stabilire durata, modalità e così via, del test, il suo superamento è una condizione sufficiente per l'intelligenza? È una condizione necessaria? O è,

#### 5. L'INTELLIGENZA UMANA

Sottoponiamo un soggetto umano a una prova molto particolare: facciamogli giocare un certo numero di partite a scacchi. Possiamo dire che se vince più della metà delle partite è intelligente? E che se ne vince meno della metà non lo è o lo è meno? Credo che pochi sottoscriverebbero queste conclusioni, al massimo potrebbero concludere che il soggetto gioca bene o male a scacchi, *relativamente al suo avversario*. Ora sottoponiamo il soggetto a un'altra prova, più elusiva: chiediamogli di scrivere un racconto. Anche in questo caso, come nel precedente, la riuscita non è condizione sufficiente per dichiararlo intelligente, ma neppure condizione necessaria. Qui giudicare della riuscita sarebbe peraltro molto più difficile.

Molte persone che sono considerate intelligenti (in base a prove indiziarie, come il successo nella vita professionale o la capacità di intessere rapporti umani soddisfacenti o la



bravura nella ricerca scientifica) non sanno giocare a scacchi e non sanno neppure scrivere un racconto. Forse potrebbero cimentarsi nella stesura di un racconto, ma la riuscita, o l'insuccesso, non aggiungerebbe, o toglierebbe, nulla alla loro intelligenza, sarebbe soltanto una prova indiziaria in più, e solo per chi ritiene importanti i racconti: di nuovo il punto di vista del giudice. Per converso molte persone che giocano bene a scacchi o sanno scrivere ottimi racconti spesso danno prova di scarso acume in molte circostanze della vita.

Il fatto è che l'intelligenza degli umani è per così dire a largo spettro (si veda la citazione di Cartesio nel riquadro 3), è multiforme, elusiva e variegata, si esplica e si accresce e si modifica con la comunicazione e più in generale attraverso quell'attività altrettanto multiforme e inafferrabile che è la vita. Da ciò segue che l'intelligenza umana è molto legata al corpo e alla sua immersione nel contesto

ambientale ed è un prodotto dell'evoluzione prima biologica e poi culturale. Inoltre il numero enorme di esseri umani viventi, operanti e comunicanti in ogni istante rende problematico l'uso di soggetti generici come "l'uomo", "la donna", o di caratteristiche sfumate come "l'intelligenza umana". L'intelligenza delle macchine, qualunque ne sia il significato e la portata è, almeno per il momento, a spettro molto stretto, è finalizzata a compiti particolari e in genere comunica con l'ambiente circostante soltanto attraverso l'esile cordone del programma. Non si estende nell'ambiente mondo né possiede la dimensione diacronica dell'evoluzione.

A rafforzare queste perplessità, si ha l'impressione che nell'articolo Turing consideri la sua proposta più un gioco (dell'imitazione, appunto) che un criterio incontrovertibile per l'intelligenza: solo in seguito, dopo la morte dell'autore, si è cominciato a parlare diffusa-

### RIQUADRO 3 - Due precedenti illustri: Cartesio e Leibniz

Debbo a Bianchini due interessanti citazioni, di Cartesio e di Leibniz, che costituiscono anticipazioni profonde e sorprendenti degli argomenti di cui si è occupata l'IA. Il passo di Cartesio espone alcuni criteri differenziali tra intelligenza umana e IA, e in particolare pone l'accento sul comportamento linguistico. Non si può non notarne l'affinità con le argomentazioni di Turing. Leibniz sembra accennare alla possibilità che esistano due tipi di spiegazioni, una per i fenomeni meccanici e una per i fenomeni dell'intelligenza (o dello spirito o della sensazione). Questa doppia spiegazione, che non sembra avere natura ontologica, ma solo epistemologica, ricorda la doppia spiegazione di Bateson in ordine ai fenomeni fisici e ai fenomeni mentali. Presi insieme, i due passi costituiscono la formulazione in nuce di uno dei problemi centrali dell'IA, cioè la valutazione dei modelli di spiegazione dell'azione intelligente.

*"Se vi fossero macchine simili ai nostri corpi, che ne imitassero le azioni quanto è praticamente possibile, avremmo sempre due mezzi certissimi per riconoscere che non per questo sarebbero dei veri uomini. Il primo è che mai potrebbero usare delle parole o di altri segni impiegandoli, come noi facciamo, per comunicare ad altri i nostri pensieri. Infatti si può senz'altro concepire una macchina costruita in modo da proferire delle parole e addirittura da proferirne qualcuna a proposito di azioni fisiche che determinino qualche mutamento nei suoi organi: per esempio, toccandola in un punto può domandare che cosa le si vuol dire; toccandola in un altro può gridare che le si fa male, e via di seguito; ma non si può concepire che essa coordini le parole diversamente per rispondere al senso di tutto ciò che si dirà in sua presenza, come possono fare anche gli uomini più ottusi. In secondo luogo, le macchine, anche se facessero parecchie cose bene quanto noi, e forse meglio di tutti noi, immancabilmente in qualche altra cosa fallirebbero, dando modo di scoprire che non agiscono in base a conoscenza, ma solo in base alla disposizione dei loro organi. Perché, mentre la ragione è uno strumento universale, che può servire in ogni sorta di occasioni, questi organi hanno bisogno, in ogni azione particolare, di una disposizione particolare; ne consegue la pratica impossibilità che una macchina ne consenta una sufficiente varietà che le consenta, in tutte le occorrenze della vita, di agire come ci fa agire la nostra ragione".*

Cartesio, *Discorso sul metodo*, Quinta parte

*"D'altronde bisogna ammettere che la percezione e quello che ne dipende sono inesplicabili mediante ragioni meccaniche, cioè mediante figure e movimenti. Se immaginiamo una macchina costruita in modo che pensi, senta e percepisca, si potrà concepire che venga ingrandita conservando le medesime proporzioni, in modo che vi si possa entrare come in un mulino. Ciò fatto, nel visitarla internamente, non si troverà altro che pezzi, i quali si spingono scambievolmente, e mai alcuna cosa che possa spiegare una percezione. Cosicché questa bisogna cercarla nella sostanza semplice e non nel composto, o nella macchina. Reciprocamente, non altro che le percezioni e i loro mutamenti si possono rinvenire nella sostanza semplice. In essi soltanto possono consistere tutte le azioni interne delle sostanze semplici".*

Leibniz, *Monadologia*, 17

mente di “test di Turing”. L’articolo presenta un’altra caratteristica: nonostante l’importanza dei temi trattati e l’acume con cui sono affrontati, non ci si può sottrarre del tutto all’impressione che il piglio dell’autore sia, appunto, un tantino giocoso e che ogni tanto egli si lasci andare a qualche divagazione bizzarra: si veda nella nota 1 la risposta data all’obiezione del paranormale.

C’è anche, curiosamente, un errore di calcolo: quando il giudice chiede al soggetto (uomo o macchina?) di sommare 34957 e 70764, la risposta è 105621 invece di 105721. Che significato ha questo errore? O è una svista?

## 6. LA PROIEZIONE COGNITIVA E LA COSCIENZA

È importante riconsiderare la funzione del giudice. Se, come ho suggerito, è C che viene messo alla prova, forse il test riguarda non tanto l’intelligenza delle macchine quanto l’atteggiamento dell’uomo nei confronti delle manifestazioni (in questo caso linguistiche) dell’intelligenza, sia essa simulata o “naturale”.

Questo rovesciamento di prospettiva è confortato dalle vicende di alcuni programmi redatti in seguito, per esempio *Eliza* di Joseph Weizenbaum (Figura 3) (riquadro 4 a p. 20). In ogni caso non si deve dimenticare che la macchina non simula il pensiero dell’uomo, ma un modello molto semplificato del pensiero dell’uomo.

Nei confronti dei programmi come *Eliza* gli umani attuano una sorta di “proiezione cognitiva” analoga alla proiezione affettiva di cui parlano gli psicologi: la somiglianza tra i segni esteriori di una certa entità E e i segni esteriori di un essere umano U ci spinge a compiere un “prolungamento”, cioè ad attribuire a E tutte le caratteristiche di U, in particolare le caratteristiche psicologiche, affettive e cognitive. Ciò accade a maggior ragione quando anche E è un essere umano, cioè quando possiede *tutti* i segni esteriori dell’umanità: l’operazione proiettiva che si compie in questo caso è tanto frequente che passa inosservata e consiste nell’attribuire all’altro sentimenti, capacità logiche e argomentative e così via, analoghe alle nostre. È solo grazie a questa sorta di immedesimazione che si supera il “problema delle

menti altrui” e si persiste nell’ostinato sforzo di comunicare senza sapere che cosa accada veramente dei nostri messaggi quando entrano nell’interlocutore.

Il problema delle menti altrui è analogo al problema della coscienza: non potendo entrare nell’altra persona, non ho modo di sapere se essa possieda la coscienza e altre caratteristiche, quindi le faccio credito operando a suo favore una fiduciosa concessione delle facoltà che possiedo io stesso. Turing conclude affermando che non si vede il motivo per cui un calcolatore dovrebbe essere trattato diversamente.

Di parere contrario era invece il chimico, filosofo ed economista di origine ungherese Michael Polanyi (1891-1976), il quale osservò che una macchina è una macchina e una mente umana è una mente umana e non c’è dato dell’esperienza che possa cambiare questo *a priori*. Polanyi tuttavia non chiarì la distinzione tra macchina e mente umana, che è il nocciolo di tutta la questione: se restiamo alla superficie, tutto sembra chiaro, ma se spingiamo le argomentazioni a fondo cominciano le difficoltà. Ciò ricorda quanto sant’Agostino diceva del tempo: “Se non me lo chiedi so che cos’è, se me lo chiedi non lo so più”. Questo fenomeno di presbiopia cognitiva riguarda anche la lingua: se ci si accontenta di una certa vaghezza legata al buon senso e all’uso ordinario va tutto bene, ma se si vuol tener conto di tutte le accezioni ed eccezioni linguistiche e definitorie si finisce nei



**FIGURA 3**  
Joseph Weizenbaum (1923-2008)

#### RIQUADRO 4 - La "psicoterapeuta" Eliza

Il programma *Eliza*, allestito nel 1964 da Joseph Weizenbaum (allora ricercatore del MIT), trae il nome dalla protagonista femminile del dramma *Pygmalion* di George Bernard Shaw, la quale diviene oggetto della scommessa del professor Higgins: si può trasformare una sciatta popolana in una raffinata dama di corte agendo su fonetica e linguaggio? Infatti *Eliza*, che consente alle persone di conversare in linguaggio naturale con un computer, può essere via via migliorato da un "insegnante" che fa la parte di Higgins. Il dialogo tra computer e utente avviene tramite telescrivente: l'utente batte una frase e passa la parola a *Eliza*, che analizza la frase e risponde battendo una frase. Ecco un frammento di dialogo tra *Eliza* e una paziente (in corsivo le frasi della paziente):

*Gli uomini sono tutti uguali.*  
In che senso?  
*In un modo o nell'altro ci tormentano sempre.*  
Può farmi un esempio?  
*Be', è stato il mio ragazzo a farmi venire qui.*  
È stato il suo ragazzo a farla venire qui.  
*Dice che sono sempre depressa.*  
Mi dispiace sentire che è depressa.  
*È vero. Sono infelice.*  
Pensa che venire qui l'aiuterà a non essere infelice?  
*Ho bisogno di aiuto, questo è sicuro.*  
Che cosa cambierebbe per lei se ricevesse aiuto?  
*Forse imparerei a sopportare mia madre.*  
Mi dica qualcosa della sua famiglia.  
*Mia madre si occupa di me.*  
Chi altri della sua famiglia si occupa di lei?  
*Mio padre.*  
Suo padre.  
*Per certi versi lei somiglia a mio padre.*  
Che somiglianze vede?  
*Lei non è molto aggressivo, ma penso che non voglia farmelo capire.*  
Perché pensa che non sia molto aggressivo?  
*Lei non litiga con me.*  
Perché pensa che io non litighi con lei?  
*Perché ha paura di me.*  
Le fa piacere pensare che ho paura dei lei?  
*Mio padre ha paura di tutti.*  
Che cos'altro le viene in mente quando pensa a suo padre?  
*I prepotenti.*  
Questo ha a che fare con il fatto che il suo ragazzo l'ha fatta venire qui?

In sintesi, il procedimento seguito dal programma è questo: *Eliza* cerca nella frase della paziente qualche parola chiave; se la trova, trasforma la frase secondo una regola associata a quella parola chiave, altrimenti fa un'osservazione acotestuale o, in certe condizioni, ripete una trasformazione precedente (in realtà il programma è molto più complesso, ma non è il caso di entrare nei particolari). *Eliza* dà all'interlocutore l'impressione di essere uno psicoterapeuta non direttivo (rogeriano), il quale è comprensivo ma piuttosto inerte: egli pone domande vaghe e parla pochissimo, come si vede dal colloquio riportato, e per un po' risulta piuttosto convincente. In *Eliza* non c'è nessun barlume d'intelligenza: il programma va semplicemente in cerca di parole come "madre" o "depresso" e poi pesca una domanda appropriata da un repertorio; se questa tattica non funziona, produce una frase generica tentando di riavviare la conversazione. Quasi tutti i programmi dialogici si basano su principi analoghi ( riquadro 5).

In realtà quello che sembra un dialogo è un monologo dell'interlocutore umano, il quale dà senso a domande e risposte solo sue, comportandosi da "animista". *Eliza* ebbe un enorme successo: chi conversava con "lei" provava spesso sollievo dopo le sedute. Uno psichiatra, Kenneth Colby, dichiarò che di lì a pochi anni sarebbero stati allestiti programmi da usare nella pratica terapeutica, e a sua volta allestì un programma, *Parry*, che simulava il comportamento linguistico di un paranoico e riuscì a ingannare parecchi psichiatri sulla propria identità. È da notare che è molto più facile simulare uno psicotico che un savio. Fu anche organizzato un dialogo tra *Eliza* e *Parry*! Weizenbaum, tuttavia, fu molto turbato dal successo del suo programma. Credo che *Eliza* sia stato determinante nella sua decisione di schierarsi contro l'intelligenza artificiale.

guai. Forse bisognerebbe appagarsi di una certa approssimazione, senza cercare una rischiosa precisione. Chi si contenta gode e chi troppo vuole nulla stringe.

Di parere contrario era anche il grande neurofisiologo e neurochirurgo britannico sir Geof-

frey Jefferson (1886-1961), di cui Turing riporta la seguente obiezione:

"Fino a quando una macchina non saprà scrivere un sonetto o comporre un concerto sulla base di pensieri e di emozioni sentite

come tali e non per una casuale combinazione di simboli - cioè finché non solo possa farlo, ma sappia di averlo fatto - noi non accetteremo mai l'idea che la macchina possa uguagliare il cervello. Non c'è meccanismo capace di provare (e non semplicemente segnalare artificialmente, il che sarebbe facile da ottenere) piacere per i successi conseguiti, dolore quando le sue valvole bruciano, gioia quando riceve complimenti, tristezza per gli errori commessi, l'incanto del sesso, la collera o la delusione perché non riesce ad ottenere ciò che desidera”.

A ciò Turing risponde:

*“Questo ragionamento sembra confutare la validità del nostro test. Secondo la versione più radicale di questa posizione, l'unico modo per essere sicuri che una macchina pensa è di essere quella macchina e sentirsi pensare. Si potrebbe allora descrivere questa sensazione al mondo, ma naturalmente nessuno sarebbe tenuto a prenderne atto. Allo stesso modo, secondo questa concezione, l'unico modo per sapere che un dato uomo pensa è di essere quell'uomo. Si tratta in effetti della posizione solipsistica. Può essere la posizione più logica da sostenere, ma rende difficile la comunicazione delle idee. Si crea questa situazione: A è convinto che “A pensa ma B no”, mentre B crede che “B pensa ma A no”. Invece di continuare ad accapigliarsi su questo punto, di solito per educazione si adotta la convenzione che tutti pensino”.*

E aggiunge che il gioco dell'imitazione con un solo interrogato è spesso usato nella pratica degli esami orali per accertarsi se lo studente abbia imparato qualcosa a pappagallo o l'abbia davvero capita. Se poi una macchina che scrive sonetti superasse un esame orale sulla composizione poetica con la stessa disinvoltura e la stessa capacità di giustificare le risposte che manifesta uno studente, sarebbe difficile sostenere che la macchina esegue soltanto delle segnalazioni artificiali o che si tratta di un facile trucco. Insomma una prestazione verbale che non si distinguesse da quella di un essere umano spingerebbe ad accettare la va-

lidità del test. Quanto alla coscienza, Turing aggiunge tuttavia:

*“Non vorrei dare l'impressione che riguardo alla coscienza non ci sia nessun mistero. Per esempio qualunque tentativo di localizzarla porta a qualche forma di paradosso. Tuttavia non penso che sia necessario risolvere questi misteri prima di poter dare una risposta ai problemi di cui ci occupiamo in questo articolo”.*

## **7. IL RUOLO DEL LINGUAGGIO VERBALE**

*“E invero tutte le qualità che un accorto novellatore di razza umana, esperto quanto si voglia di caratteri, può rilevare in un animale o attribuirgli, non sono al postutto che mere supposizioni, cui solo il nostro smodato antropomorfismo presta verosimiglianza. Fra noi: in che modo penetrare d'un brutto i pensieri, il vero significato dei suoi gesti, anche ad adottare l'accezione umana di tali termini? Un uomo di fronte a un altro uomo ha almeno una convenzione, se non altro di linguaggio, alla cui stregua commisurare gli attributi; ma riportare questa convenzione sugli animali sarebbe a dir poco arbitrario”*

Tommaso Landolfi, *Le due zittelle*

Secondo alcuni il criterio verbale di Turing non è sufficiente per l'attribuzione del pensiero e neppure necessario: un ente potrebbe essere intelligente senza essere in grado di parlare, e in effetti ciò si può dire degli animali superiori. La comunicazione, anche negli umani, non è certo solo linguistica, ma noi occidentali abbiamo privilegiato la comunicazione verbale e consideriamo le altre forme di comunicazione vaghe e ambigue, ma forse è solo questione di tradizione culturale. Ciò che conta è poter comunicare in qualche modo: un'entità intelligente che non potesse comunicare affatto sarebbe come l'albero di Berkeley, che cade nella foresta senza che nessuno ne oda lo schianto. Il criterio verbale è molto antropocentrico, tuttavia è abbastanza evidente che certi uomini non supererebbero il test, mentre certe macchine non intelligenti potrebbero superarlo (inducendo

in errore almeno un giudice per cinque minuti, come nel caso di *Eliza*). Ciò sembra indebolire di molto la portata del criterio. Il tentativo di separare la mente dal corpo, attribuendo l'intelligenza solo alla prima, solleva dunque un certo numero di problemi, che sono tipici dell'impostazione funzionalistica dell'intelligenza artificiale e che hanno portato al fallimento della sua versione forte.

Questi problemi si possono riassumere mediante alcune questioni cruciali:

Può esistere un'intelligenza senza vita e senza altri correlati della vita, per esempio le emozioni e la coscienza? Può esistere una mente senza interazione comunicativa con l'altro? Può esistere un linguaggio autoreferenziale, senza qualche forma di esistenza concreta e quindi di interazione con il mondo? E, forse la più cruciale: può esistere il pensiero senza l'esperienza (che è radicata nel corpo)?

Queste domande denunciano una chiarissima impostazione antropocentrica: noi scegliamo come modello di riferimento la nostra attività cognitiva e confrontiamo con essa ogni alternativa proposta, con il preconcetto che sia inferiore o difettosa. Allo stesso tempo, nelle domande compaiono termini indefiniti, vaghi, imprecisi, come vita, coscienza, esperienza, ai quali tuttavia noi esseri umani sappiamo dare significato operativo e di fronte ai quali adottiamo un atteggiamento di sospensione definitoria e di operosità intuitiva. Il trasferimento di questi concetti alla macchina, che non può beneficiare dell'impostazione operativa o dell'intuizione, richiederebbe invece definizioni univoche e inattaccabili. È evidente che le macchine si trovano in condizione d'inferiorità.

Non dimentichiamo che anche Ludwig Wittgenstein si poneva problemi analoghi: la lingua può essere considerata solo un gioco? Oppure ha una relazione con la vita reale? È abbastanza curioso che ancora ci siano pensatori orientati a rispondere in modo positivo alla prima domanda, dimenticando la base concreta della lingua e l'inestricabile groviglio che la lega al mondo attraverso l'attività dell'uomo. Anche le conseguenze dei teoremi di Gödel, che indicano un'apertura indefinita della matematica e una sua connessione

con il mondo dovrebbero rendere evidente l'analogia apertura sul mondo della lingua e della mente (vedi riquadro 2).

Turing oscilla tra due posizioni: una è a favore della smaterializzazione dell'intelligenza, bene espressa dal protagonista del libro *The Small Back Room* di Nigel Balchin: "È un gran peccato che non si possano abolire la marina, l'esercito e l'aviazione e continuare semplicemente a vincere la guerra senza di loro". L'altra invece, illustrata da Turing stesso nell'articolo, riconosce l'importanza del legame con la realtà per la completezza della macchina. Dopo aver elencato vari difetti che, manifestati dalla macchina, la renderebbero per alcuni incapace di pensare: essere gentile, piena di risorse, bella, cordiale, far sì che qualcuno s'innamori di lei, gustare le fragole con la panna, Turing scrive:

*"L'incapacità di gustare le fragole con la panna può sembrare frivola al lettore. Forse sarebbe possibile trovare il modo di far gustare a una macchina questo piatto delizioso, ma uno sforzo del genere sarebbe insensato. La cosa importante riguardo a questa incapacità è che essa compromette per esempio la possibilità che si stabilisca tra uomo e macchina lo stesso genere di cameratismo che può esistere tra due bianchi oppure tra due neri".*

Per concludere, si può forse dire che il cosiddetto test di Turing è un esperimento concettuale il cui valore pratico, quanto alla presenza dell'intelligenza nelle macchine, è quasi nullo. Non rappresenta una condizione necessaria e neanche una condizione sufficiente. La sua importanza risiede piuttosto nel tipo di riflessioni che ha provocato sul concetto di intelligenza, ma anche di coscienza, di computazione e via dicendo. Le stesse esitazioni di Turing tra la concezione sintattica dell'intelligenza e il suo radicamento nel corpo e nel mondo sono utili per apprezzare i limiti dell'impostazione forte dell'IA. L'ottimismo di Turing sul futuro dell'IA si è dimostrato ingiustificato, o meglio mal diretto: oggi, nel 2009, le macchine fanno cose straordinarie, ma non fanno ciò che Turing aveva immaginato (riquadro 5 a p. 23). Esse si presentano sempre più come componenti essen-

### RIQUADRO 5 - Il premio Loebner

Il test di Turing, almeno a prima vista, è di una semplicità disarmante, ma per alcuni decenni nessuno pensò di attuarlo: come aveva fatto Turing, tutti lo consideravano solo un esperimento concettuale. È vero che nel suo articolo del 1950 Turing aveva pronosticato per le macchine un avvenire luminoso, sia pure in termini piuttosto vaghi:

*“io credo che tra una cinquantina d’anni sarà possibile programmare calcolatori aventi una capacità di memoria di circa un miliardo, in modo da farli giocare così bene al gioco dell’imitazione che un esaminatore medio avrà una probabilità non superiore al 70% di compiere l’identificazione giusta dopo cinque minuti di interrogatorio. Credo che la domanda iniziale - le macchine possono pensare? - sia troppo priva di senso per meritare una discussione. Ciò nonostante credo che alla fine del secolo l’uso delle parole e l’uso corrente saranno talmente mutati che si potrà parlare di macchine pensanti senza aspettarsi di essere contraddetti”.*

Ma evidentemente i ricercatori successivi erano stati meno ottimisti e non si erano cimentati con il test. Nel 1991, tuttavia, Hugh Loebner, programmatore e uomo d’affari del New Jersey, istituì un premio di 100.000 dollari da assegnare al primo computer che superasse il test. Ritenendo piuttosto remoto questo evento, Loebner decise di assegnare un premio annuo d’incoraggiamento di 2000 dollari al concorrente migliore. Da allora la competizione è seguita con grande interesse, per quanto nessun computer si sia mai avvicinato neppure lontanamente al traguardo, sicché il concorso si è più o meno trasformato in un evento pubblicitario. Nel 1996 la gara riservò una sorpresa: tra tutte le macchine pensanti, o candidate a pensare, si era intrufolato un impostore, il programma HeX, scritto da Jason Hutchens, un dottorando dell’University of Western Australia. E, tra l’indignazione generale, HeX vinse. Dopo la proclamazione, Hutchens dichiarò che il suo programma non era più intelligente di un distributore automatico di bevande: HeX si basava su un dizionarietto di domande, risposte e frasette disinvolute, che venivano emesse quando il programma riconosceva certe parole (vedi riquadro 4). Hutchens non aveva fatto ricorso a nessuna tecnica IA e anzi le aveva evitate per dimostrare che, in rapporto all’IA, il concorso Loebner era una pura perdita di tempo. Benché in seguito avesse dichiarato di aver partecipato solo per ottenere “gloria e denaro”, non si può neppure sostenere che lo studente avesse imbrogliato, perché il premio Loebner non ha nessuna regola precisa sul contenuto o sugli intenti dei partecipanti (questa vaghezza, in fondo, riflette la vaghezza del test di Turing). Nonostante il ricorso alle tecniche IA più raffinate, non è che i concorrenti battuti avessero scritto programmi molto brillanti. Ecco uno scampolo di conversazione tra l’esaminatore e il programma Fred 16.43:

Fred: Salve esaminatore 4

Esam: Ciao. Com’è il tempo fuori?

Fred: Non sono stato molto fuori oggi. A me piacciono gli asparagi.

Esam: Quanti anni hai?

Fred: Circa trenta.

Esam: Circa?

Fred: Potrei parlare di Star Trek.

Esam: No, parliamo di te. Quanti anni hai, di preciso?

Fred: Ah, tu sei Batman.

HeX non era molto meglio. Ecco un suo scampolo:

HeX: Vado all’University of Western Australia, che è una delle più belle università d’Australia. È situata sul Fiume dei Cigni, a Perth.

Esam: Che significa “situata”?

HeX. Non ho idea di che cosa “situata” significhi essere.

Esam: Ehm... sei stanco?

HeX: Tutti sanno che sono stanco. E non faccio spesso lo spaccone. Che cosa pensa di quest’affare del concorso Loebner?

Esam: Sta diventando una cosa spassosa, no?

HeX: Eh eh eh... questa sarebbe una “domanda”!? Ha visto qualche bel film di recente?

Si consideri la conversazione che segue: il concorrente è un computer o un essere umano (test di Turing per il lettore)

Esam: Siamo in aprile o in maggio?

Conc: Ehm... credo in aprile.

Esam: Che cosa te lo fa credere?

Conc: Questa è un’ottima domanda. Perché i mesi e i giorni e così via sono arbitrari... credo.

Esam: Mi sembra di notare in te una sfumatura di agitazione. Una sfumatura umana.

Conc: Davvero? Anche lei però è un po’ reticente.

Esam: Non ho mai detto che tu sei reticente.

Conc: Lo so.

Esam: Sei compiaciuto.

Conc: Moi?

Esam: Penso che tu sia un essere umano.

Dato il livello di questi programmi, non ha molta importanza che Hutchens abbia imbrogliato. Dai brani riportati si vede quanto i computer siano ancora lontani dall’intelligenza umana: si ha l’impressione che tutte le conversazioni siano sconnesse, assurde, non *pertinenti*, appunto artificiali. Non c’è dubbio che col tempo le prestazioni miglioreranno, ma finché i programmi resteranno chiusi nel loro mondo virtuale non potranno rispondere a domande sul caldo o sul Fiume dei Cigni. Bisogna aver immerso la mano nell’acqua, aver boccheggionato di sete, essere stati lì per annegare. Turing ha basato il suo gioco solo sulla lingua, cercando di isolare la “purezza” astratta del pensiero, e ciò comporta che si possa smascherare un’intelligenza che non abbia l’esperienza del mondo: basta porre un numero sufficiente di domande ed ecco emergere la differenza *qualitativa* tra l’uomo e il computer, che deriva essenzialmente dalla formazione evolutiva e dall’immersione sistemica dell’uomo, cioè dalla presenza del corpo. In queste condizioni si rileva tutta l’inadeguatezza della locuzione “intelligenza artificiale”: l’accento andrebbe posto molto più sull’aggettivo che sul sostantivo.

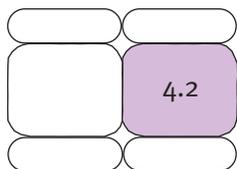




# IL COSTO DELL'IGNORANZA INFORMATICA NELLA P.A. CENTRALE

Pier Franco Camussone

Nella Pubblica Amministrazione centrale italiana i PC sono ormai molto diffusi. Sono ben utilizzati? La ricerca svolta da AICA e SDA Bocconi segnala che chi lavora al computer perde in media 1 h e 10 min ogni settimana per problemi di varia natura, un terzo dei quali dipendono dalla impreparazione informatica degli utenti. L'improduttività per ignoranza informatica ammonta a 270 milioni di € per l'intera Amministrazione centrale. Per ridurre tale onere si è cercato di aumentare la preparazione degli utenti con una formazione basata sui corsi ECDL. I risultati sono stati molto incoraggianti ed il "ritorno della formazione" davvero sorprendente.



## 1. LO SVILUPPO DELLA SOCIETÀ DELL'INFORMAZIONE

L'espressione "società dell'informazione" è un termine che connota una realtà socio-economica in cui gran parte della produzione di servizi dipende dall'elaborazione di informazioni, e anche la produzione e la distribuzione di beni materiali è notevolmente influenzata e resa più efficiente dal trattamento di dati e dalla produzione di informazioni correlate con i beni medesimi. Una società di questo tipo è fortemente dipendente dall'uso delle tecnologie informatiche e telecomunicative di tipo digitale. Per meglio comprendere la genesi del termine "società dell'informazione" bisogna risalire al 1969 quando Alain Touraine (Touraine 1969) fu tra i primi a delineare i contorni di una nuova società post-industriale, il cui assetto produttivo sarebbe stato rivolto alla produzione di beni più immateriali che fisici. Queste idee furono riprese nel 1973 da Daniel Bell (Bell 1973), professore di sociologia alla Harvard University che pubblicò un libro di grande risonanza dal titolo "*The Coming of Post-Industrial Society*", in cui pro-

spettava la nascita di una nuova società fondata sulla produzione di servizi, generati prevalentemente grazie alle nuove tecnologie informatiche e telecomunicative.

Secondo Bell l'economia dell'informazione, che l'autore poneva in contrapposizione all'economia dei beni, non avrebbe soppiantato la società industriale - così come l'avvento dell'industria non aveva distrutto l'agricoltura - ma avrebbe generato grandi cambiamenti e trasformazioni nella società. Lo spostamento delle risorse dalla produzione di beni a quella dei servizi avrebbe modificato grandemente il mondo del lavoro. Il concetto di società dell'informazione, nato dall'intuizione di Bell è stato declinato successivamente in molti modi (Negroponte 1995).

Un filone di pensiero molto diffuso in America collega questo nuovo contesto socio-economico con la realizzazione di grandi infrastrutture abilitanti, come le autostrade informatiche di Clinton e Gore. Questa concezione vede un ruolo fondamentale nella disponibilità di infrastrutture tecnologiche che semplificheranno il lavoro, ne permetteranno

lo svolgimento a distanza e ridurranno i costi di produzione dei servizi "brain intensive". Tutto ciò cambierà il lavoro e la vita degli individui, questi ultimi grazie alla disponibilità abbondante di informazioni potranno fare scelte di vita meno rischiose, trovare più facilmente ciò che desiderano e così via.

Per vivere e lavorare nella società dell'informazione saranno però necessarie nuove conoscenze e nuove competenze. Da questa osservazione si sviluppa la "visione europea". In un famoso documento del 1996, "Green Paper: Living and Working in the Information Society: People First. (COM (96) 389)" la Commissione Europea sottolinea il ruolo critico nello sviluppo della nuova società riservato alla componente umana. Sono le persone che, con il loro comportamento, determineranno l'avvento del nuovo contesto socio-economico, assai di più della disponibilità delle infrastrutture.

In opposizione con alcune previsioni catastrofiche sul futuro della economia, in particolare con le visioni pessimistiche di Jeremy Rifkin (Rifkin 1995), l'Unione Europea osserva che dal 1960 ad oggi non sono scomparsi complessivamente posti di lavoro, anzi ne sono stati creati di nuovi. Il lavoro si è ridistribuito dall'industria ai servizi, e dai settori a basso contenuto tecnologico a quelli a più alta specializzazione. In realtà la distruzione di alcune forme di occupazione c'è stata, ma essa è stata più che compensata dalla creazione di altri lavori prima sconosciuti. Il tramonto di alcune forme di lavoro, soppiantate dall'introduzione dei computer, non avverrà però contemporaneamente alla generazione di nuove opportunità lavorative. Ci potranno essere momenti in cui molte persone dovranno ricollocarsi nel mondo del lavoro. Vecchi posti di lavoro andranno certamente persi, sostituiti da altri che richiedono professionalità superiori, o differenti. Ne consegue che grande attenzione deve essere posta sulla formazione e in particolare su quella che riguarda le nuove tecnologie informatiche e le loro applicazioni. Senza questo sforzo di riqualificazione chi perde il lavoro avrà difficoltà a trovare un nuovo impiego nella società dell'informazione, ma anche chi è occupato rischia di vedere divenire obsolete le proprie competenze.

## 2. IL COSTO DELL'IGNORANZA INFORMATICA NELLA SOCIETÀ DELL'INFORMAZIONE

I lavoratori dell'era digitale dovranno essere quasi tutti alfabetizzati nell'uso del PC e delle applicazioni di uso individuale più comune quali Internet, la posta elettronica, il word processing, il foglio di lavoro elettronico, ecc.. La preparazione adeguata della forza lavoro per la "società dell'informazione", che si desidera costruire in Europa, può rappresentare però il tallone d'Achille di questa strategia di sviluppo (EU Commission 2003). La formazione in ICT non può fermarsi alla fine dei corsi scolastici. Le aziende devono favorire e incoraggiare l'aggiornamento delle conoscenze informatiche. Per i cosiddetti *generic user* l'azienda deve procedere con un intervento di formazione più estensivo, che intensivo, senza curarsi se ciò che viene insegnato possa tornare utile subito nella mansione attualmente svolta dal lavoratore. Insegnando i rudimenti dell'ICT si creano potenziali utenti, che possono più facilmente adattarsi a cambi di mansione e possono intravedere essi stessi miglioramenti organizzativi nelle modalità di svolgimento dei propri compiti (Danziger, Wang, 2000).

E qui si introduce una nuova riflessione. La formazione tecnica sugli strumenti informatici dovrebbe essere affiancata da corsi con un taglio più organizzativo, in cui si illustrino le nuove modalità di lavoro che l'ICT consente (EU Commission 2002). Come osservato da Danziger (Danziger, Jennings e Park 99) gli aumenti di produttività dipendono da come le persone sanno usare gli strumenti informatici, da come rivedono e semplificano il proprio lavoro grazie a questi strumenti, per ricavare spazi di tempo da dedicare ad attività più ricche, che possono determinare un maggior valore aggiunto.

In altri casi non è il risparmio di tempo di lavoro che ci si deve aspettare dall'uso dell'ICT, quanto piuttosto un miglioramento dell'output del lavoro medesimo che risulta più completo e più ricco di contenuti. Le misurazioni dei vantaggi tangibili hanno rilevato che entrambe queste aspettative trovano riscontro nella verifica empirica (Bryniolfsson, Hitt 2003).

Gli studi effettuati nel contesto italiano da AICA



Contesto	% del tempo di lavoro passata al computer	% del tempo di lavoro perso al computer	Costo annuo per addetto del tempo perso	Costo annuo per addetto del tempo perso per ignoranza informatica	Fonte
Aziende USA			3.400 \$	2.600 \$	Kierwin, Mieritz 2002
Aziende scandinave	60%	7,13%	3.209 €	2.567 €	Nielsen 2002
Imprese italiane			2.564 €	885 €	Camussone, Occhini 2003
Settore sanità italiana (impiegati)	67%	7,90%	2.842 €	855 €	Borgonovi, Camussone, Occhini 2004
Settore bancario italiano	80%	2%	1.124 €	237 €	Camussone, Occhini, Santececca 2006

**TABELLA 1**

*Il costo della improduttività per chi lavora al computer*

sono stati riportati i valori calcolati da AICA e SDA Bocconi nelle precedenti campagne di ricerca che hanno riguardato il mondo del lavoro italiano (Camussone, Occhini 2003; Camussone, Occhini, Santececca 2006; Borgonovi, Camussone, Occhini 2004).

Come si può vedere nella tabella 1 il costo annuo per dipendente, derivante dalla perdita di tempo sul computer, è significativo e varia dai 1.124 € del settore bancario agli oltre 2.800 € del settore impiegatizio del comparto sanitario. Anche la parte di questo onere di improduttività che si fa risalire all'ignoranza informatica è apprezzabile, dall'ordine di parecchie centinaia di euro all'anno per ogni dipendente. Proseguendo sulla strada già tracciata i ricercatori di AICA e della SDA Bocconi hanno cercato di rilevare il costo della ignoranza informatica nel settore pubblico, ed in particolare nella PA centrale, che rispetto all'intero settore si presenta con caratteristiche più omogenee.

#### **4. L'USO DEL COMPUTER NELLA PUBBLICA AMMINISTRAZIONE CENTRALE**

Anche se nell'immaginario collettivo il lavoro di un impiegato pubblico è spesso associato ad archivi polverosi, pratiche cartacee che si accumulano sulla scrivania, o passano di mano da una persona all'altra, nella realtà tutti i dipendenti pubblici che ne hanno bisogno di-

spongono di un posto di lavoro computerizzato. Nella PA il rapporto tra computer installati e dipendenti informatizzabili è di 1,3. Secondo quanto rilevato dai ricercatori, il 51% degli impiegati pubblici passa più del 60% del proprio tempo lavorativo a contatto con un computer. Addirittura un 14% della forza lavoro spende più dell'80% del proprio tempo lavorativo sul computer. Nell'insieme chi lavora nella PA centrale ha dichiarato di spendere mediamente il 61% del proprio tempo lavorativo sul PC: si tratta di un valore certamente significativo e che deve mettere in allarme circa la appropriata capacità d'uso dello strumento da parte dei lavoratori (Figura 1).

Infatti, una scarsa dimestichezza nell'uso degli strumenti informatici andrebbe ad interessare il 61% del tempo di lavoro, determinando un calo di produttività marcato.

Il campione dei dipendenti soggetti all'indagine (780 rispondenti) ha inoltre dichiarato di lavorare sia con specifiche procedure predisposte dalla Amministrazione per lo svolgimento delle proprie mansioni, sia con gli strumenti di produttività individuale quali l'elaborazione dei testi (*word processor*), il foglio di lavoro elettronico, gli strumenti grafici per le presentazioni (*effective presentation*), la posta elettronica, Internet, gli strumenti per la gestione dei dati (DB).

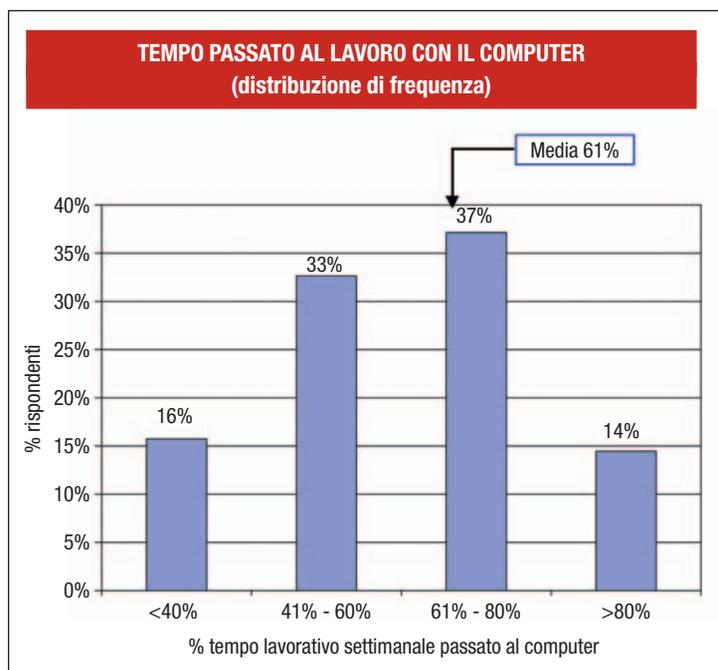
Sulla base di quanto dichiarato dagli utenti, Internet, la posta elettronica ed il *word pro-*

cessor sono usati quotidianamente da oltre il 70% degli intervistati. Le applicazioni predisposte *ad hoc* sono usate tutti i giorni da quasi il 50% degli utenti, mentre il foglio di lavoro elettronico è usato almeno qualche volta la settimana da oltre il 70% degli utenti.

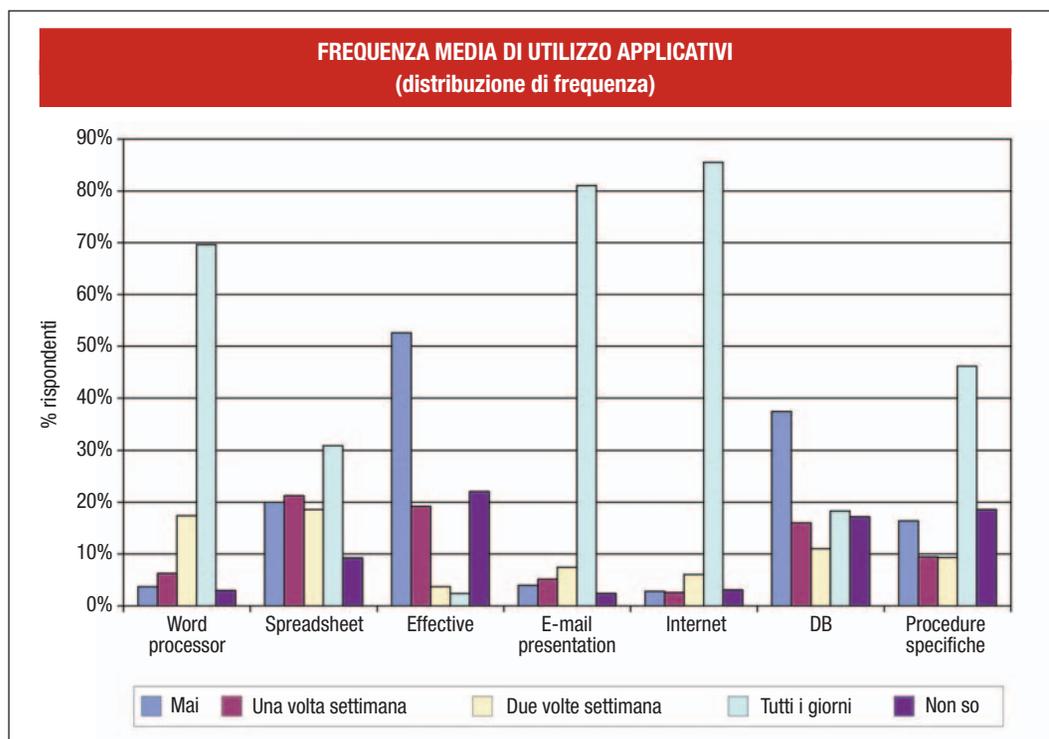
Meno diffuse nell'utilizzo, sono risultate le applicazioni di gestione dei dati e quelle di *effective presentation* (Figura 2).

## 5. L'IMPRODUTTIVITÀ DERIVANTE DA PROBLEMI INFORMATICI

A tutti gli utenti che hanno partecipato alla ricerca è stato chiesto di indicare quanto tempo essi "perdevano" durante il lavoro per problemi con i computer, o con le reti. I rispondenti al questionario sono stati 780. La figura 3 mostra che a parte una minoranza (11%) che addirittura denuncia più di 240 min persi per ogni settimana lavorativa, nella maggior parte dei casi tale perdita di tempo non supera i 60 min. Tuttavia elaborando le risposte si ricava come media pesata un valore di 69 min per settimana. L'escursione di questo valore da un minimo ad un massimo molto elevato denota scarsa omogeneità del contesto: la variabilità potrebbe derivare da condizioni di funzionamento della infrastruttura assai instabili, oppure da livelli di conoscenza informatica degli utenti molto disuguali. Il tempo improduttivo deriva fondamentalmente da tre diverse cause: motivazioni tecniche (come guasti dell'hardware e malfun-



**FIGURA 1**  
Percentuale del tempo di lavoro in cui si usa il computer nella PA centrale



**FIGURA 2**  
Frequenza d'uso degli strumenti informatici nella PA centrale

zionamenti del software), scarsa conoscenza degli strumenti da parte dell'utente e perdite di tempo per effetto di una richiesta di aiuto da parte di colleghi in difficoltà nell'uso del proprio computer. Nel nostro caso la ricerca ha indicato che le cause tecniche incidono per il 65% sul totale del tempo perso, mentre l'incompetenza informatica causa il restante 35% della perdita di produttività (Figura 4).

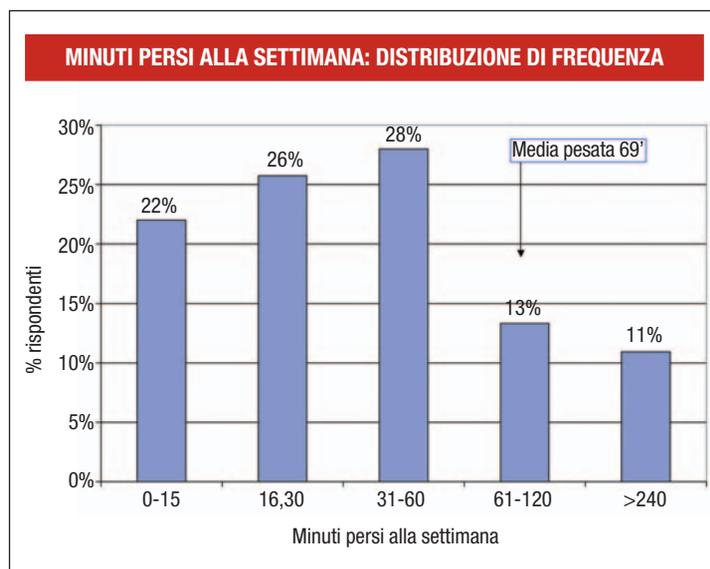
Si può ora procedere alla valorizzazione del costo della improduttività durante il lavoro con il computer. Se partiamo dal costo annuale medio di un dipendente della PA centrale (45.000 € l'anno, stando ai valori forniti dalla Ragioneria Generale dello Stato) possiamo valorizzare il tempo che i dipendenti perdono lavorando al computer. Essi perdono circa 1 ora e 10 minuti alla settimana pari al 3,2% del tempo lavorativo settimanale che è di 36 h. A tale perdita corrisponde un valore annuale di 1.439 € di costo aziendale (Tabella 2).

In realtà, tuttavia, il costo dell'ignoranza è rappresentato da una frazione del tempo perso (35%): quella che corrisponde alla non conoscenza degli strumenti da usare e all'aiuto prestato a colleghi che si trovano in difficoltà. Pertanto è possibile calcolare il costo della improduttività derivante dagli strumenti informatici e la quota parte causata dalla ignoranza informatica. Siamo ora in grado di confrontare la situazione rilevata nella Pubblica Amministrazione centrale italiana con la realtà riscontrata in altri settori economici (Tabella 3).

Si può anche calcolare il costo complessivo per l'intera PA centrale derivante dalla impreparazione informatica. Si tratta di un valore che dipende dall'area presa in considerazione: per l'area dei Ministeri degli Enti centrali la perdita è di circa 150 milioni di euro all'anno, mentre se si prende in considerazione un insieme più ampio, come quello a cui si riferi-

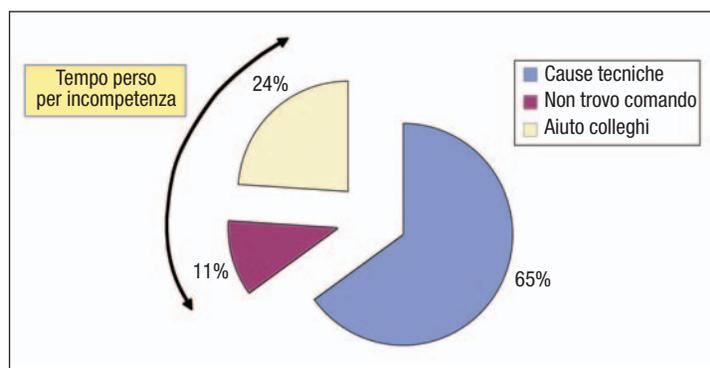
sce il CNIPA, il costo si avvicina ai 280 milioni euro l'anno (Tabella 4).

L'analisi svolta dai ricercatori ha messo in evidenza che i dipendenti della Pubblica Amministrazione centrale usano parecchio i computer: per circa il 60% del proprio tempo di lavoro essi sono alle prese con strumenti



**FIGURA 3**

*Tempo perso settimanalmente per problemi con il computer*



**FIGURA 4**

*Le cause che determinano il tempo perso settimanalmente per problemi con il computer*

Costo medio aziendale (euro)	Tempo di lavoro settimanale (ore)	% del tempo di lavoro passato al computer	Tempo di lavoro settimanale al computer (ore)	Tempo perso settimanalmente al computer (minuti)	% del tempo lavorativo perso settimanalmente al computer	Valore del tempo perso annualmente al computer per addetto (€)
45.035	36,0	61%	21,96	69	3,2%	1.439

**TABELLA 2**

*Valore del tempo improduttivo (perso) lavorando al computer*

Contesto	% del tempo di lavoro passata al computer	% del tempo di lavoro persa al computer	Costo annuo per addetto del tempo perso	Costo annuo per addetto del tempo perso per ignoranza informatica
Aziende USA			3.400 \$	2.600 \$
Aziende scandinave	60%	7,13%	3.209 €	2.567 €
Imprese italiane			2.564 €	885 €
Settore sanità italiana (impiegati)	67%	7,90%	2.842 €	855 €
Settore bancario italiano	80%	2%	1.124 €	237 €
Settore pubblico italiano (PA centrale)	61%	3,20%	1.439 €	504 €

**TABELLA 3**

Valore del tempo improduttivo (perso) lavorando al computer e stima della quota parte imputabile a ignoranza informatica: la Pubblica Amministrazione a confronto con altri contesti economici

Definizione	N° addetti	Valore del tempo perso annualmente al computer per addetto (€)	Costo per il settore dei problemi con il computer (€)	% tempo perso per incompetenza	Costo dell'ignoranza informatica per addetto (€)	Costo dell'ignoranza informatica per il settore (€)
PA intesa in senso restrittivo	300.000	1.439	431.584.417	35%	504	151.054.896
PA intesa in senso lato	550.000	1.439	791.239.931	35%	504	276.933.976

**TABELLA 4**

Il costo complessivo della improduttività del lavoro al computer nella Pubblica Amministrazione centrale italiana

informatici. Nella PA ci sono più PC dei dipendenti informatizzabili. Per migliorare l'efficienza della PA non si tratta più di investire in tecnologie nei posti di lavoro (salvo mantenere aggiornate le postazioni individuali) quanto piuttosto di assicurarsi che gli strumenti disponibili siano proficuamente utilizzati dai dipendenti.

## 6. GLI EFFETTI DELLA FORMAZIONE

L'ipotesi che i ricercatori hanno formulato all'inizio della ricerca riguarda l'esistenza di una relazione di causa-effetto tra preparazione informatica e capacità di usare meglio gli strumenti informatici. Per conseguenza ad una maggiore conoscenza degli strumenti informatici dovrebbe corrispondere una più elevata produttività individuale e aziendale. L'ipotesi alla base della ricerca potrebbe essere

formulata in modo più esplicito come segue:

**a.** la formazione informatica può aumentare il livello delle conoscenze informatiche e contribuire pertanto alla riduzione del cosiddetto costo "dell'ignoranza informatica" nel settore pubblico;

**b.** l'aumento delle conoscenze informatiche, determinato dalla formazione, può aumentare l'efficienza delle persone che lavorano al computer (riduzione del tempo di svolgimento del lavoro) aumentando la produttività individuale e - per conseguenza - l'efficienza del settore pubblico nel suo complesso.

Queste ipotesi sono già state sottoposte a vaglio nelle precedenti ricerche sul costo della ignoranza informatica (Camussone, Occhini 2003; Borgonovi, Camussone e Occhini 2004, Camussone, Occhini, Santececca 2006). Nella prima è stato misurato l'effetto della formazione informatica sui lavoratori di concetto e sui manager appartenenti ad un gruppo di imprese in-

dustriali, nella seconda invece si è ripetuta la misurazione su un campione di soggetti appartenenti al settore della sanità e nella terza ricerca ci si è focalizzati sul settore bancario.

I risultati ottenuti in precedenza hanno provato che effettivamente sussiste una relazione di causa effetto tra la formazione informatica e l'aumento di produttività di chi lavora al computer e che l'incremento delle conoscenze informatiche riduce il tempo perso al computer a causa della ignoranza informatica.

Questa volta si è cercato di ripetere la verifica delle ipotesi della ricerca con riferimento al personale della PA centrale, rilevando in modo empirico i "ritorni" della formazione informatica. A tal fine sono stati condotti degli esperimenti di formazione sul personale di alcuni Ministeri (Ministero dell'Economia e delle Finanze, Ministero dei Beni Artistici e Culturali, Ministero della Giustizia). Sono state formate più classi costituite da una quindicina, o poco più, di impiegati per un totale di 230 soggetti, a cui è stato erogato un corso preparatorio al superamento della prova per la patente europea di guida del computer (ECDL). L'analisi degli effetti della formazione sull'uso degli strumenti informatici è stata effettuata misurando la capacità dei soggetti nella risoluzione di problemi applicativi prima e dopo la frequenza al corso e il conseguimento della certificazione ECDL. Come è noto la patente ECDL riguarda la capacità di utilizzare in modo appropriato il personal computer come strumento di lavoro individuale. Si è trattato quindi di misurare i progressi dei partecipanti ai corsi nella capacità di saper lavorare con i seguenti strumenti:

- il *word processor*;
- il foglio di lavoro elettronico (*spreadsheet*);
- il software di *effective presentatio*;
- *Internet*;
- la posta elettronica (*e-mail*), e
- gli strumenti di gestione dei dati (*DB*).

Un medesimo test, costituito da una trentina di domande riguardanti la soluzione di problemi pratici nei 6 ambienti precedentemente citati, è stato affrontato dai partecipanti prima e dopo il corso. Per ogni soggetto e per ogni ambiente sono stati misurati due risultati:

**1.** il punteggio conseguito nel test, che rileva il livello di padronanza, ovvero di conoscenza, della tematica;

**2.** il tempo impiegato nella soluzione del problema prospettato dal test.

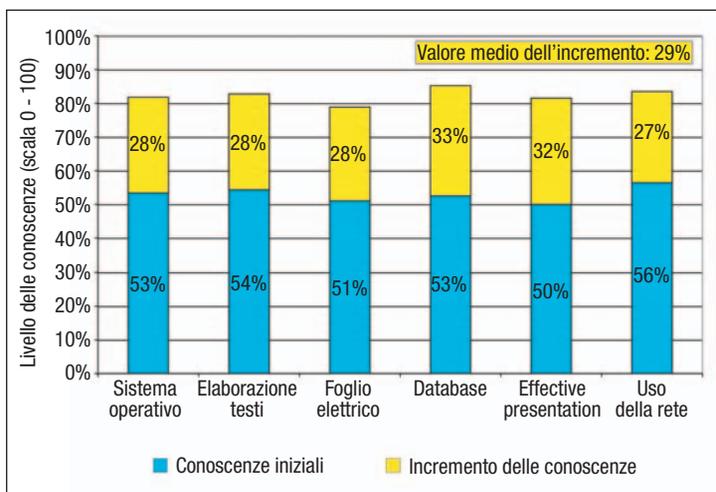
## 7. IL TEST NEI MINISTERI

Per la precisione il campione sottoposto al test della formazione è risultato formato da 29 dipendenti del Ministero dei Beni Artistici e Culturali, 21 del Ministero della Giustizia, e 180 del Ministero dell'Economia e delle Finanze.

Prima del corso, il livello delle conoscenze informatiche del campione era già piuttosto elevato, posizionandosi per i diversi moduli tra il 50% ed il 60% del valore massimo della scala, con un valor medio complessivo di 53/100. Questo valore di conoscenze iniziali sta ad indicare un elevato livello di alfabetizzazione del campione sottoposto ai test. In altri termini i dipendenti della Pubblica Amministrazione centrale dispongono già di una base di conoscenze che consente loro di usare gli strumenti informatici. In questo caso i corsi ECDL dovrebbero servire a sistematizzare e completare le conoscenze, che spesso si sono sviluppate in base ad autoistruzione e ad esperienze maturate sul campo. D'altro canto non è pensabile che persone che trascorrono circa il 61% del proprio tempo lavorativo davanti ad un computer possano essere del tutto impreparate nell'uso del software di produttività personale.

Alla fine del corso sono stati ripetuti i test (naturalmente non i medesimi, ma prove analoghe e comparabili) e il livello di conoscenza media è risultato pari a 82/100 con un incremento di ben 29,3 punti su una scala da 0 a 100 (Figura 5).

Si tratta di un risultato molto interessante soprattutto se comparato con i risultati delle precedenti campagne di ricerca. Infatti, mentre nel caso dell'industria e della sanità si partiva da un livello di conoscenze informatiche inferiore, nel caso del settore bancario il campione disponeva invece di un livello di preparazione superiore a quello della Pubblica Amministrazione. In tutti i casi precedenti l'incremento del livello di conoscenza è risultato compreso tra i 16 e i 20 punti percentuali, in dipendenza dal livello di conoscenza di partenza. Nel caso dei dipendenti pubblici si è riscontrato il maggior effetto assoluto derivante da corsi ECDL. Una spiegazione plausibile po-



**FIGURA 5**  
Stato delle conoscenze dei diversi moduli ECDL prima e dopo il corso

Settori	Livello delle conoscenze iniziali	Livello delle conoscenze finali	Variatione
Aziende Italiane	40	60	+20
Settore Sanitario	25	43	+18
Settore Bancario	77	93	+16
Settore Pubblico	53	82	+29

**TABELLA 5**  
Gli effetti della formazione sul livello delle conoscenze informatiche (scala delle conoscenze: 0 - 100)

trebbe essere la seguente: la formazione informatica (o meglio i corsi ECDL), producono gli effetti più rilevanti quando i soggetti hanno già un certo grado di padronanza degli strumenti informatici di produttività individuale. Quando il livello di partenza delle conoscenze è piuttosto basso l'incremento di conoscenze è più contenuto, perché anzitutto bisogna colmare delle lacune di base. Anche quando il livello di partenza è molto elevato (come nel caso del settore bancario) l'incremento di conoscenza è più contenuto, in quanto ci si avvicina ai massimi della scala di misurazione, per cui è difficile individuare eventuali lacune da colmare. Nel caso della pubblica amministrazione centrale i corsi ECDL producono il maggiore tra i benefici finora misurati (Tabella 5), proprio perché le conoscenze di partenza sono approssimativamente vicine al 50% del massi-

mo raggiungibile. Si ha pertanto a che fare con una "classe" di partecipanti già preparati, che quindi conosce gli strumenti oggetto di approfondimento, ma che ha ancora parecchio da imparare.

## 8. GLI EFFETTI DELLA FORMAZIONE SUL TEMPO DI LAVORO CON IL COMPUTER

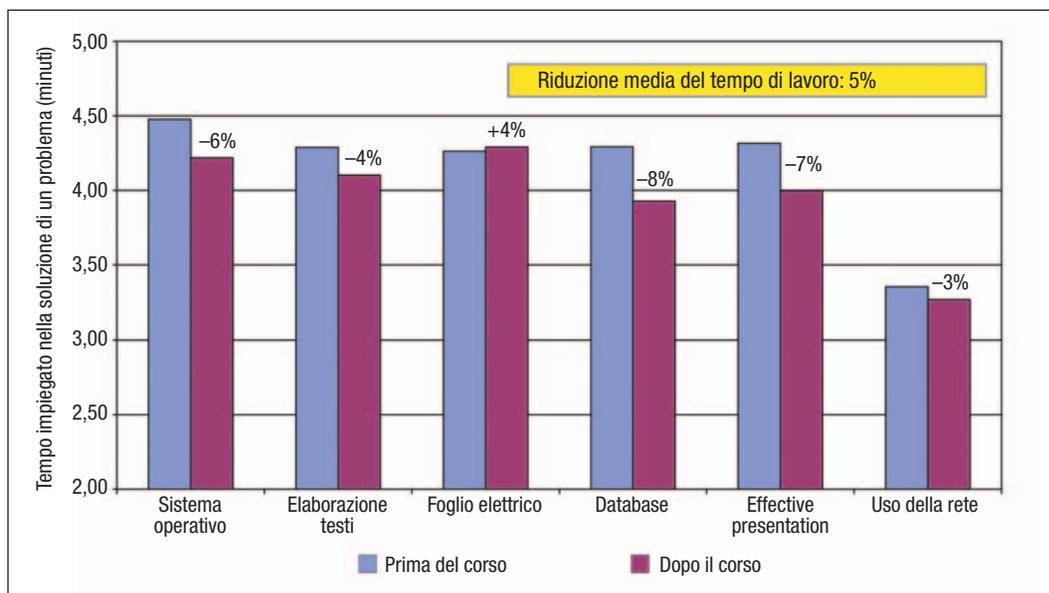
L'aumento di conoscenze informatiche, determinato dall'aver frequentato il corso ECDL, può produrre positive conseguenze sulle modalità di svolgimento del lavoro con il computer. Come osservato da Camussone (Camussone, Occhini 2003) il personale "formato":

1. fa meglio le attività che precedentemente gli creavano difficoltà;
2. svolge lavori che prima non sapeva svolgere;
3. impiega meno tempo (grazie alle nozioni apprese) a svolgere il proprio lavoro al computer;
4. perde meno tempo davanti al computer per problemi derivanti dalla sua ignoranza informatica.

Anche nel caso dell'esperimento effettuato con la Pubblica Amministrazione sorge la curiosità di sapere se i ritorni della formazione sul tempo di svolgimento del lavoro al computer siano -almeno in parte- misurabili e a quanto ammontino. Pur riconoscendo che alcuni dei vantaggi sopra citati sono di natura qualitativa, e quindi difficilmente quantificabili in termini monetari, seguiremo la strada tracciata da coloro che si sono già cimentati in questa impresa (Danzinger, Jeming e Park 1999).

Nel nostro caso è stato rilevato il tempo che i partecipanti hanno impiegato per risolvere i problemi previsti dai test prima e dopo aver frequentato il corso ECDL (Figura 6). Come si può constatare per cinque moduli su sei si è effettivamente riscontrata una diminuzione del tempo, mentre nel caso del foglio di lavoro elettronico i tempi si sono allungati. Quest'ultimo fatto può essere spiegato con una comprensione maggiore delle potenzialità dello strumento e quindi con un corrispondente aumento del tempo dedicato alla definizione di una soluzione migliore e più sofisticata rispetto a quella ipotizzata prima del corso.

Comunque dalle misure empiriche sul campione di individui che ha affrontato i test si è rilevata una riduzione del 5% nel tempo di



**FIGURA 6**  
*Tempo impiegato nella risoluzione di problemi prima e dopo il corso*

svolgimento delle varie attività sul computer per effetto del corso di formazione frequentato. A questo punto possiamo domandarci quali effetti si determinerebbero se questo risultato si potesse estrapolare alla realtà lavorativa di tutti gli addetti che usano l'informatica nella Pubblica Amministrazione centrale; in questo modo potremmo valutare gli effetti economici della formazione, ovvero il cosiddetto "ritorno della formazione".

Gli addetti del settore pubblico hanno tempi contrattuali di lavoro di 36 h la settimana, e dichiarano in media una percentuale di tempo lavorativo speso sul computer del 61%. Pertanto un aumento della produttività del lavoro con il computer del 5% incide su questa porzione della loro attività.

Volendo valorizzare su base annua tale incremento di produttività si deve tenere presente che il contratto di lavoro del pubblico impiego prevede 7,2 h di lavoro giornaliera e che, secondo i dati forniti dal Ministro per la Pubblica Amministrazione e l'Innovazione, nel pubblico impiego le giornate effettivamente passate annualmente al lavoro da un dipendente sono circa 195.

Sulla base di questi elementi e del costo medio di un dipendente pubblico si giunge a determinare in 1.374 € l'anno il valore dell'aumento di produttività derivante da un intervento di formazione informatico come l'erogazione del corso ECDL. Estrapolando al complesso dell'intera organizzazione della

Pubblica Amministrazione centrale quanto calcolato per un addetto si deduce che si potrebbe avere per la Pubblica Amministrazione centrale, intesa in senso ristretto un incremento di produttività di 412 milioni di euro, e nel caso si intenda la Pubblica Amministrazione centrale in senso più esteso (come interpretato da CNIPA) addirittura un incremento quantificabile in 755 milioni di euro ogni anno (Tabella 6).

I benefici tangibili della formazione informatica, oltre a consistere in un aumento di produttività nel lavoro svolto con il computer, riguardano anche la riduzione del tempo perso al computer per ignoranza sull'uso dello strumento, fenomeno questo che in letteratura (Camussone, Occhini 2003) è indicato con l'espressione "costo dell'ignoranza informatica". Nel caso del settore della Pubblica Amministrazione tale perdita di tempo è stata quantificata in 504 € all'anno per ogni impiegato che usa il computer (Tabella 4). Tale valore può ridursi per effetto della formazione informatica, anche se non si può pensare che tutti i dipendenti raggiungano una perfetta conoscenza degli strumenti informatici, tale da evitare qualsiasi effetto derivante dalla "ignoranza informatica".

Dal momento che i test effettuati sul campione dei dipendenti pubblici hanno mostrato un incremento medio del livello delle conoscenze informatiche del 29% (Figura 5) si può ragionevolmente ritenere che di una per-

Categorie del personale	N° addetti che usano l'informatica	Tempo medio di lavoro settimanale (ore)	Tempo di uso settimanale del computer (in ore e in percentuale del tempo lavorativo)	Aumento settimanale di produttività (+5%) in ore	Aumento annuale di produttività (giorni)	Costo annuale pro capite (€)	Costo giornaliero (€)	Valore annuale dell'aumento di produttività pro capite (€)	Valore annuale dell'aumento di produttività della categoria (€)
Personale PA centrale (in senso stretto)	300.000	36	22 61%	1,10	5,9	45.035	232	1.374	412.070.250
Personale PA centrale (nell'accezione CNIPA)	550.000	36	22 61%	1,10	5,9	45.035	232	1.374	755.462.125

NB: I costi giornalieri sono riferiti ad un valore di 194 giorni che rappresentano la media della presenza effettiva annua al lavoro nel settore della PA centrale.

**TABELLA 6**

Calcolo del miglioramento della produttività nel settore della Pubblica Amministrazione centrale per effetto della formazione informatica

Personale	N° addetti	Valore annuale del tempo perso per impreparazione informatica pro capite (€)	Percentuale del tempo perso recuperabile con la formazione	Valore annuale del tempo perso per impreparazione pro capite recuperabile con la formazione (€)	Valore annuale del tempo perso per impreparazione informatica recuperabile con la formazione nell'intero settore (€)
Personale PA centrale (in senso stretto)	300.000	504	29%	146	43.848.000
Personale PA centrale (nella accezione CNIPA)	550.000	504	29%	146	80.388.000

**TABELLA 7**

Calcolo del miglioramento della riduzione del costo della impreparazione informatica nella PA centrale per effetto della formazione informatica

centuale equivalente si riducano le occasioni di perdita di tempo a causa della ignoranza informatica. Si può quindi calcolare a che cosa equivale, in termini economici, una riduzione del 29% del tempo perso per ignoranza informatica dai dipendenti del settore pubblico (Tabella 7).

Il ritorno economico della formazione informatica, per quanto concerne la riduzione del tempo perso per ignoranza, non è comparabile con quanto si ottiene come miglioramento nella produttività del lavoro sul computer. Nel caso degli impiegati della Pubblica Amministrazione centrale si otterrebbe un beneficio circa 146 € l'anno, a fronte di un aumento di produttività di 1.374 €.

La ricerca empirica sui ritorni della formazione informatica basata sul corso ECDL porta quindi alla seguente conclusione nel caso del settore della PA centrale: gli addetti sono

già abbastanza "alfabetizzati" dal punto di vista informatico, e quindi i benefici maggiori della formazione non derivano dalla riduzione della "ignoranza informatica", quanto piuttosto dall'aumento di produttività nell'uso del computer, che consegue ad un maggior livello di conoscenza degli strumenti software. Nella tabella 8 sono riportati i valori economici dei risparmi di tempo derivanti dalla formazione e come si può vedere il risparmio di tempo nell'uso del computer (aumento di produttività pro capite) è di un ordine di grandezza di circa 10 volte superiore al valore della riduzione del tempo perso per incompetenza informatica.

Anche nel caso del settore pubblico (come nei precedenti momenti della ricerca rivolti al settore produttivo in generale e al settore sanitario e bancario in particolare) appare evidente il ritorno economico degli investimenti in forma-

Categorie del personale	N° addetti	Valore annuale dell'aumento di produttività pro capite (€)	Valore annuale del tempo perso per impreparazione informatica pro capite recuperabile con la formazione (€)	Valore annuale del ritorno complessivo della formazione per ogni addetto (€)	Valore annuale del ritorno complessivo della formazione per l'intero settore (€)
Personale PA centrale (in senso stretto)	300.000	1.374	146	1.520	455.918.250
Personale PA centrale (nella accezione CNIPA)	550.000	1.374	146	1.520	835.850.125

**TABELLA 8**

*Calcolo della ritorno complessivo della formazione informatica (ECDL) nel settore della PA centrale*

zione. A fronte di costi da sostenere una tantum per un corso di formazione ECDL, che ammontano a qualche centinaio di euro, si prospettano risparmi di costo ed aumenti di produttività di qualche migliaio di euro all'anno.

## 9. CONCLUSIONI

Sulla base dei test empirici e sulla scorta dei dati raccolti sembra possibile trarre alcune significative conclusioni circa gli effetti della formazione informatica nella Pubblica Amministrazione centrale.

**1°.** La formazione: contribuisce a migliorare il livello delle conoscenze informatiche in modo significativo (29 punti di incremento delle conoscenze su una scala da 0 a 100 quando il valore di partenza si attesta attorno a 50 punti).

**2°.** C'è indubbiamente un effetto di riduzione del tempo di esecuzione del lavoro, ma questa diminuzione non è di dimensioni vistose, secondo quanto indicato dai test empirici, che individuano in un 5% il risparmio di tempo nella soluzione del problema. Sembrerebbe logico aspettarsi che coloro che hanno raggiunto un livello più elevato di conoscenza degli strumenti informatici impieghino molto meno tempo per svolgere un lavoro che richiede l'uso degli strumenti medesimi. Quale spiegazione si può ipotizzare se non si verifica un elevato risparmio di tempo? Per prima cosa si deve sottolineare che i test non erano concepiti per premiare coloro che impiegavano meno tempo nell'esecuzione. Trattandosi di prove di esame vi era un tempo limite e i partecipanti sono portati abitualmente a usufruire di tutto l'intervallo di tempo loro concesso. In ogni caso, se vogliamo riflettere sui risultati della ricerca e trovare

una spiegazione più appropriata possiamo formulare le seguenti considerazioni.

□ Il risparmio di tempo nell'esecuzione del lavoro si verifica quando l'interessato ha sviluppato una discreta esperienza nell'uso degli strumenti, non quando ne sta apprendendo l'uso. In questa circostanza il lavoratore è ancora in fase di "metabolizzazione" delle nuove tecnologie, cerca di usarle bene, di non sbagliare e sta sviluppando le proprie modalità d'uso che, con l'esperienza, si tradurranno in automatismi di comportamento (quasi dei riflessi condizionati) che lo porteranno ad accelerare molto i processi di uso degli strumenti in questione.

□ Coloro che partono da livelli bassi di conoscenza iniziale sono molto attenti all'uso corretto delle tecnologie appena apprese e quindi il loro comportamento determina addirittura un allungamento dei tempi di test. Va tuttavia osservato che il tempo iniziale era anche influenzato dalla loro ignoranza, per cui abbandonavano velocemente un compito se non sapevano risolverlo. Mentre a fine corso si impegnano per svolgerlo fino in fondo.

□ Chi invece partiva da livelli medio-alti di conoscenza degli strumenti e quindi -probabilmente- aveva anche prima la capacità di risolvere in qualche modo il problema presentato dal test, riesce a fine corso a svolgere il compito in un tempo inferiore.

□ Per i più esperti, invece, può essersi verificato un fenomeno già noto in letteratura (Danzinger 2000). Con l'aumento delle conoscenze cresce il desiderio di rifinire meglio il lavoro di completarlo in modo elegante e così via, portando ad un allungamento dei

tempi di lavoro, cui però si accompagna un risultato finale di qualità superiore.

3°. I benefici della formazione sono apprezzabili sia sotto il profilo economico, che sotto quello motivazionale (dei dipendenti). Per il primo aspetto si stima in circa 1.500 € l'anno (Tabella 8) il valore del ritorno per ogni addetto derivante dall'investimento fatto su di lui in formazione (1374 € rappresentano il valore dell'aumento di produttività conseguente e 146 € il costo evitato per incapacità nell'uso degli strumenti informatici).

Per quanto riguarda l'aspetto motivazionale è significativo quanto dichiarato nel survey dagli utenti. Secondo la loro percezione:

- la formazione contribuisce ad aumentare l'efficacia del lavoratore nello svolgimento dei propri compiti (riduzione del tempo di esecuzione del lavoro, aumento della qualità del risultato, capacità di saper individuare nuove applicazioni degli strumenti ICT);
- la formazione crea le condizioni per una maggior soddisfazione sul lavoro (diminuisce il senso di frustrazione che deriva dall'incapacità di usare appropriatamente gli strumenti informatici, e dalla sensazione di "perdere" tempo con il computer).

Per tutte queste ragioni la formazione informatica dovrebbe essere presa in maggior considerazione da parte degli Enti e delle aziende del settore pubblico, notoriamente alla ricerca di miglioramenti di efficienza e di produttività. Si ricorda in proposito che nella PA centrale solo un addetto su quattro è in possesso di una certificazione ECDL (*European Computer driving Licence*), o di livello equivalente, circa l'uso del computer.

La ricerca empirica svolta conferma quanto ci si poteva aspettare in termini qualitativi come ritorno della formazione informatica. Valutare tale ritorno però in termini strettamente monetari non è semplice. Spesso non si approda a risultati incontrovertibili mentre talvolta si trovano risultati sorprendenti. Prendiamo ad esempio i risultati del nostro esperimento basato sul costo di un corso ECDL, che può ammontare a circa 200 – 300 € per ogni partecipante. A fronte di un costo una tantum di questo ordine di grandezza i ritorni, come illustrato nei paragrafi precedenti, sono dell'ordine dei 1.500 € per ogni addetto "formato" all'anno. Se potessimo usare le normali formule sul

calcolo del ritorno di un investimento (ROI) potremmo considerare come ritorno il valore di aumento della efficienza produttiva quantificato almeno per un triennio che andrebbe raffrontato con il costo sostenuto per ottenere tale risultato. Ne deriverebbe un valore assolutamente eccezionale pari al 1500%:

$$\frac{\text{Valore del beneficio netto}}{\text{Valore del costo sostenuto}} = \frac{1500 \times 3}{300} = 15$$

A ciò deve essere aggiunta una valutazione dei benefici di tipo "soft" derivanti dalla formazione, quali ad esempio la maggior soddisfazione sul lavoro da parte di chi usa in modo più consapevole gli strumenti informatici, la crescita del potenziale esprimibile sul lavoro da parte di chi ha accresciuto il proprio livello di conoscenza, la sensazione di sicurezza nel risolvere problemi prima ritenuti troppo difficili e così via.

## BIBLIOGRAFIA

- [1] Baily Martin Neil, Lawrence Robert Z.: *Do We Have a New E-economy*. Presented at the American Economic Association Meetings, New Orleans, LA, January 5, 2001.
- [2] Bell D.: *The Coming of Post-Industrial Society*. Basic Book, 1973.
- [3] Bemelmans T. M. A.: *Beyond Productivity: Information Systems Development for Organizational Effectiveness*. Proceedings of the IFIP WG 8.2 Working Conference on Beyond Productivity: Information Systems Development for Organizational Effectiveness, Minneapolis, Minnesota, U.S.A., 22-24 August, 1983.
- [4] Berman Evan M.: *Productivity in Public and Nonprofit Organizations*. Strategies and Techniques; Sage Publications, 1998
- [5] Borgonovi, Camussone, Occhini: *Il costo dell'ignoranza nella sanità*. ETAS 2004.
- [6] Brynjolfsson E., Hitt L.: *Productivity, profit and consumer welfare: the different measure of IT's value*. MIS Quarterly 1997.
- [7] Brynjolfsson E., Hitt L.M.: *Computing productivity: firm-level evidence*. MIT Sloan working paper June 2003
- [8] Camussone, Occhini: *Il costo dell'ignoranza nella società dell'informazione*. ETAS 2003.
- [9] Camussone, Occhini, Santececca: *Competenze digitali e produttività nel settore bancario*. McGraw-Hill 2006.

- [10] Censis 2006: *Lo stato dell'informatizzazione nella Pubblica Amministrazione*.
- [11] Danziger J.N., Jenning J.A., Park S.C.: *ICT Training; Center for research on ICT and Organizations*. University of California Irvine, 1999.
- [12] Danziger J.N., Wang Y.C.: *Enhancing end users' ICT skills in the new economy; Center for research on ICT and Organizations*. University of California Irvine, 2000.
- [13] Dedrick J., Gurbaxani V., Kraemer K.: *Information Technology and Economic Performance: A Critical Review of the Empirical Evidence. ACM Computing Surveys*, Vol. 35, n. 1, 2003.
- [14] EU Commission: *Green Paper: Living and Working in the Information Society: People First*. Commission Staff Working Document COM, Vol. 96, 1996, p. 389.
- [15] EU Commission: *Information Society jobs - quality for change - Exploiting the Information Society's contribution to managing change and enhancing quality in employment*. Commission Staff Working Document; SEC, 2002, p. 372.
- [16] EU Commission: *ICT and Social Capital in the Knowledge Society*. Report on a Joint DG JRC/DG EMPL Workshop (IPTS, Sevilla, 3-4/11/2003).
- [17] Florian L., Hegarty J.: *ICT and Special Educational Needs a Tool for Inclusion*. McGraw-Hill International, 2004
- [18] Kierwin B., Mieritz L.: *TCO and performance management in architectural choices*. Research Note, Gartner July 2002.
- [19] Jorgenson Dale, Ho Mun, Stiroh Kevin: *Lessons from the U.S. Growth Resurgence*. (January 17, 2003), presented at the First International Conference on the Economic and Social Implications of Information Technology, Department of Commerce, Washington DC, January 27-28, 2003.
- [20] March and Stiroh, K. J.: *Reassessing the Impact of IT in the Production Function: A Meta-Analysis*. Federal Reserve Bank of New York, (November 4 2002).
- [21] McKinsey Global Institute: *US Productivity Growth. 1995-2000*, Washington, DC (October 7 2001).
- [22] Negroponte N.: *Being Digital*. Knopf 1995.
- [23] Nielsen B. (Cap Gemini-Ernst Young): *Documentation of hidden cost in Norwegian workforce*. ECDL-CEO meeting Cyprus 2002.
- [24] Rifkin J.: *The End of Work: The Decline of the Global Labor Force and the Dawn of the Post-Market Era*. Putnam Publishing Group, 1995. (Trad. It: *La fine del lavoro. Il declino della forza lavoro globale e l'avvento del post-mercato*; Baldini & Castoldi, 1995).
- [25] Ross Brian H.: *The Psychology of Learning and Motivation: Advances in Research and Theory*. Academic Press, 2004.
- [26] Strassmann P.A.: *Information Productivity: Assessing the Information Management Costs of US Industrial Corporations*. Strassmann Inc. 1999.
- [27] Touraine A.: *La Société post-industrielle. Naissance d'une société*. Denoël, Paris, 1969.

PIER FRANCO CAMUSSONE è Professore Ordinario di Organizzazione e Sistemi Informativi presso la Facoltà di Economia dell'Università degli Studi di Trento. Già direttore dell'Area Sistemi Informativi della Scuola di Direzione Aziendale (SDA) della Bocconi. Autore di numerosi libri ed articoli sul ruolo strategico ed organizzativo dell'ICT. E-mail: pierfranco.camussone@unibocconi.it



# PERCHÉ È COSÌ DIFFICILE COMBATTERE LO SPAM?

Ernesto Damiani

L'invio di messaggi di posta indesiderati, il cosiddetto *spamming*, sta raggiungendo livelli preoccupanti e le previsioni per il futuro sono ben poco rassicuranti. Com'è possibile che un'intera comunità di esperti composta da aziende, utenti individuali, produttori di software e ricercatori informatici sia tenuta in scacco da un ristretto numero di malintenzionati? In questo articolo si analizzano i motivi tecnologici e organizzativi che rendono difficile combattere lo *spam*, descrivendo gli strumenti software utilizzati dagli *spammer* e quelli a disposizione di chi combatte lo *spam*.

3.7

## 1. INTRODUZIONE

**T**ra le tecnologie di Internet, la posta elettronica è forse quella che ha più radicalmente cambiato il modo di vivere e di lavorare di centinaia di milioni di persone in tutto il mondo, incluso chi scrive e quasi certamente anche chi sta leggendo quest'articolo. Purtroppo, però, l'utilizzo e la gestione del servizio di posta sono resi sempre più disagiati dalla marea di messaggi non voluti, collettivamente noti come *spam*<sup>1</sup> che ognuno di noi riceve ogni giorno. Le tecniche di difesa *anti-spam* sono state oggetto di intenso lavoro negli ultimi cinque anni, ma purtroppo la tec-

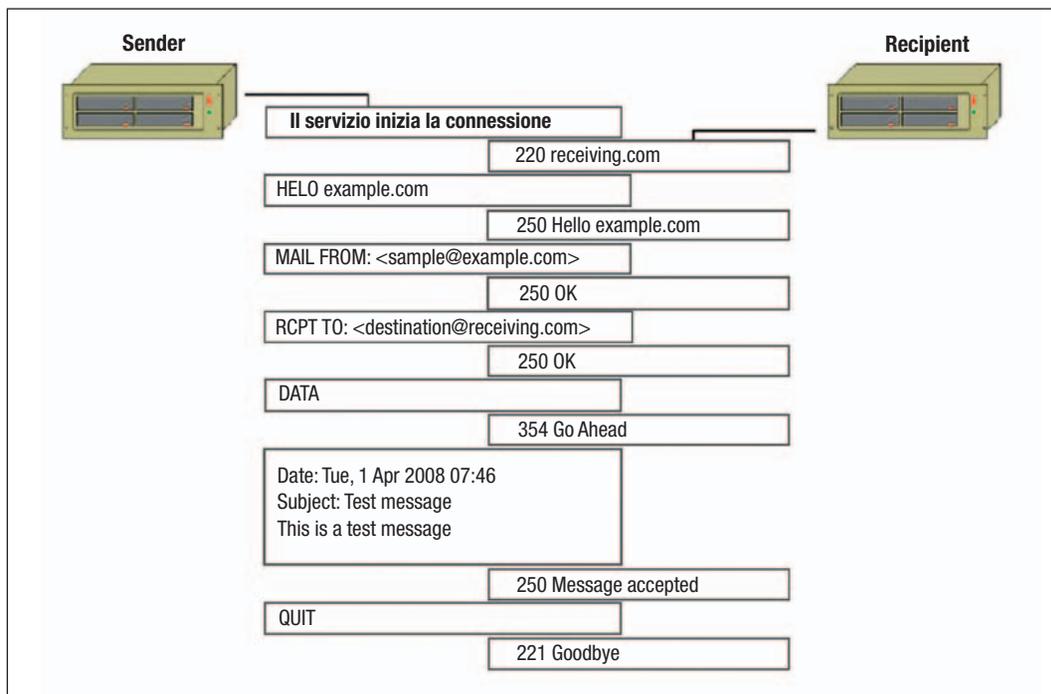
nologia per l'invio di *spam* è migliorata almeno altrettanto.

All'inizio lo *spam* consisteva soprattutto di singoli messaggi inviati attraverso server SMTP compiacenti o mal configurati. Oggi si tratta quasi esclusivamente di messaggi generati dinamicamente e inviati su vasta scala attraverso strumenti software concepiti appositamente. In questo articolo mettiamo a fuoco la natura del problema dal punto di vista tecnologico (il lettore interessato agli aspetti legali può consultare l'**Appendice 1** a p. 51) presentando l'evoluzione degli strumenti per l'attacco (gli strumenti "*malware*" per l'invio di *spam*) e le principali tecniche di difesa (i filtri) oggi disponibili.

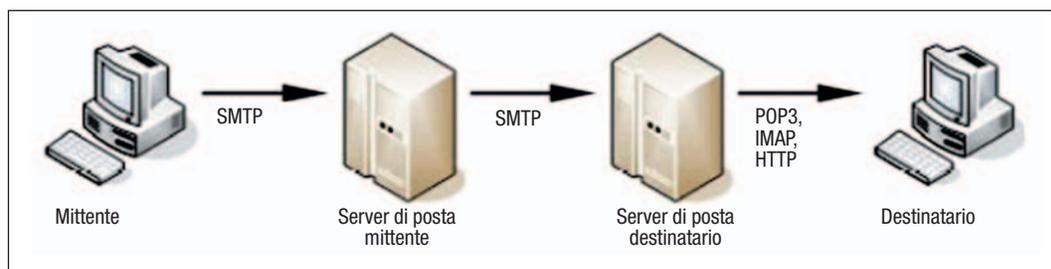
## 2. LA TECNOLOGIA DELLA POSTA ELETTRONICA

La nascita della posta elettronica risale al 1972, quando Ray Tomlinson installò su ARPANET un sistema in grado di scambiare messaggi tra le varie università connesse alla rete, ma chi realmente ne definì poi il funzionamento fu John Postel. Tutta la posta elettronica spedita su Inter-

<sup>1</sup> Il termine *spam* viene dal nome di un cibo in scatola considerato poco appetitoso e dal sapore piatto, ben poco attraente soprattutto se – come i messaggi indesiderati – viene servito sempre, a pranzo e a cena. Il tormentone *spam, spam, spam* per indicare qualcosa di noioso e ripetitivo è stato usato, tra gli altri, dal noto gruppo comico dei Monty Python nel loro storico show "Monty Python's Flying Circus", ambientato in un locale dove ogni pietanza proposta dalla cameriera era a base di *spam*.



**FIGURA 1**  
Un recapito SMTP



**FIGURA 2**  
La consegna SMTP

net viene trasferita usando un unico protocollo: lo *Standard Mail Transport Protocol* (SMTP), definito da Postel nella RFC 8219 e implementato in centinaia di strumenti software (setti spesso *Mail Transfer Agent* o MTA) come il ben noto *sendmail*.

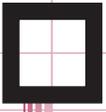
Si tratta di una tecnologia standard: ogni server Internet che utilizza SMTP è in grado di inviare e ricevere posta da qualsiasi altro server SMTP su Internet. Per capire come funziona SMTP basta dare un'occhiata alla figura 1, che mostra uno scambio di messaggi tra un MTA mittente che ha un messaggio di posta da trasmettere e un MTA ricevente che accetta il messaggio perché diretto a un indirizzo di posta da lui gestito. Inizialmente viene aperta una sessione sulla porta TCP 25, e segue una serie di messaggi, in alternanza tra client e server, che iniziano tutti con un codice numerico di tre cifre.

Ogni messaggio di posta è diviso in un'intestazione, composta dei campi *Date:* e *Subject:*

mostrati nella figura (e da altri campi come *From:* che contiene l'indirizzo del mittente, *To:* che contiene l'indirizzo del destinatario e *Return-Path:* che contiene l'indirizzo da usare per la risposta) e dal corpo del messaggio, che contiene il testo vero e proprio.

Va notato che tutti questi campi fanno parte del blocco dati del messaggio, per il quale SMTP non prevede alcun meccanismo di verifica o controllo. Naturalmente il passaggio della posta tra i due server SMTP non esaurisce il percorso di consegna del messaggio: il server mittente ha sicuramente ricevuto il messaggio da un client, e il ricevente lo consegnerà probabilmente a un altro client (Figura 2), il vero destinatario finale, attraverso appositi protocolli di consegna come IMAP e POP, su cui non ci soffermeremo qui<sup>2</sup>.

<sup>2</sup> I messaggi di posta possono anche essere recapitati dopo essere stati incapsulati in altri protocolli applicativi, quali HTTP (*HyperText Transfer Protocol*), come avviene nei sempre più diffusi servizi di Webmail.



Il protocollo SMTP è uno dei più vecchi protocolli di Internet ed è stato volutamente mantenuto semplice, visto che un server SMTP deve poter gestire decine di connessioni al secondo. Questa semplicità si traduce però in vulnerabilità, perché le due informazioni identificative che il server mittente passa al destinatario (il proprio nome e l'indirizzo e-mail a cui il messaggio è diretto) non vengono verificate da quest'ultimo e possono essere quindi facilmente falsificate.

Per chiarire questo punto, esaminiamo meglio un campo Received: facente parte dell'intestazione di un messaggio di posta:

```
from 159.149.70.1 by pollon (envelope-from <caio@crema.unimi.it>, uid 201) 08 Dec 2008 18:42:20 -0000
```

Questo campo dice che il messaggio è stato ricevuto dal server SMTP che si chiama pollon (come dice la clausola `by pollon`) e proviene da un MTA di cui non è noto il nome, ma che ha l'indirizzo IP 159.149.70.1. Osserviamo subito il campo `envelope-from`, che non è il campo `From:` all'interno del messaggio, ma quello che fa parte dell'intestazione SMTP. Contiene l'indirizzo (`caio@crema.unimi.it`) e la user-id (201) del mittente sul MTA di provenienza. Una precauzione che utilizzano molti MTA "diffidenti" è rifiutare posta elettronica in cui il contenuto del campo `envelope-from` dopo la chiocciola non è traducibile dal DNS (il servizio di traduzione nomi-indirizzi di Internet), cioè non è un *Fully Qualified Domain Name* (FQDN) ma solo un frammento non traducibile come `caio@crema`<sup>3</sup>. Questo può accadere quando l'editor di posta elettronica usato dall'utente (esempio, Outlook o Eudora) genera lui stesso i campi SMTP invece di lasciarlo fare al MTA, ma è anche un indizio che chi manda il messaggio potrebbe avere qualcosa da nascondere. Oltre a questo, vi sono altri due elementi importanti da osservare:

□ possiamo ritenere l'intero campo `Received:` affidabile solo se conosciamo e consideriamo fidato il server `pollon` che l'ha creato. Altrimenti la riga potrebbe essere falsa;

<sup>3</sup> In questo caso viene spesso generato un messaggio di errore, come:  
451 <caio@crema> ... Domain does not resolve.

□ se il campo `Received:` è considerato affidabile, la parte importante è `from 159.149.70.1`. Di questo indirizzo IP ci fidiamo, perché l'ha controllato il nostro server fidato pollon quando ha ricevuto il messaggio. Anche qui, come per `envelope-from`, si può usare il DNS per un controllo; ma in questo caso si tratta di una query DNS inversa per ricavare dall'indirizzo il nome del server che ha consegnato il messaggio a pollon, e poi usare il comando `whois` per conoscere la persona e l'organizzazione a cui l'indirizzo è stato associato. Ecco, in sintesi, il risultato di `whois` per questo messaggio:

```
# ARIN WHOIS database, last updated
2008-12-08 19:10
# Enter ? for additional hints on
searching ARIN's WHOIS database
% Information related to
'159.149.0.0 - 159.149.255.255'
inetnum: 159.149.0.0 -
159.149.255.255
netname: UNIMINET
descr: Università degli Studi
di Milano
country: IT
remarks: To notify abuse mailto:
cert@garr.it
remarks: Multiple-Lans of Milan
University
```

La conoscenza dell'indirizzo IP del mittente suggerisce l'idea di configurare il proprio MTA di ricezione in modo che possa rifiutarsi di ricevere posta da alcuni server malfamati (*blacklisting*) oppure di accettare connessioni solo da server conosciuti e fidati (*whitelisting*); ma, come vedremo, queste semplici tecniche sono tutt'altro che perfette e possono introdurre ritardi e omissioni di servizio poco graditi agli utenti.

### 3. COME NASCE LO SPAM

Fino ai primi anni Novanta, la posta elettronica indesiderata consisteva principalmente in innocui scherzi e nei messaggi delle cosiddette "catene di Sant'Antonio". Nell'insieme l'intento di chi li mandava non era criminoso e in pratica non veniva fatto alcun tentativo per falsificare la provenienza dei messaggi, che venivano inviati ai destinatari diretta-

mente dal server SMTP del loro mittente. Secondo molti osservatori, la data d'inizio dello spamming commerciale è il 1994, in cui avvenne la diffusione su tutti i gruppi di discussione Usenet del famoso messaggio "Green-card lawyers" degli avvocati Lawrence Canter e Martha Siegel (Figura 3), che più tardi divennero i primi esperti di Internet marketing. Il messaggio annunciava ai riceventi la fine della lotteria annuale per avere la Green Card, il permesso di soggiorno permanente negli Stati Uniti.

Tecnicamente, la novità del "Green-card lawyers" stava nell'utilizzo di un programma per l'invio sistematico del messaggio a centinaia di gruppi Usenet, e non nella dissimulazione del MTA mittente. Quest'ultimo obietti-

vo venne raggiunto l'anno successivo da Jeff Slaton, che divenne in breve il primo re dello spam, "the Spam King". Nella sua più che decennale attività di spammer, Slaton ha affermato di poter raggiungere fino a 8 milioni di persone i cui indirizzi erano entrati in suo possesso grazie alla raccolta su Usenet.

Oggi, lo spam è diventato uno dei più grandi problemi dell'Internet moderna e uno spreco in termini di tempo e banda. Secondo un recente rapporto rilasciato dall'agenzia specializzata Sophos (Tabella 1), il 92,3% di tutte le e-mail inviate nei primi tre mesi del 2008 è costituito da spam. Stati Uniti e Russia sono in testa alla classifica mentre l'Italia si piazza all'ottavo posto, generando il 3,6% dello spam prodotto nel mondo.

```
Path: gmd.de!urmel.informatik.rwth-aachen.de!newsserver.rrzn.uni-hannover.de!hrz-
ws11.hrz.uni-kassel.de!news.th-darmstadt.de!fauern!zib-
berlin.de!netmbx.de!Germany.EU.net!EU.net!howland.reston.ans.net!europa.eng.gtefs
d.com!hookup!news2.sprintlink.net!news.sprintlink.net!indirect.com!nike
From: n...@indirect.com (Laurence Canter)
Newsgroups: alt.bonehead.paul-hendry,alt.online-service.america-online
Subject: Green Card Lottery- Final One?
Date: 12 Apr 1994 07:40:23 GMT
Organization: Canter & Siegel
Lines: 34
Message-ID: <2odj97$25f@herald.indirect.com>
NNTP-Posting-Host: idl.indirect.com
```

Green Card Lottery 1994 May Be The Last One!  
THE DEADLINE HAS BEEN ANNOUNCED.

The Green Card Lottery is a completely legal program giving away a certain annual allotment of Green Cards to persons born in certain countries. The lottery program was scheduled to continue on a permanent basis. However, recently, Senator Alan J Simpson introduced a bill into the U. S. Congress which could end any future lotteries. THE 1994 LOTTERY IS SCHEDULED TO TAKE PLACE SOON, BUT IT MAY BE THE VERY LAST ONE.

PERSONS BORN IN MOST COUNTRIES QUALIFY, MANY FOR FIRST TIME.

The only countries NOT qualifying are: Mexico; India; P.R. China; Taiwan, Philippines, North Korea, Canada, United Kingdom (except Northern Ireland), Jamaica, Dominican Republic, El Salvador and Vietnam.

Lottery registration will take place soon. 55,000 Green Cards will be given to those who register correctly. NO JOB IS REQUIRED.

THERE IS A STRICT JUNE DEADLINE. THE TIME TO START IS NOW!!

For FREE information via Email, send request to cs...@indirect.com

--

```
*****
Canter & Siegel, Immigration Attorneys
3333 E Camelback Road, Ste 250, Phoenix AZ 85018 USA
cs...@indirect.com telephone (602)661-3911 Fax (602) 451-7617
```

FIGURA 3

Il messaggio "Green-card lawyers" di Canter e Siegel

Paese	Percentuale Spam Prodotta (primo trimestre 2008)
Stati Uniti	15.4%
Russia	7.4%
Turchia	5.9%
Cina	5.5%
Brasile	4.3%
Corea del Sud	4.0%
Polonia	3.8%
Italia	3.6%
Germania	3.4%
Gran Bretagna	3.4%
Spagna	3.3%
Francia	3.1%

**TABELLA 1**  
I Paesi maggiori produttori di Spam  
(Fonte: SOPHOS)

La raccolta degli indirizzi dei destinatari rappresenta da sempre un problema per gli *spammer* (Appendice 2 a p. 52), ma gli indirizzi dei server SMTP non sono difficili da trovare: è sufficiente consultare i campi MX presenti nei file di zona DNS. Per combattere i primi *spammer*, i gestori di server SMTP usavano strumenti di *blacklisting* molto semplici, creando in sede di configurazione dei server SMTP una lista (*killfile*) di indirizzi IP dai quali non desideravano ricevere messaggi. Ben presto però gli *spammer* scoprirono come combattere il *blacklisting* dei loro server, grazie a una funzionalità dei server SMTP chiamata *open relay*. Questa funzione esiste in tutte le implementazioni del protocollo; in questo articolo ci concentreremo su *sendmail*, l'implementazione di SMTP che discende dall'originale DeliverMail di ARPANET<sup>4</sup>. *Sendmail* è ancora oggi il più popolare MTA di Internet, sebbene stia perdendo posizioni. La sua popolarità è probabilmente dovuta al fatto che è il server SMTP standard della maggior parte delle varianti di Unix. Fino alla versione 5, *sendmail* (come molte altre imple-

<sup>4</sup> *Sendmail* è ancora oggi il mail server più usato su Internet. Secondo uno studio del 2005, il 42% circa dei mail server raggiungibili via Internet usavano *Sendmail*.

mentazioni di SMTP) inviava messaggi per conto di qualsiasi client lo richiedesse, fungendo appunto da "open relay". Invece di spedire lo *spam* direttamente al server SMTP del destinatario, gli *spammer* iniziarono a usare - alternandoli - i server SMTP di altri come intermediari. Ovviamente, l'uso dell'*open relay* non impedisce di per sé il *blacklisting* degli indirizzi IP dei server da cui proviene lo *spam*, ma l'inclusione nelle *blacklist* di server SMTP che fanno *open relay* in buona fede è molto difficile, perché impedisce anche il recapito di messaggi legittimi, che vengono bloccati insieme allo *spam* ritrasmeso.

L'unica vera contromisura sta nel disabilitare la funzionalità di *open relay* su tutti i server SMTP. Questo problema collettivo di configurazione fu oggetto di un grande dibattito nella comunità di Internet, che forse per la prima volta si accorse che un problema tecnico apparentemente banale si poteva tradurre in un incubo organizzativo. Per impedire l'*open relay* basta un semplice script di configurazione per *sendmail* come quello che segue:

```
FR-o /etc/sendmail.cR
Scheck_rcpt
# La posta che va recapitata localmente è accettata
R< $+ @ $=w > $@ OK
R< $+ @ $=R > $@ OK
# La posta che è generata localmente è accettata
R$* $: $(dequote "" $&
{client_name} $)
R$=w $@ OK
R$=R $@ OK
R$@ $@ OK
# tutto il resto è rifiutato
R$* $#error $: "550 Re-
laying Denied".
```

Risolvere il problema di quali sono i soggetti organizzativi che hanno titolo ad attivare un server SMTP pubblico e di come garantire il loro comportamento uniforme nella gestione dei server è un'impresa tutt'altro che semplice per le grandi organizzazioni decentrate come le Università. Alla fine degli anni Novanta, comunque, l'azione congiunta dei provider Internet e delle grandi organizzazioni per censire i server SMTP attivi e disabilitare l'op-

zione *open relay* aveva quasi risolto il problema dello *spam*, anche se all'interno delle Università il divieto ai singoli utenti di gestire liberamente il proprio *sendmail* suscitò parecchi malumori<sup>5</sup>. Purtroppo, però, l'evoluzione tecnica della Rete fece presto emergere tre nuove tecniche di recapito che riportarono lo *spam* d'attualità già all'inizio degli anni Duemila.

**Recapito Relay multi-hop:** il primo fattore è l'aumento di complessità dei servizi di posta elettronica gestiti dai provider, che rese possibile agli *spammer* aggirare il blocco dell'*open relay* attraverso una tecnica detta *relay multihop*. Oggi, infatti, le reti dei provider Internet e delle grandi organizzazioni si affidano a più server SMTP, alcuni usati per l'invio di posta tra utenti dello stesso dominio, ed altri MTA "di confine" usati per inoltrare la posta verso l'esterno. Ovviamente, gli MTA di confine accettano il *relay* da parte dei server interni. Se lo *spammer* ha accesso a uno dei server interni, o se quest'ultimo non è ben configurato, può inviare messaggi di *spam* tramite il MTA di confine, che (pur non facendo *open relay*) accetta di rispediti verso l'esterno perché gli sembra che provengano da un mittente autorizzato.

**Dynamic addressing e recapito No Relay:** il secondo fattore è la pratica, oggi prevalente tra i provider Internet, di assegnare ai loro clienti indirizzi IP dinamici, cioè validi solo per la durata di una connessione. Questa prassi ha dato agli *spammer* un altro modo di aggirare il blocco dell'*open relay*: lo *spammer* recapita i messaggi di *spam* direttamente ai server SMTP dei destinatari, usando il suo indirizzo IP dinamico. Periodicamente, oppure ogni volta che l'indirizzo IP dinamico dello *spammer* viene notato e elencato su una *blacklist*, lo *spammer* può semplicemente scollegarsi da Internet, riconnettersi e ricevere un nuovo indirizzo IP dinamico. Il costo di eseguire uno *spamming* di questo tipo è alto anche per lo *spammer* (soprattutto in termini di tempo), ma l'inoltro di *spam* con

questa tecnica (detta *no relay*) è molto efficace e combatterlo è estremamente difficile.

**Connection Sharing e recapito open proxy:** il terzo fattore riguarda la condivisione delle connessioni Internet. Oggi molte organizzazioni usano *proxy* sui loro server connessi a Internet per consentire ad altri computer della loro rete locale (cablata o wireless) di condividere la connessione. Come accadeva i server di posta elettronica che facevano inavvertitamente *open relay*, anche i *proxy software* sono spesso mal configurati e permettono ad *host* "parassiti" di attivare connessioni *proxy* (*open proxy*). Gli *spammer* hanno iniziato a usare i client con *open proxy* per dissimulare l'origine della posta elettronica. Se un *open proxy* non è disponibile, può essere diffuso in modo virale: già nel gennaio 2003, il noto virus *Sobig.a* installava nei computer vittime un *proxy* concepito specificatamente con l'intenzione di consentire lo *spam*.

**Tecniche ibride:** per rendere ancora più difficile prendere contromisure, gli *spammer* usano spesso una combinazione delle tecniche appena viste. Per esempio, lo *spammer* usa il server SMTP di un provider Internet poco rigoroso nei controlli o un indirizzo dinamico per raggiungere un server SMTP che fa *open relay*, tramite quest'ultimo, accede al server SMTP di un grosso provider. Il seguente frammento di *header* proviene da un messaggio di *spam* reale:

```
Return-Path: <hymcirrus@coastlinetrans.com>
Delivered-To: damiani@dti.unimi.it
Received: (qmail 9405 invoked by uid 210); 9 Dec 2008 00:00:03 -0000
Received: from 159.149.10.22 by pollon (envelope-from
<hymcirrus@coastlinetrans.com>, uid 201) with qmail-scanner-1.25st
(Clamscan: 0.94.1/8730. spamassassin: 3.2.1. perlscan: 1.25st.
Clear:RC:0(159.149.10.22):SA:-0(3.8/6.0):.
Processed in 2.340732 secs); 09 Dec 2008 00:00:03 -0000
X-Spam-Status: No, hits=3.8 required=6.0
X-Spam-Level: +++
```

<sup>5</sup> Il timore delle conseguenze dell'attivazione di server SMTP da parte di utenti ignari o inesperti è probabilmente uno dei motivi per cui anche oggi i computer Macintosh vengono consegnati agli acquirenti con *sendmail* disabilitato.

```

Received: from unknown (HELO mail-
server.unimi.it) (159.149.10.22)
  by 0 with SMTP; 9 Dec 2008 00:-
00:01 -0000
Received: from unimix1.unimi.it
([172.24.4.81])
  by ldap-s2.unimi.net (Sun Java
System Messaging Server 6.2-8.04
(built Feb 28
2007)) with ESMTP id <0KBL004-
2V1C3BAA0@ldap-s2.unimi.net> for
damiani@dti.unimi.it (ORCPT ern-
esto.damiani@unimi.it); Tue,
09 Dec 2008 01:00:03 +0100 (CET)
Received: from comercigomez.com
(unknown [123.18.210.158]) by uni-
mix1.unimi.it
(Unimi) with ESMT id EFF844A0026 for
<ernesto.damiani@unimi.it>; Tue,
09 Dec 2008 01:00:10 +0100 (CET)

```

Qui, come si vede, è stata usata la tecnica ibrida di recapito: sono presenti diversi campi `Received:` e quindi il messaggio è transitato per diversi MTA. Andando a ritroso troviamo il nostro MTA locale `pollon` e poi il MTA interno `mailserver.unimi.it`, di cui `pollon` si fida e dal quale accetta la posta. Notate che in questo caso il nome del MTA non era presente nel campo `From:` dell'-envelope SMTP ma è stato desunto dal messaggio SMTP HELO con cui `mailserver.unimi.it` si è annunciato a `pollon`. Il MTA interno `mailserver.unimi.it` a sua volta aveva accettato la mail dal server SMTP "di confine" `unimix1.unimi.it`, l'unico MTA abilitato a parlare con sever esterni.

Proseguendo l'analisi a ritroso la lista dei campi `From:` troviamo il server SMTP con indirizzo `123.18.210.158`, da cui è arrivata la mail. Ecco il potenziale colpevole, dietro cui potrebbe nascondersi lo *spammer*. La figura 4 mostra il risultato della ricerca di questo indirizzo in un database di server SMTP che eseguono *open relay* (<http://www.mail-abuse.com>). A questo punto, quindi, la caccia si interrompe: nel parlare con il server open relay, lo *spammer* può inserire i campi `Received:` che meglio crede nell'intestazione del messaggio di posta, e falsificarli liberamente.

#### 4. IL FILTRAGGIO DEI MESSAGGI

Per neutralizzare le tecniche di *spam* basate sulla tecnica *no relay* (ed alleviare quelle che ricorrono a *open proxy*) si potrebbe in linea di principio adottare il blocco completo degli indirizzi IP dinamici, cioè configurare i server SMTP in modo che non accettino connessioni da altri server che hanno un indirizzo IP dinamico. Si tratta però di un approccio poco pratico perché non esiste un semplice test per stabilire se un indirizzo IP è assegnato dinamicamente o meno<sup>6</sup>. Un'altra tecnica molto interessante è quella delle cosiddette *honeypot*, costituite da finti server SMTP poco scrupolosi e da caselle di posta non corrispondenti a utenti reali, vere e proprie trappole che catturano gli indirizzi IP dei server SMTP usati dagli *spammer*.

In pratica, però, la latenza necessaria per diffondere le segnalazioni delle *honeypot* le rende molto più indicate per attivare contro-misure legali che per reazioni in tempo reale

The IP address 123.18.210.158 does appear on the following database managed by Trend Micro's Network Reputation Services.

Database	Entry	Action
<a href="#">DUL</a>	123.18.210.158	<a href="#">Remove</a>

Please see the linked web pages for further information about the database, contact information, why the address is listed, and how to get it removed, if applicable.

Please note: These databases are based on IP addresses; they do not use host or domain names.

**FIGURA 4**  
Identificazione  
del server SMTP  
open relay

<sup>6</sup> A differenza di quanto alcuni credono, la maggior parte degli indirizzi IP pubblici attribuiti staticamente NON corrisponde a un nome nel DNS. Quindi le *query* inverse al DNS non sono purtroppo un buon test per dedurre la staticità di un indirizzo.

all'invio di *spam*. In generale, il filtraggio dinamico basato sull'IP del server SMTP mitente si è gradualmente rivelato un metodo antispam poco pratico, e all'inizio degli anni Duemila la lotta allo *spam* ha affiancato all'IP *filtering* un'altra direzione, adottando un approccio collaborativo e più orientato al contenuto, sia sui server SMTP, sia sui programmi client usati per spedire e leggere la posta.

#### 4.1. List splitting e personalizzazione dello spam

L'idea iniziale del filtraggio orientato al contenuto fu sfruttare il fatto che la maggior parte degli *spammer* inviava a tutti i destinatari una copia dello stesso messaggio. Facendo quest'ipotesi, il filtraggio collaborativo può funzionare come segue: quando abbastanza utenti di posta segnalano un messaggio sospetto, per esempio mettendolo nella cartella "*Junk Mail*" dei loro client, questi ultimi notificano la cosa al server SMTP e il messaggio incriminato (o meglio una sua rappresentazione compressa, uno *hash MD5*) viene aggiunto a una lista che è poi condivisa tra i server SMTP, con connessioni *peer-to-peer* o attraverso servizi di notifica simili a quelli usati per gli antivirus.

I server SMTP possono poi scartare i messaggi di posta in arrivo il cui *hash MD5* corrisponde a uno di quelli nella lista dello *spam*. Anche se in un primo tempo il filtraggio collaborativo fu efficace, ci si accorse subito che gli *spammer* potevano aggirarlo usando tecniche di partizionamento e di personalizzazione delle liste dei destinatari (il cosiddetto *list splitting*), in modo da aggiungere ai messaggi di *spam* delle porzioni variabili dipendenti dal destinatario.

In realtà fino ai primi anni Duemila gli strumenti più usati dagli *spammer* non comprendevano funzionalità di "*list splitting*", e molti *spammer* continuarono a inviare a tutti i destinatari gli stessi messaggi fino a quando, nel 2002, non vennero diffusi sul mercato i primi strumenti antispam che usavano classificatori di testo statistici.

L'idea di filtrare i messaggi in arrivo sulla base del loro contenuto non è concettualmente una novità; anzi, fin dagli albori della posta elettronica molti programmi per la lettura della posta sono stati dotati di filtri configurabili in base ai campi dell'intestazione dei messag-

gi. Queste regole sono in grado di individuare contenuti tipici dei messaggi di *spam*, che non appaiono nei messaggi "normali". Si possono per esempio filtrare i messaggi che non ottengono nel campo *TO*: l'indirizzo corretto del destinatario oppure il cui *Subject*: sia vuoto o tutto in maiuscolo, o contenga parole chiave specificate dall'utente. Un altro criterio di filtraggio esamina il campo *From*: operando analogamente a quanto abbiamo già visto per il campo omonimo dell'*envelope* SMTP. Se il campo *From*: è vuoto, o l'indirizzo del mittente non risponde a certe caratteristiche, il messaggio viene scartato.

Oggi gli amministratori di sistemi Unix hanno a disposizione software ben più evoluti come *procmail* (un programma che processa automaticamente i messaggi quando questi arrivano nella casella locale) per i quali si possono predisporre file di configurazione - e quindi filtri - molto complessi. Uno di questi, *Spam-Bouncer*, è in grado di generare dei falsi messaggi di errore per far credere allo *spammer* che l'indirizzo a cui si rivolge è inesistente.

#### 4.2. Tecniche automatiche di riconoscimento dello spam

La comunità della ricerca informatica - compreso chi scrive - ha versato in questi anni fiumi d'inchiostro sulle tecniche automatiche di riconoscimento dello *spam*, proponendo diversi algoritmi molto ingegnosi, in grado di classificare messaggi di testo come *spam* in modo rapido ed efficace. Queste tecniche sono in grado di ridurre i falsi positivi (cioè i messaggi che non sono *spam* ma vengono identificati come tali) anche in presenza di *list splitting* e di personalizzazione dinamica del testo dei messaggi di *spam*. Molti *spammer* hanno reagito a queste tecniche evolute di riconoscimento semplicemente spostando la parte informativa dei loro messaggi all'interno di immagini, da inviare poi come allegati MIME (*Multipurpose Internet Mail Extensions*) o agganciare ai messaggi scrivendoli in formato HTML. È noto che i computer sono molto meno bravi degli umani nel riconoscere il contenuto di immagini; anzi, il fatto che la localizzazione di caratteri all'interno di un'immagine è un problema facile per un utente umano ma difficile per un software è oggi sfruttato da molti siti Web per evitare la compilazione au-



ler. Mentre la configurazione di Dark Mailer supporta un solo *template* per messaggi di *spam*, la configurazione di Send Safe è organizzata in "campagne" e "messaggi".

Una campagna Send Safe consiste in uno o più messaggi e un insieme di *mailing list*. Un messaggio è costituito da un corpo del messaggio e da una serie di argomenti per il campo **Subject:**, indirizzi per il campo **From:** e allegati. Una campagna invia periodicamente i suoi messaggi a tutti indirizzi contenuti nei file delle *mailing list*.

Come Dark Mailer, Send Safe consente la trasmissione diretta di messaggi basata su *open proxy* e *open relay*, ma applica alcune tecniche evolute. Per eludere le *black list*, Send Safe può cambiare continuamente l'indirizzo IP che usa per collegarsi ai server di posta elettronica o ai *proxy*. Send Safe dispone anche di un *proxy* interno che è stato progettato per eludere l'individuazione tramite *honeypot*. Invece di connettersi direttamente alla lista di *proxy* specificata dallo *spammer*, si collega ad essi attraverso una serie di *proxy* intermedi considerati sicuri. Se c'è un *honeypot* nella lista di *proxy* dello *spammer*, l'indirizzo IP del sistema su cui gira Send Safe non sarà compromesso.

Un'altra tecnica interessante introdotta da SendSafe è il *proxy locking*. Partendo dall'indirizzo IP di un *open proxy*, Send Safe usa una query DNS inversa per cercare nel record **MX (Mail Exchanger)** il server SMTP usato dal *proxy*. Invece di tentare di consegnare i messaggi attraverso il *proxy*, SendSafe si rivolge direttamente al server SMTP. Questo trucco può portare i server SMTP di produzione dei provider a comparire gli uni nelle *black list* degli altri. La contromisura più evidente è attivare il filtraggio orientato al contenuto dello *spam* di cui abbiamo parlato prima anche in uscita (e non solo in ingresso) dai server SMTP interni, ma questo ha costi non indifferenti e introduce sensibili latenze nel recapito della posta.

Send Safe comprende un sistema avanzato per creare *template* di messaggi di *spam*. Si possono generare messaggi che sembrano inviati da client di posta elettronica diversi, come Microsoft Outlook Express e Mozilla Thunderbird. Quando Send Safe invia lo *spam*, alterna i *template* così che ogni messaggio suc-

cessivo che viene inviato sembra essere stato spedito usando un client diverso.

Send Safe comprende anche diverse contromisure per ingannare i filtri antispam orientati al contenuto. Per esempio, può aggiungere contenuto casuale nei campi **Subject:** e **From:**, oppure codificare la parte testuale (tipo MIME `text/html`) del messaggio usando il codice `base64` invece del `quoted-printable` standard, o ancora aggiungere in modo casuale dei tag HTML al testo del messaggio per confondere i parser HTML di alcuni filtri *anti-spam*. Ben più importante è la capacità di Send Safe di applicare algoritmi di *morphing* alle immagini per deformarle, in modo che non siano facilmente riconoscibili da eventuali algoritmi di classificazione delle bit-map. La generazione delle immagini è però lasciata allo *spammer*, e quindi Send Safe non è molto adatto per le campagne di *spam* grafico che fanno forte ricorso al *list splitting* e personalizzano i messaggi.

### 5.3. Reactor Mailer

Reactor Mailer, venduto dalla società ucraina Elphisoft, è di gran lunga il sistema di *spamming* più interessante sviluppato fino ad oggi. Mentre Dark Mailer e Send Safe generano i messaggi di *spam* localmente e poi li trasmettono attraverso una lista di *open proxy* e server SMTP che accettano *open relay*, Reactor Mailer usa un modello computazionale distribuito simile a quello dei virus. Il programma si compone di un server e di un client distribuito in forma virale, che gli antivirus Symantec conoscono come Trojan.Srizbi. I personal computer che vengono infettati dal client Reactor Mailer scaricano periodicamente *template* di messaggi e liste di indirizzi di posta elettronica, generano e trasmettono indipendentemente i loro messaggi e poi rimandano i report dei risultati al server. Questa tecnica riduce molto i costi di tempo di elaborazione e di larghezza di banda che rendono oneroso l'invio di *spam* tramite Dark Mailer e Send Safe.

Reactor Mailer usa un sistema di *template* simile al sistema di intestazioni di Dark Mailer; il *template* più usato crea messaggi quasi indistinguibili da quelli generati da Outlook Express 6.

Mentre Send Safe richiede che l'utente crei le

proprie immagini, Reactor Mailer comprende la traduzione del testo dello *spam* a immagini. Questo sistema può creare immagini basate su testo formattato HTML e può offuscare le immagini attraverso l'aggiunta di rumore *random* e rototraslazioni dei caratteri.

## 6. Un esempio

Vediamo ora una versione semplificata di un *template* di Reactor Mailer:

```
From: {rndline 008_wname.txt}-  
{rndabc 1}@{rndline  
003_domains.txt}  
Subject: {rndline 001_subject.txt}  
{rndline 005_hi.txt}  
  
{rndline 001_msg.txt}  
http://{rndline 006_sub.txt}. -  
{rndline 000_067.txt}  
{rndline 004_fin.txt}  
{rndline 002_afo.txt}, {rndline  
002_afo.txt}
```

Le intestazioni dei messaggi di *spam* generate usando questo *template* contengono un campo **From:** generato a caso, un nome di battesimo e l'iniziale di un cognome casuali come username e un **Subject:** anch'esso selezionato a caso da una lista. Il corpo del messaggio inizia con un saluto scelto a caso da una lista e poi continua con una frase scelta a caso da una terza lista. Le frasi sono seguite da un URL *random* e poi il messaggio si conclude un saluto scelto a caso. Questo *template* può produrre un numero elevatissimo di messaggi diversi, rendendo difficile il lavoro dei filtri antispam orientati al contenuto. Ecco un esempio di *spam* generato dal *template*:

```
From: LombrosoC@pollon.it  
Subject: Chi dorme non piglia pesci  
Come butta oggi?  
Le brave ragazze vanno in Paradiso,  
le cattive dappertutto.  
http://vieniacasa.org  
Grazie per l'attenzione, gente!  
La svelta volpe balza sul cane pigro,  
non aspettate tempi migliori.
```

## 7. LE CONTROMISURE

Vediamo ora le contromisure che possono essere prese contro lo *spam* usando gli strumenti di difesa basati sulle tecniche che abbiamo spiegato all'inizio dell'articolo. La soluzione di riferimento è SpamAssassin, un software che identifica automaticamente lo *spam*. Pur essendo pensato per sistemi Unix, grazie al fatto di essere open source SpamAssassin è stato proposto anche come add-in per alcuni mail server commerciali. Per identificare lo *spam* SpamAssassin esegue una serie di verifiche sull'intestazione e un'analisi del testo del messaggio. Inoltre, usa alcune blacklist di MTA inaffidabili reperibili in Rete. Dopo essere stato identificato, lo *spam* viene contrassegnato con un punteggio che si aggiunge all'intestazione del messaggio, in modo che quest'ultimo possa poi essere filtrato dal client di posta dell'utente.

Ecco un esempio dell'aggiunta generata da SpamAssassin:

```
spamassassin: 3.2.1. perlscan:  
1.25st.  
Clear:RC:0(159.149.10.22):SA:-  
0(3.8/6.0):.  
Processed in 2.340732 secs); 09 Dec  
2008 00:00:03 -0000  
X-Spam-Status: No, hits=3.8 required=6.0  
X-Spam-Level: +++
```

Per gli esempi di *spam* SpamAssassin si basa su Vipul's Razor, una rete distribuita e collaborativa di identificazione dello *spam* che opera da un paio d'anni, grazie alla quale è stato costruito un catalogo costantemente aggiornato dello *spam* in circolazione. Lo strumento Spam Arrest, invece adotta un approccio basato su *whitelist*, una lista di "amici" autorizzati a scriverti. Se qualcuno che non è nella lista scrive a una mailbox protetta da Spam Arrest, riceverà immediatamente un messaggio che lo invita a visitare un sito, da cui può iscriversi alla lista di amici. Per poterlo fare, dovrà trascrivere in un campo testo il contenuto di un'immagine che riporta caratteri testuali in posizione *random*, dimostrando così di essere una persona e non uno script utilizzato da uno *spammer*. Veniamo ora a due tecniche "storiche" che per i motivi pratici esposti fin qui non hanno risolto il

problema dello *spam*, ma risultano comunque particolarmente interessanti: il *reverse spam filtering* e i filtri bayesiani.

### 7.1. Reverse Spam Filtering

La strategia del *Reverse Spam Filtering* è diametralmente opposta a quella dei filtri orientati al contenuto. Questa tecnica infatti si propone di selezionare ciò che NON è *spam* e mandare tutto il resto in una cartella speciale, che viene controllata solo periodicamente. Anzi tutto il sistema controlla se il messaggio in entrata appartiene a qualche invio di massa sollecitato (*mailing list* o *newsletter*). In questo caso viene messo in un'apposita cartella. Altrimenti, viene controllata la provenienza: se il messaggio proviene da indirizzi approvati (cioè definiti in una lista di "amici" come quella di SpamArrest) viene posto in un'apposita cartella altrimenti il messaggio viene analizzato e quindi marchiato come *spam* con una certa probabilità, e inserito in una speciale cartella per i messaggi sospetti, il cui contenuto può essere ordinato in base alla probabilità e controllato manualmente per cercare falsi positivi. Il *Reverse Spam Filtering* necessita di un software per filtrare i messaggi, uno per analizzare e assegnare un punteggio di probabilità ai messaggi sospettati di essere *spam*, un buon client di posta che permetta di gestire più mailbox e di ordinare il contenuto delle mailbox in base a criteri personalizzati, un sistema per mantenere facilmente o automaticamente una lista di indirizzi "amici" aggiornata. In genere si usa `procmail` per filtrare i messaggi in arrivo e SpamAssassin per marciare i messaggi con un punteggio di *spam*.

### 7.2. Filtri bayesiani

La soluzione bayesiana è stata proposta inizialmente da Paul Graham ed è basata sullo studio statistico del contenuto dei messaggi. Un filtro bayesiano decide se un messaggio è *spam* o no in base alle parole contenute nei messaggi ricevuti da uno specifico utente. Prima di illustrare l'algoritmo usiamo un semplice esempio per ricordare il teorema di Bayes: abbiamo un'osservazione O (un messaggio contiene la parola "sex") e un'ipotesi H (un messaggio è *spam*).  $P(O|H)$ , cioè la probabilità che O accada dato H, ovvero la probabilità che un messaggio di *spam* contenga

la parola "sex", è facile da stimare (ad esempio esaminando la cartella "Junk Mail" in cui l'utente destinatario mette lo *spam* e contando quanti dei messaggi che vi si trovano già contengono "sex"). Per il futuro, ci interessa però sapere  $P(H|O)$ , cioè la probabilità che H accada, dato O, e cioè che un messaggio indirizzato a quell'utente e che contiene la parola "sex" sia effettivamente *spam*. Secondo il teorema di Bayes tale probabilità è:

$$P(H|O) = P(O|H) * P(H) / P(O)$$

Dove sia  $P(H)$  (la probabilità che un messaggio sia *spam*) sia  $P(O)$  (la probabilità che un messaggio contenga la parola "sex") possono essere agevolmente stimate esaminando le caselle di posta dell'utente.  $P(H)$  si stima esaminando comparativamente la cartella "Junk Mail" dove l'utente mette lo *spam* e la casella di posta generale dell'utente e contando quanti sono i messaggi di *spam* rispetto al totale dei messaggi.  $P(O)$  si stima contando quanti messaggi contengono "sex" sul totale dei messaggi (*spam* o no) ricevuti dall'utente. Va notato che queste probabilità devono essere calcolate per ogni utente perché, se i messaggi di *spam* possono essere simili per tutti (e a volte sono esattamente gli stessi), quelli personali sono invece molti diversi, e il filtro bayesiano ne tiene automaticamente conto. Le esperienze di Graham, e degli altri ricercatori che hanno lavorato nel settore, ci dicono che il suo filtro è esatto al punto di mancare solo 5 messaggi di *spam* ogni 1000, senza alcun falso positivo. Rispetto ai filtri visti in precedenza, che funzionano in base alle proprietà individuali di un singolo messaggio, l'approccio statistico su insiemi di messaggi è migliore, perché tiene conto delle specificità dei singoli utenti, esattamente come fa lo *spammer* applicando il *list splitting*. Purtroppo però questa tecnica è praticamente impotente contro lo *spam* grafico.

## 8. PROTEZIONE CRITTOGRAFICA DEGLI INDIRIZZI

Una prospettiva integralmente nuova è invece quella di togliere agli *spammer* la loro "benzina", cioè gli indirizzi di posta elettronica, attraverso nuovi schemi di generazione dinamica degli indirizzi di posta. Gli *spammer* usano programmi appositi (det-

```

<script type="text/javascript" language="javascript">
<!--
// Email obfuscator script 2.1 by Tim Williams, University of Arizona
// Random encryption key feature by Andrew Moulden, Site Engineering Ltd
// This code is freeware provided these four comment lines remain intact
// A wizard to generate this code is at http://www.jottings.com/obfuscator/
{ coded = "o8Sm8cm@oLm.McmSm.mL"
  key = "bdyPfKcQwJMAFtVNSD4Oso6pz5X0kGlahLnXUvi3W9ejrZ7EH2l8uYcRTBmgq1"
  shift=coded.length
  link=""
  for (i=0; i<coded.length; i++) {
    if (key.indexOf(coded.charAt(i))!=-1) {
      ltr = coded.charAt(i)
      link += (ltr)
    }
    else {
      ltr = (key.indexOf(coded.charAt(i))-shift+key.length) % key.length
      link += (key.charAt(ltr))
    }
  }
  document.write("<a href='mailto:"+link+"'>inviare mail al docente</a>")
}
<!-->
</script><noscript>Sorry, you need Javascript on to email me.</noscript>

```

**FIGURA 5**  
 Uno script che genera l'indirizzo damiani@dti.unimi.it

ti *harvester* o *spambot*) che scaricano le pagine Web alla ricerca di indirizzi di posta a cui mandare *spam* (vedi Appendice 2). Alcuni siti usano già oggi delle semplici precauzioni per evitarlo, pubblicando indirizzi "antispam" come `ernesto.damiani AT unimi DOT it`. Questo metodo però richiede che sia il visitatore umano a modificare l'indirizzo di posta per renderlo usabile; inoltre è facilmente aggirato dagli *spambot* più recenti. Altri siti cercano di difendersi dagli *harvester* con le loro stesse armi, ossia pubblicando immagini che mostrano gli indirizzi di mail in luogo degli indirizzi un formato testo; ma anche questa precauzione può non essere gradita ai visitatori umani, che devono ridigitare l'indirizzo da capo per poterlo usare. Infine, altri siti usano la codifica carattere per carattere HTML, per esempio usando `&#64;`; per il carattere chiocciola (@), o per tutti i caratteri dell'indirizzo. Ecco la codifica dell'indirizzo `someone@example.com`:

```

&#115;&#111;&#109;&#101;&#111;
&#110;&#101;&#64;&#101;&#120;
&#97;&#109;&#112;&#108;&#101;
&#46;&#99;&#111;&#109;

```

Questo tipo di codifica è anch'esso facile pre-

da degli *harvester*, perché qui ogni carattere corrisponde esattamente a un codice secondo una tabella ben nota. Le tecniche crittografiche invece si basano su una codifica crittata dell'indirizzo di mail. Questa codifica viene decrittata da uno script JavaScript solo al momento dell'utilizzo dell'indirizzo di posta elettronica e quindi quest'ultimo non compare da nessuna parte nella pagina. Un esempio di uno script di questo tipo per l'indirizzo `damiani@dti.unimi.it` è riportato nella figura 5 – come si vede, si tratta di un osso piuttosto duro per qualunque analizzatore di codice.

Attualmente la tecnica crittografica per la generazione dinamica degli indirizzi di mail viene complementata dalla messa a punto di strumenti innovativi per la generazione e la gestione di "short-lived alias", cioè indirizzi di posta monouso. L'idea è di dare ai diversi interlocutori che possono contattarci indirizzi di mail diversi, alcuni dei quali consentono di contattare il destinatario una volta sola. Nel futuro avremo quindi due tipi di indirizzi: quello stabile (*master*) e quelli temporanei (*alias*). Quando contatteremo qualcuno per la prima volta al suo indirizzo *master*, gli manderemo un nostro *alias*. L'interlocutore invierà la risposta al nostro *alias*, corredando

il messaggio di risposta di un suo alias, e da quel momento potremo proseguire a comunicare usando gli alias monouso allegati a ogni messaggio.

Per rimanere compatibili con il normale recapito SMTP, gli alias avranno sempre l'indirizzo master come suffisso, come segue:

ZBEF9.damiani@dti.unimi.it

Gli alias potranno così essere risolti sugli MTA di recapito. Gli alias monouso sono la nostra principale speranza di liberarci definitivamente degli *spammer*: valgono per una sola consegna e rendono molto più costoso e difficile il lavoro dello *spammer*. Ovviamente è possibile pensare a varie categorie di alias, magari accettabili più volte o da un gruppo di mittenti predefinito.

## 9. CONCLUSIONI

È abbastanza chiaro che gli attuali algoritmi di individuazione e filtraggio dello *spam* basati sul contenuto dei messaggi hanno efficacia limitata se i messaggi sono grafici e/o personalizzati rispetto ciascun destinatario. Oggi gli *spammer* hanno a disposizione gli strumenti

(se non la conoscenza) per realizzare *template* di messaggi che possono creare un numero elevatissimo di messaggi univoci. Il numero delle permutazioni che possono essere prodotte da questi strumenti è sufficiente per sovrapporre i sistemi tradizionali *antispam*, per quanto ingegnosi siano gli algoritmi di classificazione che utilizzano. A volte l'aggiunta ai sistemi *antispam* di precauzioni semplici, come proibire del tutto il recapito di immagini bitmap come allegati, può migliorarne notevolmente l'efficacia, ma non c'è alcun dubbio che – in attesa di tecniche crittografiche veramente efficaci per la generazione e la risoluzione di indirizzi monouso – il vantaggio resta, almeno per ora, dalla parte degli *spammer*. Gli strumenti per creare *spam* guidati da *template* hanno raggiunto una certa maturità, e la tecnologia *antispam* deve quindi migliorare. Per quanto riguarda il filtraggio del contenuto, allo studio ci sono nuove tecniche statistiche e di apprendimento computazionale che utilizzano la regolarità tipiche dei messaggi generati a partire da *template* invece di concentrarsi, come quelli attuali, sulle regolarità tipiche dei messaggi scritti a mano. La battaglia tra *spammer* e tecniche *antispam* non è comunque destinata a terminare tanto presto.

### APPENDICE 1 - ASPETTI NORMATIVI E LEGALI

Il primo Paese a prendere contromisure normative contro lo *spam* sono stati gli USA, che sulla base di una legge federale già in vigore contro l'abuso dei fax, diedero vita alla CAUCE (*Cohalition Against Unsolicited Commercial Email*), per porre rimedio al vuoto legislativo in materia di e-mail non richieste. Questo compito richiese molto tempo, anche per la continua controffensiva degli *spammer* che premevano per legalizzare l'*opt-out* (ossia la possibilità di negare l'invio di e-mail non richieste solo dopo averle ricevute).

Nel 2003 finalmente il Congresso americano varò la nuova legge federale "*CAN-SPAM Act of 2003*". Questa legge si fondava sul principio dell'*opt-out* e attribuiva il titolo di agire contro gli *spammer* ai soli Internet provider, e non agli utenti finali dei servizi di posta.

In Europa furono fatti vari tentativi per giungere ad una legislazione comune. Il risultato fu la Direttiva 2002/58/CE del Parlamento Europeo e del Consiglio del 12 luglio 2002, che costituì l'obbligo per gli Stati aderenti alla Comunità Europea di emanare provvedimenti legislativi sul principio dell'*opt-in* e quindi del preventivo consenso del destinatario.

In Italia la principale fonte normativa sull'argomento è la legge 675/96 sulla protezione dei dati personali. L'indirizzo di posta elettronica è considerato come un dato personale, anche se non contiene il nome del titolare. La legge sulla privacy non vieta direttamente l'invio di posta commerciale, ma limita l'uso dell'indirizzo di posta elettronica in determinati casi. Un principio importante è che gli indirizzi e-mail reperibili su internet non sono pubblici e non possono essere usati per fini com-

*segue*

mercials. Non basta quindi, per poter considerare pubblico un indirizzo di e-mail, il fatto che tale indirizzo sia conoscibile, in determinate circostanze, da una pluralità di persone come può succedere per un indirizzo pubblicato su Internet. Inoltre non possono essere considerati pubblici neanche gli indirizzi di e-mail che vengono pubblicati su forum o newsgroup. Gli indirizzi e-mail in rete possono essere utilizzati solo per le finalità che hanno portato alla loro pubblicazione. Questo principio rende pertanto non conformi alla legge né la raccolta automatica di indirizzi di e-mail presenti su internet né la loro creazione artificiosa, attività che si possono realizzare oggi con appositi software. Inoltre, la legge obbliga le persone fisiche o giuridiche a cui sono stati consegnati i dati, a fornire una descrizione chiara e precisa di quale uso ne verrà fatto: lettura, memorizzazione, trasferimento a terze parti, comunicazioni di servizio o comunicazioni commerciali; inoltre nel momento in cui si forniscono i dati, o in qualunque momento successivo, i titolari dei dati hanno il diritto di sapere entro 5 giorni dalla richiesta in quali termini verranno utilizzati o anche di limitarne o proibirne completamente l'uso.

Questo elemento è molto importante perché neutralizza la difesa degli *spammer* che si basa sulla classificazione degli indirizzi di posta elettronica reperiti sul web come pubblici. È possibile quindi perseguire contro gli *spammer* già grazie alla legge 675/96 anche se in realtà il procedimento si rivela lungo e costoso e soprattutto riguarda solo gli *spammer* italiani.

Sono state poi varate anche legislazioni più specifiche in materia. Per primo il decreto legislativo 171 del 1998, il quale sancisce che il costo pubblicitario deve essere sostenuto interamente da chi fa la pubblicità e non da chi la subisce. Da segnalare anche il decreto legislativo n.185 del 22 maggio 1999 che, quando ancora la Comunità Europea non si era espressa in materia, schierò l'Italia sul fronte *opt-in*. Dopo una serie di interventi mirati alla sospensione di attività illecite o alla denuncia all'autorità giudiziaria di talune aziende o persone fisiche il Garante della privacy è sceso in campo in maniera chiara e dettagliata per disciplinare l'argomento. Il decreto legislativo 30 giugno 2003 n. 196, denominato "Codice in materia di protezione dei dati personali", entrato in vigore dall'1 gennaio 2004, infatti, recepì nell'ordinamento italiano la direttiva europea 2002/58/CE e precisò vari aspetti legali riguardanti l'invio in Internet di e-mail promozionali o pubblicitarie.

## APPENDICE 2 - LA RACCOLTA DI INDIRIZZI

Gli *spammer* usano diverse tecniche per recuperare gli indirizzi di posta a cui inviare i loro messaggi indesiderati.

Le principali sono elencate di seguito:

- **Dictionary attack:** questa tecnica molto diffusa si basa semplicemente sull'indovinare gli indirizzi. Più precisamente lo *spammer* cerca di comporre e generare indirizzi che potrebbero effettivamente esistere. Per la parte destra della chiocciola (@) usa nomi di dominio validi e per la parte sinistra genera stringhe in base a qualche logica, per lo più nomi di persone. Per questo motivo l'indirizzo nome.cognome@dominio.it è uno dei più soggetti a questo tipo di attacco.
- **Address list:** un secondo sistema consiste nell'acquisire liste di indirizzi da soggetti che li raccolgono per poi rivenderli. Le liste di indirizzi selezionate, per esempio, sull'attività professionale del destinatario vengono vendute a prezzi elevati, che possono arrivare a diversi dollari per indirizzo nel caso di medici e commercialisti.
- **Spambot:** come abbiamo visto nell'articolo, uno *spambot* o *harvester* è un particolare tipo di *web-crawler* in grado di raccogliere gli indirizzi e-mail dai siti web, dai newsgroup, dai post dei gruppi di discussione e dalle conversazioni delle *chat-room*. Gli basano sullo stesso principio del funzionamento degli spider dei motori di ricerca, ma a differenza di questi ultimi estraggono dalle pagine web tutti gli indirizzi presenti.

## BIBLIOGRAFIA

- [1] Nancy McGough: Reverse spam filtering - Winning Without Fighting, 4 settembre 2002. In: Infinite Ink, <http://www.ii.com/internet/messaging/-spam/> (consultato il 2 dicembre 2008).
- [2] Paul Graham, A plan for spam, <http://www.paulgraham.com/spam.html>
- [3] Ernesto Damiani, Sabrina De Capitani di Vimercati, Stefano Paraboschi Pierangela Samarati, Andrea Tironi, Luca Zaniboni: *Spam attacks: P2P to the rescue*. Proceedings of the

13-th international conference on World Wide Web (WWW 2004), 2004.

## Siti interessanti

SpamAssassin, <http://eu.spamassassin.org/> .  
Vipul's Razor, <http://razor.sourceforge.net/> .  
Cloudmark, <http://www.cloudmark.com/> .  
Despammed, <http://www.despammed.com/> .  
Spamex, <http://www.spamex.com/> .  
Spam Arrest, <http://www.spamarrest.com/>

ERNESTO DAMIANI si occupa di sicurezza nei Web services, processing di informazioni semi o non strutturate, semantics-aware content engineering per il multimedia, modelli e piattaforme per lo sviluppo di codice open source, infrastrutture e protocolli di rete avanzati, design e sviluppo di ambienti di rete sicuri ad alte prestazioni. È membro di numerosi editorial boards e ha pubblicato numerosi libri e circa 200 articoli scientifici oltre a brevetti internazionali. Collabora all'organizzazione di molti congressi, conferenze e workshop.  
E-mail: [ernesto.damiani@unimi.it](mailto:ernesto.damiani@unimi.it)

# DALLA MICROELETTRONICA ALLA NANOELETTRONICA

La microelettronica, nata con l'invenzione del transistor nel 1947, ha avuto un progresso straordinario. Oggi possiamo progettare e costruire circuiti integrati che hanno fino a due miliardi di transistori per chip. Tuttavia, man mano che le dimensioni si avvicineranno al nanometro, dovremo gradualmente cambiare materiali, principi di funzionamento dei transistori e metodi di costruzione, dando così vita alla nanoelettronica – la tecnologia del futuro. Si prospetta anche la possibilità concreta che il carbonio possa sostituire il silicio come elemento di base per la nanoelettronica.

## 1. INTRODUZIONE

**L'**elettronica ha compiuto cent'anni. Nata nel 1907 con l'invenzione del triodo, la prima valvola termoionica in grado di amplificare un segnale elettrico, l'elettronica ha avuto uno sviluppo prodigioso, fornendo il principale materiale da costruzione per il sistema nervoso della società umana.

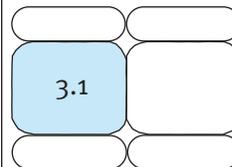
Per i primi decenni di vita, la funzione essenziale delle valvole fu l'amplificazione di segnali, permettendo così lo sviluppo della radiotelegrafia, della radio, degli amplificatori e della televisione. Durante la seconda guerra mondiale la gamma di applicazioni delle valvole si estese al radar, ai sistemi di controllo e al calcolo elettronico, culminando nel Colossus (1944), il calcolatore elettronico sviluppato dagli scienziati inglesi e specializzato nella decodifica dei messaggi segreti tra Hitler e il suo alto comando; e nell'ENIAC (1946), considerato il primo calcolatore elettronico. In queste ultime macchine la valvola fu usata come interruttore, invece di amplificatore, operando su segnali binari, e fornendo un'alternativa molto più veloce al *relais* elettromeccanico.

La ricerca per sostituire la valvola con un dispositivo più piccolo, meno costoso ed energeticamente più efficiente, iniziata negli anni trenta, portò finalmente all'invenzione del transistor nel 1947, ad opera di tre scienziati della Bell Labs (Figura 1). È con il transistor che nasce la microelettronica.

Per i primi dieci anni i transistori furono costruiti uno alla volta, usando il germanio come elemento semiconduttore. Nel 1959, l'invenzione del processo planare alla Fairchild Semiconductor, ad opera dell'ingegnere svizzero Jean Hoerni, fa cambiare tutto. Si iniziò così a costruire i transistori, un centinaio alla volta, su di una piastrina di silicio - elemento semiconduttore con caratteristiche più favorevoli del germanio - facendo scomparire il germanio nel giro di pochi anni. Con il processo planare, visto che i transistori erano costruiti uno accanto all'altro, venne naturale pensare di collegarli insieme, realizzando così il circuito integrato, una svolta decisiva per la microelettronica (Figura 2). Con l'avvento del circuito integrato, la microelettronica, nota anche come elettronica allo stato so-



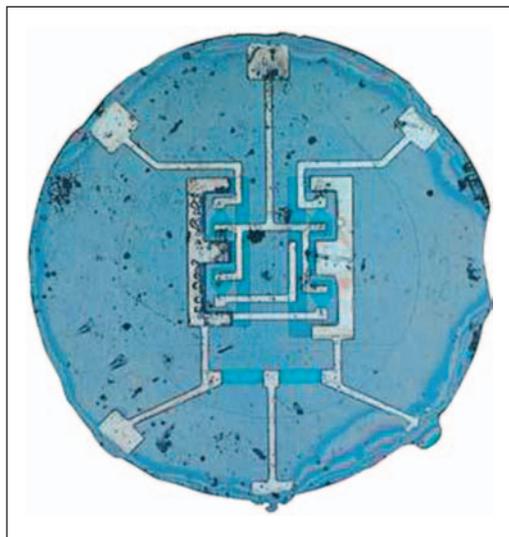
Federico Faggin





**FIGURA 1**

*Replica del primo transistor costruito nei laboratori di ricerca della AT&T Bell Labs nel 1947 ad opera di John Bardeen, Walter Brattain and William Shockley. Questo transistor a punta di contatto dimostrò per la prima volta che un dispositivo completamente allo stato solido poteva amplificare un segnale elettrico, creando la possibilità di sostituire le valvole termoioniche e motivando così la nascita della moderna microelettronica*



**FIGURA 2**

*Il primo circuito integrato costruito con il processo planare sulla superficie di una piastrina di silicio monocristallino. Il circuito integrato fu inventato da Robert Noyce alla Fairchild Semiconductor in Palo Alto, California, nel 1959. Con il circuito integrato fu possibile costruire simultaneamente sia gli elementi attivi che gli elementi passivi di un circuito elettronico, comprese le loro interconnessioni, permettendo nel tempo la costruzione di un intero computer su di un solo chip*

lido, cominciò a sostituire non soltanto i componenti attivi, ma anche quelli passivi: resistenze, condensatori, diodi ecc., nonché la porzione del circuito stampato necessaria a collegarli insieme.

Ulteriore tappa fondamentale della microelettronica fu l'invenzione di un nuovo tipo di transistor, il transistor MOS (*Metallo Ossido Semiconduttore*), che una volta perfezionato cominciò a sostituire i transistori bipolari usati in tutti i primi circuiti integrati. Con l'invenzione del processo MOS autoallineante con porta (*gate*) di silicio policristallino, fu possibile costruire le prime memorie a semiconduttore e il primo microprocessore. Il campo applicativo della microelettronica si allargò ancora una volta, rimpiazzando le memorie a ferrite magnetica fino allora usate nei computer e creando per la prima volta un intero computer allo stato solido. Nel giro di altri dieci anni i circuiti integrati MOS finirono per soppiantare quasi totalmente i circuiti integrati bipolari, contemporaneamente estendendo la loro capacità applicativa ai circuiti integrati analogici, alle memorie non volatili e ai sensori d'immagine; le ultime due applicazioni impossibili da fare sia con transistori bipolari sia con transistori MOS con *gate* di alluminio.

Nel 1965, Gordon Moore, uno dei due fondatori dell'Intel (1968), allora capo dei laboratori di ricerca della Fairchild Semiconductor, osservò che ogni anno il numero di transistori in un circuito integrato raddoppiava, prognosticando che questo comportamento sarebbe continuato nel futuro. Questa osservazione venne più tardi chiamata la legge di Moore, anche se ovviamente non è una legge fisica come la legge di Newton e la sua validità è limitata nel tempo. La longevità della legge di Moore è dovuta al principio dello *scaling*, scoperto qualche anno dopo, secondo il quale, riducendo le dimensioni critiche del transistor MOS nella stessa proporzione, si diminuisce non solo l'area del transistor e la sua potenza dissipata, ma si aumenta anche la sua velocità. A partire dalla metà degli anni settanta, lo *scaling* ha rappresentato la strategia fondamentale per migliorare le prestazioni e ridurre il costo dei circuiti integrati.

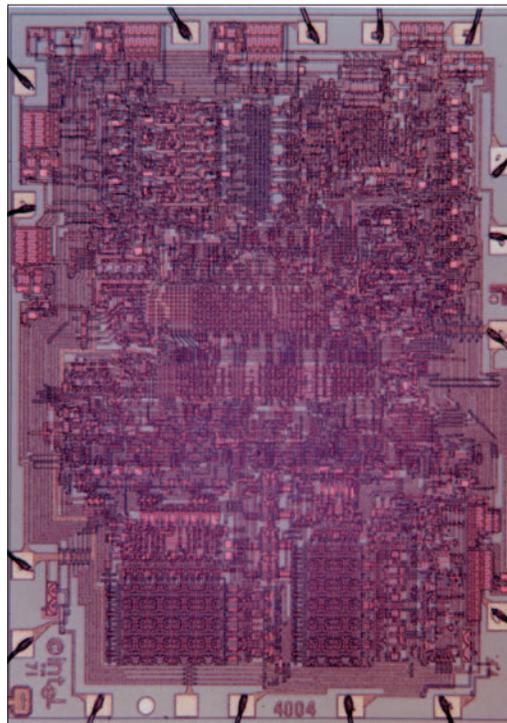
Nel 1970 la litografia più avanzata in produzione era in grado di stampare linee di 6  $\mu\text{m}$ . Oggi siamo arrivati a 45 nm, riducendo l'area di

un transistor di circa 20.000 volte in 38 anni. Nel 1970, la memoria RAM più avanzata era statica e aveva 256 bit, con un tempo di accesso di 1,5  $\mu$ s. Oggi la RAM statica più avanzata ha 64 Mbit con un tempo di accesso di 2,6 ns. Ancora più rapido è stato il progresso nel campo delle memorie RAM dinamiche e nelle memorie flash, che non esistevano ancora nel 1970. La memoria flash più avanzata oggi in produzione ha 512 milioni di celle, con un tempo di accesso di circa 50 ns, integrando circa 1,6 miliardi di transistori. Con la memoria flash è addirittura possibile immagazzinare più di un bit per cella, dosando il numero di elettroni che sono iniettati nel *floating gate* della cella di memoria e raggiungendo fino a 8 Gbit di memoria con un chip da 512 milioni di celle. Faccio notare che la differenza tra uno stato e l'altro è dell'ordine di cento elettroni in più o in meno caricati nel *floating gate*.

Nel campo dei microprocessori il progresso è stato altrettanto sorprendente. Il primo microprocessore, l'Intel 4004, introdotto nel mercato nel 1971, integrava circa 2300 transistori in un'area di 12 mm quadrati, usando 6 maschere per la sua fabbricazione (Figura 3). Se confrontiamo il 4004 con uno degli ultimi microprocessori Intel, lo Xeon L5420, introdotto all'inizio del 2008, possiamo verificare lo straordinario progresso fatto in 37 anni. Lo Xeon integra circa 820 milioni di transistori, occupa un'area di 210 mm quadrati ed è fabbricato usando circa 30 maschere - la maggioranza dei transistori sono usati per due livelli di memoria *cache* (Figura 4). Alla fine del 2008 l'Intel ha annunciato una nuova serie di microprocessori Xeon, la serie 7400, con 6 *core* e 16 MB di memoria *cache*, integrando ben 1,9 miliardi di transistori nello stesso chip.

## 2. LA MICROELETTRONICA CONTEMPORANEA E LE PROSPETTIVE FUTURE

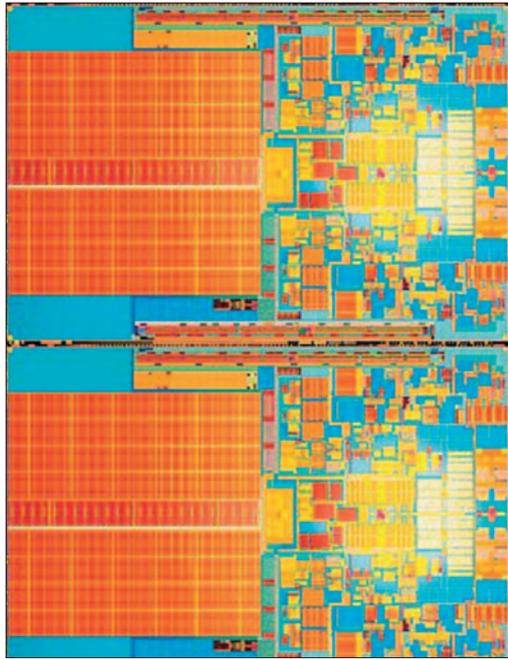
Nella storia del progresso tecnologico, la microelettronica ha superato di gran lunga qualsiasi altra disciplina nell'aumentare le prestazioni nel tempo, mantenendo lo stesso costo. Ma fino a quando sarà possibile continuare la corsa esponenziale degli ultimi 40 anni? Oggi lo *scaling* sta incontrando difficoltà sempre più grandi, mano a mano che ci avviciniamo



**FIGURA 3**

*Immagine del primo microprocessore, l'Intel 4004, introdotto nel mercato nel novembre 1971. Il 4004 faceva la funzione di un'intera unità centrale di un piccolo computer (CPU), integrando circa 2300 transistori in un'area di 12 mm quadrati, con litografia da 6  $\mu$ m, e usando 6 maschere per la sua fabbricazione. Il 4004 era in grado di eseguire circa 100.000 istruzioni al secondo, operanti su 4 bit. Ciascun ciclo di istruzione richiedeva 8 periodi di clock, alla frequenza di circa 750 kHz. Il 4004 aveva 16 pin, la tensione di lavoro era di 15 V, la potenza dissipata era 750 mW e costava \$35 in quantità di 1000 pezzi. Il 4004 faceva parte di una famiglia di 4 componenti (CPU, ROM, RAM e I/O) con cui si potevano costruire semplici computer adeguati per molte applicazioni di calcolo e di controllo. (Faggin F., Hoff M. E. Jr. Standard Parts And Custom Design Merge In A Four-Chip Processor Kit. Electronics Magazine, April 24, 1972)*

alle dimensioni molecolari. Già con la litografia da 90 nm, che risale al 2003, lo spessore dell'isolante del *gate* - fatto con biossido di silicio - fu ridotto a 1,2 nm, 5 strati molecolari in tutto. Se si riduce lo spessore a 4 strati molecolari, il *leakage* del *gate* dovuto all'effetto *tunneling*, aumenta al punto da rendere il transistor quasi inservibile. Pertanto le due generazioni successive hanno evitato di scalare lo spessore dell'ossido riducendo così



**FIGURA 4**

*Immagine di un microprocessore contemporaneo, l'Intel Xeon L5420, introdotto all'inizio del 2008. Confrontando le caratteristiche dello Xeon L5420 con le caratteristiche del 4004, possiamo verificare lo straordinario progresso fatto in 37 anni: lo Xeon L5420 integra circa 820 milioni di transistori con litografia da 45 nm (due Xeon Core 2, ciascuno con 410 milioni di transistori), occupa un'area di 210 mm quadrati ed è fabbricato usando circa 30 maschere - la maggioranza dei transistori sono usati per due livelli di memoria cache. Lo Xeon ha quattro unità centrali funzionanti in parallelo, con cui esegue fino a 20 miliardi di istruzioni per secondo, ciascuna operante su 64 bit. La frequenza di clock è di 2,5 GHz e durante ciascun periodo di clock, il chip esegue fino ad 8 istruzioni in parallelo. Lo Xeon L5420 ha 771 pin, dissipa 50 W, usa una tensione di lavoro di 1,2 V e costa \$380 in quantità di 1000 pezzi - più o meno lo stesso costo del 4004 se si considera l'effetto dell'inflazione*

l'aumento delle prestazioni dei transistori che altrimenti sarebbero state possibili. Questa limitazione è stata anticipata da tempo dall'industria dei semiconduttori che ha fatto ricerca per molti anni allo scopo di scoprire materiali ad alta costante dielettrica da sostituire al biossido di silicio. Solo così è possibile aumentare le prestazioni del transistor senza dover diminuire lo spessore dell'isolante che comporta un aumento della corrente di *leakage*. Questa ricerca si è rivelata molto più

ardua del previsto in quanto non è soltanto necessario trovare un isolante adeguato, ma l'intero sistema metallo-ossido-silicio, chiamato *stack*, deve essere compatibile a tutti i livelli richiesti per la costruzione, il funzionamento e l'affidabilità del transistor. Il sistema precedente, costituito da silicio policristallino, biossido di silicio e silicio è rimasto inalterato dal 1968, anno in cui il primo circuito integrato commerciale fatto con la tecnologia *silicon gate* - tecnologia e circuito da me sviluppati - fu introdotto nel mercato.

Finora, soltanto l'Intel è riuscita a sviluppare uno *stack* adeguato, dopo uno sforzo decennale, portando in produzione alla fine del 2007 i primi circuiti integrati a 45 nm con un isolante basato sull'elemento afnio (Hf, metallo di transizione con numero atomico 72), e due metalli diversi (uno per il transistor a canale P, l'altro per il transistor a canale N) che non sono stati specificati. Si stima che la permittività dielettrica relativa del nuovo isolante abbia un valore tra 10 e 15 - l'Intel non ha ancora pubblicato il valore esatto - che è un fattore da 2,5 a 4 volte superiore alla permittività dielettrica relativa del biossido di silicio, che è 3,9. Con questo nuovo *stack*, l'Intel ha potuto usare un isolante più spesso, circa 2,3 nm, riducendo drasticamente la corrente di *leakage* e quindi la potenza dissipata, e contemporaneamente aumentando le prestazioni dei transistori come sarebbe stato possibile fare usando 3-4 strati molecolari di ossido di silicio.

L'Intel ha dichiarato che lo stesso *stack* sarà utilizzato nel nodo litografico successivo, 32 nm, che entrerà in produzione nel 2009. Tuttavia, per la generazione a 22 nm, c'è ancora qualche incertezza sulla possibilità di usare gli stessi materiali senza conseguenze negative. Il problema sta diventando sempre più difficile e già al livello dei 22 nm il metodo usato per le due generazioni precedenti è in dubbio. Il *silicon gate*, durato per circa venti generazioni, deve essere d'ora in poi sostituito da materiali nuovi che richiedono un grandissimo sforzo economico per il loro sviluppo e che durano soltanto qualche generazione.

Malgrado queste difficoltà, penso che sarà possibile avere chip a 22 nm in produzione nel 2011-2012, allungando solo di poco la tabella di marcia che ha recentemente registrato una media di circa due anni per generazio-

ne (la legge di Moore originariamente aveva predetto un tempo di raddoppiamento di un anno; il tempo di raddoppiamento è gradualmente aumentato a circa due anni). Prevedo comunque che le generazioni successive cominceranno a richiedere tempi di sviluppo sempre più lunghi, con i 16 nm in produzione nel periodo 2015-2016, e gli 11 nm, in produzione non prima del 2019.

Gli esperti non sono ancora sicuri che sarà possibile andare sotto i 10 nm, tuttavia se il vantaggio economico è sufficiente a giustificare l'enorme costo dello sviluppo, e soprattutto se non ci sono alternative migliori, penso che l'industria riuscirà a ridurre ulteriormente le dimensioni, magari fino a 4-5 nm, usando però tempi di sviluppo sempre più lunghi. Senza dubbio comunque la strada maestra dello *scaling* del transistor MOS avrà una vita limitata e si dovranno cercare altre strutture per continuare la corsa al sempre più piccolo, più veloce e meno costoso.

Una strategia evolutiva che sta già maturando sotto i nostri occhi consiste nel fare chip con più di uno strato attivo. Il metodo più elegante, ma anche più complesso, è di costruire chip con più di un livello attivo sulla stessa fetta di silicio. Per esempio, la Foveon ha costruito sensori di immagine dove i tre fotodiodi per il rilevamento dei colori primari sono uno sopra l'altro, invece che uno accanto all'altro. Il metodo più usato, tuttavia, è quello di montare più di un chip, uno sopra l'altro nello stesso *package*, e gran parte delle memorie *flash* hanno adottato queste tecniche per aumentare il numero di bit di memoria contenuti nello stesso dispositivo. Però la strada più avanzata si baserà sulla nuova tecnologia che oggi va sotto il nome generico di *wafer-scale-packaging*. Metodi in grado di estendere lo stesso principio usato nel processo planare anche all'incapsulamento dei chip, assemblando quindi tutti in una volta i chip che si trovano in una fetta di silicio, e separandoli alla fine del processo, pronti per essere montati sui circuiti stampati.

Tutta questa attività sta creando la conoscenza necessaria per iniziare una nuova strada che va sotto il nome generico di 3D ICs, circuiti integrati a tre dimensioni. Le tecniche di base per fare 3D ICs sono l'assottigliamento delle fette di silicio ad uno spesso-

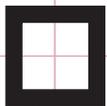
re di qualche decimo di mm, lo sviluppo di tecniche per fare *vias*, da usare per il collegamento verticale tra i circuiti integrati, e tecniche di *stacking* per sovrapporre e allineare sia chip sia wafer. Una delle ragioni pratiche che favoriscono il movimento verso le tre dimensioni è il fatto che la potenza dissipata dovuta alle interconnessioni diventa sempre più alta mano a mano che l'area dei chip aumenta. Facendo dei chip più piccoli e montandoli uno sopra l'altro si può ridurre la lunghezza media delle interconnessioni e quindi la potenza dissipata.

Non c'è dubbio che in futuro i 3D ICs, che oggi sono soltanto all'inizio del loro sviluppo, riceveranno sempre più attenzione e diventeranno di uso comune man mano che lo *scaling* diventerà sempre più difficile e costoso, prolungando così il periodo di continui miglioramenti per la microelettronica anche dopo che lo *scaling* tradizionale avrà finito il suo corso.

Faccio notare che uso il termine "microelettronica" anche quando le dimensioni dei transistori sono di qualche decina di nanometri e potrei quindi usare appropriatamente il termine "nanoelettronica." Questo perché vorrei riservare il termine "nanoelettronica" ad una nuova classe di dispositivi elettronici più piccoli e più veloci, basati su nuovi principi di funzionamento, che promettono di sostituire i transistori MOS fra circa 10 anni. La nanoelettronica offre quindi la nuova strada per continuare ad aumentare le prestazioni e ridurre le dimensioni e il costo dei circuiti integrati una volta che i transistori MOS hanno raggiunto il limite fisico dello *scaling*.

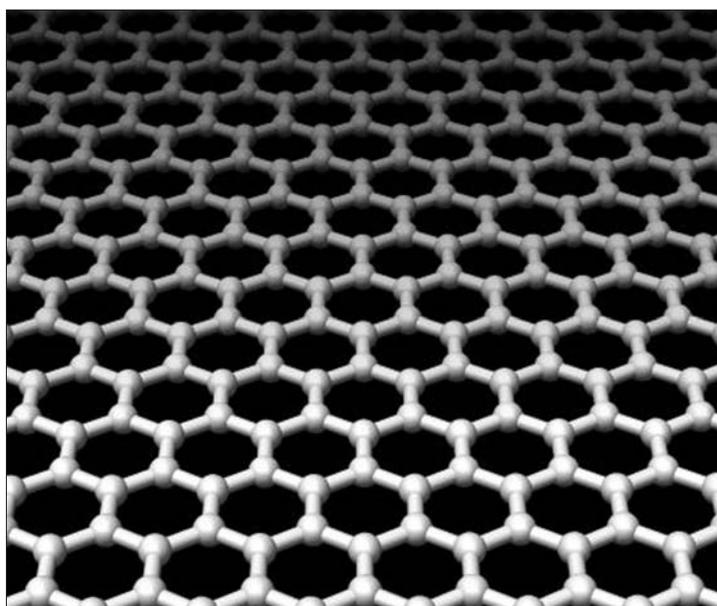
### 3. LA PROMESSA DELLA NANOELETRONICA

La nanoelettronica è oggi più scienza che tecnologia. Di fatto non ha ancora raggiunto il livello di commercializzazione benché il progresso negli ultimi dieci anni sia stato fenomenale. Alla scala del nanometro esistono proprietà elettriche e magnetiche sorprendenti in materiali sia vecchi che nuovi, che hanno dato vita a molti nuovi filoni di ricerca, ciascuno basato su una nuova proprietà o su un nuovo materiale promettente. Potrei dire che oggi c'è anche troppa dovizia di possibilità, rendendo difficile discernere la strada



maestra da strade minori o dai vicoli ciechi. Molti ricercatori stanno cercando di trovare un dispositivo molto più piccolo di un transistor MOS, senza preoccuparsi troppo dei problemi di fabbricazione. Questo va bene, purché non si dimentichi che il problema da risolvere è ben maggiore di quello di fare semplicemente un transistor più piccolo. Un transistor più piccolo è solo una condizione necessaria ma non sufficiente per sostituire il transistor MOS. È necessario anche che il nuovo dispositivo sia molto più veloce, dissipati molta meno energia, operi in una gamma di temperature equivalente ai dispositivi al silicio e sia collegabile ad altri dispositivi analoghi per fabbricare circuiti integrati con decine di miliardi di componenti a costo inferiore e con affidabilità almeno equivalente a quella dei circuiti integrati convenzionali. Un problema monumentale!

Penso che la scala del problema da risolvere sia talmente vasta che l'unica strategia possibile è quella di trovare nanomateriali e nanostrutture in grado di sostituire materiali e dispositivi tradizionali *gradualmente* - man mano che i vecchi materiali si riveleranno li-



**FIGURA 5**

*Struttura del grafene. Il grafene è composto da un solo strato molecolare di carbonio, con un atomo di carbonio su ciascun vertice di un reticolato esagonale il cui lato misura 0,142 nm. La mobilità degli elettroni nel grafene è molto più alta della mobilità degli elettroni nel silicio, dando la possibilità teorica di fare transistori delle dimensioni di pochi nanometri che possono operare alla frequenza di un Tera Hertz (1000 GHz)*

mitanti - e con processi di fabbricazione compatibili con le tecniche di fabbricazione dei circuiti integrati contemporanei. Solo così si potrà far leva sull'investimento collettivo che ha portato la microelettronica all'incredibile livello di sviluppo odierno.

Faccio ora un esempio di come l'evoluzione dalla microelettronica alla nanoelettronica potrebbe avvenire, considerando che i problemi più critici da risolvere per fare circuiti integrati sempre più avanzati sono il problema delle interconnessioni ed il problema della potenza dissipata, che è dovuto soprattutto alla capacità e alla resistenza elettrica delle interconnessioni. Oggi impieghiamo fino a nove livelli di interconnessioni, usando sottilissimi strati di rame. Però, come riduciamo lo spessore e la larghezza del nastrino di rame che collega i vari transistori, la resistenza per unità di lunghezza aumenta rapidamente, contribuendo negativamente sia alla potenza dissipata sia alle prestazioni del circuito. Se riuscissimo a scalare soltanto le dimensioni delle interconnessioni, senza però degradarne le caratteristiche elettriche, potremmo continuare a ridurre le dimensioni dei chip ancora per anni senza che sia necessario trovare un dispositivo attivo più piccolo. Per far ciò è necessario trovare un nanomateriale con proprietà di conduzione elettrica migliori del rame, quando le dimensioni sono ridotte a qualche nanometro. Recentemente pare che sia stato trovato un materiale promettente: il grafene (Figura 5).

Il grafene è una delle molte forme allotropiche del carbonio che sono state scoperte recentemente, cominciando con la fullerite, scoperta nel 1985, e nota anche con il nome di *buckyball*. La fullerite è una sfera vuota composta da 60 atomi di carbonio regolarmente disposti sulla sua superficie. Nel 1991 il nanotubo di carbonio fu riscoperto e portato all'attenzione scientifica. Si tratta di un cilindro vuoto dal diametro minimo di poco più di un nanometro, la cui superficie è formata da atomi di carbonio con disposizione esagonale. Nel 2004 il grafene fu isolato per la prima volta: si tratta di un solo strato molecolare di carbonio dove gli atomi sono disposti in un reticolato esagonale, esattamente come in un nanotubo, però su di una superficie piatta. Questo è uno sviluppo insospettato visto che la grafite, che è

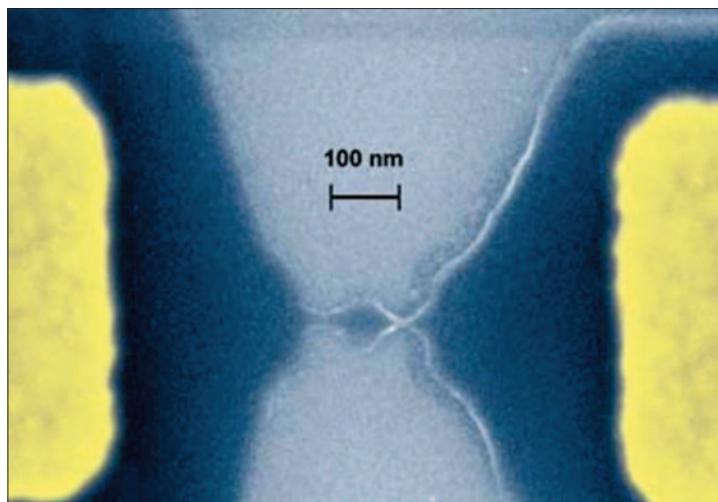
nota da sempre, è composta da tanti strati di grafene sovrapposti caoticamente.

Bene, nel giro di pochi anni gli scienziati hanno scoperto che il grafene ha proprietà insospettite, simili a quelle dei nanotubi, ma più promettenti in quanto la sua forma a due dimensioni è compatibile con la tecnologia dei circuiti integrati che sono costruiti sulla superficie di una fetta di silicio. Per esempio, un nastro di grafene della larghezza di pochi atomi può essere sia un semiconduttore sia un metallo. Non solo, le proprietà di conduzione elettrica della forma metallica sono notevolmente superiori a quelle del rame, aprendo la strada alla potenziale sostituzione del rame per le interconnessioni dei circuiti integrati. Recentemente sono stati dimostrati diversi tipi di transistori al grafene, da transistori simili al dispositivo MOS, i cui inventori sostengono che ha proprietà elettroniche superiori a quelle dei dispositivi fatti sul silicio, a transistori *quantum-dot* che fanno passare un elettrone alla volta (Figura 6). Ecco quindi prospettarsi per la prima volta una strada che non solo è compatibile con la tecnologia esistente, ma che si presta anche ad un'adozione graduale. Ciò che rende il grafene un materiale straordinario è che esso può esistere in nanoscala senza decomporsi, ossidarsi o diventare instabile; proprietà che nemmeno il silicio possiede. Si intravede quindi la possibilità reale, anche se ancora in divenire, di creare circuiti integrati in uno o due strati molecolari di carbonio - proposta impensabile soltanto dieci anni fa.

Un altro filone importante è l'uso di materiali magnetici su nanoscala in circuiti integrati basati sul silicio per fare memorie permanenti di caratteristiche superiori alle memorie flash e DRAM tradizionali. Questa nuova strada è promettente in quanto sposa il circuito integrato contemporaneo con la nuova disciplina chiamata "*spintronics*". Nata con la scoperta nel 1988 della magnetoresistenza gigante (GMR), la *spintronics* usa sia le proprietà di carica elettrica sia le proprietà magnetiche dell'elettrone che si manifestano nel suo spin quantico. I dispositivi spintronici si basano sull'abilità di creare e controllare correnti di elettroni con lo spin polarizzato in una certa direzione. Sensori magnetici basati sui principi spintronici - la *spin valve* - sono già in uso

da alcuni anni nelle memorie magnetiche rotanti. L'industria dei semiconduttori, già da molti anni, sta sviluppando memorie che usando principi spintronici che promettono di sostituire le memorie RAM statiche, le DRAM e le memorie flash con un'unica memoria universale, chiamata MRAM, che assomma le caratteristiche principali di velocità, costo per bit e non-volatilità che hanno giustificato finora l'esistenza di tre tipi diversi di memoria.

Recentemente l'IBM ha annunciato lo sviluppo di una nuova tecnologia per memorie non-volatili, chiamata "*RaceTrack*" con cui spera di soppiantare le memorie *flash* e le memorie magnetiche rotanti. Si tratta di un dispositivo allo stato solido dove le pareti dei domini magnetici si muovono all'interno di un nanoduttore, con un processo simile a un registro a scorrimento, permettendo di ottenere densità di memoria cento volte superiori alle attuali memorie *flash*, con velocità, affidabilità, costo e potenza dissipata favorevoli rispetto ad esse. L'IBM prevede che questa nuova tecnologia permetterà di fare memorie magnetiche con densità fino a 2 Gbit per mm quadrato. Benché questa tecnologia sia ancora allo stadio di ricerca e sviluppo, e potrebbe anche rivelarsi non adeguata alla commercializzazione, come minimo essa attesta l'enorme vitalità di questa nuova direzione tecnologica. Benché io abbia fatto soltanto alcuni esempi



**FIGURA 6**

*Transistore quantum-dot, fabbricato su di un nastro di grafene largo qualche nanometro all'Università di Manchester nel 2007. Il transistore è al centro dell'immagine, ma non è visibile. Le strutture visibili sono i contatti elettrici con il grafene. (A.K. Geim and K.S. Novoselov, Nature Material, 2007, 6, 183)*

delle nuove frontiere della nanoelettronica, non è ancora possibile discernere la strada evolutiva principale che sarà imboccata a livello industriale-commerciale tra circa dieci anni. Posso però dire che si intravedono già possibilità concrete di integrazione del vecchio e del nuovo che non solo permetteranno di prolungare la validità della legge di Moore ancora per molti decenni, ma che arricchiranno anche sostanzialmente le possibilità espressive del nuovo *hardware* del secolo XXI.

Concludendo, non posso fare a meno di far notare la possibilità concreta e inaspettata che l'*hardware* del futuro possa anch'esso essere basato sul carbonio, lo stesso elemento magico che ha permesso alla natura di evolvere la vita e creare la nanomacchina più avanzata del sistema solare: l'uomo.

## Bibliografia

- [1] Lojek Bo: *History of Semiconductor Engineering*. Springer-Verlag Berlin Heidelberg, 2007.
- [2] Bohr M., Mistry K., Smith S.: *Intel Demonstrates High-k + Metal Gate Transistor Breakthrough on 45 nm Microprocessors*. Intel Online Presentation, 2007.
- [3] Leduc Patrick: *What is 3D IC integration and what metrology is needed*. Conference on Frontiers of Characterization and Metrology for Nanoelectronics, March 2007.
- [4] Burr G.W., Kurdi B.N., Scott J.C., Lam C.H., Gopalakrishnan K., Shenoy R.S.: *Overview of Candidate Device Technologies for Storage-Class Memory*. *IBM Journal of Research and Development, Storage Technology and Systems*, Vol. 52, n. 4/5, 2008.
- [5] Savage Neil: *Researchers Pencil In Graphene Transistors*. IEEE Spectrum Online, June 2008.

Questo articolo costituisce la *lectio magistralis* fatta dall'Autore in occasione della consegna del diploma EUCIP *honoris causa*, avvenuta il 14 novembre 2008 a Roma nella sede della Confindustria.

*Mondo Digitale* si onora di avere tra i suoi collaboratori Federico Faggin, scienziato, inventore e imprenditore di fama mondiale.

FEDERICO FAGGIN, nato a Vicenza, si laureò in Fisica all'Università di Padova nel 1965, e si trasferì nella Silicon Valley, California, nel 1968, dove vive tuttora. Faggin è sia inventore che imprenditore, avendo fondato tre ditte di alta tecnologia. I suoi contributi più importanti sono stati la creazione della tecnologia MOS Silicon Gate (1968), il progetto del primo microprocessore (l'Intel 4004 nel 1971), e tre generazioni di microprocessori a 8 bit (8008, 8080 e Z80). Lo Z80 fu creato nella ditta Zilog da lui fondata nel 1974. Faggin attualmente Presidente Onorario della Synaptics, ditta che egli fondò nel 1986, e che diresse per molti anni. Ha ricevuto molti riconoscimenti, tra cui il Marconi Prize, il Kyoto Prize, il Lifetime Achievement Award dell'European Patent Office ed è annoverato nel National Inventors Hall of Fame in Akron Ohio.

# INFORMATICA MUSICALE DAGLI ALBORI AL PROSSIMO FUTURO

L'applicazione di metodi e tecnologie dell'informatica in campo musicale ha progressivamente interessato tutti gli aspetti della fruizione, della performance, della produzione, dell'educazione e del supporto alle attività creative. L'attenzione delle ricerche, degli strumenti e dei media più avanzati è oggi posta sull'interazione e sulla personalizzazione delle modalità di accesso e trattamento dell'informazione musicale nelle sue varie tipologie (spartiti, audio, video, testi), soprattutto se tra loro integrate e correlate.

## 1. INFORMAZIONE MUSICALE E INFORMATICA MUSICALE

L'applicazione dei metodi e delle tecnologie informatiche in campo musicale consente la rappresentazione, la conservazione, la trasmissione, la riproduzione, l'elaborazione, l'analisi e la sintesi dell'informazione musicale. È opportuno ricordare che l'informazione musicale può essere considerata a vari livelli di rappresentazione che caratterizzano le attività musicali più comuni:

### **strutturale**

composizione e analisi musicologica: strutture e relazioni strutturali, musica descritta in termini di oggetti musicali ad alto livello, forme, regole, procedimenti ecc.;

### **simbolico**

esecuzione ed interpretazione, editoria musicale: successioni e sovrapposizioni ordinate e temporizzate di eventi;

### **esecutivo**

orchestrazione, strumentazione, organologia, liuteria: generazione e controllo di vibrazioni udibili, flussi temporizzati di parametri per il controllo di modelli timbrici e interpretativi;

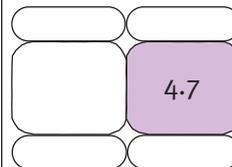
### **sonoro**

registrazione, riproduzione, produzione, postproduzione: flussi temporizzati di segnali audio digitali.

È altresì opportuno ricordare che un'informazione può essere rappresentata tanto da diverse codifiche quanto da diversi procedimenti che possono produrla, secondo il vecchio slogan dell'informatica "algoritmi+dati=programmi" [1]. Per cui è stata di volta in volta una scelta, quella di rappresentare una certa informazione musicale come tabelle di dati piuttosto che come regole per generarle. Per esempio, si pensi alla tabella dei valori di una forma d'onda sinusoidale piuttosto che all'algoritmo per calcolarli, o anche si pensi alla scrittura su partitura di un certo contrappunto piuttosto che alle regole sintattiche per generarlo. Il viaggio nel tempo che si svolge nei prossimi paragrafi mostrerà come l'applicazione di metodi e tecnologie informatici all'informazione musicale si sia estesa progressivamente, arrivando oggi a interessare praticamente tutti gli aspetti della musica.



Goffredo Haus



## 2. GLI ALBORI (DAGLI ANNI CINQUANTA AGLI ANNI SESSANTA)

Ben prima della disponibilità dei computer si era pensato alla manipolazione dell'informazione musicale a livello di simboli: si pensi per esempio ai famosi "giochi" di composizione automatica di Haydn [2] e Mozart [3] della fine del settecento. Ma è proprio la comparsa del computer che ha permesso di avere lo strumento per l'esecuzione di procedimenti automatici.

Andando oltre i simboli e arrivando al suono, sin dalla presentazione al pubblico nel 1946 dell'ENIAC, il primo computer digitale, si pensò a generare suoni sfruttando i diversi stati dei componenti. Ma le prime applicazioni del computer che possano essere propriamente etichettate come musicali risalgono alla fine degli anni cinquanta. I primi esperimenti furono orientati alla composizione automatica, al controllo di componenti oscillatori per la sintesi del suono, alla trascrizione degli spartiti per l'esecuzione automatica al computer, al computer come esecutore diretto del pensiero compositivo.

### 2.1. Generazione di testi musicali

Al 1957 si può datare il primo esperimento di applicazione musicale del computer. Il compositore Lejaren A. Hiller, aiutato dal matematico Leonard Isaacson, programmò una composizione elettronica, "Illiac Suite" [4], sul computer ILLIAC all'Università dell'Illinois a Champaign; si tratta di un quartetto

d'archi composto sulla base di un gioco probabilistico (catene di Markov e metodo di Montecarlo tra gli strumenti formali utilizzati) in quattro sezioni derivate da quattro esperimenti diversi. Il musicista Hiller, frustrato dall'imprecisione e dall'inattendibilità dei sistemi analogici, nel computer digitale trovò il mezzo per creare suoni precisamente e completamente definiti.

### 2.2. Generazione di suoni digitali

Sempre nel 1957, negli USA, presso i laboratori della Bell Telephone, Max Mathews realizzò il primo programma per la sintesi del suono MUSIC I a cui seguono poi tutte le versioni successive fino a MUSIC V [5].

MUSIC V fu poi ripreso da altri programmatori-musicisti portando alla definizione di linguaggi più recenti come – in ordine temporale - il MUSIC 360 (Barry Vercoe), MUSIC1000 (Dean Wallraff), CMUSIC (Gareth Loy), CSOUND (autori vari) e i derivati, fino ai più recenti ad MPEG4 SASL e SAOL.

### 2.3. Trascrizione al computer e composizione pseudocasuale

Pietro Grossi, il pioniere dell'informatica musicale italiana, fu un precursore tanto dell'informatica musicale quanto della personalizzazione nella fruizione e nella generazione di arte multimediale (testo, musica, visuale).

Dapprima, ebbe l'intuizione di applicare questa possibilità alla musica. Nel 1967, ai laboratori ricerca della Olivetti General Electric, programmò un computer GE-115 per controllare in frequenza e in durata un banco di oscillatori ad onda quadra al fine di produrre suoni musicali. Grossi trascrisse per GE-115 il 5° Capriccio di Paganini e altri brani classici. Inoltre con la medesima tecnica di sintesi del suono, realizzò alcune composizioni elettroniche. Tutte queste opere sperimentali furono registrate e stampate su un disco 45 giri denominato *Computer Concerto* in occasione delle festività natalizie del 1967 (Figura 1).

Grossi proseguì nelle sue attività sperimentali sviluppando principalmente il filone delle trascrizioni e della personalizzazione dell'esecuzione di brani classici e serie di composizioni basate su metodi pseudocasuali.



**FIGURA 1**  
Il lato A del 45 giri  
"Computer  
Concerto" del  
M° Pietro Grossi

## 2.4. Computer: strumento per il compositore/interprete

Nel 1970, ancora Max Mathews realizza GROOVE (*Generated Real-time Output Operations on Voltage-controlled Equipment*) [6], il primo sistema ibrido sviluppato per la sintesi del suono, utilizzando un computer Honeywell DDP-224, dotato di tastiera musicale a 24 tasti, tastiera alfanumerica, 4 potenziometri e joystick: il tutto per controllare frequenze, ampiezze, durate, evoluzione temporale delle timbriche in tempo reale. Con questo prototipo Mathews intendeva creare un sistema per la programmazione e l'esecuzione in tempo reale. Con il sistema GROOVE non ci sono intermediari, il compositore produce il suono direttamente trattando l'informazione musicale nella sua completezza.

Ancora oggi si prosegue a progredire seguendo questa linea tracciata da Mathews che a buon titolo può essere considerato uno dei padri fondatori della *computer music*.

## 3. CRESCITA DELL'INFORMATICA MUSICALE (DAGLI ANNI SETTANTA AGLI ANNI NOVANTA)

Le tecnologie informatiche degli anni cinquanta e sessanta imponevano forti limitazioni alle applicazioni musicali possibili, soprattutto per la componente multimediale - audio e video - dell'informazione musicale: non si poteva elaborare il suono in tempo reale, non si poteva interagire con i contenuti musicali di un brano non essendo possibile riconoscere in tempo reale le principali proprietà musicali, non erano disponibili memorie digitali ad accesso diretto per registrare suoni e relativi video, non erano disponibili canali per la trasmissione dei dati audio e video di banda sufficiente ecc..

I metodi, i modelli, le tecniche e le tecnologie specifiche per l'informatica musicale hanno tratto dall'evoluzione generale stimolo per evolvere a loro volta e giungere per molti aspetti a piena maturazione.

È bene ricordare che il testo musicale - partiture, spartiti ecc. - è codificabile con poche centinaia di simboli per pagina e poteva quindi essere elaborato fin da allora da processori *general purpose*, essendo la complessità computazionale richiesta paragonabile a

quella del testo in linguaggio naturale. Il suono invece richiede decine di migliaia di numeri al secondo e necessita perciò di un costo computazionale superiore di alcuni ordini di grandezza. Per questo motivo, l'evoluzione dei processori *general purpose*, dei *digital signal processor*, delle memorie, dei dispositivi per l'interazione uomo-macchina, e in generale delle architetture, seguita dall'introduzione del web e dei dispositivi mobili, ha aperto la strada al trattamento del suono e più in generale dell'informazione multimediale.

Per comprendere come questa evoluzione abbia influito sull'informatica musicale, alcuni aspetti specifici meritano di essere brevemente messi in luce: la sintesi digitale dei suoni musicali, le architetture dedicate al trattamento dei suoni musicali, l'editoria elettronica del testo musicale; questi tre aspetti sono stati di particolare rilievo, anche perchè hanno caratterizzato i principali cambiamenti nella prassi di alcune professioni musicali tipiche.

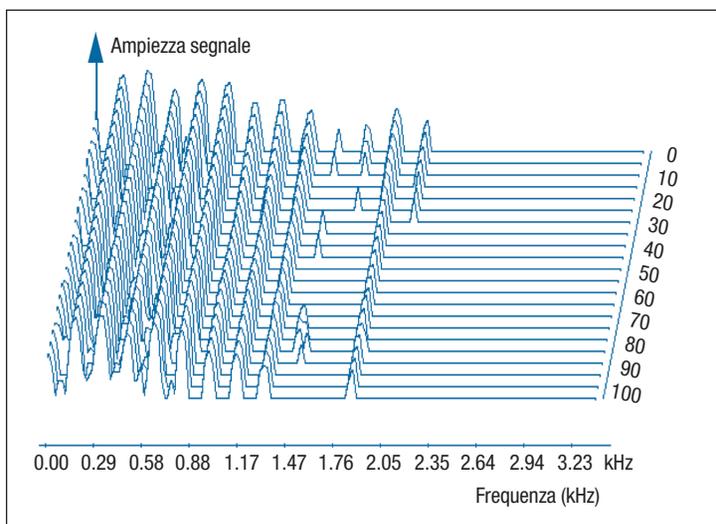
## 3.1. Modelli e tecniche per la generazione digitale di segnali audio musicali

Quello che in principio non era possibile, poi è diventato gradualmente sempre più alla portata dei laboratori di ricerca, poi degli studi di produzione, poi degli esperti e infine degli appassionati: la programmazione timbrica, meglio nota come Computer Music o Musica Informatica [7, 8].

Si possono distinguere fondamentalmente tre approcci alla generazione sintetica delle timbriche musicali:

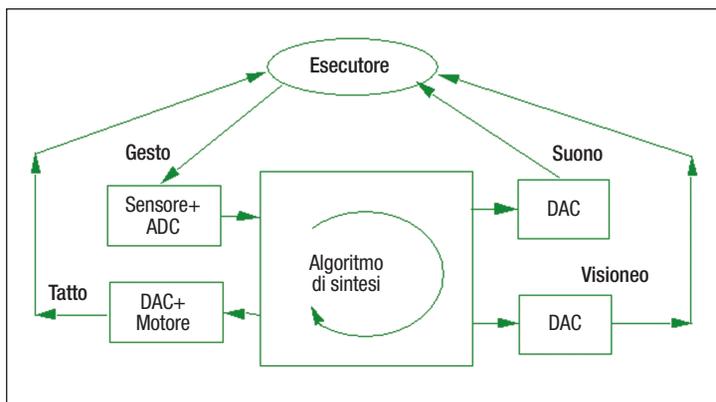
- modelli per campionamento;
- modelli matematici;
- modelli fisici.

Nel primo caso, suoni naturali di strumenti musicali od altre sorgenti sonore sono registrati digitalmente e poi riprodotti. Un approccio qualitativamente apprezzabile richiede però molta memoria, poiché molte varianti timbriche sono proprie del repertorio sonoro di uno strumento al variare dell'altezza, dell'intensità, della durata e della tecnica strumentale adoperata per produrre il suono. In ogni caso il costo computazionale è irrilevante. I limiti di questo metodo sono evidenti - principalmente staticità dei timbri registrati e scarse possibilità di controllo delle varianti - ed è utilizzato principalmente per abbattere i costi dei



**FIGURA 2**

*Modello tridimensionale di rappresentazione grafica del suono per la sintesi additiva: le tre variabili sono la frequenza, il tempo e l'ampiezza; ogni linea del grafico rappresenta la composizione del suono ad un certo istante di tempo come spettro di diverse frequenze, ciascuna dotata di una propria ampiezza relativa a quell'istante*



**FIGURA 3**

*Schema del modello di interazione tra esecutore umano e strumento musicale virtuale basato su modelli fisici dei corpi vibranti: oltre alla retroazione costituita dai suoni prodotti suonando lo strumento virtuale e alla visualizzazione grafica di forma d'onda e composizione spettrale, è presente la componente tattile, retroazione indispensabile affinché l'esecutore possa suonare adattando i propri gesti, proprio come accade con gli strumenti tradizionali*

suoni naturali prodotti “ad hoc” da musicisti. Quasi tutti i PC odierni dispongono di schede audio dotate di campionatore audio.

I modelli matematici si basano sulla simulazione di una forma d'onda data mediante una formula che la approssimi, generalmente polinomiale. Questi modelli si distinguono per la quantità di termini presenti nella formula e per la semplicità o meno nella genera-

zione di timbriche predeterminate. Da citare almeno due modelli di questa famiglia:

□ la sintesi additiva: è basata sull'idea che una forma d'onda complessa può essere costruita sommando con opportuni “pesi” – eventualmente variabili nel tempo – forma d'onda semplici, tipicamente sinusoidi (Figura 2); permette di costruire in modo accurato una forma d'onda data che, una volta analizzata per identificare le sue componenti in frequenza, può poi essere ricostruita sommando tali componenti; peraltro, richiede molti termini per essere qualitativamente apprezzabile, e da questo deriva la sua scarsa semplicità d'uso – troppi parametri per controllare il suono – e quindi il suo successo commerciale pressoché nullo;

□ la sintesi per modulazione di frequenza o di fase [9]: è basata sull'idea che una forma d'onda semplice può essere molto arricchita spettralmente modulando con termini temporari la sua frequenza o la sua fase (i due metodi sono riconducibili uno all'altro); il fatto che con poche onde sinusoidali si possano costruire timbriche musicali dinamiche e ricche ha determinato il grande successo commerciale dei sintetizzatori basati su questa tecnica; si pensi al sintetizzatore Yamaha DX7; oggi, quasi tutti i PC dispongono di librerie standard di timbriche FM.

I modelli matematici sono stati tra gli anni settanta e novanta il punto di riferimento per i compositori di computer music, ma hanno mostrato significativi limiti rispetto alla variazione delle timbriche in funzione dei gesti interpretativi, proprio per il fatto di partire dalla simulazione di una forma d'onda che è già il risultato del processo di interazione tra chi suona e lo strumento che viene suonato.

Per questo motivo, l'approccio dei modelli fisici ha subito indicato di essere quello su cui puntare per la piena soddisfazione nella generazione e nel controllo delle timbriche musicali [10]. L'idea di fondo è di simulare il comportamento dei corpi vibranti (corde, ponticelli, casse armoniche ecc.) e delle loro interazioni, sia tra di essi stessi che con il musicista. In questo modo, il modello fisico può essere “suonato bene” o “suonato male” (Figura 3), come accade con gli strumenti tradizionali; tutte le soluzioni del liutaio diventano “programmabili”, potendo andare anche oltre ciò che è fisicamente possibile o ragionevole. Si

può - per esempio - programmare un tamburo di 200 km. di diametro o una chitarra con corde di un nanometro di spessore, perfino costruiti con un materiale qualsivoglia.

I motivi per cui questo approccio non ha ancora trovato piena applicazione negli strumenti musicali offerti dal mercato sono individuabili nella grossa richiesta di capacità computazionale richiesta, e ancor più per il fatto che non è ancora matura la letteratura scientifica sui modelli fisici degli strumenti musicali. Per ora, la maggior parte delle applicazioni concrete di questo approccio si è limitata a usare modelli fisici "parziali", che simulano solo una o comunque poche componenti della catena di generazione del suono; come - per esempio - gli amplificatori digitali a modelli fisici che simulano i diversi modelli di amplificatori valvolari, o le unità di effetti che applicate ai suoni degli strumenti tradizionali ne alterano le sonorità simulando banchi di effetti analogici, o i vari tipi di chitarre (acustiche, elettriche, basse) che producono le sonorità di decine di altre chitarre di cui lo strumento digitale implementa i modelli fisici per trasformare il suono della chitarra "fisica" nel suono di una delle chitarre "virtuali" di libreria.

### 3.2. Evoluzione delle architetture hardware

Gli avanzamenti nella modellistica e nelle tecniche per il trattamento dell'informazione musicale, specialmente considerando i livelli esecutivo e sonoro, sono andati di pari passo con la disponibilità di nuove tecnologie hardware. Al tempo dei mainframe si era arrivati a considerare come prima ipotesi operativa quella di usare il computer a livello simbolico e strutturale per controllare mediante parametri una serie di dispositivi analogici esterni per la sintesi e il trattamento del suono (oscillatori, sintetizzatori, mixer, banchi di filtri ecc.): sono i cosiddetti *sistemi ibridi*, dove le quantità di informazione più modeste (strutturale e simbolico) sono gestite digitalmente e quelle invece più considerevoli (modelli timbrici e segnali audio) sono controllate da computer ma gestite da periferiche analogiche.

Il motivo di questa soluzione è che non sussisteva la disponibilità di una potenza computazionale che permettesse di trattare il segnale audio senza interruzioni o ritardi. In questa fase i compositori di computer music potevano

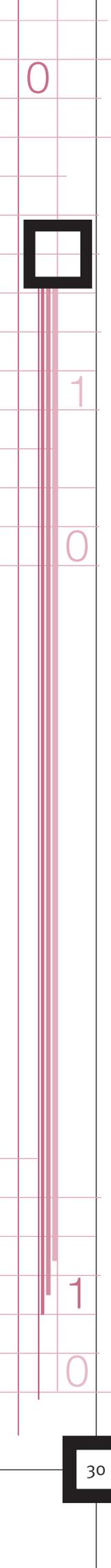
solo usare linguaggi operanti in tempo differito (tipo i linguaggi MUSIC-n); una volta date tutte le direttive al computer su come produrre i suoni corrispondenti a una composizione, veniva calcolato il segnale audio finale, in un tempo non predeterminabile e senza interazione con il compositore.

Negli anni settanta furono sviluppati i primi sintetizzatori digitali prototipali in alcuni centri di ricerca statunitensi (Stanford, Cambridge, Berkeley, Urbana), ed è nel 1978 che la Digital Music Systems, una piccola ditta spin-off del MIT, produce il primo sintetizzatore digitale - denominato DMX-1000 - programmabile da un qualunque minicomputer o mainframe [11]. Diventa realtà il controllo interamente digitale dell'informazione musicale ai vari livelli: un *computer general purpose* si occupa dell'informazione strutturale, simbolica e dei parametri che - inviati al DMX-1000 - permettono la generazione in tempo reale dei suoni. Si tratta di un'*architettura digitale mista*: tecnologia *general purpose* per controllare la tecnologia più spinta necessaria per il trattamento digitale dei segnali audio.

Solo tre anni dopo, il fisico Giuseppe Di Giugno progetta e realizza all'IRCAM di Parigi un sintetizzatore digitale - denominato 4X - capace di gestire una grande quantità di strumenti virtuali contemporaneamente e di sopportare un grande carico computazionale [12]. I modelli timbrici implementabili diventano così molto complessi, paragonabili a quelli richiesti per il controllo di un'orchestra digitale intera. Un grosso passo avanti, seguendo il tipo di architettura digitale mista introdotta con il DMX-1000.

Un cambiamento di rotta fu introdotto nel 1983 con la pubblicazione dello standard MIDI [13] e la comparsa degli strumenti musicali digitali che ne sfruttavano i modi di comunicazione: l'informatica musicale diventa repentinamente "individuale", non servono più i grandi centri di ricerca per avere a disposizione l'hardware necessario a produrre il suono digitale. Ai centri di ricerca resta da sviluppare nuove soluzioni più avanzate e da far progredire modelli e tecniche per poi tradurli in applicazioni software. L'accoppiata PC+MIDI si afferma decisamente sul mercato. Questa diffusione improvvisa porta però con sé tutti i limiti dello standard MIDI: tra-

0



missione seriale, velocità di trasmissione bassa (poco più di 30.000 bit/s), pochi canali (16 al massimo), scarsa programmabilità dei modelli timbrici.

1

Infatti è l'hardware dei sintetizzatori MIDI a determinare intrinsecamente i modelli eseguibili: Yamaha con la modulazione di frequenza, Casio con la distorsione non lineare, AKAI con i campionatori, Roland con soluzioni miste. Peraltro i ricercatori capaci di costruire modelli timbrici erano e restavano pochi; la grande utenza si è sempre accontentata dei suoni di libreria forniti dalle case costruttrici, oltre a copiare le librerie di set timbrici degli amici.

0

Quindi diffusione sì, ma progresso scientifico ben poco grazie al MIDI. Si sviluppa moltissimo la possibilità di registrare i gesti dell'interprete su dispositivi che hanno l'aspetto di strumenti musicali ma che servono a produrre dati MIDI; dati che vengono registrati con applicazioni software chiamate *sequencer*, programmi che sono registratori multitraccia di dati MIDI: si esegue e si registra, poi si può intervenire per correggere, modificare, aggiungere, togliere ovvero fare editing dei dati MIDI organizzandoli in molte tracce, anche centinaia. Si può sempre scegliere in un secondo momento quali di queste tracce usare e come combinarle tra loro; un po' come quando dalla registrazione multitraccia audio analogica si fa il mixdown al master stereo. Tant'è che nel giro di qualche anno i sequencer MIDI sono stati arricchiti delle tracce audio digitali e la registrazione multitraccia combina tracce MIDI e tracce audio, in certi casi anche tracce video.

1

Ma quello che inizialmente era arricchire il PC della possibilità di produrre suoni digitalmente con periferiche MIDI esterne, diviene poi gradualmente una soluzione unica e integrata, per cui nel PC – dopo un primo esperimento con gli Atari – viene fornita di serie l'interfaccia MIDI, librerie di campioni sonori, librerie di timbri FM, software musicali di *editing* e *sequencing* musicale, e perfino processori *special purpose* per l'audio, poi confluiti nelle moderne motherboard con una o più CPU e diversi coprocessori integrati. Oggi, CPU standard, coprocessori per l'audio, per il video, per la comunicazione interna ed esterna, MIDI e quant'altro, fanno parte di un'unica soluzione integrata. Tanto che i PC di oggi sono in grado di eseguire applicazioni software musicali che

in tempo reale eseguono tutti i tipi di tecniche per la sintesi digitale del suono; perfino i tanto dispendiosi modelli fisici, purché con opportune limitazioni, ma senza bisogno di hardware aggiuntivo. È forse il controller musicale (tipo tastiera, chitarra, strumento a fiato, percussioni o anche microfono) la parte specifica più comune che serve ancora aggiungere.

Infine, va menzionato il fenomeno delle workstation musicali: iniziato alla fine degli anni settanta con il NEDCO Synclavier e poi il Fairlight CMI, consiste nella produzione di workstation in cui sono integrate:

- tecnologie *general purpose* di tipo PC;
- tecnologie *special purpose* tipo *Digital Signal Processor* e MIDI;
- applicazioni software per l'edizione di partitura;
- applicazioni software per il *sequencing* e l'*editing* MIDI e in un secondo momento anche audio e video;
- librerie di suoni digitali di qualità professionale.

Questa tipologia di strumento è molto costosa, destinata generalmente a professionisti che vogliono trovare i problemi tecnici già risolti in una postazione fornita "chiavi in mano" per il musicista. È sempre stata da allora una nicchia ad alto valore aggiunto, come spesso accade con gli strumenti professionali. La loro diffusione è stata comunque limitata non solo dai costi alti, ma anche dal fatto che le stesse funzionalità di queste workstation integrate potevano essere implementate integrando le diverse componenti in modo anche più flessibile ed economico a patto di affrontare le problematiche tecniche del caso.

### 3.3. Il testo musicale: editoria e riconoscimento automatico

La scrittura di testi in linguaggio naturale (*text editor* e *word processor*) è stata senz'altro una delle prime e più diffuse applicazioni del computer. Il riconoscimento automatico dei testi stampati (OCR: *Optical Character Recognition*) è meno popolare, ma – nonostante l'intrinseca complessità – ha raggiunto da molti anni piena maturità, in altre parole è quasi infallibile.

Applicazioni di grande interesse per la minoranza di utenti del computer che si occupa di musica è l'equivalente capacità di scrivere spartiti (*score editor*) e di riconoscere auto-

maticamente i simboli musicali dei testi stampati (OMR: *Optical Music Recognition*). A metà degli anni ottanta sono comparsi i primi software per scrivere spartiti al computer; ben presto sia gli editori che i singoli musicisti hanno apprezzato questa tipologia di applicazioni al punto da considerarla come lo strumento prevalente per scrivere e stampare musica. Già agli inizi degli anni novanta la tecnologia (font, editor, formatter, publisher) era matura e il mercato soddisfatto. Anche i musicisti più ostili all'uso del computer se ne sono fatti una ragione, per l'affidabilità, la convenienza economica, l'autonomia acquisibile con il software per l'edizione musicale. Per quanto riguarda la lettura ottica della musica, nonostante siano comparsi negli stessi anni del software di edizione i primi applicativi, l'affidabilità nel riconoscimento automatico non è ancora così soddisfacente. Soprattutto, la percentuale di errori è fortemente dipendente dalle caratteristiche del singolo testo musicale considerato – per esempio, la densità di simboli musicali, la quantità di casi di sovrapposizione di simboli diversi nella medesima posizione - e in generale la complessità grafica del testo musicale, in combinazione con gli investimenti limitati tipici delle problematiche di nicchia, fan sì che si possa oscillare tra un 10% e un 50% di simboli riconosciuti in modo scorretto. Nei casi peggiori, è ancora più conveniente inserire il testo musicale nel computer a mano, ovvero con un music editor.

La maturazione delle tecnologie di OMR è attesa con grande interesse, per poter rendere disponibile una grande quantità di archivi musicali online, non più solo in forma grafica, ma anche in forma simbolica, editabile, disponibile per rielaborazioni, trasposizioni, estrazione delle parti strumentali, etc. Soprattutto per l'enorme patrimonio culturale di testi musicali che non sono più soggetti alla normativa sul diritto d'autore.

#### **4. INFORMATICA MUSICALE OGGI**

Nell'arco della sua ancora giovane vita, l'informatica musicale ha visto la realizzazione di numerosi progetti di ricerca i cui risultati di maggior interesse sono poi stati trasferiti al-

l'industria e all'editoria. Sono abbastanza delineate le linee evolutive che sono in corso in questi ultimi anni ed è altresì identificabile l'insieme dei problemi rimasti ancora aperti. Peraltro, non tutte le innovazioni fornite dalla ricerca sono state sfruttate a livello produttivo, generalmente per motivi economici.

Per concludere questa escursione nelle tematiche dell'informatica musicale ecco quindi lo stato dell'arte odierno e quello che è oggettivamente prevedibile nel medio periodo. Si ricordi che da almeno vent'anni un PC arricchito con opportuni dispositivi specializzati costituisce la base di quasi tutte le tipologie di stazioni di lavoro individuali, editoriali e in genere produttive. In questo contesto devono essere considerati alcuni aspetti particolarmente significativi per l'informatica musicale: l'interazione con l'informazione musicale, il riconoscimento automatico dei contenuti, internet, i dispositivi mobili. L'interazione con i contenuti multimediali è al centro dell'evoluzione informatica di questi anni, ma: cosa vuol dire interagire con l'informazione musicale?

Si consideri che l'informazione musicale può essere espressa ad uno dei livelli di rappresentazione sopra indicati in accordo con la forma più appropriata per l'interazione. Per esempio, se si vuole fornire all'utente un'applicazione capace di estrarre automaticamente le parti strumentali da una partitura orchestrale risulterà opportuno trattare l'informazione musicale mediante la forma grafica di partitura; oppure, se si vuole permettere all'utente di una libreria digitale multimediale di ascoltare un brano di musica di cui l'utente non conosce né titolo, né autore, né altri metadati, ma l'utente è in grado di canticchiarlo poiché lo ha già sentito in un'altra circostanza, allora bisognerà predisporre un'interfaccia per interagire tramite il canto dell'utente, che naturalmente funzionerà in modo efficace solo se l'utente è capace di intonare correttamente il frammento musicale che cerca. Detto questo, è necessario un piccolo approfondimento per capire cosa implica quanto ipotizzato.

Sia per il trattamento ad un certo livello, che per la trasformazione da un livello ad un altro dell'informazione musicale è necessaria la disponibilità di un insieme integrato di dispositivi software e hardware specifici capaci

di trattare (memorizzare, elaborare, trasmettere, rappresentare) i flussi informativi corrispondenti.

Per esempio, per il trattamento dell'informazione musicale a livello di suono ad alta fedeltà sono generalmente richiesti dispositivi hardware speciali per l'acquisizione e la produzione di suoni, insieme con memorie centrali di grande dimensione e memorie di massa molto capaci e con velocità di trasferimento dati molto alta. Ricordiamo infatti che un secondo di suono stereofonico di qualità Compact-Disc audio contiene una quantità di informazione pari a circa 90 pagine di testo in linguaggio naturale.

#### 4.1. Il riconoscimento dei contenuti musicali

L'analisi dell'informazione musicale può portare al riconoscimento automatico di contenuti non espressi in modo esplicito: per esempio, a livello di simboli, si può riconoscere la tonalità in cui è scritta una partitura, quali sono i frammenti tematici più importanti, qual è la forma musicale, quali sono le regole rispettate e quali quelle infrante della sintassi musicale; oppure, a livello sonoro, si può riconoscere quali strumenti stanno suonando, dove inizia e dove finisce la parte di suono che corrisponde ad un simbolo di nota, una serie di parametri fisici che permettono di avere informazioni, melodiche, ritmiche, dinamiche e anche relative all'interpretazione operata da chi ha prodotto i suoni allo strumento.

Questo tipo di analisi è caratterizzata da una mole di calcoli molto elevata, da modelli analitici complessi e non è ancora giunta a piena maturazione, ovvero ci sono ancora informazioni che non si riesce a riconoscere in modo automatico.

Già nel 1975, nella sua tesi di dottorato, Moorer [14] impostò il problema dalla base, ovvero come porsi ad osservare il segnale audio per estrarne informazioni musicali. Da allora molti progressi sono stati compiuti, e tanto i modelli di interazione uomo-macchina più avanzati quanto i metodi di classificazione automatica dei contenuti musicali nelle librerie digitali accessibili tramite il web [15] si basano proprio sulla capacità dei computer odierni di riconoscere contenuti multimediali e, in particolare, musicali.

#### 4.2. Interazione multimodale con i contenuti musicali

Già nel 1985, Roger Dannenberg aveva realizzato un prototipo di sistema capace di accompagnare automaticamente un musicista durante una performance [16]; il suo sistema conosceva tutte le parti strumentali del brano da suonare ed era in grado di riconoscere le note suonate dal musicista ed eseguire in automatico tutte le altre parti strumentali, in modo sincrono all'umano. Prototipi in questa direzione sono stati via via sviluppati da vari ricercatori ma ben poco di queste esperienze è arrivato a essere disponibile per il musicista o l'appassionato comune.

Si è così approdati alla situazione odierna, dove la possibilità più innovativa offerta dalle tecnologie informatiche per la musica combina le possibilità di riconoscimento automatico dei contenuti musicali con i formati per la rappresentazione integrata dei contenuti informativi delle diverse tipologie di materiali musicali: supporti audio, supporti video con tracce audio, pagine di musica scritta, altri materiali "fisici" (foto, video, scenografie ecc.), supporti magnetici e ottici contenenti codifiche simboliche immateriali (XML, MIDI, UNICODE, ASCII ecc.).

Infatti, l'informazione musicale può essere espressa – e di solito è prevalentemente questo che accade nel caso degli standard più diffusi come MP3, AAC, vari tipi di MPEG - ad uno specifico dei livelli di rappresentazione considerati sopra, in accordo con la forma di rappresentazione più appropriata per l'interazione. Per esempio, se si vuole fornire all'utente un'applicazione capace di estrarre automaticamente le parti strumentali da una partitura orchestrale risulterà opportuno trattare l'informazione musicale mediante la forma grafica di partitura.

Ma oggi è diventato possibile integrare i diversi livelli in un unico ambito di rappresentazione, integrando audio, timbriche, partiture, forme e strutture. Un gruppo di lavoro della IEEE Standards Association per la musica ha definito per questo uno specifico dialetto XML per la musica, denominato ufficialmente IEEE1599-2008, anche noto comunemente come "MX"; IEEE1599-2008 è uno standard internazionale approvato il 12 giugno 2008 [17], ottenuto grazie alla collaborazione tra LIM dell'Università

degli Studi di Milano (gruppo del Prof. Haus) e SUPSI di Lugano (gruppo del Prof. Baggi), con il supporto parziale del progetto CTI 8357.1 "e-Music XML Standard" e nell'ambito del progetto mondiale IMS "e-Music XML".

Il formato IEEE1599-2008 consente di navigare e interagire tra/con i contenuti musicali multimediali, come se questi materiali fossero parte di un unico ambiente di conoscenza integrata. In altre parole, tale tecnologia innovativa consente di:

- trovare tutti i file di partiture musicali e tutti i file audio digitali che contengano un dato frammento musicale, interrogando un sito Internet, un database multimediale ecc.; il frammento può essere uno spezzone audio, vocale (cantato) o strumentale (suonato), oppure una sequenza di note scritta su pentagramma;

- sincronizzare automaticamente varietà ampie a piacere di file audio, video e di partitura per una fruizione integrata delle varie forme di comunicazione musicale (CD, DVD, file su PC ecc.); ognuno dei materiali considerati per la sincronizzazione permette di "guidare" la navigazione nei vari materiali integrando di fatto i diversi player in un unico player multitraccia dove ogni traccia può essere usata per "navigare";

- interagire con l'informazione musicale, aggiungendo ai controlli standard sul volume, sulla posizione stereo o sulle frequenze un ampio repertorio di possibilità che contemplano: scorrere un brano nota per nota, battuta per battuta o frase per frase, modificare la partitura e di conseguenza i suoni corrispondenti ecc.:

Gli strumenti scientifico-tecnologici per produrre i materiali nel formato previsto dallo standard sono prevalentemente italiani: sono stati sviluppati nell'arco di trent'anni presso il LIM - Laboratorio di Informatica Musicale del Dipartimento di Informatica e Comunicazione dell'Università degli Studi di Milano ed sono stati applicati a livello prototipale con successo:

- nel sistema MAIS (*Music Archive Information System*) sviluppato dal LIM tra il 1997 e il 1999 per l'Archivio Musicale del Teatro alla Scala [18];

- nei dimostrativi tecnologici sviluppati per il Ministero per i Beni e le Attività Culturali in occasione delle mostre su tematiche musicali te-

nute a Roma (dicembre 2005-gennaio 2006), Torino (febbraio-giugno 2006), Milano (dicembre 2006-gennaio 2007), Salisburgo (maggio-giugno 2007), Lucca (novembre 2007);

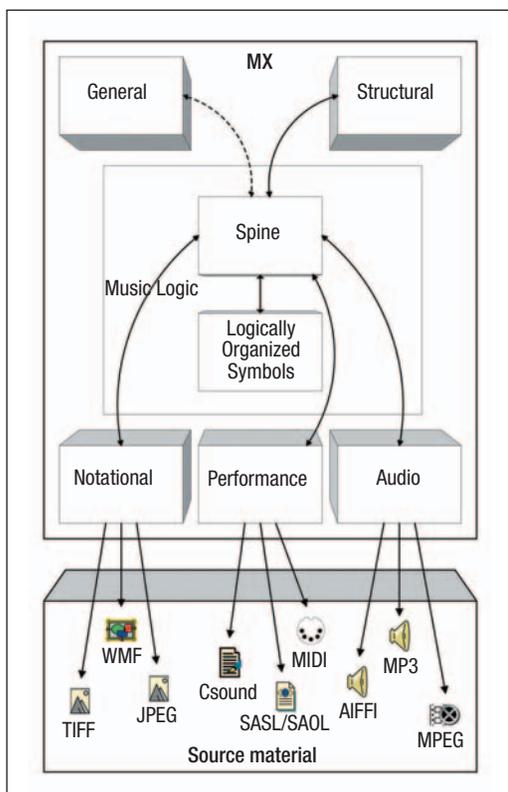
- nei dimostrativi tecnologici sviluppati nell'ambito dei succitati progetti CTI e IMS congiuntamente alla SUPSI di Lugano.

IEEE1599-2008 è una codifica XML multistrato che integra le codifiche standard delle partiture (file Finale, MusicXML, NIFF, e altri formati), dei metadati catalogafici, dell'audio compresso (MP3, AAC, WMA ecc.) e non compresso (Compact-Disc Digital Audio, WAV, DVD Audio, etc.), delle performance (Csound, MIDI, SASL/SAOL ecc.), e dei dati di sincronizzazione. Nella figura 4 si può vedere il principio che consente di trattare in modo integrato e sincrono i diversi materiali musicali considerati (nella figura di esempio si vedono audio, MIDI e partitura): una struttura di dati chiamata *spine* integra e sincronizza l'informazione spaziale delle pagine di musica scritta con l'informazione temporizzata dei processi MIDI e audio. Nella figura 5 si può vedere l'architettura



**FIGURA 4**

Esempio di schema di corrispondenze tra la posizione spaziale dei simboli di uno spartito con gli istanti di tempo dei suoni correlati e dei gesti necessari per produrre tali suoni: è la struttura portante, chiamata "spine" del nuovo standard musicale multilayer IEEE1599-2008



**FIGURA 5**

*Schema delle relazioni tra i layer del nuovo standard musicale IEEE1599-2008: i file di notazione musicale, performance e audio con eventuale video sono tra loro integrati e correlati pur restando nelle forme di codifica degli standard propri delle varie tipologie di dati multimediali*

tura che mette in relazione i diversi “strati” di informazione musicale, ciascuno in grado di recepire gli standard industriali di riferimento, nello standard IEEE1599-2008.

Molti elementi di innovazione vengono introdotti da IEEE1599-2008 nei modi di interagire tra umani e informazione musicale; tra i più evidenti e significativi, si rendono potenzialmente disponibili i seguenti:

- a.** fruizione sincrona di tutti i materiali musicali inerenti ad un brano (una o più partiture o parti strumentali, una o più interpretazioni audio ed eventualmente video, il multitraccia audio con uno o più *mixdown* disponibili, iconografia, elementi testuali);
- b.** interazione con i materiali musicali di un brano (definizione del *subset* di materiali per la fruizione, eliminazione di una o più “tracce audio” dalla fruizione);
- c.** navigazione multimodale dei contenuti: uso

di un qualunque materiale musicale per guidare la fruizione di tutti gli altri materiali sia nel testo (parole di una canzone e/o libretto d’opera), che nello spazio (materiali di partitura), che nel tempo (materiali audio e video);

**d.** “allineamento” di tracce audio non sincrone, sia rispetto ad altre tracce audio sia con i materiali di partitura;

**e.** dilatazione/contrazione di tracce audio senza modifiche timbriche e di intonazione (cambi di metronomo, sovrapposizione di tracce non sincrone);

**f.** indicizzazione dei materiali musicali (partiture, audio, video, testo letterario, iconografia) per contenuti (frammenti audio, frammenti di partitura, frammenti di testo).

Le ipotesi di nuovi media come quelle qui sopra formulate portano come effetto collaterale una potenziale forte valorizzazione dei patrimoni editoriali; i materiali degli archivi possono essere riutilizzati per le nuove iniziative editoriali con un’articolata varietà di modi e di potenziali ritorni economici.

Vale la pena sottolineare che un auspicabile effetto di valorizzazione dei patrimoni editoriali può essere ottenuto – prima ancora che pensando a nuovi media – presentando, elaborando e riorganizzando i materiali musicali degli archivi esistenti. Alcuni tipi di media, anche se già presenti sul mercato, possono essere concepiti in modo innovativo rispetto ai target del mercato di consumo, specie quando si considerino quelli professionale e formativo.

A titolo di esempio, si considerino le schermate nelle figure 6 e 7. Rappresentano casi esemplari del rappresentare e rendere fruibili i contenuti di brani musicali classici, con modalità multimodali e interattive.

Le applicazioni software – realizzate al LIM – da cui sono state prese le schermate consentono di interagire con i contenuti musicali dei rispettivi brani musicali considerati, in diversi “modi”: partitura, strumentazione, libretto, audio e video [19]. Ogni elemento informativo presente sullo schermo è attivo. La selezione con il mouse di una parola del libretto produce la riproduzione in audio o in video/audio a partire dall’istante in cui quella parola viene pronunciata, nonché il posizionamento sulla partitura di un cursore rosato che evidenzia la nota corrispondente al momento in cui la parola viene pro-

nunciata, nella parte che si è scelto di seguire per la fruizione.

Analogamente selezionare una nota sulla partitura con il mouse, posiziona automaticamente audio, video, libretto e quant'altro alla nota selezionata così come spostare il cursore del player multimediale sposta contemporaneamente e in modo sincrono i cursori sulla partitura, sul libretto e quant'altro.

In particolare nella figura 7, è mostrato un momento dell'interazione con le frasi melodiche del brano (un duetto tratto dalla Tosca di Puccini); i contenuti musicali sono percorsi mediante le schematizzazioni di frasi che possono essere selezionate nell'apposito riquadro in alto a destra nello schermo. La selezione di una frase posiziona il player multimediale, il cursore sulla partitura e il cursore sul libretto e avvia l'esecuzione della frase stessa evidenziando sulla partitura in colore giallo il segmento trasversale di partitura interessato.

La frontiera più avanzata dell'interazione consente di modificare i contenuti musicali dei brani, a partire dagli elementi informativi individuati e rappresentati. Ma questo discorso è per un futuro meno prossimo. Questa modalità di interagire con la struttura di un brano in tempo reale – l'audio e i frammenti di partitura vengono costruiti e riprodotti durante l'interazione senza ritardi – costituisce una soluzione per quando sarà disponibile una normativa sul diritto d'autore che disciplini questa forma di fruizione, oppure anche già subito, ma solo per le composizioni musicali che sono già uscite dal periodo disciplinato dalle attuali norme sulla proprietà intellettuale e industriale, o anche come strumento di produttività individuale. In questo caso la tecnologia è ben più avanti del diritto.

#### 4.4. La formazione

Accanto a nuove possibilità, nuove problematiche, nuove economie e nuove modalità di accesso alla cultura musicale, si delinea sempre più la necessità di figure professionali interdisciplinari che abbiano competenze specifiche sulle tecnologie di interesse per la musica accompagnate da competenze propriamente musicali. Se fino agli anni novanta erano sufficienti pochi professionisti e ricercatori capaci di coniugare in virtù dei loro interessi e delle loro esperienze la doppia

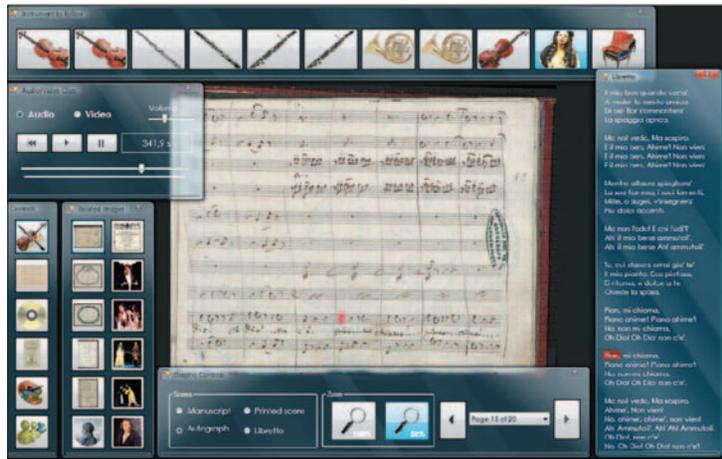


FIGURA 6

*Esempio di schermata di musica interattiva; l'autografo di un'opera di Paisiello, il libretto, due performance (una in audio e l'altra in audio e video) sono fruibili in modo sincrono, tutti gli elementi informativi possono essere usati per muoversi "nel brano", ogni nota della partitura, ogni parola del libretto, ogni istante di tempo della barra del player*

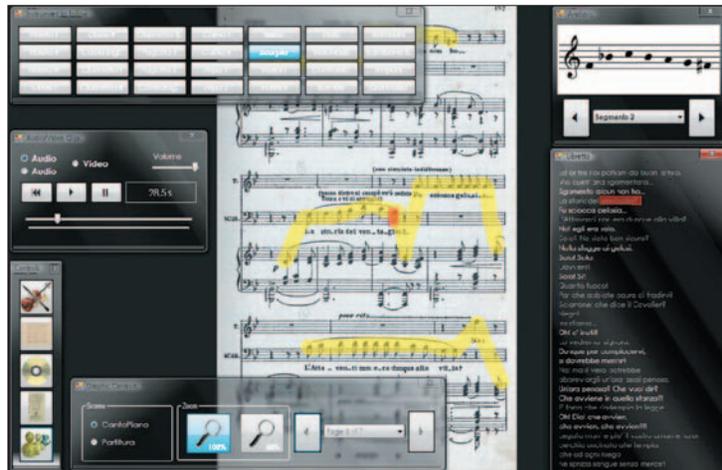


FIGURA 7

*Esempio di schermata di musica interattiva: la riduzione canto e piano di un duetto della Tosca di Puccini, il libretto, tre diverse performance (due in audio e la terza in audio e in video) sono fruibili in modo sincrono e sono percorribili oltre che "per note", "per parole" e per "istanti", anche per "frammenti melodici" grazie alla finestra in alto a destra che evidenzia il materiale tematico alla base del brano, consentendo di visualizzare con evidenziazione gialla come il singolo frammento è "mappato" nella partitura del brano ogni volta che il frammento viene presentato*

figura dell'informatico e del musicista, è ora necessario avere persone formate ad hoc e in quantità superiore. A questa esigenza cercano di rispondere alcune iniziative formative, sia a livello universitario che medio superiore. Da citare il corso di laurea triennale in "Scienze e tecnologie della comunicazione

musicale” attivo all’Università degli Studi di Milano dal 2002, che accetta 120 nuovi allievi ogni anno e riesce portarli al mondo del lavoro proprio in relazione alle specificità professionali acquisite in questo corso di studi.

## 5. PROBLEMI APERTI E PROSPETTIVE

A conclusione di questo percorso, è doveroso citare:

- cosa ancora non è possibile;
- verso dove si sta andando.

La ricerca di base deve ancora risolvere molti problemi, prima di passare le ipotesi di soluzione alla ricerca sperimentale e poi all’industria. I più rilevanti:

- maturazione delle tecniche basate su modelli fisici degli strumenti musicali e riconoscimento automatico delle relative timbriche;
- definizione di modelli di compressione multisensoriali, che sappiano sfruttare il modo in cui il nostro sistema fisiologico (cervello + sistema nervoso + organi di senso) funziona, al fine di ridurre la necessità di memoria per rappresentare l’informazione multimediale e musicale; in altre parole, integrazione degli attuali modelli percettivi monosensoriali (audio e video soprattutto) in modelli percettivi multisensoriali;
- definizione di modelli dell’interpretazione musicale capaci di eseguire automaticamente una partitura come un umano; alcune esperienze significative già sono state effettuate, ma non sono ancora pienamente soddisfacenti [20]; andando nel senso contrario, definizione di modelli interpretativi inversi, ovvero capaci di riconoscere gli interpreti e di trascrivere in partitura - fedele all’originale - le registrazioni audio;
- metodi affidabili per effettuare la separazione automatica delle tracce audio ovvero il “multitraccia” con l’audio separato per ogni voce o strumento a partire da originali mono o stereo. Alcune di queste problematiche sono più vicine a trovare soluzioni soddisfacenti, altre sono più lontane, e riguardano perciò il “domodani” dell’informatica musicale. Si può invece considerare che il domani dell’informatica musicale debba senz’altro prevedere che le prospettive di ulteriori avanzamenti, sia dal punto di vista degli arricchimenti del-

le modalità di interazione uomo-macchina che da quello della produttività e dell’economicità nella produzione di materiali musicali interattivi, passino attraverso la soluzione - o più verosimilmente la riduzione - di una serie di problemi aperti, verso cui orientare sforzi e risorse:

- incrementi di efficienza ed affidabilità nei procedimenti di riconoscimento automatico di caratteristiche dall’informazione multimediale;
- individuazione di metodiche efficaci per la separazione automatica di particolari “sorgenti di informazione”; per esempio: tracce audio di parlatori, cantanti, strumenti, personaggi statici o in movimento da immagini o video;
- ulteriori incrementi della larghezza di banda nelle telecomunicazioni e nella capacità di memoria dei supporti.

Sintetizzando al massimo, grazie al contributo dei progressi delle tecnologie informatiche specifiche, ci si può aspettare nel breve-medio periodo una sempre maggior tendenza alla personalizzazione nella fruizione musicale, sia a livello amatoriale che professionale, fino alla disponibilità globale del patrimonio di conoscenze musicali fin qui maturate, archiviate, pubblicate, suonate, scritte, discusse.

## Bibliografia

- [1] Wirth N.: *Algorithms + Data Structures = Programs*. Prentice Hall PTR, Upper Saddle River, NJ, 1978.
- [2] Haydn J.: *Musikalisches Würfelspiel*. Manoscritto originale, 1783.
- [3] Mozart W. A.: *Musikalisches Würfelspiel*. K 516f, 1787.
- [4] Hiller L. A., Isaacson L. M.: *Experimental Music: Composition With an Electronic Computer*. McGraw-Hill, New York 1959.
- [5] Mathews M.V.: *The technology of computer music*. MIT Press, Cambridge, MA, 1969.
- [6] Mathews M.V., Moore R.: *Groove—A Program to Compose, Store, and Edit Functions of Time*. Communications of the ACM Vol. 12, 1970, p. 715.
- [7] Roads C.: *The Computer Music Tutorial*. MIT Press, Cambridge, MA, 1996.
- [8] Roads C., Strawn J., eds.: *Foundations of Computer Music*. MIT Press, Cambridge, MA, 1985.
- [9] Chowning J.: The Synthesis of Complex Audio Spectra by Means of Frequency Modulation. *Journal of the Audio Engineering Society*, Vol. 21, n. 7, 1973, p. 526-534.

- [10] Computer Music Journal: Two Special Issues on Physical Modeling. *MIT Press*, Vol. 16, n. 4 & Vol. 17, n. 1, Winter 92, Spring 93.
- [11] Wallraff D.: *The DMX-1000 Signal Processing Computer*. Proceedings of the 1978 International Computer Music Conference. San Francisco: International Computer Music Association, 1978.
- [12] Di Giugno G., Kott J., Gerzso A.: *Progress Report on the 4X Machine and Its Use*. Proceedings of the 1981 International Computer Music Conference. San Francisco: International Computer Music Association, 1981.
- [13] International MIDI Association: *MIDI Musical Instrument Digital Interface Specification 1.0*. North Hollywood: International MIDI Association, 1983.
- [14] Moorer J. A.: *On the Segmentation and Analysis of Continuous Musical Sound by Digital Computer*. Ph.D. thesis, Dept. of Computer Science, Stanford University, 1975.
- [15] Wold E., Blum T., Keislar D., Wheaton J.: Content-based classification, search and retrieval of audio. *IEEE Multimedia*, Vol. 3, 1996, p. 27-36.
- [16] Dannenberg R.: *An On-line Algorithm for Real-Time Accompaniment*. Proceedings of the 1984 International Computer Music Conference. San Francisco, CA: Computer Music Association, 1984.
- [17] AA.VV. IEEE Std: *1599-2008: Recommended Practice for Defining a Commonly Acceptable Musical Application Using XML*. IEEE Computer Society Press, New York, 2008.
- [18] Haus G.: Rescuing La Scala's Audio Archives. *IEEE Computer*, Vol. 31, n. 3, IEEE CS Press, 1988, p. 88-89.
- [19] Baggi D., Baratè A., Haus G., Ludovico L.: *NINA – Navigating and Interacting with Notation and Audio*. Proceedings of the 2-nd International Workshop on Semantic Media Adaptation and Personalization, London, IEEE Computer Society Press, 2007.
- [20] Sundberg J., Askenfelt A., Frydén L.: Musical performance: A synthesis-by-rule approach. *Computer Music Journal*, Vol. 7, 1983, p. 37-43.

GOFFREDO HAUS è professore ordinario di Informatica all'Università degli Studi di Milano, dove è anche direttore di dipartimento e consigliere di amministrazione. Ha svolto ricerche di informatica applicata alla musica, all'audio e al multimedia, producendo più di 250 pubblicazioni. Goffredo Haus ha diretto e coordinato numerosi progetti, gruppi di lavoro, congressi. Tra questi, da citare i progetti per la IEEE Computer Society, il Teatro alla Scala, il Teatro Bolshoi, il Ministero per i Beni e le Attività Culturali e il conseguimento del nuovo standard musicale IEEE1599-2008.  
E-Mail: goffredo.haus@unimi.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.



## Abilis: tecnologia ICT di eccellenza made in Italy

Paolo Giacomazzi

### 1. L'AZIENDA ABILIS, CARATTERISTICHE SALIENTI

**A** bilis [1] è una piccola impresa altamente tecnologica con sede principale in Mantova e sedi distaccate di ricerca e sviluppo in Sicilia, Germania, Russia, Repubblica Ceca e Bulgaria. In più di 15 anni di ricerca industriale e sviluppo Abilis ha prodotto una propria piattaforma *hardware/software* sulla quale sviluppa diverse tipologie di apparati di telecomunicazioni a valore aggiunto che si rivolgono principalmente alla piccola-media impresa. Nelle attività sono coinvolti circa 50 giovani ingegneri, dediti al progetto e alla produzione *hardware/software*.

Nella linea dei prodotti Abilis spicca il sistema Abilis CPX. Questo sistema si presenta a prima vista come un router IP, ma in realtà offre una serie di servizi a valore aggiunto significativa che ne espandono il campo applicativo ben al di là della mera interconnessione di rete. Il sistema Abilis CPX offre le funzionalità complete di un centralino Voice over IP (VoIP) [2], di sistema di gestione integrato della telefonia fissa e mobile aziendale, di server di e-mail/fax, di *firewall*, di analizzatore di traffico. In sostanza,

Abilis è una soluzione “tutto in una scatola” che, in funzione della configurazione, è in grado di espletare una serie variegata di servizi a valore aggiunto per diverse tipologie e dimensioni di utenza.

È interessante notare che tutta la tecnologia Abilis è sviluppata in house e, di conseguenza, Abilis mantiene il controllo totale dei suoi prodotti. Infatti, l'hardware della piattaforma Abilis è basato su *mother board* con CPU Intel x86, customizzate nel corso degli anni con tecnologia standard della IT. Il software ed il firmware sono completamente sviluppati in house. Il risultato è che Abilis, al contrario della maggior parte dei vendor presenti sul territorio nazionale, non si limita a svolgere marketing, vendita e *customer care* su linee di prodotti progettati e sviluppati da case madri extranazionali e, frequentemente, extraeuropee. Il ciclo di feedback tra progetto, produzione e gestione dei rapporti con la clientela è locale e veloce e la possibilità di eseguire direttamente uno sviluppo che recepisce in modo immediato le esigenze del mercato locale. Il tutto in modo completamente indipendente dai “big player” del mercato ICT.

## 2. ALCUNI DATI TECNOLOGICI

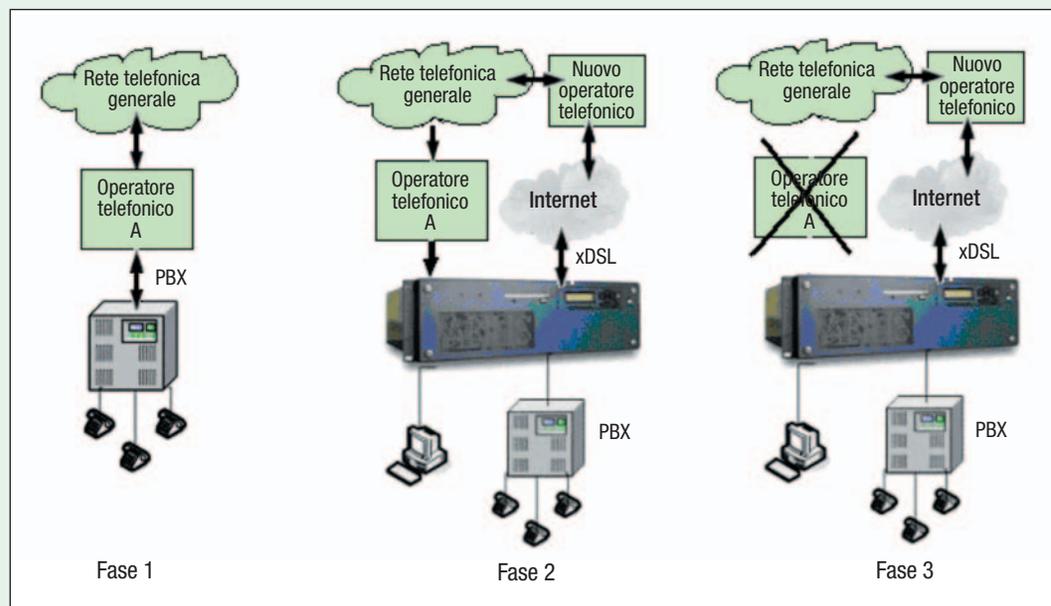
È molto interessante capire come Abilis abbia sviluppato in proprio una tecnologia "Italiana" che l'ha resa indipendente dai grandi costruttori internazionali che controllano il mercato mondiale degli apparati di telecomunicazioni. Il primo passo è stato quello di puntare ad una piattaforma *hardware* di larga diffusione, dotata di ampie possibilità di customizzazione. La scelta è caduta sulle *mother board* PC con CPU di tipo Intel x86. In tal modo, Abilis si è assicurata la possibilità di approvvigionarsi sul mercato della IT di largo consumo di componenti con potenza crescente nel tempo e con costi contenuti. La filosofia che ha guidato le scelte strategiche di base di Abilis è stata di non investire risorse nello sviluppo ex-novo di un *hardware* completamente dedicato: utilizzando come base una piattaforma IT standard si riducono i costi e si liberano risorse per la creazione di funzionalità a valore aggiunto che i dispositivi possono erogare.

La piattaforma customizzata da Abilis monta un sistema operativo *real-time*, RTOS-32 [3]: un sistema operativo *real-time multithreading* per l'esecuzione delle funzionalità con requisiti di performance più stringenti. Oltre al processore principale, è presente un secondo processore (il *watchdog*) dedicato al continuo controllo della corretta operatività del dispositivo. Il *watchdog*, non appena riscontra un'anomalia, interrompe la CPU principale, ne legge lo *stack*, memorizza le informazioni diagnostiche e quindi comanda un reset a freddo del sistema. I tempi tipici di *star-*

*tup* sono di circa 30 s. Circa il 30% del software montato sui dispositivi Abilis è applicativo e il 70% è *firmware* di gestione per il controllo automatico del dispositivo stesso. Le schede di interfaccia montate sulla *mother board* sono progettate e costruite da Abilis. Il software dei sistemi Abilis è sviluppato al 100% in house in C++ e, per le funzioni più critiche, direttamente in assembler. Anche il BIOS è stato customizzato da Abilis. In conclusione, i sistemi della Abilis sono basati su tecnologia standard a basso costo e di grande potenza di calcolo. Questi sistemi sono customizzati, in *house*, per incrementarne la gestibilità e l'affidabilità. In tal modo si è riusciti a pervenire ad una piattaforma *hardware/software* potente, versatile e totalmente sotto il controllo della Abilis.

## 3. TELECOMUNICAZIONI DELLA PICCOLA/MEDIA IMPRESA

I sistemi di Abilis [5] possono essere utilizzati dalla piccola/media impresa per ottimizzare i costi per le telecomunicazioni. In questa sezione si illustra come alcune innovative soluzioni implementate dai prodotti Abilis possono offrire nuove modalità di sfruttamento di connessioni a basso costo (Internet) per realizzare una telefonia aziendale di buon livello, a costi molto ridotti. Prendiamo in considerazione la classica situazione di una sede aziendale dotata di un classico centralino telefonico PBX, mostrata nella figura 1 (fase 1). In questa situazione, i costi telefonici consistono nei canoni mensili delle li-



**FIGURA 1**  
Migrazione della telefonia aziendale classica basata su PBX all'architettura VoIP di Abilis

nee e nel traffico telefonico tipicamente tariffato a tempo. Il primo passo consiste nell'approvvigionamento di numeri telefonici VoIP presso un operatore telefonico. Il secondo passo della migrazione (Figura 1, fase 2) consiste nell'installazione del dispositivo Abilis in azienda. L'apparato viene interposto tra le linee telefoniche già presenti ed il centralino telefonico preesistente, in modo che Abilis-CPX emuli le linee nei confronti del centralino ed emuli il centralino nei confronti del gestore. Le telefonate possono attraversare l'Abilis essendo commutate in modo trasparente (TDM/ISDN) oppure essere codificate, pacchettizzate ed instradate verso il *carrier* VoIP (o gli altri Abilis della propria rete VPN) per mezzo di Internet. In questo modo si possono azzerare i costi relativi al traffico telefonico, ma non quelli relativi ai canoni delle linee, che sono mantenute.

Il passo finale della migrazione (Figura 1, fase 3) consiste nel trasferire (*number portability*) i numeri telefonici al nuovo gestore VoIP, azione che comporta automaticamente alla cancellazione delle linee telefoniche preesistenti. In tal modo, si annullano i costi (traffico e canoni) della telefonia tradizionale e si aggiungono i costi relativi alle linee xDSL di accesso a Internet (che in generale sono molto minori).

Eventualmente, si può predisporre una linea ISDN di *backup* per superare eventuali momenti di disconnessione del collegamento Internet. Nella situazione finale, i numeri telefonici aziendali non sono cambiati e tutto avviene trasparentemente per gli utenti aziendali (non si cambia il centralino e neppure il modo di operare).

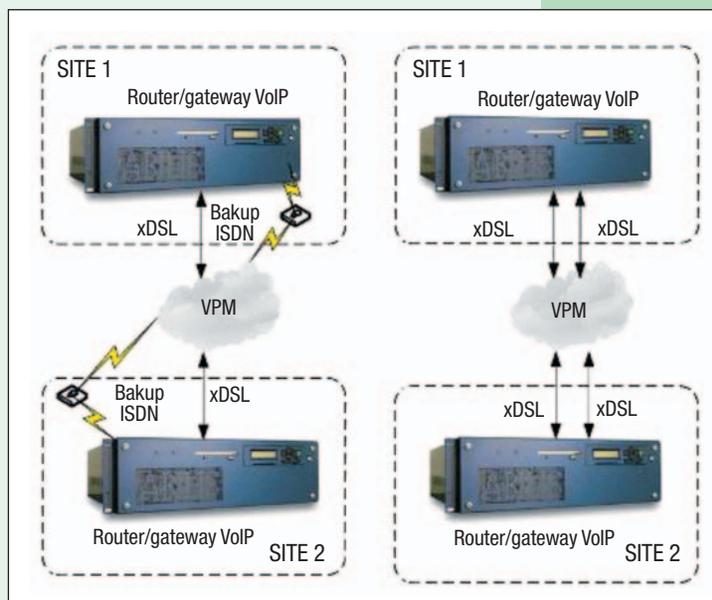
Una soluzione architetturale come quella proposta può suscitare qualche perplessità. È infatti noto che la qualità della telefonia trasportata su Internet non è in genere uguale a quella della telefonia *carrier-grade* [2]. Per fronteggiare questo problema, la piattaforma Abilis implementa una serie di soluzioni specifiche.

Per quanto riguarda l'affidabilità del sistema telefonico, si deve gestire la possibilità che il collegamento ad Internet presenti saltuari malfunzionamenti (per esempio, mancanza di connessione). La soluzione a questo problema consiste nell'adozione di alcune linee ISDN di *backup*, che intervengono nel caso di malfunzionamenti della connessione Internet. Dato che l'adozione di una linea ISDN comporta il pagamento del relativo

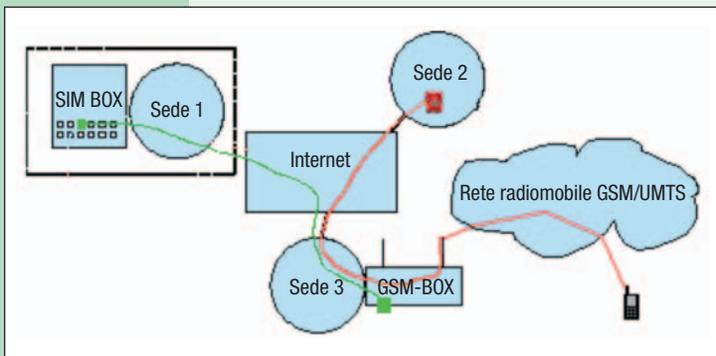
canone mensile, il numero di linee ISDN di backup deve essere minimizzato senza diminuire il traffico telefonico complessivo aziendale. La soluzione di Abilis è così strutturata:

- il sistema rileva automaticamente la disconnessione della linea Internet e commuta istantaneamente sulle linee ISDN di *backup*;
- su queste linee di *backup* interviene un protocollo di compressione proprietario che è in grado di affasciare fino a 8 chiamate telefoniche per ogni canale B-ISDN, cioè, tramite una linea ISDN standard (ISDN *basic rate*) si trasportano fino a 16 chiamate contemporaneamente. In tal modo, basta un numero molto ridotto di linee ISDN per sostenere il traffico telefonico della sede aziendale (Figura 2).

La seconda problematica relativa al trasporto della telefonia su Internet riguarda i ritardi e le perdite dei pacchetti che trasportano il segnale telefonico attraverso Internet. È infatti noto che la modalità di trasporto utilizzata attualmente in Internet è il Best-Effort, che non offre garanzie né sul ritardo dei pacchetti né sulla eventuale perdita dei pacchetti. Le perdite di pacchetti influiscono negativamente sulla qualità del segnale vocale percepito dall'utente e, similmente, pacchetti che arrivano a destinazione con un ritardo eccessivo non sono utili per la ricostruzione del segnale e comportano degradazioni analoghe a quelle introdotte dai pacchetti persi dalla rete.



**FIGURA 2**  
Configurazione per l'affidabilità



**FIGURA 3**

*Esempio di ottimizzazione dei costi per le comunicazioni radiomobili*

Per fronteggiare il problema delle perdite dei pacchetti il sistema Abilis implementa una in una soluzione proprietaria una soluzione già nota [4]: i pacchetti IP che trasportano traffico VoIP sono duplicati e vengono trasmessi sulla stessa linea xDSL o su un'altra linea, se disponibile. In questo modo si diminuisce notevolmente la probabilità che un segmento di conversazione sia perso. Affinché ciò accada, entrambi i pacchetti che trasportano lo stesso segmento vocale devono essere persi. Questa tecnica, come noto, è efficace per diminuire la degradazione dovuta alle perdite dei pacchetti.

La stessa tecnica può essere utile anche per diminuire il ritardo dei pacchetti. Il ritardo dei pacchetti nella Internet Best-Effort è dovuto ad eventuali link/nodi congestionati che i pacchetti attraversano nella rete del provider nel loro viaggio verso la destinazione. Attualmente, gli ISP adottano spesso tecniche di routing dei pacchetti basate sul *load balancing*, che distribuiscono i pacchetti su percorsi alternativi scelti casualmente. In questo scenario, due pacchetti duplicati non seguono necessariamente lo stesso percorso e, quindi, se uno di questi pacchetti subisce un ritardo eccessivo a causa di un link congestionato, non è detto che il suo duplicato subisca lo stesso ritardo. Il ritardo di un segmento vocale è determinato dal ritardo del primo dei pacchetti duplicati che arriva e questo tende a diminuire il ritardo complessivo della voce in Internet.

### 3.1.1. Diminuzione dei costi

Un potenziale problema dell'utilizzo delle linee xDSL per il trasporto del traffico VoIP è che queste linee sono asimmetriche (per esempio, 1280 kbit/s in *download* e 512 kbit/s in *upload*), mentre il traffico VoIP è simmetri-

co. Il sistema Abilis consente di affasciare le linee ADSL in un unico pool in modo da creare, per esempio con quattro linee, una capacità complessiva e aggregata di 2.048 Mbit/s. Tale configurazione consente di aumentare l'affidabilità del sistema. Per esempio, se si guasta una delle linee xDSL, le altre tre restano in funzione e sono gestite normalmente.

Con quattro linee gestite in pool, ne risulta normalmente un throughput di 5120/2048 kbit/s ad un prezzo mensile di circa 180 €, a fronte di un numero molto elevato di conversazioni telefoniche gestibili contemporaneamente. L'equivalente traffico telefonico gestito con le tradizionali linee commutate comporterebbe, solo per i canoni mensili, costi sensibilmente più elevati. Un esempio concreto di computo dei possibili risparmi tramite l'adozione dell'Internet VoIP è riportato nel paragrafo 3.

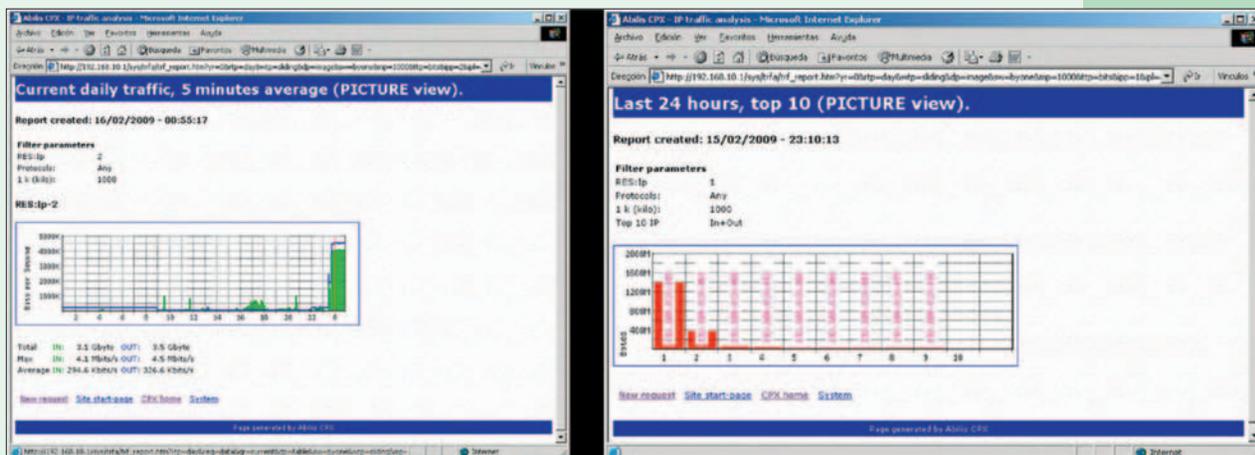
### 3.1.2. Comunicazioni radiomobili

I costi delle comunicazioni radiomobili aziendali sono tipicamente molto elevati. Il sistema Abilis offre una gamma di soluzioni per l'abbattimento di questi costi. Un esempio di applicazione, fra i molti possibili, è mostrato nella figura 3. Un utente aziendale si trova in movimento all'esterno dell'azienda ed è necessario raggiungerlo telefonicamente dalla Sede 2. Per mezzo della rete VoIP Abilis la telefonata è trasportata alla Sede 3, dove si trova un gateway GSM/UMTS che trasforma la chiamata fissa mobile in una chiamata mobile-mobile, con possibile abbattimento dei costi. Gli inconvenienti di questa soluzione sono che:

- a. il numero chiamante è quello di una SIM condivisa da molti utenti;
- b. è necessario verificare la liceità di questa soluzione rispetto ai contratti stipulati con i fornitori delle SIM.

Un'altra applicazione consiste nell'ottimizzare le telefonate che si originano dai cellulari aziendali: anziché dirigerle direttamente al destinatario desiderato si ha l'opzione di chiamare il GSM/UMTS BOX (sfruttando tariffe intra-aziendali) e poi di selezionare il numero aziendale desiderato.

I cellulari aziendali possono anche avvalersi della funzione "callback": chiamando un numero configurato con questo servizio, la chiamata viene automaticamente rifiutata ma il cellulare chiamante viene riconosciuto e richiamato dallo stesso apparato Abilis. Se il



**FIGURA 4**

Esempi di reportistica su interfaccia Web del sistema Abilis

cellulare chiamato si trova all'estero, ciò consente di trasformare il roaming attivo nel più conveniente roaming passivo. Infine, l'azienda può ospitare in una SIM-BOX più schede SIM che possono essere scelte automaticamente e dinamicamente per ottimizzare il costo della chiamata radiomobile. Si ricorda che le soluzioni descritte sono possibilità tecniche rese disponibili dal sistema Abilis. La liceità della loro effettiva applicazione deve essere verificata rispetto ai contratti stipulati con i fornitori delle SIM.

### 3.1.3. Analisi del traffico e reportistica

Il sistema Abilis offre una gamma standard di funzioni di controllo e gestione del traffico, basato su interfaccia Web, tramite un web server presente nel sistema stesso. Nella figura 4 sono mostrati due esempi di reportistica sul traffico Internet, nel primo caso complessivo e nel secondo caso disaggregato per singolo terminale. Sono disponibili le disaggregazioni per singolo protocollo e servizio. Anche i dati relativi alle chiamate telefoniche sono disponibili nel sistema di reportistica, come esemplificato nella figura 5.

## 4. UN CASE STUDY: DME S.R.L. DISTRIBUZIONE MATERIALE ELETTRICO MILANO

DME (*Distribuzione di Materiale Elettrico*) è una media impresa specializzata nella distribuzione di materiale elettrico all'ingrosso e al dettaglio che conta due sedi a Milano, sei sedi in pro-

vincia di Milano e due sedi fuori provincia, in Lombardia. DME ha adottato il sistema Abilis per la gestione delle telecomunicazioni aziendali, passando da un assetto tradizionale con telefonia su linea commutata, appoggiata ad un unico operatore, pagata a canone mensile per linea e traffico a minuti, alla soluzione VoIP su Internet di Abilis. Con la dismissione delle linee telefoniche tradizionali si sono eliminati sia i canoni mensili sia i costi per minuto del traffico telefonico tra le diverse sedi.

Per quanto riguarda i costi del traffico telefonico, DME ora si avvale dei servizi di due operatori telefonici di rete fissa e di altri due operatori di rete mobile. Grazie alla diversificazione dei fornitori, DME si approvvigiona, chiamata per chiamata, nel modo più conveniente.

Altri vantaggi indiretti riscontrati da DME sono l'integrazione dei diversi servizi (telefono fisso e mobile, rete VPN, fax, SMS) in un apparecchio unico. Ciò ha permesso a DME di semplificare e rendere più economiche le attività di gestione delle telecomunicazioni: i servizi di telecomunicazione avanzata disponibili nella sede centrale sono automaticamente disponibili presso tutte le filiali.

Il piano di migrazione è stato così strutturato:

#### 1. sede principale di Milano:

- a. per il servizio di accesso ad Internet e di VPN si sono acquisite tre linee ADSL da tre diversi operatori e si sono eliminate tutte le vecchie linee dati, in parte tariffate a consumo;
- b. per il servizio VoIP si sono utilizzate, per maggiore affidabilità, due linee ADSL, di due diversi operatori;

The screenshot shows a web browser window titled "Accounting Viewer - Calls - Mozilla Firefox" with the URL "http://192.168.0.250/viewer/". The page displays a table of call records and a detailed view of a specific call.

Client	CallId	Direction	Res.	Durat.	Start	End	Calling(IN)	Called(IN)	Calling(OUT)	Called(OUT)
cpk_pbx	490	Input (I)	OK	196	2009-03-20 00:11:45	2009-03-20 00:15:02	0289038001	0289038088	0289038001	0289038088
cpk_pbx	489	Output (O)	OK	196	2009-03-20 00:11:45	2009-03-20 00:15:02	0289038087	00289038088	0289038001	0289038088
cpk_pbx	487	Output (O)	OK	39	2009-03-20 00:09:45	2009-03-20 00:10:25	0289038087	00289038088	0289038001	0289038088
cpk_pbx	488	Input (I)	OK	39	2009-03-20 00:09:45	2009-03-20 00:10:25	0289038001	0289038088	0289038001	0289038088
cpk_pbx	485	Output (O)	OK	200	2009-03-19 22:53:10	2009-03-19 22:56:30	0289038086	00289038088	0289038001	0289038088

Agent	Orig. cluster	Dest. cluster	Orig. port	Dest. port	Orig. side	Dest. side
192.168.0.232		EDISON2	560 (CTI IAX)	306 (Cluster)	User	Network-private

(IN) Called	Num. Type	Num. Plan	SubAddr.	SubAddr. Type	SubAddr. Ind.
0	1				

(IN) Calling	Num. Type	Num. Plan	Num. Pres.	Num. Screen	SubAddr.	SubAddr. Type	SubAddr. Ind.
0	0		0	1			

(OUT) Called	Num. Type	Num. Plan	SubAddr.	SubAddr. Type	SubAddr. Ind.
0	1				

(OUT) Calling	Num. Type	Num. Plan	Num. Pres.	Num. Screen	SubAddr.	SubAddr. Type	SubAddr. Ind.
2	1		0	1			

Audio Codec	Codec	Law	Silent Suppr.	Bitrate	Bandwidth
	Spirit	n		9600	12000

Fax Codec	Codec	Relay	Bypass	Bitrate	Bandwidth

Data Codec	Codec	Relay	Bypass	Bitrate	Bandwidth

Conn. type	Disc. direction	Disc. cause (hex)	Disc. cause ITU decoded	Bearer	Lost records
ItoC	Input (I)	80 9f	U - Normal, unspecified	Speech 64 kbit/s A-law	

Ext. ConnId	Ext. Parent CallId (IN)	Ext. Parent CallId (OUT)	Parent CallID (IN)	Parent CallID (OUT)
200903192256300485			0	

cpk_pbx	486	Input (I)	OK	199	2009-03-19 22:53:10	2009-03-19 22:56:30	0289038001	0289038088	0289038001	0289038088
cpk_pbx	483	Input (I)	OK	196	2009-03-19 22:37:28	2009-03-19 22:40:45	0289038001	0289038088	0289038001	0289038088
cpk_pbx	482	Output (O)	OK	196	2009-03-19 22:37:28	2009-03-19 22:40:45	0289038087	00289038088	0289038001	0289038088

Completato

**FIGURA 5**  
Reportistica  
delle chiamate  
telefoniche

c. si è eliminata un'ulteriore linea HDSL;  
d. si è installato un sistema Abilis ridondante (due apparati in VRRP, dotati di commutatore elettronico per la deviazione automatica delle linee WAN e dei collegamenti interni);

## 2. filiali

a. si sono cessate le vecchie linee di trasmissione dati;  
b. si sono aggiunte due linee ADSL;  
c. si è mantenuta una linea ISDN di backup;  
d. si sono eliminati i centralini telefonici, sostituiti dai moduli telefonici analogici, DECT e VoIP-SIP dei router Abilis;  
e. per la gestione del traffico fisso-mobile sono state acquisiti due gateways GSM/UMTS.

Il costo di acquisizione del sistema Abilis è consistito in due componenti. In primo luogo, i sistemi Abilis della sede centrale e delle otto sedi periferiche sono stati acquisiti tramite noleggio con canone mensile complessivo pari a 1200 €. A conti fatti, il risparmio complessivo netto è di circa 25mila euro/anno. Il conto economico complessivo è sintetizzato nella tabella 1. Per la corretta interpretazione del conto economico riportato, si deve notare che DME non ha ancora completato la migrazione (vi è una sede migrata solo recentemente) e

che quindi il risparmio a regime potrà essere maggiore. Inoltre, DME non ha ancora adottato soluzioni particolarmente spinte sulla telefonia radiomobile che potrà offrire ulteriori risparmi.

Altre funzioni disponibili grazie al sistema Abilis sono:

- posto operatore unico: trasferimento chiamate, gestione code;
- contabilizzazione dei dati di tutte le telefonate entranti ed uscenti, andate a buon fine e fallite (con indicazione dell'eventuale motivo di fallimento della chiamata: nessuna risposta, occupato, numero inesistente, guasto di rete);
- controllo attivo (*blacklist*) delle chiamate: a ciascun telefono aziendale possono essere autorizzate o vietate le telefonate a determinati numeri o prefissi telefonici;
- risponditore automatico interattivo (IVR), gestore delle code; quando il centralinista è occupato, le ulteriori chiamate possono essere prese dallo IVR e tenute in parcheggio o commutate in funzione dei tasti premute dal chiamante;
- VoiceMailbox (collegamento tra la segreteria telefonica e la posta elettronica);
- audioconferenza.

	Dimissioni risparmi mensili	Acquisizioni nuovi costi mensili
<b>Sede centrale</b>		
Canone linea dati principale	-500.00 €	
Canone linee frame relay	-750.00 €	
Canone linea HDSL	-150.00 €	
Canone cinque linee ADSL		225.00 €
<b>Sedi periferiche</b>		
Canone linee dati	-891.00 €	
Canone linee ADSL		720.00 €
Canone linee ISDN backup		248.00 €
<b>Canoni telefonici</b>		
Linea ISDN PRI	-395.00 €	
Nuove linee ISDN BRI con selezione passante	-567.00 €	
Undici linee ISDN BRI	-341.00 €	
Quattro linee ISDN BRI in sede centrale	-68.00 €	
Tre linee telefoniche business	-54.00 €	
<b>Riduzione costi telefonici</b>		
Traffico su rete fissa	-471.00 €	
Traffico fisso-mobile	-810.00 €	
Due SIM GSM		320.00 €
<b>Implementazione del sistema Abilis</b>		
Canoni di acquisizione del sistema Abilis		1,200.00 €
Attività startup e help-desk		200.00 €
<b>Totale dimissioni (Euro/mese)</b>	<b>-4,997.00 €</b>	
<b>Totale spese acquisizione e canoni Abilis</b>		<b>2,913.00 €</b>
<b>Risparmio netto (Euro/mese)</b>	<b>2,084.00 €</b>	
<b>Risparmio annuo</b>	<b>25,008.00 €</b>	

**TABELLA 1**

Conto economico relativo alla migrazione di DME verso la telefonia VoIP su Internet del sistema Abilis

## 5. CONCLUSIONI

La telefonia IP su Internet offre alla piccola/media impresa un *tradeoff* tra prestazioni e costo diverso da quello della classica telefonia a livello *carrier*. La telefonia *carrier-grade* presenta alti livelli di qualità, ma i costi possono essere elevati. Al contrario, la telefonia VoIP su Internet offre un livello di qualità non garantito, spesso inferiore al livello *carrier-grade*. D'altra parte, i costi possono essere tagliati in modo talmente decisivo che l'azienda potrebbe trovarsi propensa a migrare verso una soluzione VoIP Internet.

In questo settore, Abilis rappresenta una realtà italiana molto interessante. Abilis ha sviluppato una piattaforma hardware/software completamente in house, basandosi su tecnologia ICT di largo consumo dotata di un sistema operativo leggero e veloce e di una pila software, di base e applicativo, sviluppata completamente all'interno dell'azienda. Questo offre una serie di vantaggi:

- Abilis è indipendente dai grandi vendor di apparati di telecomunicazioni;
- l'utilizzo di tecnologia ICT di largo consumo permette ad Abilis di praticare una politica di *pricing* molto aggressiva;

0



• lo sviluppo dei dispositivi è molto più vicino all'utente finale che nel caso dei vendor maggiori: Abilis può intervenire su tutte le pile software per implementare nuovi servizi e customizzazioni.

### Bibliografia

[1] [www.abilis.net](http://www.abilis.net)

[2] Daniel Collins: *Carrier Grade Voice Over IP*. McGraw-Hill Networking Professional.

[3] [www.on-time.de](http://www.on-time.de)

[4] Teck-Kuen Chua, David C. Pheanis: QoS evaluation of sender-based loss-recovery techniques for VoIP. *IEEE Network Magazine*, Vol. 20, n. 6, Nov/Dec 2006, p. 14-22.

[5] <http://www.abilis.net>

1

0

1

0

PAOLO GIACOMAZZI si è laureato in Ingegneria Elettronica presso il Politecnico di Milano nel 1990 ed ha conseguito il Master in tecnologia dell'informazione al CEFRIEL. Dal 1992 al 1998 è stato ricercatore con il Politecnico di Milano dove ora è professore associato di telecomunicazioni. L'attività didattica e la ricerca riguardano la qualità del servizio nella rete Internet multimediale, le reti radiomobili B3G e la sicurezza nelle reti di telecomunicazioni. È editor del IEEE Network Magazine ed è editor della Book Reviewing Feature del IEEE Network Magazine.

E-mail: [giacomaz@elet.polimi.it](mailto:giacomaz@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.

## ISO/IEC 20000: la Norma per la qualità dell'erogazione dei Servizi IT

Attilio Rampazzo, Antonio Piva, David D'Agostini

### 1. INTRODUZIONE

La gestione di un'organizzazione moderna è un'attività complessa: deve rispondere in maniera sistematica e documentata del rispetto di leggi, di norme settoriali e volontarie, di impegni contrattuali, di attese della proprietà, dei clienti e della comunità e lo deve fare operando in un contesto sempre più dinamico; deve gestire e far coesistere interessi economici con requisiti relativi alla sicurezza delle persone e dei prodotti, alla qualità e rispetto di tutte le parti interessate.

Tra le varie incombenze la gestione e l'erogazione dei servizi IT (conosciuta come *IT Service Management*) si dimostra sempre più come una necessità per tutte le organizzazioni, pubbliche e private, sia che si tratti di servizi destinati all'interno dell'organizzazione, sia che i servizi vengano rivolti al cliente/utente.

Tra le varie proposte introdotte da IBM, HP, Microsoft e altre organizzazioni del mondo IT, il modello *IT Infrastructure Library* (ITIL) ha dimostrato la sua validità ed efficacia rappresentando un framework e un approccio internazionalmente condivisi per una corretta gestione ed erogazione dei servizi IT. Partendo da questo modello è nata una Norma Internazionale, l'ISO/IEC 20000, che costi-

tuisce un valido strumento per l'ottimizzazione degli aspetti di gestione dei servizi per l'IT e per la progettazione dei processi di business correlati<sup>1</sup>.

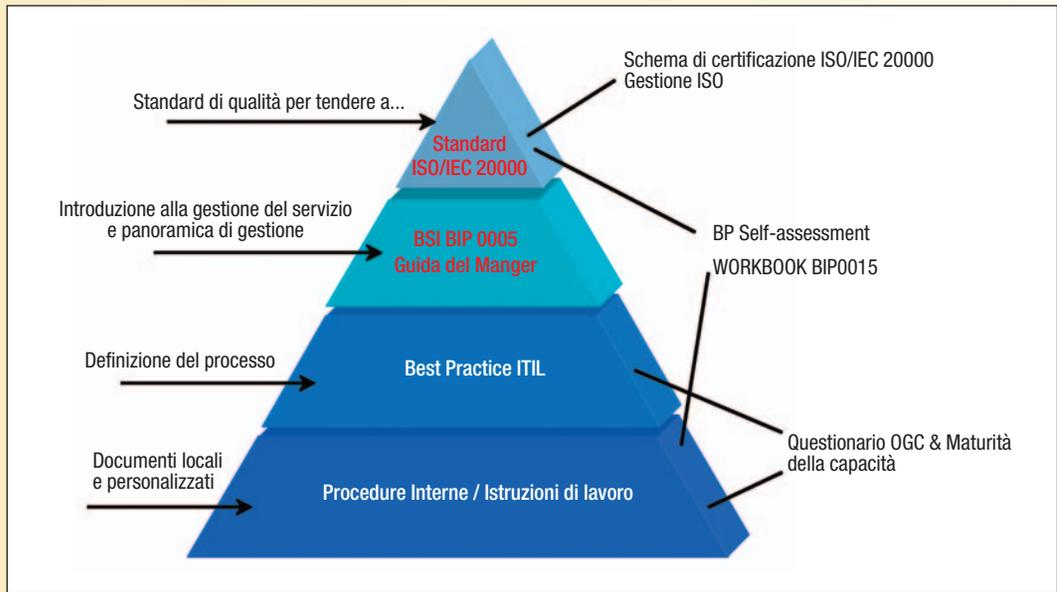
ITIL è un insieme di consigli pratici (*best practice*), mentre la ISO/IEC 20000 è un insieme formale di specifiche di cui un fornitore del servizio deve tendere per essere in grado di fornire un'elevata qualità dei servizi<sup>2</sup> (nella Figura 1 sono illustrate le relazioni fra la ISO/IEC 20000 e ITIL).

Prima d'ora, chi erogava servizi IT per manifestare la sua capacità si certificava ISO 9001 settore EA 33 (Servizi IT), mentre adesso ha una nuova e valida possibilità di dimostrare il controllo efficace e il miglioramento continuo dell'intero complesso di prestazioni dell'IT Service Management grazie alla Norma ISO/IEC 20000, che prevede l'implementazione di un **Sistema di Gestione dei Servizi IT** (SGSIT). Infatti l'ISO/IEC 20000 è una norma dedicata alla valutazione delle organizzazioni che erogano servizi IT che riconosce l'importanza dei servizi IT, ne individua le specificità e stabilisce l'esigenza di una risposta adeguata ai problemi che le tecnologie dell'informazione comportano nella impostazione e nell'esercizio di un Sistema di gestione del servizio.

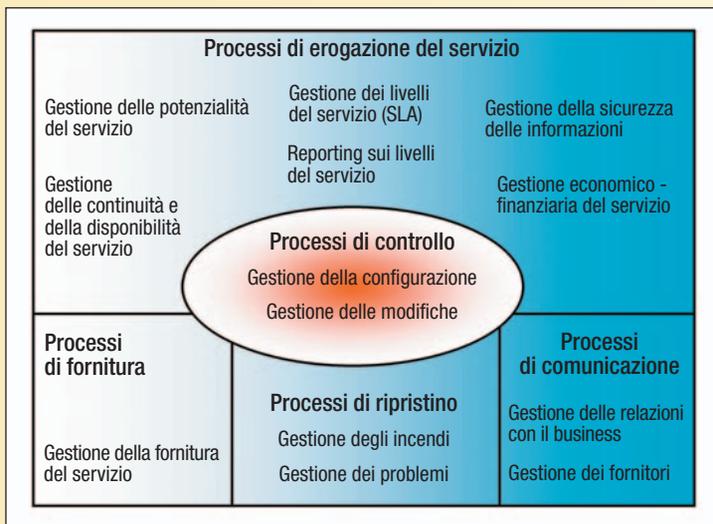
La Norma ISO/IEC 20000, applicabile a organiz-

<sup>1</sup> In merito alle normative ISO/IEC si ricorda che questa rubrica ha già preso in esame la famiglia ISO 27000 nell'articolo pubblicato su Mondo Digitale di settembre 2008.

<sup>2</sup> L'applicazione dei "consigli pratici" ITIL potrà supportare un fornitore del servizio nel raggiungere la qualità del servizio così come richiesto dalla ISO/IEC 20000.



**FIGURA 1**  
Relazioni fra  
la ISO/IEC 20000  
e ITIL



**FIGURA 2**  
Processi del Service Management

zazioni di tutte le dimensioni, si compone di due parti:

□ parte 1 *Specification*: ovvero la ISO/IEC 20000-1:2005. Contiene le specifiche (o requisiti) per l'implementazione di un Sistema di Gestione per l'Erogazione dei Servizi IT;

□ parte 2 *Code of practice*: ovvero la ISO/IEC 20000-2:2005. Contiene linee guida, raccomandazioni utili per la corretta implementazione di quanto specificato nella parte 1.

L'ISO/IEC 20000 si configura, quindi, come una nuova famiglia di Norme.

La prima parte della Norma (ISO/IEC 20000-1) si pone l'obiettivo di promuovere un approccio integrato alla gestione di un'organizzazione erogatrice di servizi IT<sup>3</sup> con l'idea che una buona qualità del processo di erogazione implichi una buona qualità del servizio erogato. La capacità dell'organizzazione consiste nel saper coniugare in modo equilibrato efficienza ed efficacia, profitto economico, vantaggi per il cliente e miglioramento continuo.

La ISO/IEC 20000-1 impone una lista di obiettivi e controlli<sup>4</sup> ed evidenzia i processi strettamente collegati alla gestione del servizio, in sinergia a ITIL (la Figura 2 rappresenta lo schema dei processi). Esprime i requisiti dei processi di erogazione attraverso precise prescrizioni elementari<sup>5</sup>. La seconda parte della Norma (ISO/IEC 20000-2) si pone l'obiettivo di affiancare la prima parte della Norma fornendo linee guida e raccomanda-

<sup>3</sup> L'organizzazione erogatrice è un insieme complesso di processi, di relazioni, di flussi, di competenze, di risorse, di tecnologie. L'approccio integrato proposto dalla norma consente l'impostazione del controllo sui flussi ciclici di erogazione del servizio e delle opportunità di miglioramento continuo.

<sup>4</sup> Possono risultare non esaustivi: ogni organizzazione, in base alle proprie necessità, può considerare obiettivi e controlli che si aggiungono a quelli previsti dalla norma.

<sup>5</sup> *Shall*: i requisiti espressi devono applicarsi al sistema di erogazione del servizio IT quando l'organizzazione erogatrice si ponga l'obiettivo di realizzare un servizio gestito con una predefinita (e accettabile) qualità per i propri clienti.

zioni, descrizione di buone pratiche e consigli pratici connessi alle possibili applicazioni dei requisiti espressi nella prima parte. La norma individua le aree critiche della progettazione e dell'erogazione dei servizi IT e richiede che persone, processi e tecnologie (ICT) interagiscano e si mantengano costantemente coordinati sui fini stabiliti per l'organizzazione. Tra i fattori che determinano il successo di un'organizzazione erogatrice di servizi IT si riconosce la preponderanza del fattore umano, ovvero le persone che governano i processi di erogazione dei servizi<sup>6</sup>.

L'uso congiunto di seconda e prima parte della norma mette l'organizzazione erogatrice in condizione di progettare e realizzare il miglioramento della qualità del servizio erogato in modo strutturato e proattivo anziché sporadico e reattivo<sup>7</sup>.

Al di là del regolare uso combinato, si potrebbe anche affermare che l'ambito di applicazione della seconda parte della norma sia progettazione ed amministrazione del sistema di gestione piuttosto che la valutazione di conformità (di prima, seconda e terza parte<sup>8</sup>) trattata e resa praticabile dalla prima parte della norma.

La Norma ISO/IEC 20000, come già accennato, è sostanzialmente allineata con l'*Information Technology Infrastructure Library* (ITIL), nonostante non ne includa formalmente l'approccio. Il contenuto della norma è comunque tale da poter supportare altri frameworks e approcci simili (per esempio *MOF Microsoft Operational Framework* (Microsoft), *HP ITSM Reference Model* (Hewlett Packard), *IT Process Model* (IBM) ecc.).

## 2. ORIGINE ED EVOLUZIONE DELLA FAMIGLIA DI NORME ISO/IEC 20000

Le origini di ISO/IEC 20000:2005 sono abbastanza recenti e affondano le proprie radici negli anni '80 quando nasce il progetto *Information Technology Infrastructure Library* (ITIL) per iniziativa del Governo Britannico.

ITIL v.1 (1989 – inizialmente "*Guide for Government*") è stato sviluppato sotto gli auspici del CCTA *Central Computer and Telecommunications Agency* (CCTA) oggi conglobata nell'*Office of Government Commerce* (OGC) in risposta alla crescente difficoltà di azione delle agenzie governative britanniche nella gestione dell'innovazione della tecnologia dell'informazione. I lavori del gruppo ITIL, inizialmente orientati sulla costruzione di un metodo formale, è decisamente cambiata in corso d'opera concentrando l'attenzione verso le "best practices". In tal modo i consigli e linee guida di ITIL sono stati implementati in diverse organizzazioni ed utilizzati nei settori più disparati (amministrazioni pubbliche locali e centrali, industrie e imprese di servizi). Sebbene sviluppato durante gli anni 80, ITIL ha raggiunto grande popolarità solo dopo la metà degli anni 90.

Il *British Standards Institute* (BSI) organizzazione britannica fondata nel 1901 e operante oggi nello sviluppo di norme nazionali e internazionali, nella certificazione di terza parte indipendente di prodotti e sistemi di gestione e nella formazione manageriale - ha pubblicato "*A Code of Practice for IT Service Management - DISC PD0005:1998*". La norma documenta l'approccio alla gestione del servizio IT basato sui principi stabiliti da ITIL e sulle esperienze consolidate nell'industria. Nella norma si può trovare una prima definizione dei processi chiave della gestione dei servizi: *service design & management, control, release, resolution, supplier*.

Nel 2001 OGC ha rilasciato ITIL v.2 concentrando il focus sull'approccio per processi; *Service Support & Service Delivery* costituiscono la macro area fondamentale per la gestione e l'erogazione dei servizi IT e sono il cuore della seconda versione la cui struttura si completa con altre sei aree:

- *ICT Infrastructure Management;*
- *Planning To Implement Service Management;*
- *Applications Management;*
- *The Business Perspective;*

<sup>6</sup> L'applicazione delle prescrizioni della norma alla gestione del servizio ed alla evoluzione del sistema di gestione è affidata all'attenzione di persone competenti.

<sup>7</sup> Naturalmente l'adozione della norma nella progettazione organizzativa richiede all'Alta Direzione le capacità interpretative che consentono la costruzione di un sistema di gestione in funzione delle specifiche caratteristiche dell'organizzazione erogatrice (dimensioni, struttura, competenze, personale, strategia ecc.).

<sup>8</sup> Di prima parte mediante Audit interni dell'organizzazione; di seconda parte mediante verifiche effettuate dall'organizzazione nei confronti dei fornitori; di terza parte indipendente per esempio nel processo di certificazione.

- *Software Asset Management*;
- *Security Management*.

Nel Novembre 2000 BSI ha pubblicato la prima norma che stabilisce i requisiti del sistema di gestione dei servizi IT: la BS15000:2000 è stata sviluppata sulla base della precedente pubblicazione DISC PD005:1998. La seconda edizione è uscita nel 2002: l'aggiornamento è stato fatto raccogliendo l'esperienza degli utilizzatori della prima edizione. La struttura della seconda edizione anticipa la 20000 con una prima parte che definisce i requisiti (BS15000-1 IT service management Part 1: *Specification for service management*) e una seconda parte che fornisce raccomandazioni (BS15000-2 IT service management Part 2: *Code of practice for service management*).

Nel 2004 l'*International Organization for Standardization* (ISO) - organismo internazionale per la definizione delle norme a livello mondiale - e l'*International Electrotechnical Commission* (IEC) - Commissione Elettrotecnica Internazionale - attraverso il comitato tecnico congiunto JTC1/SC7, responsabile dello sviluppo delle norme nell'a-

rea dell'ingegneria del software e dei sistemi, hanno accolto la norma nazionale BS15000 e l'hanno promossa il 15 dicembre 2005 a Norma Internazionale ISO/IEC 20000:2005<sup>9</sup>.

Questa breve cronologia mostra come la nascita e l'evoluzione della Norma ISO/IEC 20000:2005 si intreccino con l'evoluzione di ITIL. Norma e ITIL si sviluppano in maniera complementare e non costituiscono alternativa<sup>10</sup>. La norma formalizza in termini di requisiti del sistema di gestione gli elementi chiave descritti nelle "Buone Pratiche" raccolte in ITIL. Le organizzazioni che adottano lo schema ITIL pubblicizzano tale scelta per distinguersi nello scenario competitivo di interesse. Da qui nasce l'esigenza di poter verificare in modo obiettivo e imparziale un'affermazione di conformità di un'organizzazione allo schema dichiarato: in altre parole nasce l'esigenza di una norma che possa funzionare da metro per la verifica in analogia a quanto avviene con la Norma ISO 9001. Nella tabella 1 possiamo trovare una sintesi dei principali eventi che hanno caratterizzato l'evoluzione della normativa su IT Service Management.

**TABELLA 1**  
Sintesi dei principali eventi che hanno caratterizzato l'evoluzione della normativa su IT Service Management

Anno pubblicazione	Norma/Proposta pubblicata
1998	BS DISC PD0005:1998 - "A Code of Practice for IT Service Management"
2000	BS 15000:2000 - "Specification for IT Service Management"
2002	BS 15000-1:2002 - "IT service management. Specification for Service Management"
2003	BS 15000-2:2003 - "IT service management. Code of practice for Service Management"
2005	ISO/IEC 20000-1, Information technology — Service Management — Part 1: Specification - Stabilisce i requisiti del sistema di gestione
2005	ISO/IEC 20000-2, Information technology — Service Management — Part 2: Code of Practice - Raccomanda buone pratiche per la gestione
2008	ISO/IEC CD TR 20000-3 - Information technology — Service Management — Part 3: Guidance on compliance with ISO/IEC 20000-1
2008	ISO/IEC CD TR 20000-4 - Information Technology - Service Management — Process Reference Model

<sup>9</sup> Secondo la procedura "fast track" (percorso abbreviato - tre passi invece dei sei canonici, che mediamente consente in poco meno di un anno di lavoro la pubblicazione di una norma internazionale a partire da una consolidata norma nazionale).

<sup>10</sup> Il primo giugno 2007 OGC ha pubblicato la nuova versione ITIL v3 concentrando il focus sull'intero ciclo di vita del servizio e sull'allineamento tra la gestione del servizio e la gestione dell'impresa. I nuovi volumi coprono i seguenti temi: *Strategy, Design, Transition, Operation, Continual Improvement*.

### 3. CONTENUTI DELLA NORMA ISO/IEC 20000-1:2005

La Norma ISO/IEC 20000-1 è composta da 10 capitoli in cui sono specificati i requisiti che possono essere classificati in tre macro categorie:

- i requisiti di carattere generale dell'organizzazione – capitoli 3, 4, 5;
- i requisiti dei processi primari di erogazione – capitoli 6, 7;
- i requisiti dei processi di supporto – capitoli 8, 9, 10.

La parte iniziale della norma fino al capitolo 2 ha un carattere introduttivo come in tutte le norme. Di particolare interesse sono le ultime tre definizioni tratte dal glossario, parte integrante della norma, in quanto anticipatrici del principio della norma ovvero il concetto di livello di servizio:

□ **SLA - Service Level Agreement** - Accordo sui livelli di servizio: accordo documentato fra il fornitore di servizi e il cliente che descrive il servizio e ne definisce i livelli concordati.

□ **Service Management** - Gestione del servizio: le attività di organizzazione/pianificazione/controllo che mirano a soddisfare le esigenze di business (fornire valore ai clienti) attraverso un servizio.

□ **Service Provider** - Erogatore del servizio (Fornitore): organizzazione che fornisce servizi a uno o più clienti (interni o esterni).

I caratteri generali della norma prendono corpo nei capitoli 3, 4 e 5.

Nel capitolo 3 sono riportati i requisiti generali

per implementare un Sistema di Gestione che permetta la progettazione-erogazione di tutti i servizi IT.

La realizzazione del sistema di gestione si basa su tre fattori chiave:

- responsabilità dell'Alta Direzione;
- documentazione;
- competenza, consapevolezza e formazione del personale.

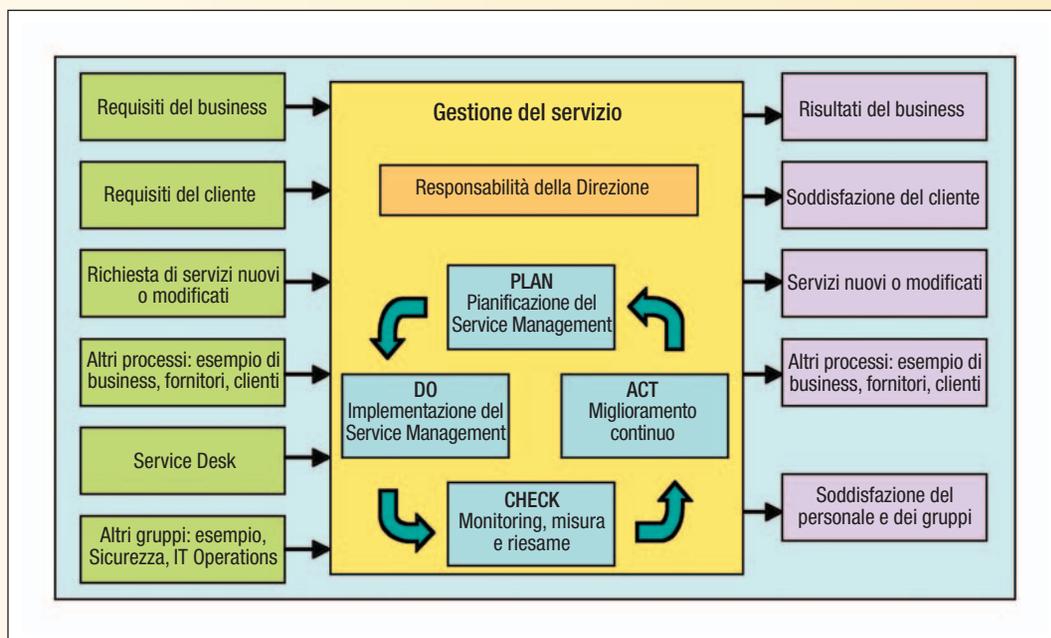
Nel capitolo 4 sono riportati i requisiti per la pianificazione e l'implementazione della gestione del servizio e si adotta la metodologia "Plan-Do-Check-Act" (PDCA – ciclo di DEMING) per tutti i processi organizzativi e gestionali (Figura 3).

Il processo direzionale viene messo in relazione con il ciclo evolutivo del sistema di gestione:

- pianificazione della gestione del servizio (*Plan*);
- implementazione della gestione ed erogazione dei servizi (*Do*);
- monitoraggio, misurazioni e revisioni (*Check*);
- miglioramento continuo (*Act*).

Nel capitolo 5 sono riportati i requisiti per la pianificazione e l'implementazione dell'innovazione (servizi nuovi o cambiamenti a servizi attivi); la gestione corrente del servizio, impostata secondo il metodo PDCA, consente un primo livello di governo orientato al miglioramento continuo.

L'Organizzazione deve mantenere attivo un sistema per il monitoraggio delle situazioni di crisi (opportunità e minacce) cui rispondere



**FIGURA 3**  
Ciclo PDCA del Sistema di Gestione dei Servizi IT

con un approccio all'innovazione radicale e settoriale<sup>11</sup>.

In questo contesto si intende per innovazione quel cambiamento del servizio e del sistema di gestione del servizio finalizzato alla massimizzazione dei benefici attesi per una data configurazione di risorse (finanziarie e di conoscenza). Così come per il miglioramento anche per l'innovazione l'iniziativa deve essere gestita come un programma-progetto i cui elementi devono costantemente essere allineati alla strategia e obiettivi dell'Organizzazione.

Nei successivi capitoli vengono esposte le specificità di un Sistema di Gestione dei Servizi IT. Nel capitolo 6 sono riportati i requisiti per l'erogazione del servizio, vengono trattati tutti quei processi che negoziano, definiscono e concordano gli effettivi livelli di servizio e che riportano le prestazioni raggiunte rispetto a tali obiettivi. Rientrano a far parte dell'ambito di erogazione del servizio i seguenti processi:

- *Service Level Management;*
- *Service Reporting;*
- *Capacity Management;*
- *Service Continuity and Availability Management;*
- *Information Security Management;*
- *Budgeting and Accounting for IT services.*

Nel capitolo 7 sono riportati i requisiti relativi alla gestione dei subfornitori e alla gestione delle relazioni di business. Rientrano a far parte dell'ambito di gestione delle relazioni del servizio i seguenti processi:

- *Business Relationship Management;*
- *Supplier Management.*

Sia il subfornitore che il cliente possono essere sia interni che esterni all'organizzazione del fornitore del servizio, ma mentre le relazioni esterne vanno formalizzate anche attraverso contratti di subfornitura, per le relazioni interne bastano accordi sui livelli operativi o di servizio. Al fine di stabilire e mantenere buone relazioni dovrebbero essere definite le figure di interfaccia per tutte le parti, che dovrebbero definire e concordare i bisogni di business, l'ambito, i ruoli e le responsabilità delle relazioni e i canali e la frequenza di comunicazione.

Nel capitolo 8 sono riportati i requisiti relativi

alla risoluzione attraverso i processi di risoluzione, incident e problem management, processi separati, benché strettamente collegati. Rientrano a far parte dell'ambito di risoluzione del servizio i seguenti processi:

- *Incident Management;*
- *Problem Management.*

Nel capitolo 9 sono riportati i requisiti che permettono al fornitore del servizio di controllare i componenti del servizio dell'infrastruttura e di mantenere informazioni accurate sulla configurazione. Questa accuratezza delle informazioni è un requisito fondamentale per il processo decisionale interno al processo di *change management* così come per tutti gli altri processi dell'organizzazione di servizio IT. Nell'alveo del controllo del servizio vi sono i processi:

- *Configuration Management;*
- *Change Management.*

Nel capitolo 10 sono riportati i requisiti per il rilascio dei cambiamenti pianificati. Alla messa in produzione del servizio appartiene il processo:

- *Release Management.*

La norma evidenzia l'importanza che il processo di *release management* sia integrato con i processi di *configuration* e *change management*, al fine di assicurare la sintonia e l'accordo dei rilasci e dei cambiamenti eseguiti.

#### 4. FUTURO ED EVOLUZIONE

La versione attuale della norma internazionale riflette la struttura della BS15000:2002 (originaria norma britannica) la quale era già in fase di sperimentazione nel contesto anglosassone. Si può ritenere, pertanto, che il livello di penetrazione della certificazione sia in una fase evolutiva e ciò è testimoniato anche dal numero di certificazioni a livello mondiale (Tabella 2). Le prospettive della norma sono, poi, delineate all'interno dei progetti del Comitato Tecnico congiunto ISO & IEC JTC1/SC7: riguardano una revisione per adeguamento a ITIL v.3 e la redazione della terza e della quarta parte della Norma ISO 20000.

La parte 3 (*Guidance on Compliance with ISO/IEC 20000-1*) fornirà linee guida addizionali per la definizione dei contenuti (campi di applicazione) della gestione dei servizi IT con il duplice scopo di sostenere l'approccio sia alla pianificazione dei sistemi, sia alla verifica di conformità da parte degli auditors (di prima, seconda e terza parte). La ISO/IEC 20000-3

<sup>11</sup> In quanto l'approccio graduale del miglioramento continuo potrebbe non essere sufficiente: la risposta ad una aggressività tecnologica potrebbe arrivare in ritardo.

sarà indirizzata alle organizzazioni erogatrici di servizi e prenderà la forma di una raccolta di spiegazioni, raccomandazioni ed esempi sulla definizione del campo di applicazione della gestione. Si prevede l'uso congiunto delle parti terza e prima, senza escludere l'utilità di un affiancamento con la seconda.

La parte 4 (*Process Reference Model for Service Management*) fornirà un modello di processo utilizzabile come riferimento per la valutazione del livello di capacità dei processi di gestione dei servizi implementati da un'organizzazione erogatrice che ha adottato la ISO/IEC 20000-1. ISO/IEC 20000-4 sarà fedele alla impostazione data alla nuova Norma ISO/IEC 12207 (pubblicata nel marzo 2008) e consentirà sia di stabilire l'organizzazione dei processi del ciclo di vita dei servizi IT, sia di classificare la capacità del processo di gestione dei servizi IT sulla scala dei livelli di maturità.

## 5. COLLEGAMENTI CON ALTRE NORME

La Norma ISO/IEC 20000-1, attraverso le evoluzioni seguite alla pubblicazione nel 2002 della prima release del predecessore BS 15000, ha visto una migliore armonizzazione dei processi, inglobando anche il noto paradigma *Plan-Do-Check-Act* (PDCA) e, quindi, risulta allineata alle altre norme che adottano un Sistema di Gestione (vedi ISO 9001, ISO 14001, ISO 27001). Molte organizzazioni IT<sup>12</sup> sono certificate secondo la ISO 9001 nel settore EA 33, quindi la ISO/IEC 20000-1 potrebbe sostituire o meglio integrare molte di queste certificazioni.

Uno dei processi del *Service Delivery* della ISO/IEC 20000-1 è l'*Information Security Management*. La ISO/IEC 27002:2005<sup>13</sup> fornisce delle indicazioni o *best practice* sulla Gestione della Sicurezza delle Informazioni atte a implementare questo processo. Le organizzazioni già certificate ISO/IEC 27001:2005 possono soddisfare i requisiti di sicurezza propri della ISO/IEC 20000-1, sempre che gli scopi di certificazione coincidano o comunque siano nel perimetro di certificazione.

Ulteriore processo del *Service Delivery* della

<sup>12</sup> Al giorno d'oggi se ne contano circa 2800 (Fonte: sito SINCERT al 31 dicembre 2008).

<sup>13</sup> Rinumerazione della ISO/IEC 17799:2005.

Nazione	Nr. Aziende Certificate
UK United Kingdom	52
Japan	50
India	42
China	35
South Korea	35
Germany	20
Taiwan	14
Switzerland	12
USA	18
Hong Kong	7
Austria	8
Czech Republic	8
France	4
Poland	4
Australia	3
Denmark	3
Netherlands	3
Spain	3
Thailand	3
Brazil	2
Ireland	2
Malaysia	2
Philippines	2
Saudi Arabia	2
Singapore	2
Slovakia	2
Turkey	2
United Arab Emirates	2
Botswana	1
Colombia	1
Finland	1
Kuwait	1
Italy	1
Latvia	1
Liechtenstein	1
Qatar	1
Russia	1
Sri Lanka	1
<b>Totale</b>	<b>352</b>

**TABELLA 2**

Aziende certificate ISO 20000-1 suddivise per Nazione (Fonte: [www.isoiec20000-certification.com](http://www.isoiec20000-certification.com) al 10/marzo/2009)

ISO/IEC 20000-1 è il *Service Continuity Management*. La BS 25999-1:2006 del British Standard fornisce delle indicazioni o *best practice* sulla Gestione della *Business Continuity*, nota come BCM. A dicembre 2007 è stata pubblicata la BS 25999-2:2007 che permette alle organizzazioni di certificare il Sistema di Gestione della *Business Continuity* aziendale soddisfacendo così anche i requisiti propri della ISO/IEC 20000-1 in questo particolare argomento<sup>14</sup>.

La normativa ISO/IEC esaminata, in tutta evidenza, presenta tratti di forte connessione con la disciplina legislativa in materia di trattamento dei dati personali e della sicurezza informatica di cui al d.lgs. 196/03 (cosiddetto Codice della privacy) e relativo Allegato B.

Tali disposizioni (già trattate approfonditamente in questa rivista nel numero di giugno 2007) hanno naturalmente forza imperativa di legge, a differenza della ISO/IEC a cui si può liberamente scegliere di sottostare; tuttavia, considerato che quest'ultima contiene previsioni a tratti più rigorose di quelle dettate dal legislatore, appare evidente che l'ottenimento della certificazione implica l'ottemperanza (ovvero favorisce il rispetto) di determinate norme del Codice della privacy o, quantomeno, una rivisitazione dei relativi adempimenti in ambito aziendale (si pensi alle misure minime di sicu-

rezza in riferimento ad accessi abusivi o a danneggiamenti informatici).

## 6. CONCLUSIONI

Nel business di oggi giorno è evidente la necessità di dimostrare la capacità di fornire servizi che soddisfino i requisiti dei clienti sia interni che esterni all'organizzazione: ciò può essere ottenuto attraverso la standardizzazione dei processi di gestione del servizio. L'approccio è teso a enfatizzare i benefici per l'utente finale e rappresenta un risultato condiviso relativamente agli standard di qualità per i processi di gestione dei servizi IT.

La ISO/IEC/IEC 20000-1 rappresenta un grandissimo balzo in avanti, che supera i precedenti tentativi di gestione dei servizi IT come attività individuali separate e distinte. Il nuovo approccio si concentra invece sulla fornitura di servizi *end-to-end* mediante l'applicazione dei modelli e *best practice* o consigli pratici per l'implementazione e gestione dei processi.

## Bibliografia

- [1] ISO/IEC 20000-1:2005 *Information technology – Service management – Part 1: Specification*.
- [2] ISO/IEC 20000-2:2005 *Information technology – Service management – Part 2: Code of practice*.
- [3] ISO/IEC 20000 *Pocket Guide* – ITSMF ITALIA – Van Haren Publishing.

<sup>14</sup> Sempreché gli scopi di certificazione coincidano.

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

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

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



# 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.

## Competenze per la Governance dell'ICT

### Rafforzare le strategie di outsourcing dei servizi ICT e gestire il cambiamento della PA

**Marco Gentili**

#### 1. IL CNIPA

Il Centro Nazionale per l'Informatica nella Pubblica Amministrazione (CNIPA) opera presso la Presidenza del Consiglio dei Ministri per l'attuazione delle politiche formulate, a nome del Governo, dal Ministro per la Pubblica Amministrazione e l'Innovazione. La missione è quella di contribuire alla creazione di valore per cittadini e imprese, fornendo supporto alla Pubblica Amministrazione nell'uso innovativo dell'informatica e, più in generale, dell'ICT (*Information and Communication Technology*).

In questo contesto, le modalità di azione del CNIPA sono:

- l'attività di consulenza e proposta relativamente a strategie e azioni puntuali rivolte ai decisori politici, alle Pubbliche Amministrazioni e agli operatori del settore;
- l'emissione di normative tecniche a livello secondario, quali linee guida e guide tecniche, sia emesse dal CNIPA, sia da parte delle Amministrazioni, con il supporto del CNIPA;
- la valutazione - ex ante (di coerenza con le strategie nazionali di innovazione del Governo), in itinere (durante l'attuazione dei progetti pro-

grammati), ed ex post (sui risultati raggiunti) - delle attività ICT della Pubblica Amministrazione centrale, sia a livello strategico, sui Piani pluriennali di sviluppo dei sistemi informativi, sia a livello di singoli interventi, che per legge sono sottoposti a parere (obbligatorio, ma non vincolante) di congruità tecnica ed economica da parte del CNIPA;

□ la definizione e gestione (a livello di committente e con riferimento in particolare alle specifiche funzionali) di progetti dimostrativi, ad alto impatto innovativo, nell'uso dell'ICT nella P.A. centrale, nelle Regioni e negli Enti locali.

Il CNIPA è quindi un generatore di conoscenza, condivisa con tutte le parti interessate (*stakeholders*), siano esse Amministrazioni Centrali, Regioni o Enti Locali, o il mercato rappresentato dai fornitori ICT, dai mondi associativi riferibili a Confindustria e Confcommercio, come anche alle numerose associazioni che presidiano i diversi aspetti afferenti all'ICT e alla sua *governance e management* (*best practices, standard, project management, quality management, management consulting, process engineering, software engineering* ecc.). Questa conoscenza, in relazione al governo dell'innovazione su cui il CNIPA è impegnato, è stata codificata nelle "Linee guida sulla qualità dei beni e servizi ICT per la definizione e il governo dei contratti della Pubblica Amministrazione". Le Linee Guida sono utili per definire le modalità per ottenere la qualità necessaria dai fornitori ICT, allo scopo di creare valore per i cittadini e le imprese interpretati come clienti dell'Amministrazione acquirente.

Queste Linee Guida, in costante evoluzione e miglioramento a cura dell'area "metodologie per la qualità e l'innovazione organizzativa", sotto la responsabilità del Dott. Marco Gentili, costituiscono un'opera enciclopedica che copre tutte le tematiche relative all'acquisizione di beni e servizi ICT con particolare riferimento all'interazione tra stazioni appaltanti e fornitori ICT, utili per:

- supportare la gestione del cambiamento della pubblica amministrazione;
- rafforzare le strategie pubbliche di outsourcing dell'ICT;
- rendere la qualità delle forniture ICT un concetto concretamente e pragmaticamente utilizzabile.

Per perseguire questi obiettivi le Linee Guida propongono un approccio pragmatico al tema della qualità, che fornisce indicazioni concrete

e facilmente attuabili, per la gestione operativa della relazione cliente-fornitore relativamente all'intero ciclo di vita dell'acquisizione di forniture ICT, in particolare identificando come:

- **correlare la qualità alla missione istituzionale** e alle esigenze dell'Amministrazione, nell'impostazione di una strategia di acquisizione delle forniture ICT;
- **definire la qualità necessaria e i relativi costi**, effettuando un'analisi di fattibilità;
- **richiedere la qualità**, nell'appalto pubblico di forniture ICT definendo adeguatamente i criteri di selezione delle offerte;
- **valutare la qualità offerta**, in fase di aggiudicazione delle gare;
- **formalizzare la qualità richiesta**, rappresentandola senza ambiguità nel contratto;
- **verificare la qualità ottenuta**, in fase di attuazione e governo del contratto mediante la valutazione in-itinere e ex-post.

Le Linee Guida sono scaricabili gratuitamente dal sito del CNIPA ([www.cnipa.gov.it](http://www.cnipa.gov.it)) a partire dal menù intitolato "Qualità delle forniture ICT", nella sezione denominata "Attività" posta sulla sinistra dell'*home page*.

La qualità viene indissolubilmente rapportata, in una visione unitaria e integrata, alla *governance e management* dell'ICT che l'Amministrazione deve attuare e alle vaste e intersecanti tematiche ad essa afferenti che in letteratura vanno sotto il nome di *policy* (politiche), *sourcing strategies* (strategie di acquisizione), *feasibility analysis* (analisi di fattibilità), *public procurement* (appalto pubblico), *deal negotiation* (negoziato del contratto), *management* (gestione operativa), *monitoring & control* (valutazione in itinere), *impact analysis* (valutazione ex-post).

## 2. GOVERNARE L'INNOVAZIONE

Il rapporto tra Amministrazioni Pubbliche e cittadini e imprese si è modificato nell'ultimo decennio in conseguenza a due fenomeni concomitanti. Da un lato è emersa una maggiore consapevolezza che l'efficienza e l'efficacia della Pubblica Amministrazione è un elemento indispensabile per la competitività del sistema economico nazionale. Dall'altro è cresciuta la domanda di servizi più avanzati in grado di soddisfare le esigenze connesse:

- alla **crescente integrazione europea**, in un'ottica di globalizzazione dei mercati internazionali;

□ alla **semplificazione legislativa**, perseguita, sia delegando e deregolamentando mediante redazione di testi unici, che ricorrendo all'autocertificazione;

□ al **federalismo amministrativo**, basato su uno Stato più leggero ma più efficace, attuato secondo i principi della sussidiarietà orizzontale, ottenuta dismettendo le funzioni non essenziali per concentrare le risorse sui compiti basilari, e verticale, raggiunta trasferendo i compiti, funzioni e risorse dalle amministrazioni dello stato agli enti locali.

Tutto ciò si può tradurre, utilizzando il linguaggio della norma ISO EN 9001:2000, nell'affermazione che la Pubblica Amministrazione ha innanzitutto scoperto di avere dei clienti, i cittadini e le imprese, e in secondo luogo di dover offrire a questi clienti dei servizi di qualità in termini di modalità di contatto con le amministrazioni, accesso ai servizi, tempestività di risposta, professionalità del personale degli uffici pubblici.

Questo profondo processo di cambiamento della Pubblica Amministrazione non è attuabile esclusivamente nei termini normativi delle leggi e dei decreti. Le riforme si possono attuare solo con il contributo di milioni di amministratori e dipendenti pubblici. Evidentemente all'interno di questa categoria esistono resistenze culturali legate a vecchie logiche burocratiche, formalistiche, giuridiche e centralistiche. Lo sforzo di cambiare cultura, di acquisire e diffondere la mentalità dell'innovazione tecnologica e organizzativa, della semplificazione, della qualità dei servizi e delle prestazioni, della soddisfazione dei cittadini e delle imprese, della valorizzazione delle professionalità e del merito, è un compito di tutti, dal dirigente all'impiegato pubblico. Questo pone fortemente l'accento sulle professionalità e le competenze.

Nel recente passato l'innovazione si traduceva soprattutto nell'acquisizione di dotazioni tecnologiche (hardware e software applicativo, cioè reti di personal computer, PC, con programmi gestionali e di produttività individuale); oggi queste infrastrutture possono essere considerate semplici *commodities*, ovvero beni ICT di larga diffusione dei quali la PA non può fare a meno ma che non costituiscono, di per sé, un diretto vantaggio competitivo.

Nell'ultimo decennio con l'avvento di Internet e con la conseguente spinta alla globalizzazione che la rete ha impresso, l'innovazione tecnologi-

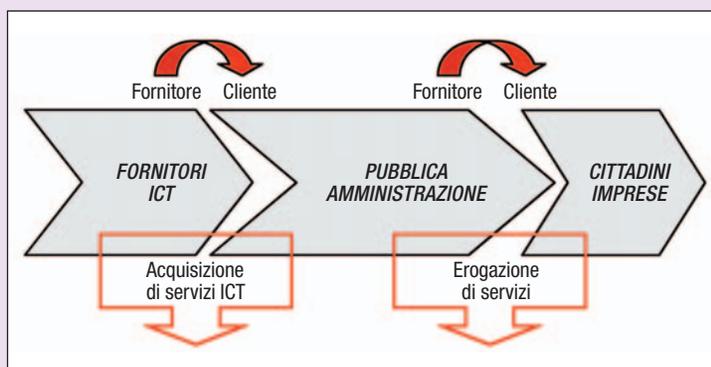
ca ha assunto una valenza assai più pervasiva investendo ogni dimensione della PA: i processi interni, la progettazione e produzione, la distribuzione, la stessa capacità di costruire relazioni (con i cittadini, con le imprese e con i fornitori).

La PA deve essere in grado di governare il proprio fabbisogno tecnologico, costruendo le relazioni logiche ottimali tra la propria missione istituzionale, la struttura organizzativa, i processi produttivi e le opportunità che il mercato ICT offre. Se la PA restasse priva degli strumenti per interpretare e guidare le scelte tecnologiche, rinunciando a governare pienamente la propria crescita.

Nella costruzione del valore per la PA la relazione che si instaura con il fornitore ICT è centrale, sia sul piano dei processi e dell'organizzazione interni sia su quello dell'erogazione di servizi da parte della PA ai cittadini e alle imprese. La qualità è l'attributo principale della costruzione del valore sull'intera catena cliente-fornitore rappresentabile come un unico flusso di processi scomponibili in specifiche attività da gestire con qualità.

La posizione centrale della PA fa sì che essa giochi simultaneamente entrambi i ruoli di cliente e di fornitore: cliente di un *outsourcer* di beni e servizi ICT e fornitore dei propri clienti finali. La qualità della relazione con il fornitore ICT si riflette sulla qualità dei beni o servizi che la PA fornisce ai propri clienti (Figura 1).

La PA che governa le proprie scelte tecnologiche agisce sulla relazione con il fornitore ICT e, coerentemente, razionalizza i processi produttivi e distributivi, migliora l'efficienza, ottimizza i costi: in breve, tiene saldamente in mano le leve della crescita. La gestione della relazione tra PA e fornitore ICT – intesa come la capacità della PA di fare leva sulle tecnologie per la propria



**FIGURA 1**  
Catena cliente/fornitore per la creazione del valore per cittadini ed imprese

crescita – è la parte più consistente della *governance*, cioè del sistema di organizzazione della PA sull'ICT proprio perché si trova al centro del processo della qualità.

### 3. COMPETENZE

L'estrema criticità dell'attività di governo dell'ICT si evidenzia semplicemente scorrendo l'elenco dei principali compiti da svolgere:

- modulare l'erogazione dei servizi tecnologici in funzione degli obiettivi istituzionali;
- fornire servizi di qualità ai cittadini e alle imprese;
- valutare l'impatto che gli investimenti tecnologici generano;
- ottimizzare l'efficacia dell'introduzione di nuove tecnologie attraverso la reingegnerizzazione e il monitoraggio dei processi e il *change management*;
- garantire la continuità e la disponibilità di servizi basati su complicate architetture tecnologiche e l'integrazione di vasti portfolio applicativi, attraverso oculati studi di fattibilità tecnica e la selezione di qualificati fornitori ICT;
- tenere sotto controllo la spesa in informatica, che, per le amministrazioni, rappresenta un dato in continua crescita a dimostrazione di un'aumentata consapevolezza sulle opportunità che offre l'innovazione tecnologica;
- garantire l'integrità e la riservatezza dei dati trattati dai propri sistemi informativi sia attraverso l'uso della tecnologia, sia attraverso la predisposizione di specifiche procedure organizzative.

Le Pubbliche Amministrazioni sono chiamate a perseguire tali obiettivi avvalendosi di una struttura organizzativa che tipicamente ha anche il compito di gestire il sistema informativo, le interazioni con le altre strutture interne, la relazione e il coordinamento con le altre Amministrazioni, la gestione dei fornitori ICT.

La scelta strategica da parte di un'Amministrazione di operare in regime di *outsourcing* ha rilevanti riflessi sull'organizzazione interna che deve occuparsi di nuove attività derivanti dal fatto che alcuni compiti, svolti in precedenza all'interno dell'Amministrazione, sono affidati a un fornitore esterno. Ciò impatta, a differenti livelli, sui ruoli, sui compiti e sulle responsabilità delle professionalità che operano nel settore con ricadute riferibili all'organigramma, alla

pianta organica, al funzionigramma, ai meccanismi operativi e al sistema dei valori condivisi all'interno dell'organizzazione.

Occorre elaborare, quindi, strategie per modificare profondamente la funzione ICT interna anche in termini riconversione o allocazione delle risorse umane allo scopo di mantenere il controllo sui servizi ICT dati in *outsourcing* allo scopo di dotarla di adeguati strumenti e competenze.

### 4. PROFILI PROFESSIONALI ICT

Le Linee guida concentrano l'attenzione sulla delega operativa relativa alle forniture ICT che una Amministrazione concede a un fornitore quando demanda all'esterno le attività di esercizio del sistema informativo, cosa di cui si parla in termini di esternalizzazione (*outsourcing*). Questo passaggio di attività dal cliente al fornitore è estremamente delicato per i rischi strategici, organizzativi, economici e operativi che comporta, e, al tempo stesso, ha delle fortissime rilevanze in termini di impatto organizzativo sulla struttura del cliente.

L'esternalizzazione di attività operative o altamente specialistiche cambia la natura e la valenza delle competenze interne, poiché la delega operativa che una Amministrazione concede a un fornitore, relativamente alla fornitura di beni e servizi ICT, ha come immediata conseguenza la necessità da parte dell'Amministrazione Pubblica di dotarsi di competenze di governo delle realizzazioni e dei contratti ICT, adeguate alla complessità del rapporto con il fornitore.

D'altra parte, dal punto di vista del cliente, specificatamente nel caso in cui esso sia una Pubblica Amministrazione, un problema che si pone al centro dell'attenzione nell'utilizzo di una strategia di acquisizione basata sull'esternalizzazione dei servizi ICT è che le competenze acquisite tramite il contratto, siano adeguate ai compiti ad esse affidati.

Due nuovi manuali delle Linee Guida, rispettivamente intitolati "Organizzazione della funzione ICT" e "Dizionario dei profili professionali ICT" forniscono un contributo operativo alla definizione delle competenze necessarie.

Il modello europeo EUCIP di descrizione dei profili professionali ICT rappresenta la base di partenza: sia per definire la professionalità delle risorse umane all'interno dell'Amministrazione; sia per supportare l'Amministrazione in ca-

Profili Professionali EUCIP	Classi di Fornitura CNIPA																					
	Responsabile commerciale	Revisore di Sistemi Informativi	Resp. di Sistemi Informativi	Cons. vendite e app. di TI	Cons. di Soluzioni Aziendali	Cons. di Logistica e Automazione	Analista di Business	Capoprogetto di Sistemi Inf.	Analista di Sistemi Informativi	Analista Programmatore	Tec. Collaudo e Integ. Sistemi	Exp. App. Web e Multimediali	Progettista di Sistemi Informativi	Progettista delle Telecomunicazioni	Consulente per la Sicurezza	Resp. Config. e Centro Dati	Resp. Basi di Dati	Responsabile di Rete	Formatore IT	Sup. Centro di Assistenza	Sistemista multipiattaforma	
Sviluppo e MEV di software ad hoc				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Personalizzazione e riuso di prodotti esistenti				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Sviluppo e MEV mediante soluzioni commerciali				X		X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Gestione applicativi e Basi Dati				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Manutenzione correttiva ed adeguativa (MAC)				X		X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Migrazione e conversione applicazioni				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Assistenza in remoto e in locale				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Formazione e addestramento				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Integrazione di prodotti software e basi dati				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Integrazione di sistemi e infrastrutture				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Servizi applicativi in modalità ASP				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Posta elettronica				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Posta elettronica certificata				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Servizi Internet				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Gestione contenuti WEB				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Certificazione delle firma digitale				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Gestione di Carte per l'Accesso ai Servizi				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Sviluppo Reti				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Gestione e manutenzione reti				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Sviluppo sistemi				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Gestione sistemi				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Manutenzione sistemi				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Gestione della sicurezza logica				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Gestione della sicurezza fisica				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Continuità operativa				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Trattamento documentale e acquisizione dati				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Gestione elettronica dei documenti				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Controllo dei livelli di servizio				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Gestione e manutenzione delle postazioni di lavoro				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Consulenza				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Direzione lavori				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Misura della Customer Satisfaction				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Prodotti Hardware e Software				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Documentazione				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Gestione della Configurazione				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Assicurazione della Qualità				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Gestione e Processi Organizzativi				X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

TABELLA 1 Profili professionali utilizzati per le diverse tipologie di fornitura ICT

so di affidamento in *outsourcing* di specifiche funzioni ICT.

Nel manuale "Organizzazione della funzione ICT" viene fornito un modello organizzativo della funzione ICT interna ad un'Amministrazione senza proporre una rigida struttura organizzativa ma scomponendo la funzione ICT per aree di attività, descrivendone obiettivi e compiti, e rappresentandone le principali interazioni tra le stesse.

Sono inoltre trattati gli aspetti relativi alle risorse umane da allocare all'interno della struttura organizzativa individuando capacità, competenze ed esperienze necessarie a svolgere i diversi compiti precedentemente determinati.

L'utilizzo delle descrizioni dei profili professionali e, in particolare, delle competenze, può consentire di evidenziare eventuali carenze del personale addetto e di intervenire attraverso piani di formazione e training individuali, spostamento delle risorse allocate in attività non corrispondenti alle relative competenze, programmazione delle carriere.

Nel manuale "Dizionario dei profili professionali ICT" i profili professionali descritti aiutano nella stesura dei contratti di *outsourcing* per vincolare il fornitore all'utilizzo di professionalità adatte allo tipo di servizio da erogare riducendo così i rischi di degrado della fornitura.

Proprio per questo scopo all'interno del manuale i profili professionali ICT sono correlati alle tipologie di servizi ICT (denominate "classi di fornitura") tipicamente affidati a fornitori esterni, descritti nel manuale "Dizionario delle forniture ICT" in termini di attività, *deliverables* contrattuali e indicatori di qualità.

All'interno di ogni classe di fornitura i profili professionali di pertinenza sono associati alle attività che devono essere svolte per l'erogazione del servizio, dando un'indicazione di massima dell'impegno percentuale nell'attività. Queste informazioni sono utili, sia per stimare i costi d'acquisto del servizio che per dimensionare eventuali gruppi di lavoro interni (Tabella 1).

MARCO GENTILI, nato nel 1958, laurea con lode in Fisica (La Sapienza 1982), master in R&D Management (SDA Bocconi 1990) e Quality Management (EOQ 1996). 1983/86 Datamat, progettista di sistemi di office automation ed esperto di sw engineering. 1986/97 AED Group, dal 1991 Dirigente resp.: R&D, Education, Quality Assurance, Marketing; gestisce progetti IT nei settori Finance, Industry, Government, anche in ambito internazionale (Eureka e Esprit, Banca Mondiale degli Investimenti). 1997/oggi CNIPA (prima AIPA), Dirigente resp. Area "Metodologie per la qualità e per l'innovazione organizzativa". Esperto di ICT governante & management, ideatore, curatore e coautore delle "Linee guida sulla qualità dei beni e servizi ICT per la definizione ed il governo dei contratti della PA". Partecipa ai comitati direttivi delle associazioni, AICA, AICQ, CDTI, ISIPM, PMI.

E-mail: gentili@cnipa.it