Editoriale

Cari lettori,

con questo numero lascio la direzione di Mondo Digitale, incarico che ho avuto il privilegio e il piacere di ricoprire dalla nascita della rivista, ormai più di dieci anni fa.

Subentra nel compito Viola Schiaffonati, docente al Politecnico di Milano, che già collabora alla pubblicazione e alla quale faccio i più cordiali auguri di buon lavoro.

In questi ultimi dieci anni il mondo dell'ICT è cambiato in modo radicale, con una accelerazione straordinaria. Mondo Digitale ha testimoniato questa evoluzione in termini rigorosi, ma accessibili anche ai non specialisti, una linea editoriale che ha caratterizzato la rivista e riscosso il consenso dei lettori.

Colgo l'occasione per ringraziare tutti coloro che hanno contribuito a questo risultato: la Redazione, il Comitato Scientifico, gli Autori.

Buona lettura e cordiali auguri per il nuovo anno.

Franco Filippazzi

E' davvero un onore per me assumere, dal prossimo numero, la direzione di Mondo Digitale, una rivista che ho avuto il privilegio di conoscere negli anni prima come lettrice, poi come saltuaria collaboratrice e, solo recentemente, come membro del comitato scientifico.

Sono grata al direttore uscente, Franco Filippazzi, e al coordinatore del comitato scientifico, Giulio Occhini, per la fiducia che mi hanno accordato proponendomi di dirigere Mondo Digitale in un momento denso di rapidi cambiamenti e sfide per il futuro.

Mi auguro che le mie competenze e la mia pratica accademica fra "le due culture" possano rappresentare un punto di forza per Mondo Digitale che continua a contribuire alla diffusione di una cultura informatica intesa nel senso più ampio del termine e al di là degli aspetti meramente tecnici che, pur costituendo elemento portante della disciplina, non debbono farci dimenticare il suo respiro più ampio di impresa scientifica, culturale e umana.

Buona lettura e arrivederci a presto,

Viola Schiaffonati

Computer Quantistici

Alessandra Di Pierro - Oliver Morsch

Come tutte le grandi scoperte, il computer quantistico nasce da un'idea visionaria. Nel caso specifico, l'idea fu avanzata dal premio Nobel Richard P. Feynman che in uno dei suoi più famosi articoli lo suggerì come una possibile soluzione al problema: `Can physics be simulated by a universal computer?'. In questo articolo, racconteremo l'evoluzione della ricerca teorica e sperimentale in quell'area che oggi è nota come `computazione quantistica', dall'idea di Feynman ai nostri giorni, descrivendo i risultati ottenuti sia nell'ambito della realizzazione fisica del computer quantistico, sia riguardo ad aspetti più prettamente informatici relativi alla teoria della calcolabilità e degli algoritmi e complessità.

Keywords: Quantum Physics; Quantum computation; Quantum algorithms

1. Introduzione

Quando Charles Babbage intorno al 1830 ebbe l'idea di un dispositivo meccanico in grado di eseguire compiti generici non ristretti a puri calcoli matematici, il progresso tecnologico non aveva ancora messo a disposizione gli strumenti specifici necessari per realizzare il suo prototipo. Valvole e transistor arrivarono solo cent'anni dopo per permettere la costruzione del calcolatore programmabile che tanto aveva in comune con la Macchina Analitica di Babbage.

Ritornando ai nostri giorni, il computer è innegabilmente un'icona dell'era in cui viviamo, imprescindibile per le sue capacità e in continua evoluzione diventando ogni anno più veloce, più piccolo e più economico secondo un processo di crescita che sembra non avere limiti.

In una situazione così diversa dai tempi di Babbage dobbiamo tuttavia di nuovo supporre che le tecniche attualmente esistenti, seppur avanzatissime per le nostre conoscenze attuali, non siano sufficienti a realizzare quella nuova rivoluzione nei sistemi di computazione prospettata dalla *computazione quantistica*.

Come vedremo in questo articolo, la realizzazione di un computer quantistico avrebbe conseguenze pratiche di portata enorme. Un computer quantistico avrebbe infatti una velocità di calcolo che supera di ordini di grandezza quella realizzabile con i computer tradizionali; questo renderebbe possibile la soluzione di molti problemi che i computer odierni non possono risolvere in modo effettivo, come ad esempio la fattorizzazione di numeri grandi, la cui rilevanza nella crittografia è ben nota. Vedremo anche che l'interesse in questa nuova forma di computazione va ben oltre le applicazioni pratiche.

2. Teoria della computazione quantistica

Il computer quantistico non è semplicemente il prossimo passo nel processo evolutivo dei computer ma anche e, soprattutto, il rappresentante di un paradigma di computazione non classico il cui studio ha dato origine ad un nuovo settore della ricerca teorica in informatica e fisica che va sotto il nome di computazione quantistica. Ad iniziare questa linea di ricerca fu Richard Feynman che per primo si rese conto di un problema fondamentale dei computer classici: non sono in grado di simulare la realtà quantistica. Nel suo famoso articolo `Simulating Physics with Computers' [8], Feynman descrive il problema con estrema semplicità concludendo che solo con un computer quantistico sarebbe stata possibile una simulazione efficiente. Il punto cruciale è che i sistemi fisici quantistici esibiscono un comportamento probabilistico che non corrisponde a quello implementabile sui computer classici.

La computazione classica probabilistica acquistò enorme importanza in informatica soprattutto negli anni `70 del secolo scorso, dopo l'introduzione ad opera di Solovay e Strassen di un algoritmo randomizzato per determinare se un numero intero è primo o no. L'algoritmo, che usa un generatore di numeri casuali, dà una risposta corretta solo con una certa probabilità. Questo fu il primo algoritmo efficiente per risolvere il problema della primalità, per il quale in quegli anni non si conoscevano ancora soluzioni deterministiche[1].

Un algoritmo *probabilistico* usa la casualità (tipicamente il lancio di una moneta) per rappresentare l'incertezza di proseguire una computazione lungo una direzione o un'altra tra le tante possibili. Il modo in cui le scelte fatte ad ogni passo di computazione si combinano per determinare la probabilità del risultato finale è dettato dalla teoria classica della probabilità; in particolare la regola di Bayes stabilisce che la probabilità di un evento che si può verificare in due o più modi distinti è la somma delle probabilità di ciascun modo considerato

[1] *Un algoritmo deterministico che effettua il test di primalità in tempo polinomiale fu poi introdotto nel 2002 da Agrawal, Kayal e Saxena, tre ricercatori dell'Indian Institute of Technology, Kanpur*

separatamente. L'albero in Figura 1a esemplifica una computazione probabilistica con risultato ½ R + ½ B, dove R rappresenta il risultato raggiunto negli stati rossi 5 e 7 mentre B quello raggiunto negli stati blu 4 e 6.

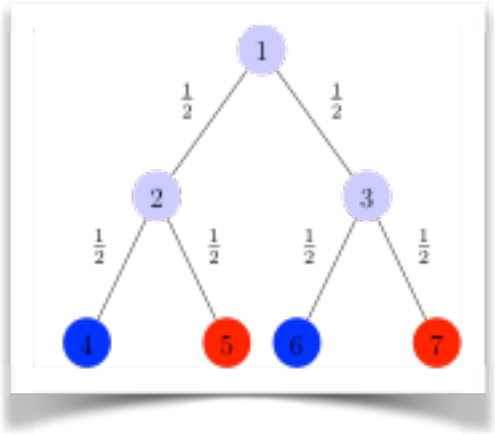


Figura 1a
Una computazione probabilistica in cui ogni cammino di esecuzione ha probabilità 1/4; poiché ciascuno dei due risultati viene raggiunto da due cammini diversi, entrambi i risultati si ottengono con la stessa probabilità 1/4+1/4=1/2.

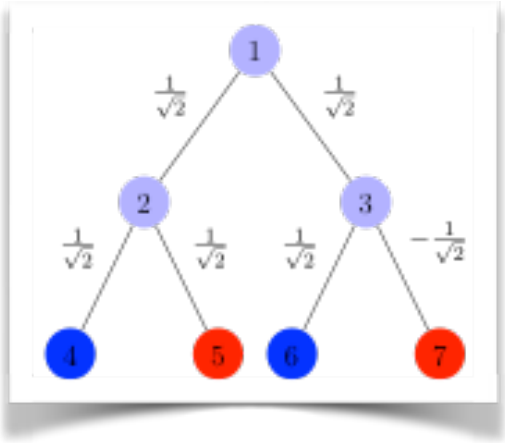


Figura 1b
Una computazione quantistica in cui il risultato blu ha ampiezza di probabilità 1/2+1/2, mentre il risultato rosso ha ampiezza di probabilità 1/2-1/2, da cui si ricava la probabilità 1 per il risultato blue e 0 per il risultato rosso.

L'elemento che rende la computazione quantistica nello stesso tempo più generale (essa comprende la computazione probabilistica come una particolare istanza) e distinta (essa consente nuovi modi di calcolo che non hanno alcuna controparte classica) deriva dalla considerazione di *numeri complessi* al posto dei numeri reali usati nella computazione probabilistica classica. Tali numeri, che al contrario delle probabilità possono essere anche negativi, stabiliscono *ampiezze di probabilità*; da queste si può ottenere la probabilità in senso classico di un certo evento (o cammino di esecuzione), calcolandone il quadrato del modulo. La regola di Bayes viene ora sostituita da un'altra regola che stabilisce come combinare ampiezze di probabilità e che, in onore del famoso fisico promotore della computazione quantistica, è chiamata in [14] regola di Feynman: l'ampiezza di probabilità di un evento che si può verificare in due o più modi distinguibili è la somma delle ampiezze di probabilità di ciascun modo considerato separatamente.

La conseguenza di questa sostituzione è che i diversi percorsi di una computazione possono ora *interferire* distruttivamente gli uni con gli altri. Questo si verifica per esempio quando le ampiezze associate ai due percorsi hanno

modulo uguale ma segno opposto. Una tale situazione è esemplificata dalla computazione quantistica rappresentata in Figura 1b, dove i due numeri complessi corrispondenti alle ampiezze di probabilità per le due derivazioni del risultato rosso (cioè $1/\sqrt{2}\times 1/\sqrt{2}$ e $1/\sqrt{2}\times(-1/\sqrt{2})$) si annullano dando probabilità zero a questo risultato.

Possiamo dunque aspettarci che la nozione di probabilità quantistica abbia grande influenza sul modo di costruire e sulle possibilità computazionali degli algoritmi quantistici. Proprio per l'uso del nuovo concetto di probabilità, questi algoritmi si possono descrivere come algoritmi randomizzati con un generatore quantistico di probabilità al posto del generatore di numeri `pseudo-casuali' tipicamente usato in quelli classici. Usando la fisica quantistica, la generazione di numeri `genuinamente random' si può ottenere oggi mediante un semplice dispositivo quantistico che opera su fotoni (particelle di luce) mediante specchi semi-argentati (*beam-splitter*) e *detector*. Dispositivi di questa natura (e loro varianti) sono stati realizzati e messi sul mercato dalla ditta IDQ (http://www.idquantique.com) e attualmente possono essere acquistati online a prezzi che variano dai 1000 ai 2500 euro.

Da un punto di vista più propriamente teorico degli algoritmi eseguibili su un computer quantistico, questo dispositivo implementa un'operazione chiamata operazione di *Hadamard* (e indicata con H) che compare in tutti gli algoritmi quantistici rappresentandone, come risulterà chiaro in seguito, un elemento imprescindibile.

2.1 Computazione quantistica e multiverso

Capire il funzionamento di un computer classico è relativamente semplice: esso è un sistema fisico che obbedisce alle leggi della fisica classica, cioè le leggi che regolano la nostra esistenza nell'universo macroscopico in cui viviamo e che ci sono quindi familiari. Tuttavia, secondo una particolare interpretazione della teoria quantistica, queste leggi rappresentano solo una particolare istanza o approssimazione della realtà fisica nella quale siamo immersi e che va al di là dell'universo contingente di cui abbiamo esperienza diretta. Tale interpretazione fu proposta nel 1957 da un fisico dell'Università di Princeton, Hugh Everett III, ed è oggi nota col nome di *interpretazione a molti mondi*. Essa spiega i fenomeni quantistici osservati sperimentalmente sulla base dell'esistenza di un'infinità di universi che coesistono mantenendo ciascuno la propria individualità. La sovrapposizione, dunque, altro non è che la visione completa di un oggetto in tutti i suoi possibili stati in tutti gli universi che compongono la realtà fisica o, con un termine coniato da David Deutsch (uno dei promotori della computazione quantistica e grande sostenitore dell'interpretazione di Everett), il *multiverso*.

Nel suo libro `The Fabric of Reality' [6], Deutsch ci spiega tutte le implicazioni di una tale visione del mondo e come la teoria classica della computazione, che possiamo identificare con la teoria delle Macchine di Turing, si può considerare un'approssimazione della teoria quantistica della computazione, proprio come la fisica classica è un'approssimazione della fisica quantistica. L'approssimazione classica della teoria della computazione è comunque sufficiente a descrivere quello che i computer oggi disponibili sono in grado di fare. Quello che essi non sono in grado di fare dipende dal fatto che, nel definire il loro funzionamento e modi di calcolo, si trascura la possibile interazione con gli altri universi. Quest'ultima si rivela negli esperimenti fisici come interferenza quantistica e dà

luogo a comportamenti osservabili ben lontani dalla nostra comune esperienza di vita quotidiana. Analogamente, un computer quantistico potrebbe sfruttare l'interazione tra le diverse computazioni parallele che avvengono nel multiverso per effettuare operazioni che nessun computer classico sarebbe in grado di svolgere.

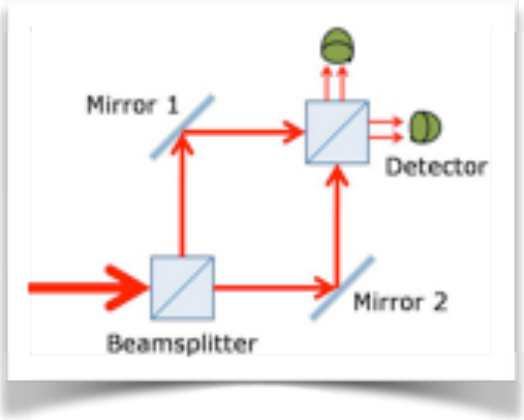
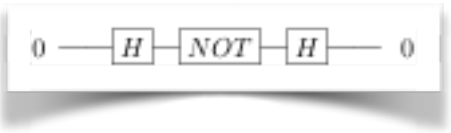


Figura 2a
Un esperimento che rivela l'interferenza quantistica: un singolo fotone passa attraverso uno specchio semi-argentato (beam splitter) che ne cambia la direzione da orizzontale a una sovrapposizione di orizzontale e verticale. Dopo aver attraversato nuovamente il beam splitter il fotone viene rilevato solo dal detector sulla direzione orizzontale

Figura 2b
Circuito quantistico che esegue una computazione equivalente all'esperimento del fotone e del beam splitter.



La corrispondenza tra sistema fisico e computazione quantistica è molto forte. Per rendersene conto basta pensare a un esperimento come ad un processo di calcolo (e viceversa). Si prenda ad esempio il famoso esperimento raffigurato in Figura 2a e consistente in un fotone che viene lanciato contro uno specchio semi-opaco o *beam splitter*. L'effetto del *beam splitter* è quello di creare uno stato in cui il fotone potrà essere rilevato con il 50% delle probabilità nella direzione iniziale e con il 50% nella direzione opposta, cioè il fotone si troverà in metà degli universi a viaggiare in un verso (direzione orizzontale nella figura) e nell'altra metà nel verso opposto (direzione verticale nella figura). Come già accennato, l'effetto del *beam splitter* è quello dell'applicazione dell'operazione di Hadamard allo stato computazionale rappresentato dal fotone. Questo è rappresentato nel circuito in Figura 2b dal qubit 0 in input, il quale viene sottoposto alla componente del circuito che implementa l'operazione H. La presenza dei due specchi in Figura 2a esprime il fatto che il fotone deve poter rimbalzare in tutti gli universi in cui si viene a trovare per effetto del *beam splitter*. L'effetto dello specchio è dunque di invertire la direzione del fotone, rimandandolo indietro verso il *beam splitter*. L'azione dello specchio corrisponde quindi ad applicare un'operazione di NOT allo stato del qubit ottenuto dopo l'applicazione di H (cfr. Figura 2b). Dopo aver incontrato lo specchio, il fotone ritorna quindi al *beam splitter* e a questo punto si osserva che i due *detector* posti nella due direzioni segnalano la presenza del fotone solo nella direzione orizzontale: il *beam splitter* ha agito ora come `*beam joiner*' riportando la direzione del fotone ad essere solo quella

iniziale. Analogamente, la seconda applicazione di H nel circuito di Figura 2b riporta la computazione nello stato di partenza realizzando di fatto una computazione analoga all'esperimento in Figura 2a (ed equivalente a quella rappresentata nell'albero in Figura 1b). Il risultato inaspettato dell'annullamento del secondo risultato creato da Hadamard (o il fatto di non rilevare la presenza del fotone nella direzione verticale creatasi dopo l'impatto con il *beam splitter*) è l'evidenza dell'interazione avvenuta tra le controparti dello stato del fotone (visto come oggetto multiversale) nei due universi corrispondenti alla direzione verticale e a quella orizzontale: esse hanno interferito distruttivamente le une sulle altre riportando il fotone ad assumere la stessa direzione iniziale, cioè quella orizzontale, in tutti gli universi. La computazione corrispondente, d'altra parte, è un modo non classico per calcolare la funzione identità.

2.2 Problemi computazionali

La sfida tra computer quantistici e computer classici si svolge sul piano della complessità computazionale, cioè la teoria che si occupa di stabilire le risorse (tipicamente tempo o spazio) necessarie per risolvere un dato problema mediante un dato algoritmo.

Molti problemi che si presentano nella vita reale possono essere formulati in modo astratto come problemi di ricerca: si cerca tra tutti i possibili candidati quello che soddisfa un certo criterio che lo caratterizza come soluzione al problema dato [4]. In questa formulazione, un metodo, o algoritmo, per risolvere un problema computazionale è migliore di un altro se è in grado di esplorare l'intero spazio di ricerca in modo più *efficiente* dell'altro. Per misurare l'efficienza di un algoritmo si guarda alla crescita asintotica del tempo impiegato (ad es. numero di passi dell'algoritmo) in funzione della dimensione n dell'input (ad es. numero di variabili utilizzate, numero di bit necessari a codificare ogni istanza del problema, ecc.). La teoria della complessità identifica un algoritmo efficiente con uno per cui tale funzione è polinomiale in n, come ad esempio le funzioni n, n^2, n^3, ecc. La classe di tutti i problemi per cui esiste un algoritmo efficiente è notoriamente la classe **P** (= Polynomial-time). In pratica, dopo l'avvento degli algoritmi randomizzati, la classe che identifica i problemi con soluzione efficiente è più realisticamente identificabile con **BPP** (*Bounded-error Probabilistic Polynomial-time*), cioè la classe dei problemi che possono essere risolti da algoritmi probabilistici in tempo polinomiale e con una probabilità di errore minore di un 1/3. Risolvere un problema di ricerca in modo efficiente è quindi un compito non banale se si tiene conto del fatto che tipicamente, per un input espresso su n bit, lo spazio di ricerca contiene un numero di candidati dell'ordine di 2^n; di conseguenza, un semplice algoritmo che fa la ricerca esaustiva in questo spazio avrà necessariamente una complessità limitata superiormente da una funzione esponenziale in n. Questo significa, per esempio, che se le nostre istanze si potessero codificare con n=64 bit, la ricerca esaustiva di tutte le possibili soluzioni potrebbe richiedere fino a 18.446.744.073.709.551.615 test!

Classicamente, la possibilità di progettare algoritmi efficienti dipende da quanto il problema si presta ad una strutturazione conveniente dello spazio delle possibili soluzioni; ad esempio, se quest'ultime si possono disporre ai nodi di un albero binario secondo un criterio discriminante tra soluzioni e non, allora la ricerca procederà velocemente potendo scartare ad ogni passo una metà dei candidati superstiti. In molti casi, tuttavia, l'unico criterio di ricerca risulta quello di controllare i candidati uno per uno, con conseguente necessità di esplorare, nel

peggiore dei casi, l'intero spazio di ricerca. Questa categoria di problemi computazionali è chiamata in teoria della complessità la classe dei problemi **NP-completi** o *difficili*. Esponenti ben noti di questa classe sono il problema del commesso viaggiatore (TSP), il problema della soddisfacibilità (SAT) e il problema dello zaino (Knapsac problem). **NP** sta per *Nondeterministic Polynomial-time*, nome scelto per indicare che una soluzione ad ogni problema di ricerca si può sempre trovare in tempo polinomiale purchè si usi un algoritmo non deterministico, cioè un algoritmo ideale che sfrutta la potenza di un numero indeterminato di macchine parallele per cercare la soluzione e che una volta trovata può verificarla in modo efficiente. L'aggettivo "completi" si riferisce alla proprietà di questi problemi di poter rappresentare qualsiasi problema in NP in modo che ogni soluzione per un problema NP-completo può essere adattata in modo efficiente per risolvere un qualsiasi problema in NP (ma non viceversa). Le relazioni tra queste classi di complessità classiche sono rappresentate nel diagramma A.

La computazione quantistica e, in particolare, l'introduzione di un algoritmo quantistico che risolve in modo efficiente un problema classicamente in NP ha modificato questo scenario rendendo necessaria l'introduzione di una nuova classe di complessità, cioè la classe **BQP**, che descriveremo nel seguito insieme al più famoso algoritmo che la rappresenta, cioè appunto l'algoritmo di Shor per la fattorizzazione di un numero. La collocazione della nuova classe all'interno dello scenario classico può essere approssimativamente raffigurata come nel diagramma B, anche se la relazione tra BQP e NP non è attualmente nota.

La domanda cruciale che da decenni impegna gli algoritmisti e che ancora non ha trovato risposta è: esistono algoritmi efficienti per i problemi **NP-completi** oppure è vero che **P≠ NP**?

La computazione quantistica non è servita finora a dare una risposta a questa domanda. Infatti, sebbene l'algoritmo di Shor riduca drasticamente il tempo di esecuzione richiesto dal più potente computer classico oggi esistente per risolvere il problema della fattorizzazione, questo problema non rappresenta la complessità di tutti i problemi in NP (non è un problema NP-completo) e risolverlo efficientemente non fornisce nessuna evidenza al fatto che NP possa collassare a P. Un altro esempio dove il computer quantistico avrebbe un vantaggio sugli odierni processori è la simulazione del comportamento degli atomi e delle molecole. Tuttavia, ancora una volta, questi risultati non hanno conseguenze sulla famosa questione **P≠ NP**? e, in generale, nonostante gli innumerevoli progressi nello studio della computazione quantistica, non esiste al momento alcuna evidenza che la potenza di calcolo del computer quantistico potrà essere dirimente nello sforzo di risolverla.

Se quindi nei prossimi decenni gli studi di fisici e informatici porteranno alla realizzazione del computer quantistico e dell'informatica quantistica, le ricadute saranno innanzitutto di natura pratica, provocando cambiamenti rivoluzionari almeno nei campi della crittografia, della nanotecnologia e della medicina.

Se invece falliranno, le conseguenze saranno ancora più interessanti sotto molti punti di vista; ad esempio questo potrebbe evidenziare errori in una teoria, quella quantistica, la cui validità è rimasta senza rivali per ormai un secolo, perché "*... not only can physics determine what computers can do, but what computers can do, in turn, will define the ultimate nature of physical laws*" (Rolf Landauer [11]).

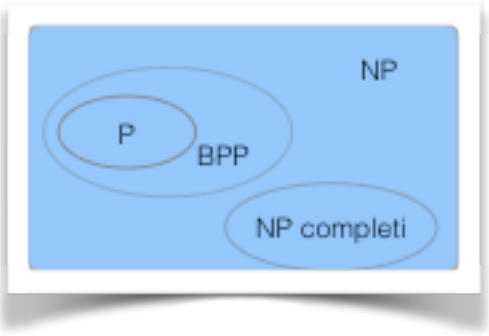


Diagramma A

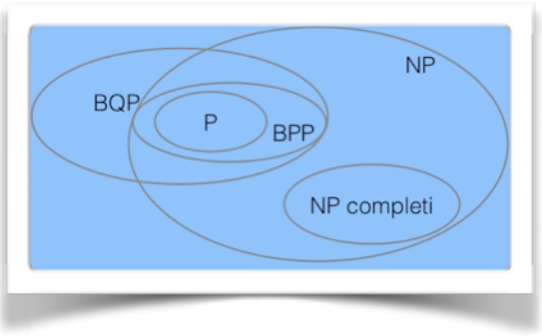


Diagramma B

2.3 Il qubit

L'abilità di manipolare ampiezze di probabilità è un aspetto nuovo che non può essere catturato con il semplice lancio di una moneta. Poiché un bit di informazione non è sufficiente a rappresentare questa nuova abilità, Benjamin Schumacher, fisico teorico e studioso della teoria dell'informazione quantistica al Kenyon College (Ohio) inventò una nuova parola, *qubit*, per indicare l'unità di informazione che rispetta la regola di Feynman al posto della regola di Bayes della teoria classica della probabilità. Il qubit corrisponde al più semplice tra tutti i sistemi fisici quantistici. Per capire la natura di questa entità e la differenza con la sua istanza classica, il bit, dobbiamo fare riferimento alle leggi che regolano il comportamento e l'evoluzione di un sistema fisico reale e di conseguenza l'elaborazione dell'informazione in esso contenuta. Dal punto di vista computazionale i postulati della meccanica quantistica ci permettono di allargare il campo d'azione di un calcolatore ad uno spazio che oltre alle dimensioni classiche corrispondenti alle sequenze binarie (registri di bit) ne include anche tutte le infinite combinazioni (principio di sovrapposizione degli stati) con le loro interazioni non classiche (fenomeno dell'interferenza) e gli effetti sui risultati finali (principio di misurazione).

Sovrapposizione

Le architetture basate sui microchip standard dei computer odierni rispettano rigorosamente la dicotomia del bit classico, cioè il loro funzionamento dipende in modo essenziale dalla codifica binaria dell'informazione (un bit può solo essere 0 oppure 1). La potenziale superiorità di un computer quantistico in quanto a capacità di calcolo viene dal fatto che esso opera su informazione codificata in qubit, cioè oggetti che possono assumere un'infinità di stati oltre agli stati 0 e 1 del bit. Infatti, i postulati della meccanica quantistica identificano il qubit con un vettore v nello spazio complesso bidimensionale C^2 (spazio di Hilbert), cioè con un oggetto

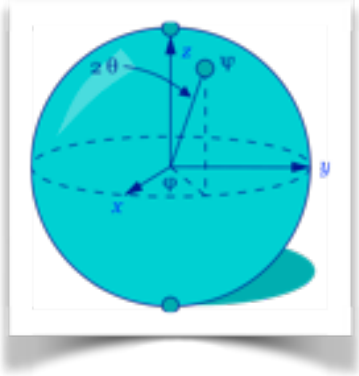


Figura 3
Sfera di Bloch. Il punto ψ sulla sfera rappresenta lo stato α0+β1 di un qubit. Le sue coordinate sferiche (φ, 2θ) sono codificate nelle ampiezze α e β (e viceversa).

della forma v=α0+β1, dove gli scalari α e β sono numeri complessi che esprimono la percentuale di probabilità che lo stato risulti essere effettivamente 0 e 1, rispettivamente. Questi stati intermedi del qubit, che implicano una coesistenza degli stati classici 0 e 1 in determinate proporzioni, vengono chiamati *sovrapposizioni* di stati e possono essere interpretati come situazioni di incertezza sullo stato *interno* del qubit sul cui valore non abbiamo garanzie assolute ma solo una probabilità che questo possa essere 0 oppure 1. Sull'interpretazione del principio di sovrapposizione non c'è tuttavia una tesi universalmente accettata come valida. Come verrà chiarito in seguito, il problema è strettamente legato al problema della misurazione sul quale il dibattito, iniziato alla nascita della teoria quantistica, continua a dividere i fisici teorici assumendo inevitabilmente risvolti filosofici che non affronteremo in questa sede.

Una rappresentazione del qubit che aiuta ad avere un'intuizione visiva di questa entità è la sua identificazione con i punti sulla superficie di una sfera unitaria nello spazio reale a tre dimensioni, nota come la sfera di Bloch (Figura 3) dal nome del fisico svizzero Felix Bloch che stabilì tale corrispondenza.

I punti corrispondenti al polo nord e al polo sud sono tipicamente associati ai due stati corrispondenti agli stati classici[2] 0 e 1, ma oltre a questi ogni altro punto della superficie sferica rappresenta un possibile stato del qubit [9,15]. Questa rappresentazione permette di visualizzare qualsiasi operazione su un qubit come una rotazione di un punto sulla sfera di Bloch. Così come una rotazione può essere applicata al contrario per riportare il punto nella sua posizione originale, un'operazione sul qubit può essere invertita in modo da riportare il qubit dallo stato finale nello stato iniziale, annullando di fatto il suo effetto. Questa *reversibilità* computazionale caratterizza la computazione quantistica e riflette il modo in cui un sistema fisico evolve nel tempo secondo i postulati della meccanica quantistica.

Il problema della misurazione quantistica

Nell'esperimento del *beam-splitte*r rappresentato in Figura 2a, la presenza del *detector* corrisponde ad una misurazione dello stato finale del qubit sottoposto alla computazione in Figura 2b. L'effetto che sperimentalmente si osserva dopo la misurazione di uno stato quantistico è un effetto distruttivo sulla *coerenza* del sistema che fa *collassare* una sovrapposizione in uno stato classico.

Il concetto di sistema coerente si riferisce all'interazione del sistema con un'entità esterna, come ad esempio uno strumento di misura, che determina una fuoriuscita (di parte) dell'informazione contenuta in esso. Questo processo, noto come decoerenza, determina inevitabilmente per le leggi della meccanica quantistica un cambiamento del sistema stesso e, come sarà chiarito in seguito, rappresenta uno dei più grossi ostacoli per l'implementazione pratica della computazione quantistica e la realizzazione di un computer quantistico *general-purpose*.

Il collasso di uno stato quantistico per effetto di una misurazione rappresenta un fenomeno la cui spiegazione diede luogo ad accesi dibattiti sin dalla nascita della teoria quantistica negli anni `20 del secolo scorso, facendo emergere soluzioni

---

[2] *Questa è la base standard dello spazio degli stati di un qubit, ma una qualsiasi coppia di punti antipodali sulla sfera sarebbe una scelta altrettanto adeguata.*

varie e contrastanti ancora oggi in discussione[3]. John von Neumann introdusse il primo trattamento assiomatico rigoroso della meccanica quantistica nel 1955, intervenendo in maniera decisiva sul problema della misurazione e fornendo una spiegazione ai vari paradossi che erano stati introdotti per sostenere l'inadeguatezza della teoria quantistica. Secondo la formalizzazione di von Neumann il processo di misurazione avviene in due fasi. Nella prima fase l'operatore che rappresenta l'osservabile (cioè la proprietà del sistema che si intende misurare) viene applicato allo stato del sistema. In una seconda fase avviene la ``riduzione di stato'', cioè il passaggio dallo stato di sovrapposizione coerente allo stato corrispondente ad uno dei risultati osservabili (corrispondenti agli autovettori dell'operatore lineare che rappresenta l'osservabile). Questa riduzione è nondeterministica e conseguentemente non c'è modo di prevedere quale dei risultati sarà ottenuto prima che il processo di misurazione abbia termine. In altre parole, per un osservabile non è mai possibile stabilire in maniera definita il valore che verrà misurato. La meccanica quantistica fornisce tuttavia delle informazioni statistiche sui possibili risultati di una misurazione secondo quella che è nota come l'*interpretazione statistica di Born*[4]. Attraverso misurazioni fatte su copie del sistema opportunamente preparate, è possibile stabilire la distribuzione probabilistica dei risultati. Il significato di probabilità di un risultato va inteso secondo l'interpretazione data in teoria delle probabilità come frequenza relativa: la probabilità di un risultato è il rapporto tra il numero delle volte che l'esperimento ha successo (cioè si ottiene quel risultato) e il numero totale degli esperimenti fatti, purché si ripeta l'esperimento un numero sufficientemente grande di volte.

2.4 Deutsch, parallelismo e interferenza

Il principio di sovrapposizione implica un enorme potenziale di capacità di calcolo. In particolare, la possibilità di identificare lo stato di una computazione con uno tra gli infiniti vettori dello spazio di Hilbert, cioè con una sovrapposizione qualsiasi di due stati classici, permette di codificare in un qubit molta più informazione di quella che può essere memorizzata in un singolo bit: tutti i numeri complessi nel primo caso, al posto dei soli 0 e 1 nel secondo. È dunque lecito pensare che, dal momento che operare sullo stato di un qubit corrisponde a operare contemporaneamente sia su 0 che su 1, la quantità di passi necessari per effettuare una computazione si possa ridurre notevolmente, tanto più quanto più grande è il numero n di qubit utilizzati: lavorare con n qubit significa in pratica considerare uno spazio di computazione infinito di dimensioni 2^n. In termini di complessità algoritmica questo significa passare da una complessità asintoticamente esponenziale ad una polinomiale in n. Sfruttare le potenzialità offerte dal parallelismo quantistico non è tuttavia immediato perché operare contemporaneamente su tutti i possibili input non produce direttamente tutti i possibili output ma solo una sovrapposizione di essi; sarà poi necessario effettuare una misurazione e quindi la selezione probabilistica di uno solo dei

---

[3] *Per un approfondimento su questo dibattito si può consultare il testo Quantum Theory and Measurement, Wheeler and Zurek, Princeton University Press, 1983, che contiene articoli originali e commenti sul problema della misurazione.*

[4] *Max Born, fisico e matematico tedesco, fu tra i fondatori della teoria quantistica e premio Nobel per la fisica nel 1954.*

possibili risultati. Il parallelismo deve essere quindi sfruttato in modo adeguato nella costruzione di algoritmi quantistici che siano più efficienti di quelli classici.

Un algoritmo che usa il parallelismo quantistico si può vedere come un insieme di computazioni classiche che si svolgono in parallelo su una molteplicità di input, tipicamente tutte le possibili configurazioni iniziali per un dato problema. Tuttavia, questo è solo un passo preliminare verso la costruzione di un algoritmo quantistico. Il passo successivo è identificare il processo di computazione vera e propria che interessa lo stato in sovrapposizione cioè quello che avviene quando il sistema si trova in uno stato di coerenza.

Perché una computazione quantistica possa aver luogo è infatti necessario che il sistema sia coerente. Il modo in cui essa avviene dipende dal tipo di problema che si intende risolvere. I risultati che si sono ottenuti fino ad ora dimostrano che l'unica tecnica algoritmica in grado di abbattere la complessità esponenziale di alcuni problemi considerati classicamente in NP è quella che sfrutta il parallelismo combinato con l'interferenza quantistica. Questa tecnica altro non è che la traduzione in termini computazionali dell'esperimento del fotone e del *beam splitter*

(cfr. Paragrafo 2.1). Verso la metà degli anni '80 del secolo scorso quando ancora non c'erano evidenze su come il parallelismo quantistico potesse essere usato per ottenere un vantaggio computazionale, David Deutsch descrisse un problema che, nonostante la sua semplicità, risultava intrattabile con un computer classico. Deutsch è uno dei pionieri della computazione quantistica e fu probabilmente il primo a mettere in atto l'idea di Feynman descrivendo in [5] un insieme di circuiti universale per la computazione quantistica[5]. L'algoritmo da lui ideato, noto appunto come l'algoritmo di Deutsch[6], introduce una tecnica che verrà poi sviluppata in quella che oggi è nota come la Trasformata di Fourier Quantistica ed è l'ingrediente fondamentale di tutti gli algoritmi quantistici ad oggi conosciuti che esibiscono un vantaggio esponenziale rispetto agli omologhi classici.

Il problema di Deutsch si può descrivere come quello di stabilire se una data funzione booleana f: {0,1} {0,1} darà lo stesso risultato su ogni input oppure risultati distinti su input distinti. La funzione f ci viene data attraverso un oracolo, cioè una scatola nera di cui non conosciamo il funzionamento interno ma che possiamo solo interrogare sui possibili input.

Chiaramente il problema, generalizzato a funzioni booleane di n bit, si può formulare come un problema di ricerca su uno spazio di dimensione 2^n e risulta

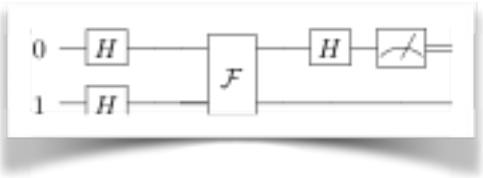


Figura 4
Un circuito quantistico che realizza l'algoritmo di Deutsch. Le linee singole rappresentano qubit, mentre la doppia linea dopo la misurazione rappresenta un bit classico (il risultato).

[5] *A.C. Yaho dimostrò nel 1993 che questi circuiti potevano simulare la macchina di Turing quantistica universale a qualsiasi grado di accuratezza.*

[6] *L'algoritmo fu poi migliorato nel 1998 da Richard Cleve, Artur Ekert, Chiara Macchiavello e Michele Mosca.*

classicamente un problema NP: sarà necessario nel caso più sfortunato un numero di interrogazioni esponenziale in n per poter stabilire con certezza che tipo di funzione è f. Il circuito in Figura 4 è la soluzione proposta da Deutsch che invece risolve il problema con una sola invocazione dell'oracolo. Il circuito opera su due qubit di cui il primo viene preparato nello stato 0 mentre il secondo, destinato a contenere il risultato, viene preparato in una opportuna sovrapposizione di stati ottenuta applicando H allo stato classico 1. L'operazione di Hadamard (*beam splitte*r) applicata a 0 produrrà una sovrapposizione di tutti gli input cosi che un'invocazione dell'oracolo produrrà una sovrapposizione di tutti i risultati (parallelismo). A questo punto entra in gioco l'interferenza realizzata mediante una nuova applicazione di H (*beam joiner*). Per effetto di quest'ultima l'informazione su f si ritroverà nell'output codificata nella fase relativa dello stato finale. Sarà ora sufficiente misurare il primo registro per ottenere la risposta al problema con probabilità 1: se il risultato sarà 1 allora la funzione è certamente costante, se invece sarà 0 allora vorrà dire che sicuramente $f(0) \neq f(1)$.

2.5 Shor e la classe BQP

Quando nei primi anni '80 del secolo scorso, Richard Feynman suggeriva nelle sue lezioni all'università di Caltech (*California Institute of Technology*) l'idea del computer quantistico, Peter Shor, allora studente in matematica che seguiva il corso di meccanica quantistica tenuto da Feynman, fece oro di quelle lezioni e circa tredici anni dopo arrivò ad una scoperta sensazionale. Nel 1994 Shor ideò un algoritmo che, usando un computer quantistico, risolveva il problema di trovare i fattori primi di un numero intero molto grande in maniera efficiente. Questa scoperta portò un grosso cambiamento nel campo della complessità computazionale facendo diventare `facile' un problema che fino ad allora si riteneva di complessità NP. D'altra parte, l'algoritmo prevedeva dei passi di computazione quantistica e quindi non poteva essere considerato al pari degli algoritmi polinomiali classici.

Si introdusse quindi una nuova classe di complessità destinata a contenere tutti i problemi che possono essere risolti con un algoritmo polinomiale *su un computer quantistico* a meno di un errore che si verifica con probabilità minore di 1/3. Questa classe, chiamata **BQP** (*Bounded-error Quantum Polynomial-time*), include anche i problemi che sono classicamente risolvibili con algoritmi deterministici o probabilistici in tempo polinomiale, cioè P e BPP (cfr. diagramma B). Tuttavia non è ancora chiara la sua relazione con i problemi difficili, cioè NP-completi. Una risposta a questa domanda darebbe una risposta anche al problema di stabilire l'effettiva superiorità o meno della computazione quantistica rispetto a quella classica, oltre a risolvere la ben nota questione **P = NP**?

Ritornando all'idea di Shor, la sua tecnica per ottenere lo *speed-up* esponenziale rispetto ai migliori algoritmi classici finora implementati per il problema della fattorizzazione è ancora una volta basata sul parallelismo e l'interferenza quantistica. Rispetto al problema di Deutsch la differenza è nell'uso consistente di risultati matematici (in particolare di teoria dei numeri e aritmetica modulare) per la formulazione del problema. L'ingrediente matematico costituisce la parte classica dell'algoritmo e si può riassumere nel seguente fatto: se N è il numero da fattorizzare, x un numero casuale tra 1 e N-1 e r l'ordine di x modulo N (cioè $x^r=1 \bmod N$), allora basta calcolare il massimo comune divisore tra N e $x^{r/2}-1$ e tra N e $x^{r/2}+1$ per ottenere con alta probabilità due fattori primi di N. La parte

“difficile” dell’algoritmo è quindi ridotta al calcolo dell’ordine r. Per far questo l’algoritmo di Shor usa un circuito quantistico molto simile al circuito di Deutsch, dove però l’oracolo è sostituito da un’opportuna funzione definita in modo da produrre nell’output una sovrapposizione corrispondente alla Trasformata di Fourier Quantistica (TFQ, una generalizzazione di Hadamard) del valore cercato. Questo potrà quindi essere ottenuto mediante un’applicazione della trasformazione inversa.

Lavorando su una sovrapposizione di tutti i possibili input, e sfruttando l’interferenza creata dalle operazioni quantistiche H e TFQ, l’algoritmo permette di calcolare i fattori di N ad un costo dominato da quello di operazioni classiche come il calcolo del massimo comun divisore, mentre il più efficiente algoritmo classico ad oggi noto per lo stesso problema, il “*number filed sieve*”, o crivello del campo di numeri, ha complessità superpolinomiale.

Per avere un’idea di cosa significhi questo dal punto di vista pratico, si pensi che con questo algoritmo si può attualmente fattorizzare un numero di 193 cifre usando una rete di qualche centinaio di computer ad altissima prestazione impiegati esclusivamente a questo scopo; il risultato si ottiene dopo qualche mese di lavoro ininterrotto. Utilizzando lo stesso hardware, per fattorizzare un numero di 500 cifre dovremmo aspettare un tempo più lungo dell’età dell’universo. Se avessimo a disposizione un computer quantistico in grado di effettuare lo stesso numero di operazioni al secondo del supercalcolatore classico descritto prima, potremmo utilizzare l’algoritmo di Shor per ottenere la fattorizzazione di un numero di 193 cifre in appena 0.1 secondi e di un numero di 500 cifre in soli 2 secondi.

A chi importa un tale risultato? In primo luogo, Ron Rivest, Adi Shamir e Len Adleman vedrebbero cadere la congettura che garantisce la sicurezza dello schema crittografico a chiavi pubbliche alla base del loro famoso cifrario RSA (dalle loro iniziali). Nel 1978, quando questo cifrario venne pubblicato sulla famosa rivista "Communications of the ACM", sembrava infatti ragionevole assumere che la fattorizzazione di numeri interi molto grandi fosse un problema difficile. In effetti, sebbene abbia attratto l'interesse di numerosi crittoanalisti, l'RSA è rimasto sino ad oggi sostanzialmente inviolato. Se a questo si aggiunge una grandissima semplicità strutturale, si può capire perché questo metodo di cifratura abbia avuto così tanto successo. Per la sua semplicità l'RSA è ampiamente utilizzato nelle applicazioni pratiche (dalle transazioni finanziarie effettuate in Internet alla protezione della privacy e autenticità dell’email e alla maggior parte delle applicazioni di sicurezza dei dati digitali, informatici e telefonici), e numerose sono le sue realizzazioni in hardware presentate nel corso di questi anni.

Tutto ciò dovrà quindi essere pesantemente rivisto quando/se il computer quantistico sarà realizzato e messo in commercio.

2.6 Grover e la complessità della ricerca algoritmica

Nel 1996 Lov Kumar Grover, originario di Delhi e attualmente ricercatore ai Bell Labs in New Jersey, ideò un metodo quantistico per risolvere problemi di ricerca (notoriamente *difficili*, cioè NP-completi) che migliora di un fattore quadratico le prestazioni degli algoritmi classici fino ad oggi proposti per questi problemi. Il metodo consiste essenzialmente nella definizione di un operatore che ha come sub-routine un oracolo O in grado di stabilire se un certo candidato è soluzione

oppure no al problema dato. Combinato con altre appropriate operazioni, un'invocazione di O sulla sovrapposizione di tutti i possibili candidati determina un'amplificazione dell'ampiezza di probabilità associata alla soluzione, rendendo di conseguenza quest'ultima di gran lunga più probabile come risultato di una misurazione. Questa tecnica si può visualizzare come una serie di rotazioni (applicazioni dell'operatore di Grover) applicate successivamente al vettore iniziale preparato nella sovrapposizione di tutti i candidati soluzione. La sequenza di rotazioni ha l'effetto di avvicinare il più possibile il vettore iniziale al vettore soluzione.

L'algoritmo di Grover è stato dimostrato *ottimale*, cioè nessun algoritmo classico o quantistico potrebbe effettuare una ricerca esaustiva più velocemente dell'algoritmo di Grover. Malgrado il titolo in un certo senso fuorviante dell'articolo in cui Grover introduce la sua tecnica di ricerca quantistica ("A fast quantum mechanical algorithm for database search[7]"), l'algoritmo di Grover potrebbe risultare di scarso vantaggio per le ricerche in ambito puramente databasista, dove una base di dati viene tipicamente realizzata come oggetto *custom-built* mediante memorie di sola lettura. La costruzione di questi oggetti come memorie quantistiche (così come per quelle classiche) richiederebbe di per sè un numero di operazioni esponenziale nel numero dei dati e l'algoritmo di Grover potrebbe aumentare la velocità di ricerca di un fattore al più costante. Inoltre questo vantaggio verrebbe con alta probabilità annullato dalla complessità tecnologica di mantenere la coerenza dello stato di computazione quantistica.

Il grande vantaggio nel poter utilizzare l'algoritmo di Grover, cioè il vantaggio di avere a disposizione un computer quantistico, è invece legato alla impraticabilità dell'alternativa classica della ricerca esaustiva come unica tecnica per i problemi NP-completi. Un campo di applicazione tipico è la criptoanalisi e il problema di decifrare un testo crittografato. Come è noto gli algoritmi di cifratura usano una chiave la cui segretezza è l'unica garanzia per impedire ad una persona non autorizzata di carpire l'informazione contenuta nel testo cifrato. La lunghezza in bit di una chiave dipende dal particolare algoritmo utilizzato ma deve essere sempre fissata in modo da assicurare tale segretezza. Questo significa che non deve essere possibile svelarla mediante un cosiddetto *attacco di forza bruta*: una chiave di n bit avrà 2^n chiavi distinte e non conoscendo quale chiave sia stata usata bisognerà provarle tutte fino ad individuare la chiave giusta. Se un ipotetico nemico avesse a disposizione un computer quantistico il problema rimarrebbe per lui ancora difficile per il fatto che la sua complessità asintotica non è cambiata (rimane NP-completo anche in campo quantistico), ma i progettisti del sistema crittografico dovrebbero ricorrere a chiavi di dimensioni notevolmente maggiori per tener conto del fatto che la potenza di calcolo a disposizione del nemico è aumentata di un fattore quadratico. Per avere un'idea delle ricadute sul piano pratico del guadagno in termini di tempo di ricerca si pensi ad una chiave di lunghezza 10^{30} e si supponga di trovarsi di fronte ad un nemico classico estremamente potente con a dispozione un processore in grado di effettuare 100

[7] *Proceedings of 28th Annual ACM Symposium on Theory of Computing (STOC), pages 212-219, May 1996.*

milioni di test al secondo[8]. Si calcola facilmente che nonostante la sua potenza egli dovrà mettere in conto la necessità di effettuare al più un numero di chiamate dell'ordine di 10^{29} e quindi di impiegare circa 10^{21} secondi, un tempo cioè molto vicino all'età dell'universo. A un suo omologo quantistico con un processore della stessa velocità basterebbero invece 10^{7} secondi, cioè circa quattro mesi, perché il suo algoritmo di ricerca effettuerà nella peggiore delle ipotesi "solo" 10^{15} test.

3. Realizzazione sperimentale di un computer quantistico

3.1 Sovrapposizione quantistica e decoerenza

Le promesse e le possibili implicazioni di un computer quantistico, come abbiamo visto, sono tante. Vediamo ora fino a che punto è stato possibile creare un apparato "reale" in grado di eseguire computazioni quantistiche. Vista l'importanza in questo contesto di alcuni concetti della fisica quantistica, come il principio di sovrapposizione e la decoerenza, iniziamo con un breve riepilogo di quanto già introdotto sopra, con l'aiuto di un altro esperimento fisico leggermente diverso da quello del *beam splitter*.

La realizzazione di un computer quantistico richiede, innanzitutto, l'implementazione fisica del suo elemento chiave, il qubit. Per rendersi conto dell'importanza – e della difficoltà – di questo compito, guardiamo prima al suo equivalente classico, il bit. Esso rappresenta "0" oppure "1" nel sistema binario e può essere realizzato in tantissimi modi: meccanicamente, come pallina su un abaco; elettricamente, con un interruttore che fa passare una corrente elettrica oppure la blocca; e infine elettronicamente, con capacità e transistor, che nei computer moderni sono miniaturizzati fino a raggiungere delle dimensioni di molto meno di un millesimo di un millimetro, permettendo di metterne milioni o addirittura miliardi su una superficie di pochi centimetri quadri. Tuttavia, queste realizzazioni hanno una cosa in comune: seguono le leggi della fisica classica, secondo la quale un qualsiasi oggetto può trovarsi, in un determinato momento, soltanto in uno stato ben specifico. Nel caso del bit, questo vuol dire che esso – per esempio, lo stato logico di una capacità caratterizzato dalla presenza oppure assenza di cariche – può essere nello stato "0" oppure nello stato "1". Non è ammesso nessun altro stato del sistema

Come abbiamo accennato sopra, nella fisica quantistica invece questo non è più vero. Qui vale il principio della sovrapposizione, secondo il quale un oggetto fisico può esistere simultaneamente in due o più stati possibili (cioè, ammessi dalle leggi della fisica quantistica) del sistema. Per semplicità, illustriamo questo principio con un esempio che, secondo Richard Feynman, "porta dentro di sé il cuore della meccanica quantistica" (vedi Figura 5). Un elettrone (oppure un fotone o un'altra particella microscopica) si trova a sinistra di uno schermo con due aperture e si muove verso lo schermo. A destra dello schermo si trova un rivelatore – per esempio, una pellicola fotografica che è sensibile all'arrivo di elettroni. Ogni volta che un elettrone viene lanciato, dopo essere passato attraverso le aperture nel primo schermo, raggiungerà la pellicola e lascerà una traccia su di esso.

[8] *Nella realtà si usano chiavi di di almeno 128 bit per cifrari simmetrici e di 1024 bit per cifrari asimmetrici, anche se questi numeri vengono aggiornati continuamente per far fronte al rapido aumento della velocità degli odierni processori.*

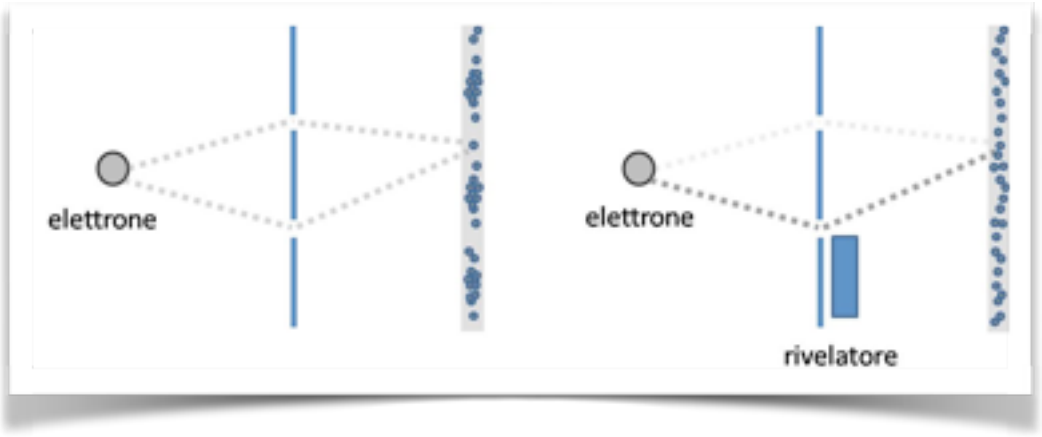


Figura 5
Interferenza quantistica di un elettrone che passa attraverso due aperture in uno schermo. A sinistra: L'interferenza dei due cammini (per l'apertura in alto e quella in basso) crea una variazione di intensità (cioè, della probabilità di trovare l'elettrone in quella posizione) sulla pellicola fotografica. A destra: Se un rivelatore di elettroni indica che l'elettrone è passato per l'apertura in basso, l'effetto di interferenza sparisce.

Il risultato sorprendente di un tale esperimento è che lanciando ripetutamente degli elettroni attraverso lo schermo in questo modo, sulla pellicola fotografica si formerà una struttura periodica molto simile a quella che si crea nell'interferenza di onde sull'acqua. L'interpretazione della fisica quantistica è che gli elettroni, infatti, si comportano come delle onde – e in più, un singolo elettrone in un tale esperimento percorre simultaneamente due cammini, uno che passa per l'apertura superiore e un altro che passa per quella inferiore. Durante il tragitto, quindi, l'elettrone si trova in una sovrapposizione di due stati. Il risultato del processo di interferenza che accade quando l'elettrone arriva alla pellicola e viene rilevato dipende, dunque, da quello che succede in entrambi i cammini.

Questa sovrapposizione, però, può essere compromessa se viene effettuata una misura in grado di rivelare per quale delle aperture l'elettrone è passato. In tal caso la coerenza tra i due cammini viene distrutta, e la struttura periodica sulla pellicola non si forma più.

Il principio illustrato sopra vale per qualsiasi sistema quantistico, e in particolare per un qubit. Per esempio, gli stati logici di un qubit possono essere rappresentati da due stati energetici di un atomo (più precisamente degli elettroni che orbitano intorno al nucleo dell'atomo). Le leggi della fisica quantistica permettono che l'atomo si trovi in una sovrapposizione dei due stati energetici. Una qualsiasi operazione logica – una porta NOT, per esempio – viene quindi effettuata su entrambi gli stati simultaneamente. Nello stesso modo, usando più atomi l'operazione viene effettuata su tutte le combinazioni degli stati 0 e 1 di ciascun atomo e quindi su un numero molto elevato di stati logici. Questo è il principio del parallelismo quantistico.

Come nell'esempio dell'elettrone che percorre allo stesso tempo due cammini diversi, anche nel caso di un qubit una misura fatta durante l'evoluzione

distrugge la coerenza e quindi la sovrapposizione dei due stati quantistici. Per "misura" qui si intende non soltanto una misura fatta volontariamente dall'operatore, ma anche una qualsiasi interazione con l'ambiente che provoca lo stesso effetto di una misura (si può anche interpretare come una "fuga di informazione" dal qubit verso l'esterno). Nell'esempio dell'atomo usato come qubit, una collisione con un altro atomo può provocare la decoerenza, cioè la perdita parziale o totale delle proprietà della sovrapposizione (per questo motivo in molti esperimenti le particelle usate vengono tenute sotto vuoto). Anche l'interazione tra l'atomo e un campo elettrico o magnetico esterno può provocare decoerenza. Appena questo accade, l'integrità del calcolo quantistico è compromessa e il risultato della computazione non è più affidabile.

3.2 I criteri di Di Vincenzo

Una condizione importantissima per realizzare un computer quantistico, dunque, è che il qubit mantenga la coerenza durante tutto il tempo necessario per effettuare un'operazione logica. Questa condizione si può esprimere con la seguente disugualianza: $t_{coh} >> t_{porta}$, dove t_{coh} è il tempo di coerenza del qubit e t_{porta} è il tempo impiegato per effettuare l'operazione di porta logica.

Alla condizione di coerenza si aggiungono altri criteri per la scelta di un sistema fisico adatto per implementare un computer quantistico, elencati nel 2000 da Davide Di Vincenzo [7]:

- Identificazione di qubit ben definiti. Questo criterio richiede che il sistema fisico scelto permetta di identificare delle entità ben distinte – per esempio, gli spin di nuclei atomici oppure singoli atomi preparati dentro un reticolo ottico (vedi sotto) – che possano svolgere il ruolo di qubit.
- Preparazione affidabile dello stato iniziale del computer quantistico. Evidentemente, deve essere possibile inizializzare il computer in uno stato ben noto, per esempio "00.....000", dal quale può partire l'algoritmo quantistico.
- Operazioni precise di porta quantistica. Bisogna poter controllare lo stato dei qubit in maniera accurata per implementare le varie porte quantistiche – rotazione di fase, gate CNOT – sia per singoli qubit che per coppie di qubit. Questo, in pratica, richiede sia un controllo preciso dei campi magnetici, impulsi laser ecc. sia un'ottima conoscenza delle proprietà fisiche del sistema.
- Possibilità di misurare in maniera accurata lo stato quantistico dei qubit. Al termine dell'algoritmo, è necessario "leggere" lo stato del sistema per conoscere l'esito della computazione.

Vedremo in seguito alcuni sistemi fisici che sono stati realizzati in laboratorio e che potrebbero soddisfare i criteri di Di Vincenzo. Questi sistemi si possono suddividere in due classi: sistemi *naturali*, quali atomi, molecole e fotoni, e sistemi *artificiali*, ovvero fatti dall'uomo, come le giunzioni di Josephson.

Visto il numero ormai molto elevato di sistemi sotto considerazione, nella maggior parte dei casi ci limiteremo a descrivere brevemente le loro caratteristiche essenziali. L'approccio basato sugli ioni intrappolati [12], invece,

verrà spiegato più in dettaglio per illustrare le difficoltà che si incontrano nella realizzazione di computer quantistici e per presentare il tipo di tecniche sviluppate per superare tali difficoltà.

3.3 Ioni intrappolati

Per molti anni il metodo degli ioni intrappolati è stato considerato come il candidato più promettente per realizzare un computer quantistico, e tuttora la ricerca sta andando avanti a pieno ritmo, anche se le aspettative sono state leggermente ridimensionate. Comunque, le tecniche sviluppate per controllare e manipolare gli stati quantistici di singoli ioni da sole rappresentano un importante progresso scientifico, e nel 2012 i fisici David Wineland e Serge Haroche sono stati onorati col premio Nobel per la fisica per le loro scoperte in questo campo.

La trappola di Paul

La storia della computazione quantistica con ioni intrappolati iniziò nel 1995, quando il fisico teorico austriaco Peter Zoller e il suo collega spagnolo Ignacio Cirac pubblicarono un articolo di ricerca di sole quattro pagine ma che diede inizio a un campo di ricerca che negli anni successivi avrebbe attratto centinaia di ricercatori. Nel loro articolo i due fisici proposero un approccio alla computazione quantistica nel quale si usano degli atomi carichi, anche noti come ioni, che vengono tenuti fermi nello spazio da campi elettrici. Gli stati quantistici necessari per realizzare dei qubit sono quelli degli elettroni dello ione, simili a quelli di un atomo neutro. Con fasci laser i singoli ioni possono poi essere controllati per eseguire delle porte quantistiche (vedi Figura 6).

La scelta di ioni come qubit sembra quasi naturale: grazie alla loro carica possono essere tenuti fermi e manipolati da campi elettrici e quindi non è necessario fisicamente “toccarli”, cosa che li disturberebbe e in ultima analisi distruggerebbe gli stati quantistici. Inoltre, scegliendo i livelli quantistici giusti dello ione si possono realizzare qubit con tempi di coerenza lunghissimi, fino a diversi secondi. C’è infine anche una considerazione pratica a favore di questa scelta e cioè che le tecniche per intrappolare particelle cariche sono state studiate per anni, e quindi si sa che in principio l’approccio deve funzionare.

Nel computer quantistico con ioni intrappolati, essi vengono tenuti fermi dentro una trappola di Paul. Una trappola di Paul è basata sul principio della “sella ruotante”: una pallina posata su una sella è intrappolata nella direzione longitudinale, nella quale la sella è rialzata, ma può facilmente scappare lateralmente. Se la sella viene montata su un palo ruotante, invece, ogni volta che la pallina inizia a scappare lateralmente, la posizione della sella è già cambiata ed a questo punto la pallina vede davanti a se la parte rialzata della sella. In questo modo, la pallina è effettivamente intrappolata in tutte le direzioni.

Un effetto analogo può essere usato per intrappolare delle particelle cariche in campi elettrici. Mentre una legge di James Clark Maxwell, lo scopritore delle leggi omonime dell’elettrodinamica, vieta l’intrappolamento stabile di una carica in un campo elettrico statico, è possibile creare delle trappole stabili usando campi oscillanti. L’equivalente di una sella nel caso di campi elettrici è un cosiddetto campo quadrupolare, formato per esempio dentro una configurazione di quattro elettrodi con segno alternato del potenziale. La “rotazione” della sella è provocata da una variazione periodica dei segni del potenziale, con il risultato che una carica posta in una tale trappola dinamica rimane ferma (a parte delle piccole oscillazioni

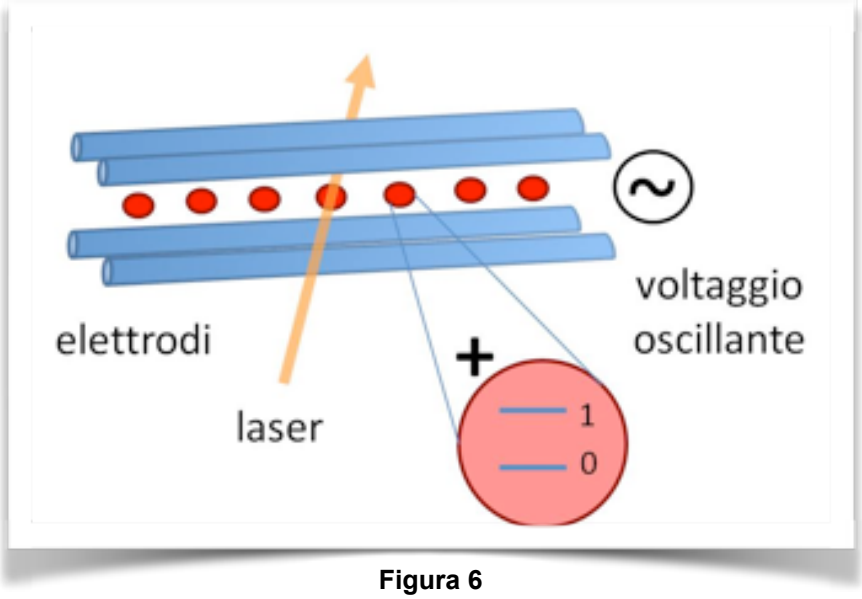


Figura 6
Una trappola di Paul (lineare) per intrappolare ioni. Il fascio laser può essere puntato su un singolo ione per controllarlo (per esempio, per effettuare una porta quantistica su un singolo qubit, rappresentato per i due livelli energetici 0 e 1).

note come micromoto). In questo modo si possono intrappolare e studiare piccoli oggetti come grani di polvere, ma anche elettroni o – appunto – ioni.

Una volta intrappolato in una trappola di Paul, uno ione non può essere usato subito come qubit per la computazione quantistica. Tipicamente gli ioni vengono intrappolati da un gas a temperatura ambiente, il che vuol dire che possiedono un'elevata energia cinetica e che una volta intrappolati oscilleranno dentro la trappola. Questo moto incontrollato rende difficile un controllo accurato dello ione, ed è quindi necessario ridurlo. Ridurre la velocità media di un gas, ovviamente, significa raffreddarlo

Tecniche di raffreddamento

Il raffreddamento viene effettuato tramite dei fasci laser. In un primo passo viene usato il "raffreddamento Doppler", così chiamato perché sfrutta l'effetto Doppler – quell'effetto che si conosce bene dall'apparente variazione in frequenza della sirena di un ambulanza che si avvicina oppure allontana. Per rallentare il moto degli ioni viene usato un effetto simile: un fascio laser che incide su un ione con una frequenza leggermente più bassa di quella risonante tra due stati quantistici dello stesso viene assorbito più facilmente da uno ione che si sta muovendo verso il fascio. In questo caso, lo ione "vede" una frequenza un po' più alta e quindi più vicina a risonanza. Vice versa, uno ione che si sta allontanando dal fascio laser vede una frequenza più bassa e quindi ancora più lontana dalla risonanza. Se ora teniamo conto del fatto che ogni volta che uno ione assorbe un fotone del fascio laser subisce un piccolo urto nella direzione opposta, vediamo subito che se illuminiamo lo ione da tutte le direzioni spaziali esso sentirà una forza rallentante dovunque si muova, come se si muovesse in un liquido viscoso (si parla anche di "melassa ottica").

Risultati ottenuti con ioni intrappolati

In questo modo, e usando altre tecniche simili e ancora più sofisticate, è possibile raffreddare uno ione fino a raggiungere lo stato quantistico più basso possibile. Partendo da un tale stato, già nel 1995 David Wineland riuscì a realizzare una porta CNOT con un singolo ione intrappolato (usando come primo qubit lo stato quantistico dell'elettrone e come secondo qubit quello del moto dello ione). Da allora numerosi gruppi di ricercatori hanno sviluppato delle tecniche sempre più avanzate. Nel 2003 all'università di Innsbruck in Austria fu realizzata una porta CNOT seguendo l'approccio di Zoller e Cirac del 1995. In quell'esperimento vennero usati due ioni dentro una trappola di Paul che si potevano indirizzare separatamente con due fasci laser. Da allora, i fisici sono riusciti anche ad implementare dei semplici algoritmi quantistici, come per esempio l'algoritmo di Deutsch. Per poter implementare algoritmi più complicati, ad esempio quello di Shor per numeri grandi, uno degli ostacoli principali è la scalabilità. Mentre fino ad una mezza dozzina di ioni possono essere intrappolati e manipolati dentro una trappola di Paul senza problemi, non è per niente banale aumentare questo numero fino a decine o centinaia di ioni. Un approccio studiato negli ultimi anni risolve questo problema usando delle trappole suddivise in vari segmenti, ciascuna dei quali contiene circa dieci ioni. Il calcolo quantistico può essere effettuato agendo su un segmento alla volta, con trasporto degli ioni tra i segmenti nelle varie fasi dell'algoritmo.

3.4 Atomi neutri in reticoli ottici

Come abbiamo visto, la realizzazione di un computer quantistico con ioni intrappolati ha come limitazione principale la scarsa scalabilità. Questa limitazione può essere superata in parte usando atomi neutri che vengono intrappolati dentro dei "cristalli di luce" creati da fasci laser sovrapposti che creano interferenza. Tali reticoli ottici intrappolano gli atomi sfruttando la forza della luce (nota come forza dipolare), e la scelta libera della geometria dei fasci laser permette di creare delle strutture spaziali a piacere, per esempio cristalli tridimensionali. In queste strutture possono essere intrappolati milioni di atomi che possono essere usati come qubit. Il problema principale, al momento attuale, sta nel realizzare porte quantistiche con due o più qubit. Per implementare tali porte è necessaria una interazione tra i qubit, che nel caso di ioni è data dall' interazione elettrostatica ma è assente per atomi neutri. Un approccio studiato negli ultimi anni usa stati altamente eccitati, noti anche come stati di Rydberg, per indurre un'interazione forte tra atomi adiacenti nel reticolo [16].

3.5 Qubit in superconduttori

I superconduttori sono dei materiali che conducono la corrente elettrica senza nessuna resistenza. Una corrente che scorre in un anello fatto di un superconduttore girerà dentro l'anello per sempre. La superconduttività fu scoperta dall'olandese Kammerlingh Onnes nel 1911 e spiegata teoricamente nel 1957 da tre scienziati americani che usarono il concetto del condensato di Bose-Einstein, nel quale delle particelle bosoniche si aggregano tutti nello stesso stato quantistico, benché gli elettroni che conducono la corrente elettrica siano dei fermioni che non possono condividere lo stesso stato quantistico. La teoria di Bardeen, Cooper e Schrieffer risolvette questo problema introducendo il concetto della coppia di Cooper – cioè, una coppia fatta da due elettroni (che sono

fermioni) che, insieme, si comportano come un bosone e quindi possono formare un condensato di Bose-Einstein insieme ad altre coppie di Cooper.

Anche se la corrente che scorre dentro un superconduttore è fatta da miliardi di elettroni (o meglio coppie di Cooper), la si può considerare come un singolo stato quantistico. Di conseguenza, tale corrente può anche esistere in uno stato di sovrapposizione di due o più stati quantistici - per esempio, in una sovrapposizione di correnti che scorrono allo stesso tempo in senso orario e antiorario dentro un anello superconduttore. Nello stesso modo una corrente che scorre avanti e indietro in un oscillatore può esistere in una sovrapposizione di stati quantistici dell'oscillatore. È lecito, quindi, pensare che i superconduttori possano essere dei candidati possibili per l'implementazione di un computer quantistico.

Vi sono alcuni vantaggi ovvii di un tale approccio rispetto all'uso di ioni o atomi come abbiamo visto fino ad ora. Innanzitutto, i superconduttori sono delle strutture fatte dall'uomo, e quindi è possibile adattare le loro proprietà alle esigenze del problema. Gli ioni e gli atomi, d'altro canto, sono dati dalla natura, e le loro caratteristiche non possono essere modificate. Inoltre, i circuiti di superconduttori possono essere fabbricati usando i processi ben noti che vengono già applicati alla produzione di circuiti integrati e microchip per computer classici.

Nonostante questi vantaggi non è affatto semplice usare i superconduttori per implementare qubit [13,18]. Un problema fondamentale è che gli stati quantistici di un oscillatore superconduttore sono equidistanti, cioè la differenza in energia tra uno stato e gli stati adiacenti è la stessa per tutti gli stati. Questo fa sì che diventi impossibile individuare due stati da usare come 0 e 1 del qubit. Esiste, però, un rimedio, la cosiddetta giunzione di Josephson. Tale giunzione consiste, sostanzialmente, in una lastra isolante molto sottile inserita tra due fili superconduttori (Figura 7). Nel mondo classico, per definizione un isolante non fa passare nessuna corrente. Nella fisica quantistica, invece, le coppie di Cooper di

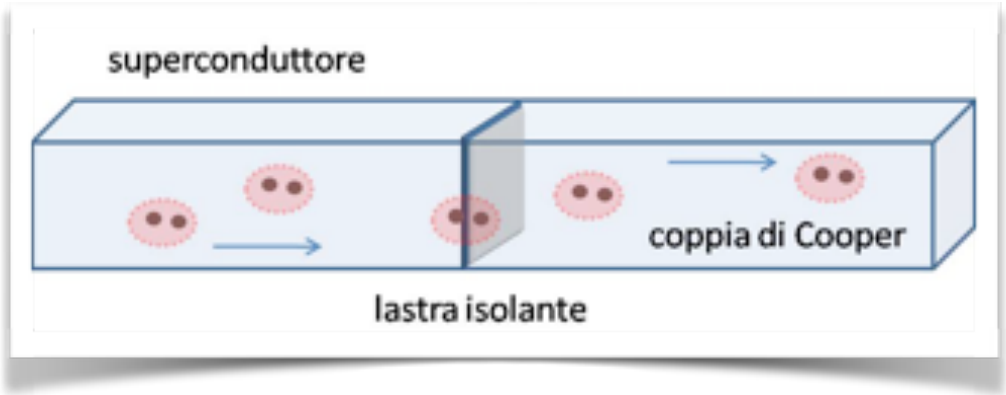


Figura 7
Schema di una giunzione di Josephson. Le coppie di Cooper possono attraversare la lastra isolante tramite il tunneling quantistico.

un superconduttore possono attraversare la barriera isolante tramite l'effetto tunnel. Una conseguenza di questo effetto è che si possono distinguere due livelli energetici del sistema, che a loro volta sono utilizzabili come qubit. Recentemente sono state sviluppate delle tecniche per una lettura efficace dello

stato di questi qubit, ed i tempi di coerenza sono stati allungati fino ad essere abbastanza lunghi da permettere l'esecuzione di alcune porte quantistiche.

3.6 Elettroni in quantum dots

I *quantum dots* sono delle strutture a semiconduttore in grado di confinare singoli elettroni in gabbie minuscole le cui dimensioni tipicamente sono inferiori a un micrometro. Dentro un tale *quantum dot*, l'energia di un elettrone è quantizzata, permettendo all'elettrone di muoversi soltanto in determinate "orbite". Per questo motivo i *quantum dots* vengono anche chiamati "atomi artificiali". Negli ultimi anni i ricercatori hanno investigato la possibilità di usare tali *quantum dots* come qubit. In particolare, lo spin di un elettrone si presta in maniera naturale alla realizzazione degli stati logici 0 e 1 di un qubit.

Per poter fare computazione quantistica con gli elettroni in *quantum dots*, bisogna dimostrare che è possibile implementare la porta CNOT oppure un'altra porta logica universale a due qubit oltre alla manipolazione di un singolo qubit. In esperimenti recenti questo è stato fatto, e sono stati misurati dei tempi di coerenza di alcuni microsecondi.

Una delle sfide più importanti nell'uso di quantum dots come qubit è la lettura dello stato finale della computazione. Per una lettura efficiente il quantum dot deve essere molto vicino al dispositivo di lettura; se invece è troppo vicino quest'ultimo può causare una decoerenza che interferisce con la computazione stessa.

3.7 Risonanza magnetica

Uno dei primi approcci sperimentali alla computazione quantistica è stata la risonanza magnetica (NMR). Questa tecnica [17], molto diffusa nella diagnostica, usa dei campi elettromagnetici per invertire la direzione dello spin del nucleo di un atomo dentro una molecola. La frequenza caratteristica per questa inversione dipende sia dallo spin stesso che anche dagli spin nucleari di altri atomi nella molecola (tramite accoppiamento mediato da elettroni). Mentre in diagnostica questa dipendenza viene sfruttata per distinguere diverse molecole, nell'ambito della computazione quantistica si può usare per eseguire porte quantistiche sia su un qubit che su due qubit. In questo modo nel 2001 è stato implementato l'algoritmo di Shor per fattorizzare il numero 15 usando molecole in cui i spin nucleari di 5 atomi di fluoro (simbolo chimico "F") e 2 due di carbonio (simbolo "C") fungevano da qubit (vedi Figura 8). Per fattorizzare numeri più grandi bisognerebbe aumentare di ordini di grandezza il numero di spin contenuti in una molecola, il che richiederebbe la sintesi controllata di molecole contenenti migliaia di atomi. Al momento attuale questa non sembra essere possibile, e per questo motivo la risonanza magnetica non viene più considerata una tecnica adatta per la realizzazione di computer quantistici.

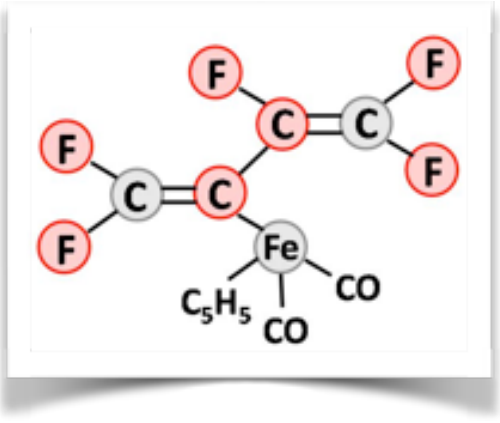


Figura 8
Molecola usata per implementare l'algoritmo di Shor. Gli atomi usati come qubits sono evidenziati in rosso.

3.8 Fotoni

Nella prima parte di questo articolo abbiamo parlato più volte di fotoni e del fatto che un *beam-splitter* può essere considerato un'implementazione di una porta Hadamard. Sembrerebbe, quindi, possibile usare fotoni per creare un computer quantistico. In linea di principio questo è vero, ma l'assenza di interazioni tra due o più fotoni rende complicata la realizzazione di porte a più qubit. Comunque, negli ultimi anni sono state sviluppate tecniche in grado di implementare tali porte tramite delle misure quantistiche e post-selezione [9].

4. Conclusioni

Nonostante considerevoli sforzi e notevoli progressi negli ultimi vent'anni, ad oggi non è ancora stato possibile costruire un computer quantistico in grado di fare cose "utili"- cioè, computazioni che non si possano eseguire sui più potenti computer oggi in commercio. La ditta canadese D-Wave da qualche anno sta sviluppando computer quantistici che si basano sulla tecnica del *quantum annealing*, ma al momento non è chiaro se il computer prodotto dalla D-Wave veramente esegua computazioni quantistiche e se, quindi, potrebbe superare le capacità di un computer classico[9].

In parallelo allo sviluppo di computer quantistici, più recentemente si è manifestato un altro filone di ricerca che si concentra sui cosiddetti simulatori quantistici. Nello spirito di Richard Feynman, che negli anni '80 del secolo scorso si chiedeva fino a che punto la fisica quantistica potesse essere simulata su dei computer classici, l'idea del simulatore quantistico è piuttosto semplice. Evitando di simulare un sistema quantistico su un computer, si crea nel laboratorio un sistema quantistico ben controllato in grado di simulare il fenomeno di interesse. A tutti gli effetti, quindi, un simulatore quantistico si può definire come computer quantistico "a singolo uso".

Scott Aaronson afferma nel suo libro [1] che fortunatamente oggi possiamo dire di aver fatto grandi passi avanti grazie a decenni di studio e di lavoro nel campo della computazione quantistica e dei fondamenti della teoria quantistica.

Tuttavia, visto che al momento non possiamo ancora contare su un computer quantistico per svolgere quei compiti che siamo soliti eseguire su un normale pc (e soprattutto per quelli che non possono essere svolti neanche dai più potenti computer oggi disponibili) una domanda che potrebbe venire spontanea al lettore è: ma allora ha un senso studiare la computazione quantistica?

Certamente la ricerca nel campo della computazione quantistica è puramente teorica e altamente speculativa; pur tuttavia la risposta al lettore è (in breve) `sì, ha senso'. Una risposta più dettagliata potrebbe elencare varie motivazioni, ma la più importante è che la ricerca teorica in generale (e quindi nel caso specifico dell'informatica e dell'informazione quantistica) contribuisce in modo fondamentale alla nostra comprensione dell'universo e ci permette di acquisire conoscenze che, indipendentemente dall'esistenza fisica del computer quantistico, rappresentano un arricchimento culturale e scientifico dell'umanità.

---

[9] *Si veda a tal proposito il recente articolo di Jeremy Hsu su IEEE Spectrum (http://spectrum.ieee.org/computing/hardware/dwaves-year-of-computing-dangerously) e il video del seminario di Matthias Troyer, fisico dell'ETH, su http://nitsche.mobi/2013/troyer/.*

Qualunque sarà il risultato degli sforzi rivolti alla costruzione del computer quantistico, il successo della computazione quantistica si può comunque affermare già da ora come il raggiungimento di un risultato non meno importante: l'aver messo in relazione molte questioni fondamentali della fisica e dell'informatica in uno sforzo scientifico comune.

Bibliografia

[1] Aaronson, S., *Quantum Computing since Democritus*, Cambridge University Press, 2013

[2] Bohr, N., Discussion with Einstein on Epistemological Problems in Atomic Physics, in Albert Einstein: Philosopher-Scientist, Cambridge University Press, 1949

[3] Das, A., Chakrabarti, B.K., *Colloquium*: *Quantum annealing and analog quantum computation,* Reviews of Modern Physics, vol. 80, p. 106, 2008

[4] Dasgupta, S., Papadimitriou, C., Vazirani, U., *Algorithms*, Mcgraw Hill Book Co., 2006

[5] Deutsch, D., Quantum Theory, the Church-Turing principle and the universal quantum computer, Proceedings of the Royal Society of London, vol. 400, p. 97,1985

[6] Deutsch, D., *The Fabric of Reality*, Penguin Books, London, 1997

[7] Di Vincenzo, D.P., The Physical Implementation of Quantum Computation, Fortschritte der Physik, vol. 48, p. 771, 2000

[8] Feynman, R.P., *Simulating Physiscs with Computers*, International Journal of Theoretical Physics, vol.21, p. 467, 1982

[9] Kaye, P.R., Laflamme, R., Mosca, M., *An Introduction to Quantum Computing*, Oxford University Press, 2007

[10] Kok, P., Munro, W.J., Nemoto, W., Ralph, T.C., Dowling, J.P., Milburn, G.J., *Linear optical quantum computing with photonic qubits*, Reviews of Modern Physics, vol. 79, p. 135, 2007

[11] Landauer, R., Computation and Physics: Wheeler's Meaning Circuit, Foundations of Physics, vol. 16, 1985

[12] Leibfried, D., Blatt, R., Monroe, C., Wineland, D., *Quantum dynamics of single trapped ions,* Reviews of Modern Physics, vol. 75, p. 281, 2003

[13] Makhlin, Y., Schön, G., Shnirman, A., Quantum-state engineering with Josephson-junction devices, of Modern Physics, vol. 73, p. 357, 2001

[14] Milburn, G.J., *The Feynman Processor*, Perseus Books, 1998

[15] Nielsen, M.A., Chuang, I.L., Quantum Computation and Quantum Information, Cambridge University Press, 2000

[16] Saffman, M., Walker, T.G., Mølmer, K., *Quantum information with Rydberg atoms*, Reviews of Modern Physics, vol. 82, p. 2313, 2010

[17] Vandersypen, L.M., Chuang, I.L., NMR techniques for quantum control and computation, Reviews of Modern Physics, vol. 76, p. 1037, 2005

[18] Ze-Liang X., Ashhab, S., You, J.Q., Nori, F., Hybrid quantum circuits: Superconducting circuits interacting with other quantum systems, Reviews of Modern Physics, vol. 85, p. 623, 2013

Biografie

Alessandra Di Pierro ricopre il ruolo di Professore Associato presso il Dipartimento di Informatica dell'Università di Verona, dove insegna Informatica Quantistica nel corso di laurea magistrale in Ingegneria e Scienze Informatiche. La sua attività di ricerca si svolge nell'ambito della semantica e dell'analisi dei linguaggi probabilistici e in quello della computazione quantistica, dove recentemente si è rivolta allo studio del paradigma di computazione topologica.

Email: alessandra.dipierro@univr.it

Oliver Morsch Primo Ricercatore presso l'Istituto Nazionale di Ottica (INO-CNR) a Pisa. Dopo la laurea in fisica all'Università di Oxford (Inghilterra) nel 1995, ha conseguito il dottorato di ricerca, sempre a Oxford, nel 1999. La sua ricerca sperimentale si concentra sugli studi di atomi freddi, condensati di Bose-Einstein, controllo quantistico e atomi di Rydberg.

Email: morsch@df.unipi.it

Software Defined Network:

Un nuovo paradigma per il piano di controllo delle reti

Simone Mangiante - Pierpaolo Baglietto - Massimo Enrico

Negli ultimi anni il paradigma Software Defined Network (SDN) si è imposto come tema di grande interesse nell'ambito delle reti di telecomunicazione. SDN è una nuova architettura di rete in cui piano di controllo e piano di instradamento dei pacchetti sono disaccoppiati, l'intelligenza di rete è logicamente centralizzata e l'infrastruttura fisica sottostante è astratta dai servizi e dalle applicazioni di rete offerte dai provider. L'evoluzione di SDN è cominciata dagli ambienti dei datacenter per poi estendersi anche alle reti WAN/MAN dei grandi operatori, promettendo risparmi e facilità di gestione. Questo articolo presenta il paradigma SDN, descrivendo tecnologie e protocolli coinvolti ed evidenziandone benefici e criticità.

Keywords: Software Defined Network; Next Generation Networks; Network Funtion Virtualization

1 Introduzione

La diffusione di dispositivi mobili, la crescita costante dei contenuti multimediali, la virtualizzazione dei server e la comparsa di servizi cloud, l'affermazione delle reti fisse ad alta velocità e delle reti mobili 4G stanno guidando l'industria delle telecomunicazioni verso un cambiamento delle architetture di rete tradizionali. Le reti attuali sono costruite gerarchicamente, con molti livelli organizzati in una struttura ad albero. Ciò aveva un senso nell'era del paradigma client-server, ma un'architettura così rigida non può più far fronte alla necessità di flessibilità ed efficienza di reti odierne.

Si deve far fronte, infatti, alle seguenti problematiche:

- Cambiamento dei modelli di traffico: al contrario delle applicazioni client-server, dove praticamente tutto il traffico avviene tra i client e il server, le applicazioni odierne accedono a diversi database e server, creando un traffico “est-ovest” di tipo machine-to-machine prima di rispondere ai client nel classico modello “nord-sud” (come evidenziato in ***Figura 1***). Servizi ed applicazioni sofisticate fanno uso di molte più risorse (server, database, applicazioni web distribuite) durante la fase di elaborazione di una richiesta da parte di un client (si pensi a servizi complessi come applicazioni di suggerimenti personalizzati e geolocalizzati). Il classico modello di traffico nord-sud, da client a server, rappresenta ormai la parte minore del traffico generato da una richiesta ad un servizio, ed inoltre non è più prevedibile come nel passato, poiché i client sono in mobilità e cambiano dinamicamente la banda utilizzata e la qualità del servizio richiesta (si pensi a clienti dotati di abbonamenti premium o alla differenza di parametri di qualità tra lo streaming video su un tablet e la lettura di mail su smartphone). Anche le stesse nuove architetture di rete prevedono che il traffico dati possa passare peer-to-peer direttamente da un nodo di accesso ad un altro.
- Proliferazione di dispositivi di accesso: gli utenti incrementano l’utilizzo di dispositivi mobili personali come smartphone e tablet per accedere alle reti aziendali: si pensi al paradigma “Bring your own device” (BYOD, uso del dispositivo personale in ambito aziendale) sempre più adottato all’interno delle aziende. Il reparto IT deve quindi controllare finemente questi accessi in mobilità sia in termini di accesso sia di larghezza di banda, proteggendo sicurezza, continuità di servizio e integrità dei dati aziendali e, al contempo, offrire flessibilità nel rispondere ad esigenze di traffico in costante evoluzione.
- Aumento dei servizi cloud. Le aziende hanno definitivamente adottato servizi cloud sia pubblici sia privati; le varie funzioni aziendali desiderano accedere ad applicazioni, infrastrutture e risorse “on demand”. L’ambiente cloud deve essere messo in sicurezza e soddisfare i requisiti aziendali. Tutto ciò richiede un insieme di macchine, spazio disco e risorse di rete scalabile dinamicamente.
- Big Data: questo nuovo paradigma richiede una maggiore larghezza di banda: la gestione degli attuali enormi volumi di dati delle grandi aziende del web richiede processori in parallelo su migliaia di server, tutti interconnessi. Gli operatori di grandi data center devono quindi scalare la rete ad un livello non immaginato prima, mantenendo una corretta connettività globale.

Attualmente le aziende cercano di utilizzare al massimo la rete attraverso la gestione e configurazione manuale dei singoli apparati. Gli operatori devono affrontare problemi simili poiché la domanda di servizi mobili e larghezza di

banda aumenta in modo esponenziale; i margini vengono erosi dall'aumento di CAPEX (costi fissi) e OPEX (costi operativi) indotti dai nuovi servizi e dalla loro manutenzione a cui si contrappongono guadagni che non crescono abbastanza da compensare i maggiori costi.

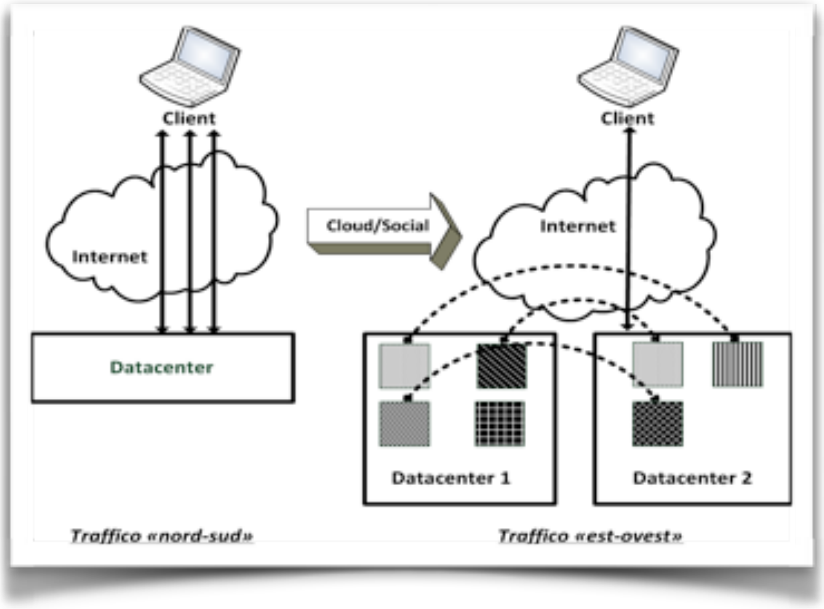


Figura 1
Cambiamento del pattern di traffico nei moderni datacenter
sviluppo prodotto

Le limitazioni delle reti attuali si possono riassumere in:

- Complessità. La tecnologia di rete finora si è evoluta come un enorme insieme di protocolli progettati per connettere apparati in modo consistente attraverso collegamenti arbitrariamente disponibili e veloci e organizzati in svariate topologie. Questi protocolli, pur migliorando efficienza, stabilità e sicurezza della rete, sono spesso definiti in modo isolato, ognuno risolve uno specifico problema senza introdurre nessun tipo di astrazione. La natura rigida delle attuali architetture di rete è in forte contrasto con la natura dinamica dei server e dei data center, sempre più virtualizzati senza fissa posizione fisica e spesso spostati dinamicamente tra i data center per ottimizzarne il carico. Inoltre, molti operatori forniscono una rete convergente per servizi dati, voce e video: se da un lato gli attuali apparati riescono a differenziare il traffico e la qualità del servizio associata a ciascun tipo di servizio, la configurazione delle risorse è ancora per la maggior parte manuale. A causa della sua staticità, la rete non riesce ad adattarsi dinamicamente al cambiamento di traffico, applicazioni e richieste degli utenti.

- Configurazioni inconsistenti. Per implementare una configurazione che tenga conto dell'intera rete, gli amministratori devono modificare una miriade di apparati e protocolli, spesso manualmente e in un modo altamente soggetto ad errori. La complessità delle reti attuali rende difficile l'implementazione di politiche di accesso, qualità del servizio (QoS) e sicurezza consistenti.
- Incapacità di scalare. Se la domanda degli utenti cresce, anche la rete cresce, ma diventa sempre più complessa con l'aggiunta di centinaia o migliaia di apparati da configurare e gestire. Gli operatori spesso si affidano al sovradimensionamento dei collegamenti basato su modelli di traffico prevedibili e statistiche dell'utilizzo della rete. Tuttavia, negli odierni data center virtualizzati il traffico diventa sempre più dinamico e imprevedibile. Grandi operatori, detti Over The Top (OTT), come Google, Yahoo! e Facebook, impiegano algoritmi ed enormi insiemi di dati largamente distribuiti su migliaia di macchine virtuali, perciò hanno bisogno di reti ad altissime scalabilità e prestazioni, possibilmente a costi non elevati. Questo livello di scalabilità non è raggiungibile attraverso configurazioni manuali.
- Dipendenza dai fornitori. Operatori e aziende cercano di rilasciare nuove funzionalità e servizi rapidamente per tenere il passo con le richieste degli utenti. Tuttavia la loro abilità di rispondere agilmente alla crescente domanda è legata ai cicli di produzione dei loro fornitori, che spesso ammontano a tre anni o più: mancano standard e interfacce aperti che favoriscano l'innovazione.

2 Software Defined Network

Negli ultimi anni una nuova architettura di rete si è imposta come possibile soluzione alle problematiche appena esposte: il paradigma di Software Defined Network (da qui in avanti SDN). Fin da subito ha ottenuto il consenso da parte del mondo accademico e dell'industria, comprovato da ingenti investimenti economici, e le previsioni di mercato la dipingono come ciò che rivoluzionerà le reti di telecomunicazioni nei prossimi anni.

2.1 Definizione

Non esiste una definizione di SDN che metta d'accordo mondo industriale e mondo accademico, né una precisa definizione di quanta parte di una rete impatti. L'ONF (Open Networking Foundation, organismo nato dagli operatori per promuovere SDN) si concentra sul concetto di un piano di controllo centralizzato; parallelamente, IETF (Internet Engineering Task Force, organismo di standardizzazione Internet) propone una visione più ampia circa la programmabilità della rete.

La definizione finora più accreditata è quella dell'ONF [13]:

"SDN è una nuova architettura di rete in cui il piano di controllo è disaccoppiato dall'instradamento dei pacchetti ed è direttamente programmabile. Lo spostamento del controllo, da localizzato all'interno degli apparati di rete a macchine centralizzate e facilmente accessibili, permette ad applicazioni e servizi

di sfruttare un'astrazione dell'infrastruttura di rete sottostante, che può quindi essere considerata come un'unica entità logica o virtuale. L'intelligenza della rete è logicamente centralizzata in uno o più controllori ('controller' in terminologia SDN) software, che mantengono una visione globale della rete stessa. La rete appare quindi alle applicazioni e ai meccanismi di policy come un singolo switch logico."

Questa definizione si concentra su tre caratteristiche principali:

- separazione del piano di controllo dal livello di instradamento dei pacchetti;
- controller centralizzato e visione globale della rete;
- programmabilità della rete da parte di applicazioni esterne.

ONF pone l'accento sul trasferimento dell'intelligenza di rete dai router/switch verso server esterni, lasciando negli apparati distribuiti solamente il motore di instradamento dei pacchetti con un meccanismo di tipo "lookup table" su cui effettuare velocemente la decisione di routing.

IETF dichiara una visione alternativa di SDN (a partire dall'acronimo) [1]:

"Software Driven Network è un approccio alle reti che permette alle applicazioni di programmare e manipolare il software di controllo degli apparati e delle risorse che compongono l'infrastruttura di rete. Le SDN comprendono applicazioni, software di controllo e interfacce ai servizi ospitati in una rete logica/virtuale così come gli stessi componenti della sottostante infrastruttura. Le applicazioni possono trarre vantaggio dalla conoscenza delle risorse disponibili e dalla possibilità di richiederne l'attivazione in modi specifici."

IETF presenta una definizione più ibrida: la programmabilità del piano di controllo può coesistere con gli attuali protocolli di controllo distribuiti; presuppone l'utilizzo di interfacce di programmazione di alto livello ("northbound API") che permettano alle applicazioni di avere accesso alle informazioni sullo stato della rete e di inviare comandi agli apparati per influenzare configurazioni, instradamento dei pacchetti e larghezza di banda.

Il primo passo dell'ONF è stata la standardizzazione del protocollo OpenFlow, il componente di più basso livello su cui costruire reti SDN. Il protocollo OpenFlow rappresenta l'interfaccia di un controller SDN verso gli apparati sottostanti ("southbound API"), e risulta necessario poiché si assume che apparati e controller siano fisicamente separati e geograficamente distribuiti. Parallelamente, l'obiettivo primario di IETF è quello di sviluppare standard per applicazioni e servizi che soddisfino il bisogno emergente di coordinare la configurazione di rete con la configurazione di macchine fisiche e virtuali e applicazioni, attraverso reti e datacenter differenti. IETF si concentra quindi sull'interfaccia "northbound" di un controller SDN, che permette ai software di gestione di configurazione, sicurezza, ingegneria del traffico e risparmio energetico di modificare le decisioni del controller. Il lavoro di IETF sembra quindi ortogonale a quello dell'ONF, che si focalizza su come realizzare una rete SDN senza indagarne l'interfaccia verso i livelli più alti né quali servizi SDN debba offrire.

Indipendentemente dalle dispute sulla definizione, con SDN aziende e provider possono avere un controllo non dipendente dal fornitore sull'intera rete da un

punto logicamente centralizzato che rende la progettazione, l'aggiornamento e la manutenzione della rete più semplici. SDN semplifica anche gli apparati di rete, poiché non devono implementare migliaia di protocolli e standard, ma solamente accettare comandi da un controller SDN attraverso un protocollo standardizzato e aperto.

Operatori e amministratori di rete possono configurare un'astrazione della rete attraverso interfacce software piuttosto che integrare moduli di configurazione in ogni singolo apparato. Avendo lo stato della rete centralizzato in un controller, l'architettura SDN fornisce più flessibilità nella gestione, manutenzione e configurazione delle risorse di rete attraverso applicazioni software dinamiche.

Gli operatori di rete e i provider possono sviluppare programmi di controllo personalizzati senza dover attendere che i fornitori degli apparati introducano nuove caratteristiche proprietarie nei loro prodotti. Oltre all'astrazione della rete fisica, l'architettura SDN si basa su interfacce standard che rendono più semplice la realizzazione e la personalizzazione di servizi comuni, gestiti da piattaforme di orchestrazione e capaci di sfruttare l'astrazione della complessa rete sottostante.

2.2 Funzionamento

Una rete SDN funziona essenzialmente creando una rete virtuale indipendente dalla rete fisica sottostante.

La virtualizzazione è resa possibile aggiungendo un livello software tra il tradizionale livello hardware e il software esistente che lo controlla. Questo nuovo strato, evidenziato in ***Figura 2***, che separa il piano di instradamento dei pacchetti dal piano di controllo e permette all'utente di controllare i flussi di traffico nella rete, fa credere ad ogni singola applicazione di possedere l'intera rete, mentre in realtà la sta condividendo con molte altre. Il risultato è che – ad esempio – più server possono utilizzare in modo efficiente la rete: la separazione dei due piani permette ai controller di manipolare la rete più facilmente e un operatore di rete può spostare liberamente componenti della rete senza intaccare la visione globale fornita alle applicazioni dallo strato software aggiuntivo. SDN trasforma la rete esattamente come le macchine virtuali hanno trasformato i server.

Lo scopo principale di una rete è trasportare dati da un punto ad un altro. Durante il percorso, i dati attraversano diversi componenti hardware che ospitano software di controllo: oltre a trasmettere i dati, gli apparati ne controllano il flusso, scegliendo il percorso più veloce o più economico, assegnando priorità a seconda del tipo di traffico e mantenendo tutto in sicurezza.

Quando un pacchetto arriva ad un router in una rete tradizionale, regole compilate all'interno del *firmware* proprietario del router stabiliscono dove trasmettere il pacchetto. Il router trasmette tutti i pacchetti che hanno la stessa destinazione lungo lo stesso percorso e li tratta allo stesso modo; all'interno delle reti degli operatori esistono router più sofisticati che possono riconoscere diversi tipi di pacchetti (a seconda della sorgente, del protocollo trasportato, ecc.) e trattarli in modi differenti, ma sono molto costosi.

In una rete SDN un amministratore può gestire il traffico da un controller centralizzato senza dover intervenire sui singoli router distribuiti lungo i nodi della rete; può cambiare le regole di instradamento di ogni singolo apparato, quando necessario, con un livello fine di granularità. Ciò consente di gestire il traffico in

modo più dinamico, efficiente e flessibile, utilizzando apparati più economici e mantenendo un controllo preciso sui flussi dei pacchetti all'interno della rete.

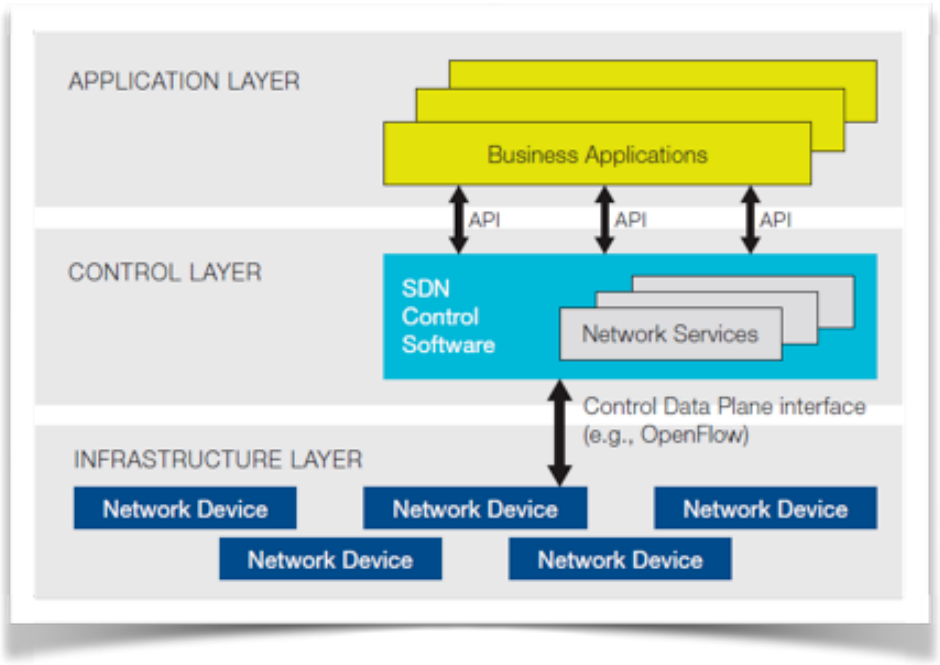


Figura 2
Architettura SDN secondo ONF (fonte: [13])

SDN astrae la rete come un sistema operativo disaccoppia le applicazioni dall'hardware su cui vengono eseguite [5]: si pone l'obiettivo di trasformare le reti così come i sistemi operativi hanno trasformato i *mainframe*. Questi ultimi avevano hardware specializzato e le applicazioni dovevano essere altamente integrate con esso, formando un'integrazione verticale chiusa e proprietaria. I sistemi operativi introdussero l'astrazione di funzioni hardware e permisero ai programmatori di sviluppare applicazioni ad un livello più alto, attraverso interfacce standardizzate e spesso aperte.

SDN cerca di far evolvere la rete in questa direzione: gli apparati fisici non devono più essere specializzati e altamente complessi, ma possedere interfacce aperte e una maggiore interoperabilità.

2.3 OpenFlow

OpenFlow è il primo protocollo standard definito per l'interfaccia tra gli strati di controllo e di instradamento dei pacchetti dell'architettura SDN. Esso consente l'accesso diretto e la modifica del piano di instradamento degli apparati (sia fisici sia virtuali) abilitati. L'assenza di un'interfaccia aperta e standard ha portato alla situazione attuale in cui gli apparati di rete sono oggetti monolitici, chiusi e proprietari. Ciò è stato conseguenza della specializzazione dell'hardware, in termini di funzionalità e caratteristiche, che necessitava interfacce e protocolli appositamente progettati, creando di fatto un ambiente chiuso e proprietario. OpenFlow invece separa il piano dei dati di uno switch, che rimane all'interno

dello switch stesso, dalle attività del piano di controllo, che in SDN risiedono tipicamente su un server centralizzato, esponendo un insieme di interfacce che permettono di programmarne via software il comportamento: questo protocollo rappresenta il primo passo verso lo spostamento dell'intelligenza di controllo dagli apparati ad un controller logicamente centralizzato.

OpenFlow può essere paragonato all'insieme di istruzioni di un processore: specifica le operazioni primitive utilizzabili da un software esterno per programmare l'instradamento dei pacchetti degli apparati, come le istruzioni di un processore programmano l'hardware di un calcolatore. Il protocollo OpenFlow deve essere implementato da entrambi i lati dell'interfaccia tra apparati della struttura di rete e software di controllo SDN.

OpenFlow utilizza il concetto di flusso (flow) per identificare il traffico a seconda di regole di identificazione che possono essere staticamente o dinamicamente programmate dal software di controllo SDN. Permette anche di definire come un flusso di traffico deve attraversare la rete basandosi su parametri globali come l'utilizzo di banda, lo stato delle applicazioni e le risorse disponibili. Un'architettura SDN basata su OpenFlow fornisce un controllo della rete estremamente fine, riuscendo a controllare il traffico a livello di singoli flussi, in modo da rispondere agilmente ai rapidi cambiamenti delle applicazioni, degli utenti e delle sessioni. Il routing IP attuale non consente un tale livello di controllo, poiché tutti i flussi tra due punti devono seguire lo stesso percorso all'interno della rete, senza possibilità di distinguere requisiti differenti.

OpenFlow nasce dalla necessità di eseguire esperimenti su protocolli di rete velocemente e in maniera efficiente nei campus universitari di Berkeley e Stanford [12]. Switch e router commerciali non forniscono né una piattaforma di programmazione aperta né strumenti efficaci per virtualizzare il loro hardware e software. Ogni apparato è diverso a seconda del produttore, con nessuna interfaccia standard a disposizione di ricercatori che vogliano sperimentare nuovi algoritmi di controllo. OpenFlow estrae un insieme comune di funzionalità di rete dalle tabelle di routing contenute in switch e router ed espone un protocollo aperto per programmare tabelle di flussi negli apparati [14]. La ***Figura 3*** mostra uno switch compatibile con OpenFlow, che consta di 3 parti:

- una tabella di flussi con un'azione associata ad ogni flusso che definisce la modalità con cui elaborare il dato flusso;
- un canale sicuro di comunicazione con un controller su cui ricevere comandi ed inviare dati statistici e pacchetti;
- il protocollo OpenFlow che espone API standard utilizzabili dal controller.

OpenFlow è il protocollo chiave per la creazione di una SDN ed attualmente è l'unico standard ONF per la gestione del piano dei dati degli apparati di rete. Gli apparati di rete possono supportare OpenFlow parallelamente ai sistemi di forwarding tradizionali: lo sforzo maggiore di ONF e dei suoi organi di standardizzazione è rivolto ad assicurare l'interoperabilità tra apparati di rete e software di controllo provenienti da diversi fornitori. Tra i benefici ottenibili attraverso OpenFlow citiamo:

- Prestazioni e costi: gli switch, senza più componenti di controllo, si specializzano nel gestire l'instradamento dei pacchetti nel modo più efficiente e veloce possibile, con costi più contenuti.
- Facilità di implementazione e test di nuove caratteristiche nella rete: grazie alle interfacce software, una modifica nel piano di controllo centralizzato viene istantaneamente propagata a tutti gli switch controllati.
- Visione unificata della rete: gli amministratori di rete possono avere una visione d'insieme di tutti gli apparati della rete, che comunicano attraverso il protocollo OpenFlow permettendo una più facile gestione della rete stessa.

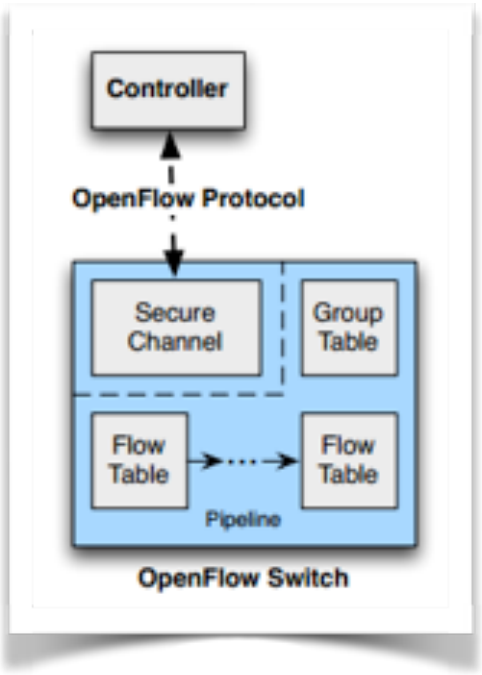


Figura 3
Componenti di uno switch OpenFlow (fonte: [14])

2.4 Vantaggi dell'SDN

Aziende e operatori, adottando il paradigma SDN, possono ottenere benefici sia nella gestione di rete sia nella gestione dei servizi.

Gestione di rete

- Controllo centralizzato di ambienti "multi-fornitore". Il software di controllo di SDN può gestire qualsiasi apparato conforme al protocollo OpenFlow di qualsiasi fornitore, inclusi switch, router e apparati virtuali. Invece di gestire separatamente gruppi di apparati provenienti dallo stesso fornitore, gli amministratori di rete possono

sfruttare le applicazioni SDN di orchestrazione e gestione della rete per configurare e aggiornare in maniera efficiente l'intera rete, abbattendo i costi operativi.

- Aumento della disponibilità e della sicurezza della rete. SDN rende possibile definire policy ad alto livello per l'intera rete, che successivamente vengono inserite all'interno degli apparati fisici sottostanti attraverso il protocollo OpenFlow. L'architettura SDN elimina la configurazione manuale dei singoli apparati e di conseguenza gli errori e le inconsistenze delle policy. La visione globale della rete disponibile al controller SDN è indispensabile per sviluppare meccanismi di IDP (*Intrusion Detection and Prevention*)/ IDS (*Intrusion Detection System*) efficaci: il controller riesce a rilevare intrusioni in ogni parte della rete e proteggere il resto dell'infrastruttura fisica.
- Controllo più fine. Il modello di controllo basato su flussi del protocollo OpenFlow consente di applicare regole ad un livello molto fine, includendo i concetti di sessione, utente, applicazione, dispositivo, in un ambiente astratto e in modo automatico. La rete può operare migliori decisioni di ingegneria del traffico poiché l'ottimizzazione è centralizzata e globale e non distribuita e locale.
- Risparmio energetico. SDN può creare un percorso dinamico tra due punti, a seconda dello stato globale della rete: in questo modo alcuni apparati possono essere automaticamente spenti o messi in stand-by per risparmiare energia.

Gestione dei servizi

- Elevato livello di astrazione. Un operatore può semplicemente aggiungere un nuovo cliente o una nuova macchina virtuale in una rete: SDN è responsabile di prendere questo comando di alto livello e tradurlo in configurazioni adeguate per l'infrastruttura fisica di rete, in modo intelligente attraverso meccanismi sofisticati di orchestrazione, riducendo costi operativi ed aumentando flessibilità e agilità.
- Elevato grado di innovazione. L'adozione di SDN accelera l'innovazione poiché consente agli operatori di programmare la rete in tempo reale soddisfacendo requisiti utenti e di business nel momento in cui sorgono. Inoltre, la conoscenza globale della rete permette di simulare e testare nuovi servizi prima di rilasciarli, isolando segmenti di rete per utilizzarli come ambiente di test.
- Migliore esperienza di utilizzo. Attraverso il controllo centralizzato e la visione globale della rete, un'architettura SDN può adattarsi meglio alle richieste dinamiche degli utenti. Per esempio, un operatore può introdurre un servizio video che offre ad utenti premium la migliore risoluzione possibile in modo del tutto trasparente e automatico. Generalizzando, le applicazioni possono chiedere al controller i requisiti di larghezza di banda e connettività, consentendo decisioni più intelligenti circa il routing del traffico.

- Inserimento e concatenazione di servizi. SDN può fornire un'interfaccia "northbound" di tipo RESTful che consente l'inserimento di servizi. Grazie ad interfacce RESTful un operatore riesce a sviluppare le proprie applicazioni ad un alto livello di astrazione, creando servizi velocemente. I servizi possono inoltre essere concatenati dal software di orchestrazione del controller SDN in modo da sfruttare al meglio l'infrastruttura di rete nello stato in cui si trova.

3 Storia e motivazioni di SDN

3.1 Ricerca

All'interno del mondo accademico, i concetti di SDN sono emersi negli ultimi anni grazie alle pubblicazioni dei professori Shenker (Berkeley) e Casado (Stanford) [2]. Secondo la loro visione, il fattore vincente di Internet è stato il modello a livelli (ISO OSI), che ha chiaramente scomposto il problema di trasportare dati in componenti fondamentali. I livelli sono indipendenti ma compatibili e l'innovazione può avvenire in ciascuno di essi. Hanno rappresentato un grande successo industriale, ma un fallimento dal punto di vista accademico, poiché hanno creato un insieme di regole piuttosto che una disciplina ben definita. La tecnologia di rete è un insieme di protocolli, difficili da gestire, che evolvono lentamente, se paragonati ad altri campi dell'IT, come sistemi operativi e database, che si fondano su principi generali di base, sono facili da padroneggiare e sono in continua evoluzione.

Le reti nacquero come protocolli semplici, ma nuovi requisiti di controllo nel tempo hanno portato ad una grande e spesso ingovernabile complessità. Le infrastrutture odierne funzionano solamente perché esistono esperti abili nel padroneggiare questa complessità. La gestione della complessità si oppone all'estrazione della semplicità, che rappresenta una via desiderabile per costruire sistemi di facile utilizzo e comprensione. Nella programmazione, i linguaggi macchina non possedevano nessuna astrazione e gestire la complessità era cruciale per sviluppare applicazioni. Con l'invenzione dei sistemi operativi, alcune funzioni furono astratte dall'hardware, come il file system, la memoria virtuale, i tipi di dato, e nacquero linguaggi di più alto livello. L'astrazione quindi è la chiave per estrarre la semplicità e costruire interfacce verso la modularità. Il piano di controllo delle reti non possiede astrazioni ma queste possono essere derivate da 3 principali requisiti per il problema di controllo delle reti:

- operare senza garanzia di connettività;
- effettuare la configurazione di ogni singolo apparato;
- operare all'interno di un protocollo ad un dato livello della rete.

Le tre astrazioni che ne derivano sono:

- *astrazione delle risorse fisiche*: evita di dover interagire con ogni singolo apparato creando uno stato globale della rete, trasforma il meccanismo di controllo da un insieme di protocolli distribuiti in operazioni su un grafo;

- *astrazione della rete*: definisce un modello semplificato della rete, una virtualizzazione della rete stessa, in modo da specificare un comportamento desiderato senza preoccuparsi di implementarlo in una particolare configurazione fisica;
- *astrazione dell'instradamento dei pacchetti*: definisce un modello flessibile per l'instradamento dei pacchetti. Non dovrebbe rappresentare vincoli per il programma di controllo ma supportare qualsiasi comportamento richiesto. Dovrebbe nascondere i dettagli degli apparati sottostanti ed evolvere indipendentemente dalle specifiche soluzioni prodotte dai fornitori di apparati.

SDN è definita esattamente da queste tre astrazioni, mostrate in ***Figura 4***.

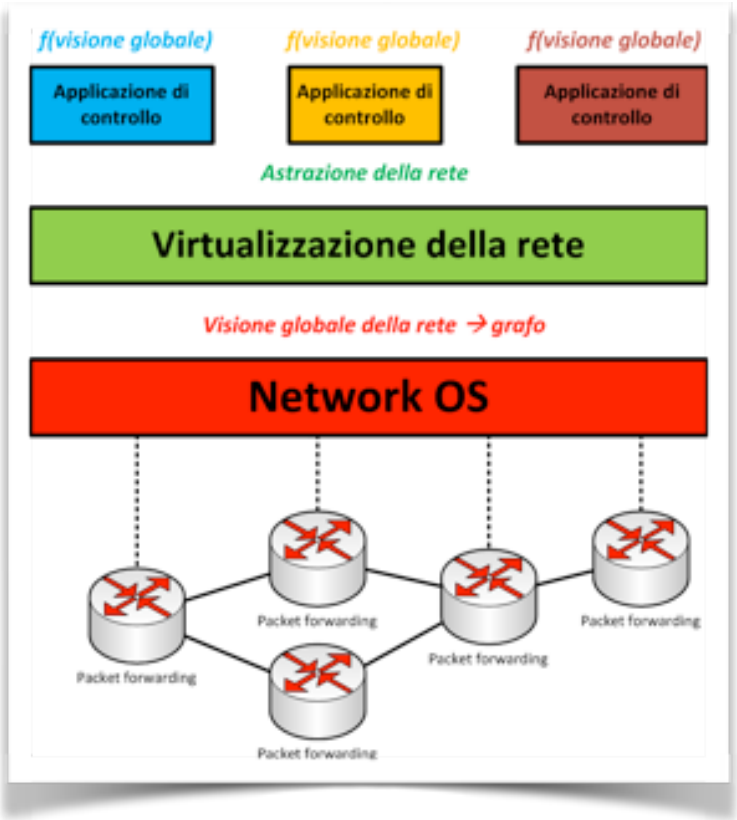


Figura 4
Astrazioni di SDN

Per realizzare le astrazioni citate, il cambiamento principale è rendere il controllo da distribuito a logicamente centralizzato, progettando uno strato intermedio che rilevi informazioni dalla topologia di rete e distribuisca i comandi ai singoli apparati: un vero e proprio sistema operativo di rete. Non c'è più bisogno di protocolli di controllo distribuiti, la configurazione è una funzione centralizzata dello stato globale della rete. Questo cambiamento può dare il via a un'era

software delle reti, diventate così programmabili, opposta ad un'era in cui si definiva "un nuovo protocollo per ogni nuovo problema".

Al di sopra dello strato di sistema operativo di rete, SDN necessita uno strato di controllo condiviso che implementi l'*astrazione della rete*: un'applicazione di controllo configura un modello astratto della rete e questo strato mappa la configurazione virtuale sulla corrispondente configurazione fisica.

Per implementare l'*astrazione dell'instradamento dei pacchetti* bisogna realizzare un'interfaccia comune verso switch veloci e poco intelligenti, rappresentata ad oggi dal protocollo OpenFlow.

Ad una simile architettura consegue una chiara separazione di funzionalità:

- le applicazioni di controllo specificano un comportamento su un modello astratto a seconda dei requisiti dettati dall'operatore;
- la virtualizzazione della rete traduce da modello astratto a visione globale della rete;
- il sistema operativo di rete traduce da stato globale della rete ad apparati fisici.

3.2 Datacenter

Il cloud [15], paradigma nuovo per le infrastrutture ICT e le applicazioni operanti su di esse, rende infrastrutture e applicazioni disponibili come servizi, richiedibili on-demand dagli utenti finali. Nei prossimi anni, sempre più applicazioni e software saranno forniti secondo un modello basato su cloud a smartphone, tablet, console di gioco, tv e nodi della "Internet of Things". La tendenza nel cloud è stata quella di virtualizzare gli apparati ICT fisici, come server e dischi di memorizzazione, rappresentandoli via software: i componenti ICT possono quindi essere gestiti e configurati da remoto, molto più velocemente; diventano elastici, capaci di espandersi e ridursi dinamicamente, in modo da reagire in maniera più efficiente ai cambiamenti in tempo reale delle applicazioni e ai guasti.

Il problema principale dell'infrastruttura cloud è rappresentato dalla rete: sempre più aziende vorranno possedere un data center nel cloud che metta in comunicazione apparati e uffici geograficamente distribuiti in modo semplice e gestibile da remoto. Il fatto che la virtualizzazione della rete non sia così matura come la virtualizzazione dei server ICT rende la gestione di essa, ancora prevalentemente manuale, troppo lenta ed onerosa in termini di costi operativi: per raggiungere elevati livelli di utilizzazione della rete stessa, con conseguente abbattimento dei costi, i service provider hanno bisogno di una gestione più flessibile. Un ulteriore problema è rappresentato dalla capacità di gestire, da parte di un service provider, molti clienti attraverso la stessa infrastruttura di rete, ognuno con macchine virtuali che si accendono e spengono in base all'operatività delle applicazioni, mantenendo ciascun dominio isolato e fornendo differenti livelli di qualità del servizio ad ogni cliente: se la rete è realizzata completamente in hardware statico, come avviene attualmente, la soluzione è assai ardua e costosa. Da qui la necessità di introdurre un'architettura SDN, in cui un controller centrale gestisca dinamicamente apparati di rete virtuali utilizzando uno standard di comunicazione come OpenFlow e possa ridurre drasticamente i costi operativi anche su larga scala.

SDN non trova applicazione concreta solamente all'interno di un data center, ma anche per la connettività tra data center, in cui bisogna ottimizzare i flussi di comunicazione tra i data center stessi e gli utenti finali, bilanciare il traffico tra i data center, operare politiche di risparmio energetico e di manutenzione e continuità del servizio.

3.3 Operatori di telecomunicazioni e fornitori di apparati

Per gli operatori di telecomunicazioni il problema attuale maggiore è dato dagli elevati costi operativi (OPEX) necessari per mantenere la gestione degli apparati il più semplice possibile. I moderni switch e router sono oltremodo complessi e di conseguenza molto costosi da mantenere: implementano più di 6000 RFC tradotti in milioni di righe di codice solo per la parte riguardante il piano di controllo.

Gli oepratori vogliono diminuire i costi derivanti dall'installazione e dalla gestione delle reti, cercando al contempo di acquistare la capacità di fornire servizi in tempi brevi e gestirli dinamicamente mentre sono attivi. L'architettura della maggior parte degli apparati presenti sul mercato oggigiorno è rigida, integrata verticalmente, e basata su soluzioni proprietarie: il software proprietario è fortemente legato all'hardware specializzato ed è ottimizzato per sfruttarlo al meglio. Ciò limita l'innovazione poiché l'introduzione di nuove capacità dipende dai cicli di rilascio, solitamente lunghi, dei produttori (quantomeno rispetto alle esigenze del mercato). Il costo di una rete viene ulteriormente aumentato dalla compresenza di apparati provenienti da differenti produttori, dotati dei propri sistemi di gestione che comportano uno sforzo elevato (in termini di costi e di tempo di sviluppo) di integrazione. In questo scenario, il tempo per fornire un nuovo servizio è lungo non solo per la necessità di configurare manualmente numerosi apparati fisici, ma anche per la necessità di effettuare i relativi cambiamenti nei diversi sistemi di gestione. SDN mira ad appiattire le differenze tra apparati di produttori differenti, poiché si basa su protocolli di gestione e funzionalità standard, con l'obiettivo di eliminare la dipendenza da sistemi proprietari esistenti, ridurre i costi di gestione ed accelerare l'innovazione nei servizi.

I produttori di apparati vedono in SDN l'opportunità di ridurre la complessità dei loro prodotti, poiché possono concentrarsi solamente sull'hardware e sul miglioramento delle prestazioni dei sistemi di instradamento dei pacchetti, sviluppando apparati più efficienti, più velocemente e ad un costo ridotto. Non avrebbero più l'incombenza di aggiungere sofisticati algoritmi di controllo su ogni apparato, ma trasferirebbero la loro conoscenza nell'implementazione di controller in grado di gestire qualsiasi apparato conforme all'architettura SDN. D'altra parte i produttori sviluppano e vendono apparati con software di controllo a bordo, che devono essere supportati per lungo tempo, e spesso questi software forniscono un vantaggio competitivo. Ciò prefigura un'evoluzione industriale simile a quella del mondo IT, in cui hardware e software di base sono diventati commodity appannaggio di pochi fornitori che sfruttano vantaggi di economia di scala. Vi è quindi un delicato compromesso per i fornitori di apparati riguardo alla transizione verso un'architettura SDN:

- tenere il piano di controllo sugli apparati/portarlo su un controller SDN esterno;

- tenere gli apparati chiusi e proprietari/implementare interfacce e protocolli standard di interoperabilità come OpenFlow.

Una soluzione intermedia di transizione graduale sembra essere NFV (*Network Functions Virtualization*, vedere riquadro).

4 Impatto economico di SDN

Recenti studi di mercato e previsioni economiche mostrano in modo concorde che SDN sarà la prossima tecnologia di riferimento nell'industria delle telecomunicazioni:

- Nel 2016, secondo IDC [7], il valore totale del mercato SDN sarà di 3.7 miliardi di dollari.
- Entro il 2014 il 90% dei maggiori produttori di apparati supporterà OpenFlow [11].
- 85% di 315 aziende intervistate sta effettuando ricerche su SDN [9].
- 80% di 21 operatori intervistati sta includendo OpenFlow nei futuri acquisti [3].
- IDC prevede che nel 2016 il fatturato globale dei servizi cloud ammonterà a 73 miliardi di dollari.
- Infonetics [8] nel 2012 ha realizzato un sondaggio chiedendo agli operatori le motivazioni per introdurre SDN: 52% ha indicato "semplificazione del processo di creazione dei servizi"; 48% "creare servizi attualmente non erogabili"; 48% "creare reti virtuali su apparati provenienti da fornitori diversi".

Startup	Acquirente	Periodo	Valore acquisizione (in dollari)
Nicira	VMWare	Luglio 2012	1.26 miliardi
Xsigo	Oracle	Luglio 2012	Non dichiarato
vCider	Cisco	Ottobre 2012	Non dichiarato
Meraki	Cisco	Novembre 2012	1.2 miliardi
Cariden	Cisco	Novembre 2012	141 milioni
Vyatta	Brocade	Novembre 2012	Non dichiarato
Contrail	Juniper	Novembre 2012	176 milioni

Tabella 1
Le maggiori acquisizioni di aziende correlate con SDN

La ***Tabella 1*** evidenzia le ultime e maggiori acquisizioni di startup operanti nell'ambito di SDN con sviluppo di tecnologie e prodotti da parte delle principali aziende del mondo ICT, a ulteriore conferma di un mercato in fermento e di una volontà da parte degli attori più influenti di investire in questa tecnologia.

5 Sviluppi e criticità

SDN è ancora in uno stato di sviluppo iniziale e consolidamento, sia per quanto riguarda la standardizzazione sia per quanto riguarda l'implementazione industriale. Rimangono perciò alcuni dubbi sulla direzione che potrà prendere il suo sviluppo.

5.1 NaaS (Network as a Service)

Un obiettivo a lungo termine di SDN è abilitare gli operatori di rete nell'offrire la rete come servizio (NaaS): come accaduto per gli apparati ICT, gli apparati fisici di rete sono virtualizzati e convertiti in un unsieme di risorse dinamicamente utilizzabili.

5.2 Transizione

I produttori di apparati stanno cominciando ad abbracciare i concetti di SDN, ma è ancora non definito quante funzionalità di controllo rimarranno distribuite sugli apparati stessi. Inoltre gli operatori non possono costruire una rete SDN pura, poiché possiedono già una base installata e funzionante. Un approccio ibrido, con funzionalità SDN introdotte gradualmente su reti esistenti, schematizzato in ***Figura 5***, può essere una soluzione percorribile.

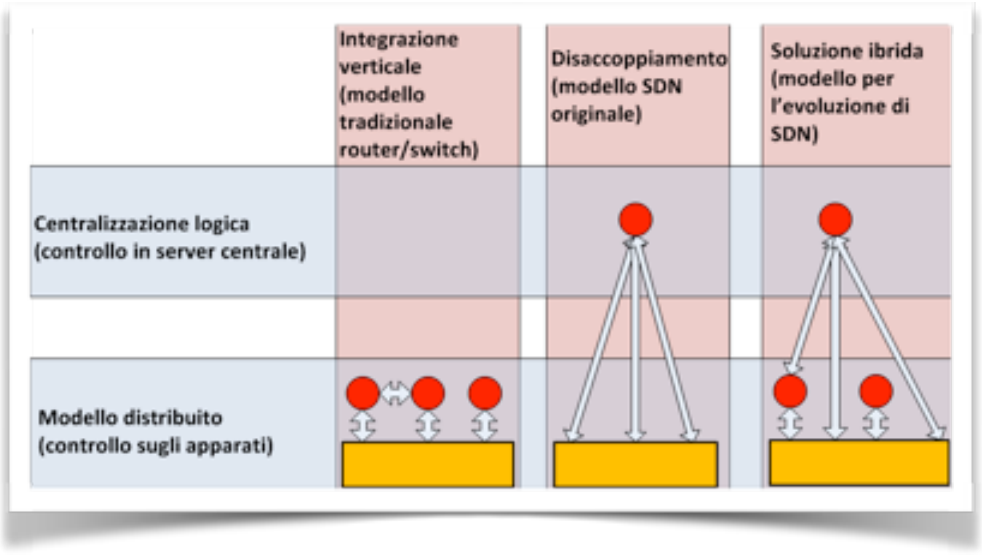


Figura 5
Architettura SDN ibrida

5.3 Estensione alle reti di telecomunicazione

L'interesse iniziale di SDN è stato sollevato dai problemi dei data center. L'architettura SDN può sicuramente essere estesa anche alle reti di telecomunicazione. Tuttavia una rete di un data center ha alcune caratteristiche peculiari:

- la banda interna può essere effettivamente considerata infinita e libera;
- l'incremento di capacità non è costoso e solitamente non comporta l'introduzione di nuove installazioni, dal momento che nella pratica la rete è sovradimensionata;
- l'ambiente è relativamente omogeno, i server interagiscono secondo modalità ben dominate;
- SDN ed OpenFlow possono essere facilmente introdotti.

Una rete di telecomunicazione è molto diversa:

- la banda dipende dall'infrastruttura presente in campo, installata a costi molto elevati;
- la connettività non è completamente magliata né sovradimensionata, per problemi di costi;
- l'ambiente presenta diversi tipi di tecnologie (reti a pacchetti, reti radio);
- SDN e OpenFlow non possono facilmente soppiantare sistemi già esistenti come NETCONF, SNMP, GMPLS.

Se si accetta la consueta divisione di una rete di un operatore di telecomunicazione in *core*, *metro* e *accesso*, si evidenzia che:

- la rete di accesso è caratterizzata da una singola connessione da un nodo cosiddetto "*edge*" (di bordo) a un cliente, in cui non vi sono particolari risorse da dover controllare in maniera efficiente;
- la rete *core* è invece fortemente aggregata e trasporta il traffico di tutti i clienti, il quale è relativamente prevedibile e controllabile in quel punto della rete;
- la rete *metro* è soggetta a fluttuazioni di traffico e picchi di richieste non prevedibili e solitamente vi risiedono risorse IT condivise quali server di memorizzazione dei dati e applicazioni.

SDN può quindi essere di grande utilità particolarmente in quest'ultimo settore della rete di un operatore, la cui gestione è per sua natura più complessa e costosa.

5.4 OpenFlow

Non è al momento consolidato, tra service provider e fornitori di apparati, il ruolo di OpenFlow come protocollo principale di comunicazione tra switch/router e controller.

A causa della sua natura di origine accademica, è stato inizialmente implementato per rispondere ad esigenze di studio e sostanzialmente semplificate rispetto alle esigenze operative delle reti di telecomunicazione. Negli ultimi anni si è quindi assistito ad un processo di sua evoluzione, teso a svilupparne le caratteristiche necessarie per renderlo un protocollo maturo per esigenze industriali, quali scalabilità e sicurezza [10]. È pensiero comune dei produttori la necessità di far evolvere OpenFlow per coprire alcune funzionalità attualmente implementate da hardware specializzato ed allo stesso tempo sviluppare hardware a caratteristiche sempre più simili, come già accaduto durante l'affermazione ed il consolidamento della virtualizzazione dei calcolatori.

Non c'è accordo sulla possibile collocazione del protocollo OpenFlow:

- come interfaccia dal controller SDN verso gli apparati fisici;
- come un controller diretto degli apparati fisici;
- come un controller di *Element Manager Systems* (sistemi di gestione di famiglie di apparati omogenei) esistenti, trasformando il controller SDN in un gestore di gestori.

5.5 Il ruolo del controller

Il controller SDN presenta delle criticità riguardo a scalabilità, ridondanza e prestazione. Non è ancora definito se il controller sarà un'applicazione all'interno degli OSS esistenti o andrà a sostituire totalmente un OSS [4].

6 Conclusioni

Il futuro delle reti sarà sempre più basato sul software, che permetterà di accelerare il tasso di innovazione nello stesso modo in cui ha agito per il dominio dei server. SDN promette di trasformare le reti statiche odierne in piattaforme flessibili e programmabili dotate di "intelligenza" per allocare risorse dinamicamente, scalare per supportare un'enorme quantità di nodi e consentire una virtualizzazione ottimale per costruire ambienti cloud dinamici, altamente automatizzati e sicuri. Con i tanti vantaggi individuati dal mondo accademico e i grandi investimenti dell'industria, SDN vuole diventare il nuovo paradigma per le reti di telecomunicazione: l'elevato potenziale di SDN di facilitare la gestione di reti virtuali, supportare il cloud computing, ridurre i costi ed incrementare l'agilità degli operatori di rete sta convincendo fornitori e service provider ad adottare ed investire su SDN e OpenFlow.

Il potenziale impatto del paradigma SDN sulle reti di prossima generazione può anche abilitare lo sviluppo di innovativi servizi ICT, costituenti opportunità di sviluppo e vantaggio competitivo per gli operatori in grado di sfruttare al meglio la nuova architettura.

La sua potenzialità è quella di portare lo stesso livello di innovazione e sviluppo che l'introduzione dei sistemi operativi ha portato nel mondo ICT.

Le soluzioni SDN consentono alle applicazioni di rete di ottenere una vista astratta globale della rete, fornita da un piano di controllo logicamente centralizzato; diventa quindi possibile implementare politiche e logiche di controllo evitando la complessità fisica data dalla moltitudine e dall'eterogeneità

degli apparati. SDN si configura come un ecosistema di moduli software di controllo, interoperabili e capaci di facilitare ed ottimizzare le gestione delle risorse di rete; esistono tuttavia alcune criticità, quali prestazioni, affidabilità, scalabilità, stabilità e sicurezza, dovute alla non completa maturazione della tecnologia, ancora in una fase di definizione, accettazione globale e consolidamento.

L'interesse verso il paradigma SDN è confermato dal crescente numero di iniziative di carattere industriale ed accademico (conferenze, simposi, joint venture, ecc.), dalla nascita di alcune grandi comunità open source di sviluppatori di controller SDN (OpenDaylight, NOX/POX, Floodlight) e dalla costituzione di diverse start-up di rilievo. SDN è una tecnologia ormai accettata dall'industria come prossima evoluzione delle reti e trainata da una serie di esigenze ineludibili (sviluppo del cloud, virtualizzazione delle risorse, ecc.).

Il successo di SDN dipenderà da quanto i vantaggi auspicati saranno trasformati in realtà e da come sarà gestita la fase di transizione dalle reti attuali: deve esserne confermata la fattibilità tecnologica, a seguito di una maturazione degli standard ed un ampliamento delle funzionalità per coprire tutti i requisiti odierni degli operatori di rete.

Riquadro 1 – I piani del processo di routing nelle reti

In una rete di telecomunicazioni a pacchetto, ogni apparato (router/switch) che operi la funzione di instradamento dei dati (*routing*) viene abitualmente suddiviso in tre livelli (o piani) funzionali:

- piano dei dati o di instradamento dei pacchetti, in cui l'apparato processa il traffico che lo attraversa e decide come trattare ogni singolo pacchetto ricevuto;
- piano di controllo, in cui l'apparato interagisce con la rete, creando una mappa della rete circostante e partecipando ai protocolli di *routing* per determinare l'algoritmo di instradamento dei pacchetti;
- piano di gestione, in cui interagisce con sistemi di manutenzione, attraverso protocolli standard quali SNMP, SSH, o altri protocolli proprietari, fornendo comandi e allarmi.

Riquadro 2 – NFV (Network Functions Virtualization

Un nuovo gruppo interno all'ETSI, guidato da operatori di rete, sta lavorando alla definizione di NFV [6]. NFV mira alla virtualizzazione di molte funzioni di rete in server e servizi standard da collocare in data center, nodi di rete e apparati degli utenti. Presuppone l'implementazione di funzioni di rete in moduli software eseguibili su qualsiasi server standard e trasferibili dinamicamente su qualsiasi risorsa all'interno di una rete, senza il bisogno di installare nuovi apparati fisici. NFV può essere considerata complementare a SDN in quanto fornisce gli elementi di base su cui un controller SDN può essere costruito. NFV è applicabile a qualsiasi funzione del piano dei dati o del piano di controllo, in reti fisse e mobili. Potenziali esempi di funzionalità di rete virtualizzabili sono: nodi di rete mobile, analisi del traffico, monitoraggio e gestione di SLA (Service Level Agreement), accounting, controllo delle configurazioni, firewall, sistemi di rilevamento di intrusioni, protezione dallo spam. I benefici di NFV includono:

- riduzione dei costi degli apparati, legata alla loro semplificazione;
- aumento della velocità di innovazione e di erogazione dei servizi;
- apertura del mercato ad attori puramente software;
- migliore ottimizzazione della rete

Riquadro 3 –L'esperienza di Google

Nell'aprile 2012 Google ha realizzato una rete WAN SDN. Google necessitava una rete con migliori prestazioni, migliore tolleranza ai guasti e facilmente gestibile come una singola entità e non come un insieme di apparati distribuiti. La rete WAN di Google è organizzata in due settori principali: una rete (I-scale) esterna verso Internet e gli utenti e una rete interna (G-scale) che trasporta il traffico tra i vari data center, con caratteristiche e requisiti ben diversi. Google ha testato una rete SDN basata su OpenFlow in un parte della rete interna: non essendoci, al momento dei primi sviluppi, prodotti compatibili con OpenFlow sul mercato, ha costruito nuovi switch che supportassero il protocollo OpenFlow. Il controller SDN è stato replicato per evitare di avere un singolo punto di vulnerabilità ed è stato sviluppato anche un servizio centralizzato di ingegneria del traffico: esso raccoglie in tempo reale i dati di utilizzo della rete e la richiesta di banda delle applicazioni, programmando di conseguenza i migliori percorsi di rete utilizzando OpenFlow. Tale rete SDN è stata installata in una porzione di WAN che serve 12 datacenter in tutto il mondo collegati da link a 10Gbps e il rate di utilizzo è arrivato quasi al 100%. I benefici rilevati da Google sono stati:

- visione unificata della rete;
- alto rate di utilizzo della rete, grazie al servizio di ingegneria del traffico centralizzato;
- gestione dei guasti più veloce;
- aggiornamenti (software) possibili senza intaccare il funzionamento della rete;
- ambiente di test molto fedele alla rete in produzione;
- elasticità delle risorse di rete.

I principali problemi incontrati sono stati:

- l'immaturità del protocollo OpenFlow, in termini di sicurezza, funzionalità e scalabilità, comunque sufficiente per installare il primo prototipo;
- tolleranza ai guasti del controller SDN, che deve essere replicato con l'introduzione di meccanismi complessi quali algoritmi di elezione del master e partizione della rete;
- programmazione dei flussi: per reti di grandi dimensioni, programmare singolarmente ogni flusso di pacchetti è temporalmente oneroso.

Bibliografia

[1] Software Driven Networks Problem Statement, IETF NetworkWorking Group Internet-Draft, 2011

[2] Casado, M., Koponen, T., Ramanathan, R., Shenker, S., Virtualizing the network forwarding plane, Proc. of the Workshop on Programmable Routers for Extensible Services of Tomorrow, 2010

[3] Duffy, J., Many Cisco carrier customers planning SDNs, NetworkWorld, 2012

[4] Fernandez, M. P., Evaluating OpenFlow Controller Paradigms, ICN, 2013

[5] Gude, N. et al., NOX: Towards an Operating System for Networks, ACM SIGCOMM, 2008

[6] http://portal.etsi.org/NFV/NFV_White_Paper.pdf

[7] http://www.idc.com

[8] http://www.infonetics.com

[9] Kerravala, Z., SDNs provide a solution for Extreme Networks' XOS, NetworkWorld, 2012

[10] Kobayashi, M. et al., Maturing of OpenFlow and Software Defined Networking through Deployments, August 14, 2012
[11] MacDonald, N., The impact of Software Defined Data Centers on Information Security, Gartner, 2012
[12] McKeown, N. et al., OpenFlow: Enabling Innovation in Campus Networks, ACM SIGCOMM, 2008
[13] ONF, Software-Defined Networking: The New Norm for Networks, ONF white paper, April 13, 2012
[14] OpenFlow Switch Specification. Version 1.3.2, ONF Specification, April 25, 2012
[15] Sher DeCusatis, C. J., Carranza, A., DeCusatis, C. M., Communication within Clouds: Open Standards and Proprietary Protocols for Data Center Networking, IEEE Communications Magazine, September 2012

Biografie

Simone Mangiante ha conseguito il titolo di dottore di ricerca in Ingegneria Informatica all'Università di Genova nel 2013, all'interno del Centro Interuniversitario per le Piattaforme Informatiche (CIPI), occupandosi principalmente di reti di telecomunicazione e Software Defined Networking. Ha collaborato con M3S, azienda spin-off dell'Università di Genova, dove ha maturato esperienze nell'ambito delle reti, dei suoi protocolli di controllo e gestione e dei sistemi distribuiti. Attualmente collabora con il CIPI e occupa una posizione di ricercatore post doc all'Università di Genova.

E-mail: s.mangiante@m3s.it

Pierpaolo Baglietto si è laureato in Ingegneria Elettronica presso l'Università di Genova dove nel 1994 ha conseguito il Dottorato di Ricerca in Ingegneria Elettronica ed Informatica. È attualmente Professore Associato presso la Facoltà di Ingegneria dell'Università di Genova dove è docente dei corsi di Calcolatori Elettronici e di Engineering of Network and Computing Platforms. Dal 2013 è Direttore del Centro Interuniversitario di Ricerca sull'Ingegneria delle Piattaforme Informatiche (CIPI). I suoi interessi di ricerca sono centrati sui sistemi distribuiti, i sistemi di monitoraggio e gestione delle reti di nuova generazione e le piattaforme per lo sviluppo e l'erogazione di servizi.

E-mail: p.baglietto@cipi.unige.it

Massimo Enrico si è laureato in Ingegneria Elettronica presso l'Università di Genova nel 1993. È attualmente un dirigente dell'Ericsson presso il centro di R&D di Genova dove si occupa dello sviluppo di sistemi di gestione per reti di telecomunicazione. Prima di assumere questa posizione, ha lavorato diversi anni nel Product Management e Marketing in diversi ruoli quali Sales & Marketing e Ricerca & Innovazione in ambito internazionale. Ha iniziato la sua carriera in Marconi, dove ha ricoperto il ruolo di Vice Direttore Vendite EMEA & Customer Marketing Engineering. Collabora da diversi anni con l'Università di Genova in progetti di innovazione legati alle reti di telecomunicazione.

E-mail: massimo.enrico@ericsson.com

Dispositivi mobili: nuovi problemi di sicurezza

Mauro Migliardi - Alessio Merlo

L'evoluzione tecnologica degli ultimi anni ha visto molti e diversi fattori cambiare radicalmente il panorama della sicurezza informatica. Tra questi fattori sono certamente degni di citazione la radicalizzazione nell'uso di dispositivi capaci di lavorare in mobilita' quali smartphone e tablet sempre più evoluti e complessi, la diffusione pervasiva sia nello spazio che nel tempo di tecnologie di connessione ad alta velocità come le reti WiMax, il 3G e il nascente LTE e, ultimo ma non meno importante, la crescente dipendenza della forza lavoro altamente tecnologizzata da dispositivi che sono sempre meno controllati dai reparti IT aziendali. In questo articolo analizzeremo come cambiano le prospettive e i problemi della sicurezza informatica a fronte di questo cambio epocale, discutendo le nuove sfide che tale settore di ricerca dovrà affrontare anche a fronte della crescente attenzione verso il consumo e la dipendenza energetica delle componenti l'infrastruttura stessa.

Keywords: Mobile computing, Security, BYOD, GSM/UMTS networks

1 Introduzione

La sicurezza è, da sempre, uno degli aspetti critici e fondamentali nella progettazione e costruzione dei sistemi informatici. Storicamente, il modello di riferimento per i sistemi di sicurezza è stato quello che si riferiva metaforicamente alle fortezze medioevali, ovvero uno in cui un insieme di strutture difensive posizionate staticamente tra le informazioni da proteggere ed il resto del mondo permetteva di controllare chi e come accedesse a quelle stesse informazioni.

Un primo cambiamento epocale per questo modello si è certamente verificato in passato con l'avvento di Internet; tuttavia, oggi la sicurezza informatica si trova ad affrontare una nuova e probabilmente ancora più impegnativa sfida. Alla base di questa nuova sfida si possono identificare tre fattori fondamentali. Il primo è la

mobilità dei punti di accesso; ogni modello di difesa perimetrale è messo in crisi dalla possibilità di trasportare facilmente all'interno di ogni linea di difesa un sistema che può sia facilmente interconnettere le reti interne con quelle pubbliche, sia rappresentare un vettore di "infezione". Un secondo aspetto è dato dalla crescente dipendenza delle pratiche quotidiane di lavoro da strumenti informatici mobili come smartphones e tablet PC; questi dispositivi, infatti, avendo una capacità energetica intrinsecamente limitata, forniscono una sfaccettatura per la loro superficie d'attacco che gli strumenti correnti non sono pronti a difendere adeguatamente. Infine, un ultimo, ma non meno significativo aspetto, è dato dalla natura ibrida, tra telefono e computer, dei nuovi smartphone e dal fatto che essi convivono con una rete nata per dispositivi "stupidi". La diffusione di questi telefoni intelligenti, infatti, mette in contatto l'infrastruttura di rete telefonica con dispositivi anche di notevole capacità computazionale e capaci di essere controllati a distanza da entità malevole fornendo quindi la possibilità di realizzare campagne di attacco nei confronti della rete telefonica impensabili sino a pochi anni fa.

In questo articolo vogliamo analizzare questi recenti aspetti dell'evoluzione informatica nella loro capacità di rivoluzionare i modelli della sicurezza e discutere come la comunità scientifica si stia dedicando ad adeguare le nostre conoscenze ed anziché e a produrre strumenti per evitare di essere sorpresi dalla rivoluzione in atto.

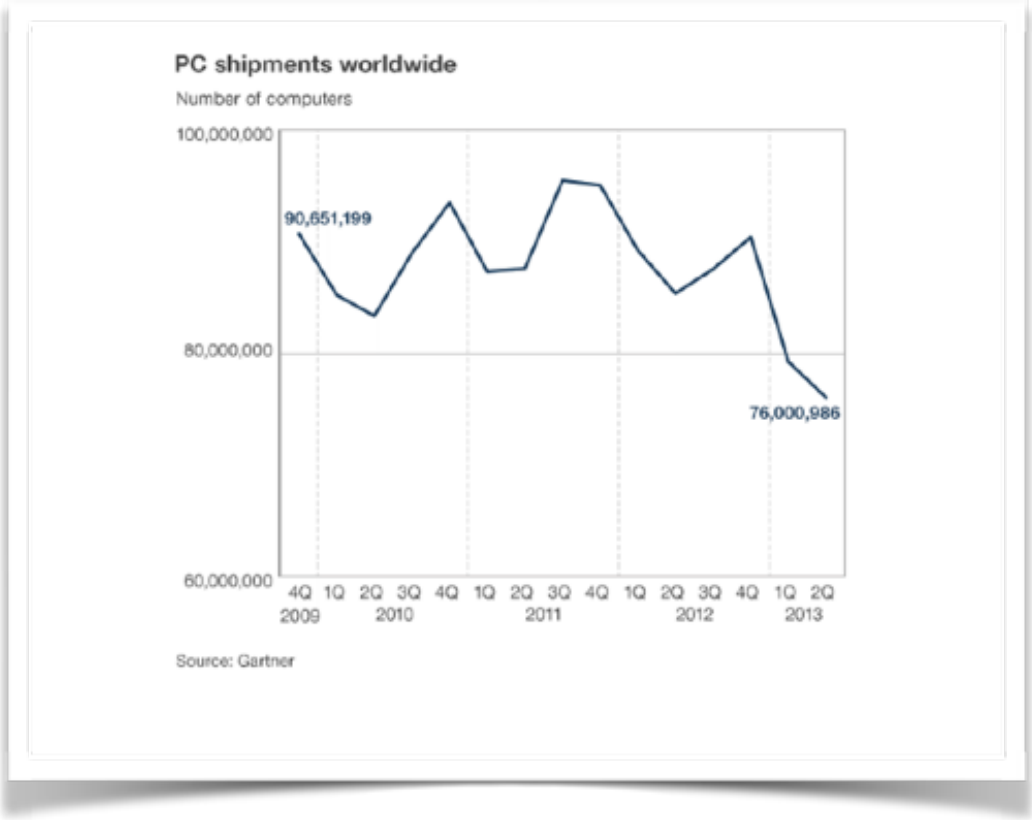


Figura 1
Evoluzione delle vendite di PC negli ultimi 4 anni (fonte Gartner)

2 La crisi del modello a fortezza

Alla base del modello metaforicamente definito "a fortezza" per la sicurezza informatica vi è l'idea di restringere il più possibile i punti di accesso alla dimensione da proteggere (i *cancelli* che permettono l'accesso al cortile) e presidiare queste con adeguate strutture di controllo.

2.1 Identificazione delle difese perimetrali tradizionali

In un sistema informatico questo si traduce in un insieme di difese perimetrali (firewall, IDS, filtri per i contenuti) uniti a un controllo degli accessi.

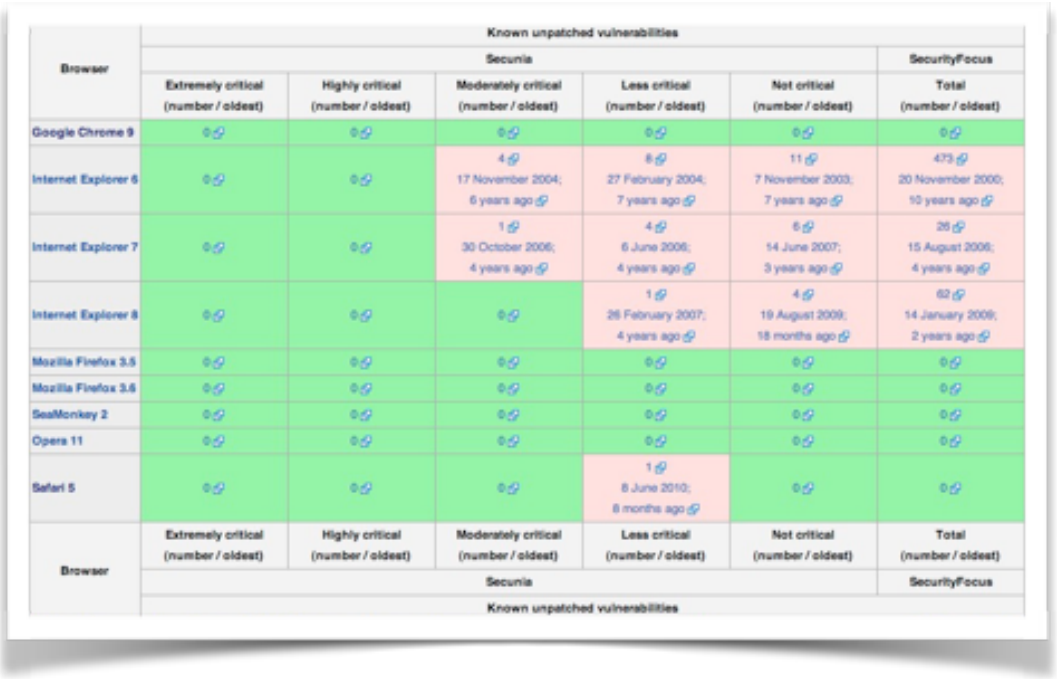
La recente evoluzione dei sistemi di autenticazione con meccanismi biometrici atti a superare la classica “coppia” login e password (si pensi ad esempio all'introduzione di un lettore d’impronte digitali sul nuovo smartphone di Apple, l’iPhone 5s) va oltre lo scopo di quest’articolo, ma vale comunque la pena di citare il fatto che il rafforzamento di questo aspetto, pur essendo fondamentale, non è sufficiente, se utilizzato secondo la classica logica di difesa “perimetrale”, a mettere i sistemi informatici al riparo dalle nuove minacce. Come analizzeremo nel prosieguo dell'articolo, infatti, queste nuove minacce tendono a sgretolare il perimetro rendendolo poroso e permeabile a molteplici vettori di intrusione.

3 La mobilità e le sue implicazioni

Mobilità è una parola chiave del mercato informatico recente. Ormai da alcuni anni le vendite di PC fissi sono in declino (vedi Figura 1) a favore prima di laptop, poi di netbook ed infine di tablet. Questa tendenza del mercato, sebbene non sia assimilabile come alcuni vogliono portare a credere con la morte del Personal Computer così come lo conosciamo, implica comunque una significativa espansione del numero di utenti dotati di un dispositivo mobile. Inizialmente, i laptop erano appannaggio dei livelli superiori delle aziende e dei cosiddetti “esterni”, ovvero di quegli impiegati che passavano la maggior parte del loro tempo al di fuori delle pareti aziendali. Il numero limitato di queste figure (all'interno della singola azienda) permetteva l'adozione di soluzioni ad-hoc, di limitato impatto sui sistemi aziendali, quali l'introduzione di *Virtual Private Network* (VPN) per simulare la presenza del dispositivo mobile all'interno del perimetro aziendale dotandolo di un'interfaccia virtuale che appariva essere parte della Intranet aziendale. Le implicazioni di sicurezza dovute alla presenza di questa VPN rispetto al modello di difesa perimetrale erano piuttosto limitate in quanto si poteva tranquillamente considerarla una porta di accesso secondaria con utenza ben delineata e soggetta a tutti i meccanismi di controllo degli accessi considerati necessari.

3.1 Una nuova tipologia di utenti, di dispositivi, di connettività

La situazione, tuttavia, cambia sostanzialmente nel momento in cui l'evoluzione tecnologica, provocando un significativo calo dei prezzi, ha portato ad avere sempre più utenti dotati di un computer portatile. Dagli “esterni”, utenti tendenzialmente forzati all'uso dei sistemi portatili e, in molti casi, atti a vedere il laptop come un male necessario allo svolgimento del loro lavoro, si passa quindi ad un gran numero di utilizzatori che vedono nel computer portatile una comodità atta a permettergli sia di "portare il lavoro a casa", sia di avere un ambiente integrato di lavoro e di uso domestico.



Browser	Known unpatched vulnerabilities					
	Secunia					SecurityFocus
	Extremely critical (number / oldest)	Highly critical (number / oldest)	Moderately critical (number / oldest)	Less critical (number / oldest)	Not critical (number / oldest)	Total (number / oldest)
Google Chrome 9	0	0	0	0	0	0
Internet Explorer 6	0	0	4 17 November 2004; 6 years ago	8 27 February 2004; 7 years ago	11 7 November 2003; 7 years ago	473 20 November 2000; 10 years ago
Internet Explorer 7	0	0	1 30 October 2006; 4 years ago	4 6 June 2006; 4 years ago	6 14 June 2007; 3 years ago	26 15 August 2006; 4 years ago
Internet Explorer 8	0	0	0	1 26 February 2007; 4 years ago	4 19 August 2009; 18 months ago	62 14 January 2009; 2 years ago
Mozilla Firefox 3.5	0	0	0	0	0	0
Mozilla Firefox 3.6	0	0	0	0	0	0
SeaMonkey 2	0	0	0	0	0	0
Opera 11	0	0	0	0	0	0
Safari 5	0	0	0	1 8 June 2010; 8 months ago	0	0
Browser	Extremely critical (number / oldest)	Highly critical (number / oldest)	Moderately critical (number / oldest)	Less critical (number / oldest)	Not critical (number / oldest)	Total (number / oldest)
	Secunia					SecurityFocus
	Known unpatched vulnerabilities					

Figura 2
Quadro delle vulnerabilità non risolte nei principali web browsers al Marzo 2011 (fonte: http://en.wikipedia.org/wiki/Comparison_of_web_browser#Vulnerabilities)

Questo fatto costringe il reparto IT delle aziende a rinforzare considerevolmente le politiche di utilizzo ammissibile per i dispositivi e porta, nella grande maggioranza dei casi, a non permettere l'installazione di alcun pacchetto software al di fuori di quanto previsto dall'azienda. Nonostante questo, l'utenza continua generalmente a percepire il computer portatile come uno strumento di accesso generico ad Internet non ristretto al solo uso aziendale e le limitazioni imposte centralmente non possono quindi essere considerate sufficienti a garantire l'integrità dello strumento di lavoro. Infatti, anche la sola navigazione su Internet può portare all'inquinamento della stazione di lavoro nel caso in cui si visitino siti malevoli capaci di sfruttare le vulnerabilità dei browser [1] come grimaldello per penetrare il sistema sottostante. L'importanza dei browser nel panorama informatico moderno fa sì che la stragrande maggioranza delle falle presenti venga rapidamente corretta non appena queste sono scoperte (si osservi in Figura 2 una valutazione delle vulnerabilità rimaste non gestite al marzo 2011), tuttavia, esiste sempre una finestra di pericolo in quanto nuovi punti attaccabili sono scoperti con una frequenza non nulla come osservabile sulle basi di dati dedicate a tenere traccia di questi eventi (si consideri ad esempio il caso di Mozilla Firefox [2]). Nonostante questo, il pericolo rappresentato da una stazione di lavoro compromessa ma dotata di sola connettività in rete locale (con o senza fili) è controllabile in quanto, nel momento stesso in cui si introduce all'interno dei perimetri di difesa aziendali, ha bisogno di utilizzare l'infrastruttura dell'azienda stessa per contattare qualsiasi altro nodo ed è quindi osservabile, identificabile e, potenzialmente, neutralizzabile dalle difese perimetrali. Si può ovviamente pensare a vettori ad hoc, in grado di distinguere in base alla

configurazione di rete la situazione corrente e adattare in modo automatico il proprio comportamento; tuttavia questo tipo di attacco ha costi sensibili e richiede competenze che lo pongono decisamente al di sopra delle minacce più comuni [3].

Una situazione completamente diversa si pone nel caso di smartphone e tablet PC dotati di connettività 2G/3G/4G. Questi ultimi, infatti, non solo non richiedono l'utilizzo delle risorse di rete aziendali per accedere a risorse esterne, ma possono costituire essi stessi dei punti di ingresso diretto alle aree protette. Infatti, mentre un dispositivo dotato di sola connettività in rete locale (con o senza fili) non è, nella maggior parte dei casi, in grado di connettersi a reti esterne all'azienda per aprire una connessione non controllata tra dentro e fuori, ogni smartphone recente, così come ogni tablet dotato di connettività 2G/3G/4G, dal 2011 in poi nasce con la predisposizione ad agire come router. In questa situazione, il modello a fortezza diventa completamente poroso: ciascun utente dotato di uno smartphone o di un tablet diventa il portatore di un punto di accesso alternativo (collegando appunto la rete locale con la rete cellulare) a quelli più o meno fortemente ed efficacemente sorvegliati dai sistemi di sicurezza aziendali e rende quindi del tutto insufficienti le metodologie di sicurezza tradizionali. In analogia con il modello a fortezza, ove si controllano con attenzione un numero fisso e noto a priori di punti porte di comunicazione tra dentro e fuori (non è un caso che il termine con cui si indicano le porte di una fortezza sia appunto *gateway*), la presenza di smartphone e tablet che connettono la rete locale alla rete cellulare 2G/3G/4G si configura l'inattesa comparsa di un insieme di stazioni mobili di teletrasporto sparse all'interno dell'area da proteggere: una situazione che vanifica completamente i controlli posti in essere nei punti di scambio predefiniti.

Appare quindi chiaro, a questo punto, come non sia più possibile parlare di difese perimetrali ma diventi invece necessario affrontare il problema della sicurezza in modo più capillare: è necessario controllare ciascun elemento secondo una strategia di "sicurezza del terminale" [4] in quanto ciascun elemento, appunto, diventa un "terminale" attaccabile direttamente o quasi.

Le nuove tecnologie legate "al contesto" forniscono un elemento in più all'arsenale a disposizione della sicurezza aziendale permettendo di introdurre un ulteriore elemento di selezione dei meccanismi di controllo degli accessi [5]; allo stesso tempo, gli elementi di sensorizzazione che permettono di costruire il contesto in cui lavorare e le informazioni stesse che costituiscono il contesto rappresentano un elemento in più da proteggere e richiedono soluzioni ad-hoc [6].

Un primo esempio applicato di sicurezza legata al contesto è rappresentato dal *geo-fencing*, ovvero dalla creazione di un perimetro virtuale sensibile al transito di determinati dispositivi. Il perimetro può essere determinato in modo semplice dalla distanza radiale da un singolo punto (per esempio calcolando la potenza con cui si riceve il segnale di un "faro"), oppure assumere forme più elaborate sfruttando tecnologie aggiuntive per il posizionamento (il GPS) o dispositivi di prossimità (RFID e NFC). Con la determinazione di un perimetro virtuale, è possibile modificare il comportamento sia dei singoli terminali che dei dispositivi con cui questi interagiscono. Si consideri ad esempio il caso in cui un'azienda abbia istituito un *perimetro* al limite del suo edificio principale: in un settaggio come questo sarebbe possibile sia ai singoli terminali mobili di percepire quando entrano all'interno dell'edificio, sia alle strutture di sicurezza dell'edificio stesso di percepire quali dispositivi mobili entrino o escano. Questo controllo della

Ingegneria Sociale

Si definisce Ingegneria Sociale la costruzione sistematica e metodica di un profilo del bersaglio dell'attacco a partire dai suoi comportamenti quotidiani, allo scopo di identificare i suoi punti deboli e le modalità più efficaci per carpire informazioni non direttamente disponibili. Si tratta di ingegneria sociale la classica analisi dei rifiuti prodotti dal bersaglio esemplificata in tante pellicole d'azione, ma anche l'analisi del traffico Internet di un soggetto volta alla costruzione di attacchi di tipo phishing più evoluti e plausibili.

posizione può essere utilizzata in termini di sicurezza in diversi modi.

Sarebbe possibile, ad esempio, richiedere ai dispositivi in ingresso all'azienda di disabilitare automaticamente la connettività 2G/3G a fronte di una fornitura contestuale di connettività in rete locale senza fili tramite le infrastrutture dell'azienda. In questo modo si verrebbe ad eliminare l'effetto portale incontrollato che pone a rischio elementi interni della rete aziendale.

Un secondo esempio di azioni di sicurezza attivate dall'attraversamento del *perimetro* (da fuori a dentro) è il costringere ogni dispositivo a superare un controllo di integrità prima di essere ammesso all'interno della rete aziendale. Una pratica di questo genere riduce sostanzialmente la pericolosità dei dispositivi stessi in qualità di vettori di programmi malevoli (*malware*) e agenti aggressivi in generale.

Infine, un ultimo esempio di controllo di sicurezza attivato dall'attraversamento della *geo-fence* (questa volta da dentro a fuori) è la scansione dei contenuti di ogni dispositivo in uscita per assicurarsi che non vi siano dati incompatibili con il livello di sicurezza del dispositivo e/o dell'utilizzatore del dispositivo.

3.2 La sfida del "Bring Your Own Device"

Se inizialmente il rischio principale di "inquinamento" degli ambienti di lavoro era rappresentato dall'utilizzo degli strumenti aziendali anche per scopi diversi (per esempio navigazione in Internet con il laptop aziendale), la massiccia diffusione di dispositivi come smartphone e tablet PC ha portato oggi gli utenti ad essere sempre più dipendenti da essi e a ribaltare la logica di utilizzo precedente: non si usa più il laptop aziendale per scopi personali, ma si vuole utilizzare lo smartphone o il tablet privato anche in ambito lavorativo in quanto il dispositivo è comunque ormai assolutamente capace di fornire i servizi richiesti anche dalle necessità aziendali.

Questa tendenza è etichettata con il termine *Bring Your Own Device* (BYOD).

I centri IT aziendali tendono a resistere a questa richiesta vedendo con allarme la perdita di controllo del parco dispositivi che essa rappresenta. In effetti, è del tutto palese che il proliferare di soluzioni tecnologiche e configurazioni diverse renda impossibile al momento mantenere un vero controllo e la capacità di standardizzare le procedure di intervento sul parco dispositivi aziendale.

Tale resistenza, comunque, rappresenta una battaglia persa in partenza in quanto gli utenti data la loro ormai conclamata dipendenza dai nuovi dispositivi non sono disponibili a rinunciarvi: il risultato di un irrigidimento da parte dell'azienda è quindi, al più, quello di avere utenti con due o più dispositivi similari, in uso contemporaneo con la tendenza a replicare dati e processi

operativi ed una minore capacità di gestione degli stessi dato l'aumento della complessità.

Il superamento di questo empasse ci sarà solo quando si troveranno modalità per inglobare completamente il BYOD nelle politiche di sicurezza aziendali [7] tuttavia, anche utilizzando tecniche di sicurezza basate sul contesto delineate in precedenza, implementare una politica sicura per il BYOD in ambito aziendale rimane una sfida significativa.

Il primo passo è sicuramente rappresentato da una fase di responsabilizzazione degli utenti. Infatti, se il rischio più grande per la sicurezza aziendale è sempre rappresentato dall'utente disattento facile preda dell'ingegneria sociale, il livello di libertà offerto dal BYOD deve essere accompagnato da una parallela crescita della responsabilità dei singoli nella gestione dei loro dispositivi e dei dati.

A tale proposito, le guide pubblicate da diversi istituti dedicati alla sicurezza informatica pongono specifico accento sulla necessità di formulare in modo chiaro e completo quali siano gli usi accettabili per i dispositivi coinvolti nel BYOD, quali siano le politiche di uso dei dati aziendali con i dispositivi coinvolti nel BYOD e, non ultimo, quali siano i diritti dell'azienda nei confronti dei dati privati contenuti nel dispositivo stesso. Un esempio classico di problema è quello dato dai sistemi di *Mobile Device Management* (MDM). Per limitare i rischi di fuga di notizie riservate, in caso di furto di uno dei dispositivi coinvolti nel BYOD, gli stessi sono registrati in un sistema di MDM in grado di tracciarne la posizione e cancellarne il contenuto tramite un comando remoto. Questo rende possibile al reparto IT aziendale riconoscere lo spostamento di un dispositivo e, nel caso questo non coincida con lo spostamento del legittimo proprietario, eliminare i dati contenuti. Tuttavia, è ovvio che questo implica anche la capacità del reparto IT di tracciare gli spostamenti dei dipendenti, con una potenziale grave violazione della loro privacy; allo stesso tempo, rende altresì possibile la cancellazione di tutti i dati privati di un dipendente per mano dell'azienda. Al momento, non esiste una soluzione tecnologica a questo problema e l'unica possibile via rimane quella di definire in modo chiaro quali siano i diritti delle parti coinvolte, quali siano le garanzie fornite da entrambe ad entrambe e quali politiche di comportamento dovranno essere tenute.

4 Dispositivi affamati di energia

Un ulteriore aspetto rivoluzionario con cui si deve confrontare la sicurezza informatica odierna è quello energetico. Una differenza sostanziale tra il personal computer "vecchio modello" e gli attuali smartphone e tablet è, infatti, il consumo energetico. Se nel vecchio modello il consumo energetico è un problema molto poco rilevante nell'ottica della produttività del sistema, con l'avvento di queste nuove tecnologie la questione energetica diventa, di fatto, predominante. Infatti, i tablet e gli smartphone hanno una batteria molto più limitata rispetto ai laptop e ai netbook e la durata della batteria è intrinsecamente legata all'utilizzabilità e quindi alla produttività, del dispositivo mobile.

A titolo di esempio, consideriamo il grafico in Figura 3.

Tale grafico mostra la durata media della batteria di diversi modelli di smartphone di pochi anni fa quando questi utilizzano la rete 3G. Tutti i modelli hanno una durata, in connessione, tra le 3 e le 5 ore prima che la batteria si scarichi totalmente, rendendo il dispositivo mobile inutilizzabile se non si è vicini ad una

fonte di corrente elettrica. I nuovi dispositivi tendono ad avere un consumo energetico sempre più elevato data la dotazione hardware (Multicore e GB di Ram) e di rete (le nuove reti 4G che sostituiranno le 3G e le GPRS), mentre gli avanzamenti in termini di tecnologia delle batterie procedono più a rilento.

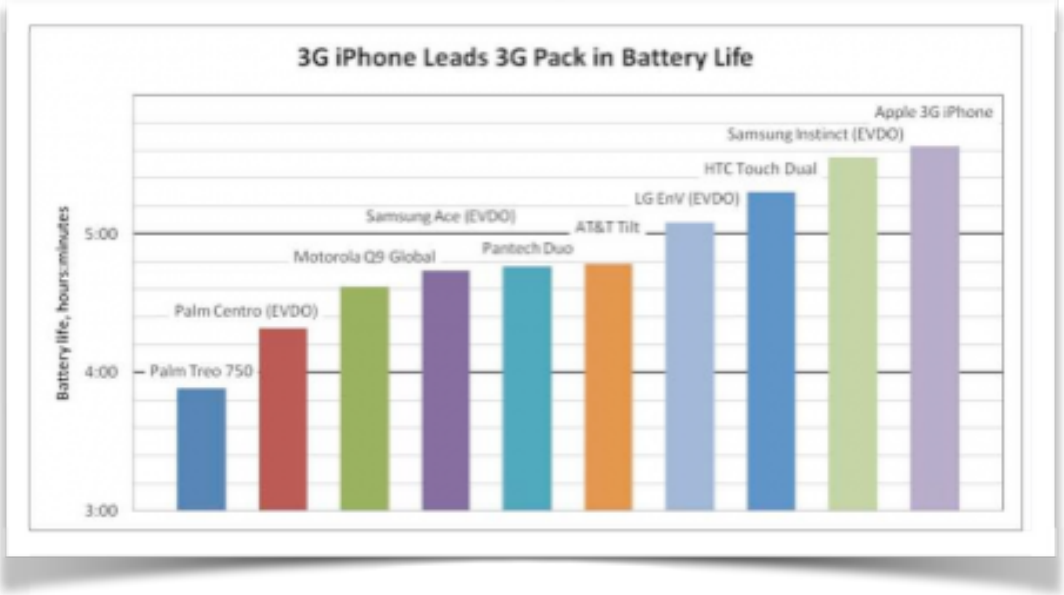


Figura 3
Grafico della durata della batteria in alcuni dispositivi mobili

Questo fatto porta a continue ricerche da parte dei produttori di dispositivi mobili allo scopo di migliorare l'efficienza energetica sia del dispositivo che della batteria. Ma il problema dell'efficienza energetica non è limitato a questioni di design di architetture e soluzioni software. Al contrario, è un problema che tocca anche la sicurezza informatica. Sempre con riferimento al grafico precedente, i valori ottenuti si basano sull'utilizzo del dispositivo da parte di applicazioni fidate, dove per fidate in un contesto energetico si possono intendere tutte le applicazioni che usano la CPU solo al bisogno per svolgere le proprie funzioni. Dall'altra parte possiamo definire una applicazione come maliziosa se volutamente o forzatamente spinge al consumo energetico speculativo (ad esempio attivando ed usando la CPU a vuoto, senza eseguire computazioni utili). In questa prospettiva, la domanda fondamentale è: che impatto avrebbe un insieme di applicazioni maliziose che non fanno danni ai dati dell'utente né al sistema operativo, ma consumano forzatamente la batteria tramite l'uso della CPU o l'attivazione ingiustificata dello schermo o delle periferiche di rete sulla durata della batteria stessa? Inoltre, i sistemi operativi attualmente installati sui dispositivi mobili sono in grado di distinguere tra applicazioni sane e applicazioni che maliziosamente sprecano la carica della batteria fino a rendere il dispositivo inutilizzabile? In caso contrario, che impatto può avere tutto questo sull'utilizzabilità del dispositivo?

La risposta a queste domande è oggetto di indagine di una nuova branca di ricerca nel settore della sicurezza informatica: la *Green Security* [8][9].

Presenteremo ora i primi risultati di questo nuovo tipo di studi e la pletora di problematiche non risolte che possono (e devono) essere oggetto di futuri sviluppi e ricerche.

4.1 Vulnerabilità ad attacchi battery drain

Come si può facilmente osservare, la batteria è l'obiettivo primario delle applicazioni maliziose che abbiamo introdotto prima. Tali applicazioni perpetrano nuovi tipi di attacchi, non rivolti (come classicamente accade) a violare la confidenzialità e la privacy dell'utente e dei suoi dati, ma hanno come scopo "abusare" della batteria [10]. Questo nuovo tipo di attacchi, detto appunto *battery-drain* (letteralmente, scarico della batteria), è volto ad accelerare il consumo della carica della batteria portando il dispositivo mobile ad eseguire operazioni non richieste sulle diverse periferiche hardware. Gli attacchi battery-drain possono essere perpetrati dall'interno del dispositivo (le applicazioni maliziose di cui abbiamo accennato in precedenza) o dall'esterno, attraverso tentativi di forzare il consumo energetico sollecitando le interfacce di rete dello smartphone. Lo scopo finale di tali attacchi è realizzare un *Denial-of-Device*, ovvero, in analogia con gli attacchi *Denial-of-Service* (DoS), rendere il dispositivo indisponibile al suo legittimo utente fino ad una nuova ricarica della batteria.

Attualmente, gli attacchi *battery-drain* sfruttano la limitata capacità dei sistemi operativi mobili di discriminare tra operazioni legali (richieste dall'utente del dispositivo) da operazioni non legali (forzate da attaccanti esterni) e apre a nuovi approcci all'analisi delle intrusioni dove anche la componente del consumo energetico diventa una metrica fondamentale per riconoscere gli attacchi alla batteria.

4.2 L'analisi di consumo energetico come identificatore di malware

In quest'ottica, nuove ricerche si stanno concentrando sulla misurazione del consumo energetico legato agli attacchi battery-drain, allo scopo di definire sistemi di sicurezza residenti su dispositivi mobili che siano in grado di riconoscere un attacco in funzione della sua caratteristica "impronta energetica". In particolare, recenti lavori si sono focalizzati sulla misurazione del consumo energetico di attività lecite e illecite su dispositivi Android [11] allo scopo di fare un primo passo verso la catalogazione dei "comportamenti" in funzione del consumo energetico della singola attività. Tale studio ha proposto un nuovo modello per la misurazione del consumo energetico delle attività di un dispositivo e ha eseguito le prime misurazioni di consumo effettivo su dispositivi reali. Le componenti per tale tipo di misurazioni sono state definite ad hoc e implementate su dispositivi Android, dal momento che questo tipo di misurazione energetica non è prevista in modo analiticamente accurato né su Android né su altri sistemi operativi mobili (come iOS o Windows Phone).

Le componenti di misurazione sono state implementate su diversi dispositivi mobili ed utilizzate per costruire un profilo energetico delle attività legali (applicazioni fidate) e attacchi battery-drain provenienti da applicazioni interne e fonti esterne. A titolo di esempio, nella Figura 4 viene mostrato il risultato di una profilazione energetica di un comportamento di rete legale (chiamata con la applicazione di Skype) e illegale (un attacco battery-drain esterno basato su Ping Flood).

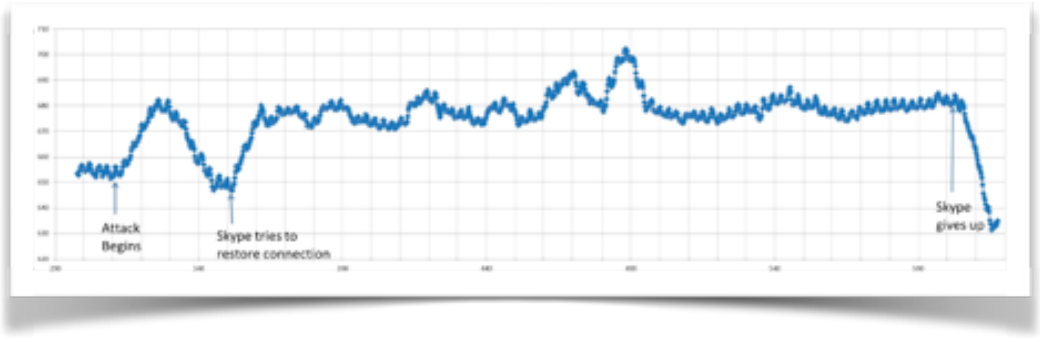


Figura 4
Grafico del consumo energetico di un dispositivo mobile durante un attacco Ping Flood e una chiamata Skype

In particolare, il grafico mostra il consumo energetico legato alla periferica WiFi rilevato sulla batteria nel caso in cui un attacco esterno (PingFlood) abbia inizio durante un'attività legale (Skype). Senza entrare nei dettagli tecnici, questo grafico è sufficiente a sottolineare come sia possibile catalogare e rilevare gli attacchi (battery-drain, ma non solo) in termini di consumi energetici. Tuttavia, l'analisi del medesimo grafico fornisce una significativa intuizione relativamente ad alcune problematiche correlate a questo nuovo campo di ricerca. Ad esempio, si nota che il consumo energetico di due attività contemporanee non è uguale alla somma algebrica dei consumi delle singole attività: questo richiederà la definizione di opportuni modelli avanzati e non lineari che permettano di isolare le componenti energetiche di attività contemporanee all'interno del dispositivo. Inoltre, il consumo energetico istantaneo di una singola attività può avere una fluttuazione rilevante. Al momento, il problema è affrontato considerando valori medi (in determinati periodi di tempo) del consumo energetico dell'attività (come nel caso del grafico in questione). Tuttavia, occorre anche qui determinare quanto tale variazione si debba all'attività in sé, alla latenza del sistema operativo oppure ad imprecisioni dei meccanismi software di misurazione.

Tutto questo mostra che, sebbene l'approccio sia promettente e perseguibile, molto lavoro debba essere fatto sia in termini di modellazione che di misurazione del consumo energetico per poter sviluppare meccanismi *energy-based* di rilevazione e analisi di attacchi.

5 Una rete progettata per dispositivi diversi

Un ultimo aspetto di questa rivoluzione della sicurezza che vogliamo trattare qui è la commistione di diverse generazioni tecnologiche in una delle infrastrutture di telecomunicazione più pervasiva al mondo: la rete cellulare.

Nella rete cellulare sono oggi contemporaneamente presenti la tecnologia di seconda generazione (GSM), la tecnologia di terza generazione (UMTS) e quella di quarta generazione (LTE). Questa rete complessa è nata per gestire dispositivi sostanzialmente "stupidi" (*dumb terminals*) e in grado di fornire due soli servizi (la voce ed i messaggi di testo), ma si è via via trovata a fornire servizi sempre più evoluti a terminali sempre più intelligenti (smartphone) con bande trasmissive crescenti. Ad oggi, una singola cella 2G/3G si basa sulla struttura mostrata in

Figura 5. Come si può notare, mentre le porzioni radio sono separate e distinte, le porzioni di infrastruttura dedicate alla gestione e al trasporto sono in comune. Si consideri inoltre che, mentre i terminali mobili si sono evoluti sino a diventare veri e propri calcolatori con capacità computazionali molto superiori a quelle dei PC degli anni '90 (quando la rete venne progettata), la necessità di mantenere la compatibilità con i terminali più semplici ha fatto in modo che tutte le operazioni di segnalazione e gestione del servizio richiedessero una quantità di risorse molto maggiore da parte della rete di quella richiesta ai terminali.

Questo fatto, rende la rete vulnerabile ad attacchi combinati da parte di terminali completamente programmabili che possono agire in maniera coordinata.

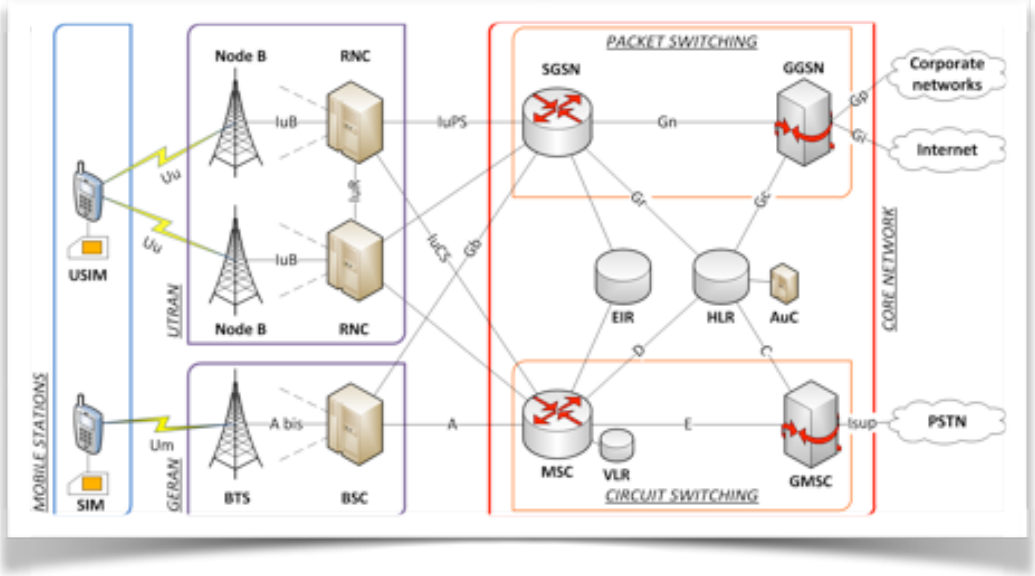


Figura 5
Architettura base per una cella di rete voce/dati

5.1 Un esempio di vulnerabilità

Si consideri, ad esempio, uno dei due elementi condivisi dall'intera architettura di Figura 5, per la precisione lo *Home Location Register* (HLR). Questo componente dell'architettura si occupa di tenere traccia delle informazioni relative agli utenti della rete: sia di quelle permanenti, ad esempio le credenziali di accesso, sia di quelle transitorie quali le deviazioni di chiamata, la locazione geografica del terminale o i settaggi GPRS. Dal punto di vista logico, HLR è una struttura distribuita e una rete può avere diversi HLR; tuttavia, un singolo utente è associato in modo univoco con un singolo HLR e le sue informazioni non sono replicate su altri HLR. Ovviamente, questo non implica che HLR non sia intrinsecamente una struttura resistente ai guasti, implica solo che è sempre possibile associare un utente con il "suo" HLR di riferimento.

L'utilizzo di HLR è richiesto da molti dei servizi forniti dalla rete, tuttavia questa base di dati viene interrogata solo in fase di segnalazione e non durante la fase di trasferimento del traffico (sia esso voce o dati, indifferentemente); in particolare, HLR viene interrogato per autenticare inizialmente gli utenti (fase di

attach alla rete), per instradare le chiamate entranti e per instradare i messaggi SMS.

Per questo motivo HLR non è dimensionato in base all'effettivo traffico di rete, ma piuttosto in base al numero di richieste di servizio effettuate. D'altro canto, essendo coinvolto nella fase iniziale delle chiamate e nella consegna dei messaggi di testo, il suo mancato funzionamento avrebbe effetti disastrosi sulla rete. La combinazione di questi due condizioni lo rende un bersaglio assai interessante per attacchi di tipo *Denial of Service* (DoS).

5.2 Un attacco Denial of Service

In passato, la possibilità di effettuare attacchi DoS alla rete cellulare era fortemente limitata dalla mancanza di dispositivi capaci di utilizzare la rete in modo programmato. E' del tutto evidente che la possibilità di organizzare una folla di migliaia di utenti distribuiti sul territorio in grado di effettuare la stessa operazione sul telefono cellulare nello stesso istante con precisione al millisecondo è nulla. Inoltre, la natura fisica dei terminali "stupidi" disponibili sino a pochi anni fa rendeva altrettanto impossibile una ripetizione rapida delle operazioni al fine di superare il requisito del perfetto sincronismo. Per questo motivo la progettazione della rete cellulare non ha mai posto particolare enfasi sulla sua resistenza ad attacchi di tipo DoS se non nella forma di *jamming* radio delle singole antenne.

Questi problemi, tuttavia, sono stati superati dalla disponibilità dei moderni smartphone, veri e propri computer il cui comportamento può essere completamente controllato da programmi. Inoltre, mentre in passato l'incapacità dei telefoni cellulari di essere programmati per un comportamento complesso li rendeva impervi ai tentativi di attacco da parte di *malware*, la completa programmabilità dei moderni telefoni cellulari li ha resi vulnerabili esattamente alla stregua dei comuni PC [12]. La vulnerabilità all'infezione da *malware*, rende possibile quindi la costruzione da parte di malintenzionati di vere e proprie *botnet* di telefoni cellulari, attivabili con comandi remoti per agire in modo coordinato e ripetitivo anche senza il benché minimo coinvolgimento dell'utente possessore del telefono stesso.

La *botnet* è, come ampiamente dimostrato da quanto avviene su Internet [13], lo strumento perfetto per perpetrare un attacco di tipo DoS, e la possibilità di allestirne una in grado di interagire con la rete cellulare costituisce quindi lo strumento di scasso che non era stato preso in considerazione in fase di progettazione della rete cellulare stessa.

Jamming

Il Radio Jamming è la trasmissione deliberata di segnali radio che interferiscano e rendano inintelleggibile i segnali radio di un canale di comunicazione. Il Jamming è una delle forme più comunemente note di attacco alle infrastrutture di telecomunicazione wireless.

BotNet

Si definisce botnet un insieme di dispositivi informatici connessi in rete ed infettati con un programma di controllo che li rende controllabili da un singolo terminale remoto (il botmaster). Un uso comune delle botnet è lo scatenare un attacco di tipo Distributed Denial of Service particolarmente insidioso in quanto proveniente da un insieme molto numeroso di nodi mai precedentemente associati a comportamenti malevoli

Recenti studi [14] dimostrano infatti che, tramite una botnet di poco più di undicimila telefoni cellulari sarebbe possibile rendere inutilizzabile la rete su un'area geografica significativa (per esempio una regione italiana). E' ovvio che la cattura di oltre undicimila telefoni cellulari e la disponibilità di tutti i terminali nello stesso istante e secondo una distribuzione geografica precisa rende questo attacco di non banale realizzazione. Tuttavia, successivi sviluppi di questi stessi studi [15] mostrano che, con l'ausilio di dispositivi dedicati, il numero di risorse richieste può anche essere diminuito drasticamente (sino a raggiungere le millecinquecento unità) rilassando contemporaneamente la dipendenza dal comportamento dell'utente possessore del telefono.

Il rischio di un attacco DoS alla rete cellulare, dunque, non è più oggi solo uno scenario da fantascienza e richiede l'attenzione della comunità scientifica.

6 Conclusioni

In questo breve excursus abbiamo voluto introdurre una serie di nuove problematiche di sicurezza informatica che la recente evoluzione tecnologica, assieme alla commistione tra nuovi paradigmi emergenti ed i limiti intrinseci all'ambiente mobile, ha portato alla luce. Come spiegato nei tre casi esemplari utilizzati, e precisamente mobilità e BYOD, dispositivi energeticamente vincolati e rete cellulare multi-generazionale, il campo della sicurezza informatica si trova oggi costretto ad abbandonare un modello di lavoro stabilizzato. Infatti, il modello a fortezza, quello in cui l'accesso è limitato a pochi passaggi fortemente sorvegliati, si dimostra incapace di catturare le caratteristiche peculiari che novità ormai capillarmente diffuse - come connettività pervasiva e dispositivi mobili personali - introducono nel panorama IT.

Per superare questo empasse esistono diverse nuove linee di ricerca che vanno dall'applicazione delle metodologie dipendenti dal contesto per il controllo della connettività e del comportamento dei dispositivi, all'applicazione di metodologie legate al consumo energetico per l'identificazione di software malevoli sui dispositivi mobili, sino a studi sulla vulnerabilità della rete cellulare ad attacchi di tipo Denial of Service impensabili sino a pochi anni fa. Queste nuove linee, tuttavia, non possono essere considerate avulse dalla necessità di responsabilizzare gli utenti e di fornire loro politiche di utilizzo chiare e prive di ambiguità, in modo da limitare i rischi non solo dal punto di vista del supporto tecnologico, ma anche da quello della componente umana.

Bibliografia

1. Šilić, M., Krolo, J., Delac, G., *Security vulnerabilities in modern web browser architecture*, MIPRO, 2010 Proceedings of the 33rd International Convention , pp.1240-1245
2. CVE: *Firefox Vulnerability Statistics*, aggiornato al 14/10/2013 da http://www.cvedetails.com/product/3264/Mozilla-Firefox.html?vendor_id=452
3. Colombini, C.M., Colella, A., Mattiucci, M., Castiglione, A., Cyber Threats Monitoring: Experimental Analysis of Malware Behavior in Cyberspace, CD-ARES Workshops 2013, pp. 236-252
4. Kuper, P., *A warning to industry - fix it or lose it*, Security & Privacy, IEEE, 2006, 4(2), pp.56-60

5. Mokdong, C., Jaehyuk, C., Seokhwan, Y., Shi-Kook, R., *Context-Aware Security Services in DAA Security Model*, Advanced Language Processing and Web Information Technology, 2008. ALPIT '08. International Conference on ,pp . 424-429, doi: 10.1109/ALPIT.2008.97

6. Al-Rabiaah, S., Al-Muhtadi, J., *ConSec: Context-Aware Security Framework for Smart Spaces*, Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2012 Sixth International Conference on , pp. 580-584, doi: 10.1109/IMIS.2012.41

7. Armando, A., Costa, G., Merlo, A., *Bring Your Own Device, Securely*. In Proceedings of the 28th Annual ACM Symposium on Applied Computing (ACM SAC 2013), pp. 1852-1858

8. Caviglione, L., Merlo, A., Migliardi, M., *What Is Green Security?*, Proc. of the 7th International Conference on Information Assurance, Malacca (Malaysia) 5 - 8 December 2011, pp. 366-371

9. Caviglione, L., Merlo, A., Migliardi, M., Green Security: risparmio energetico e sicurezza, Mondo Digitale, 44, Dicembre 2012

10. Caviglione, L., Merlo, A.,. The energy impact of security mechanisms in modern mobile devices, Network Security, 2012, 2, pp. 11-14

11. Curti, M., Merlo, A., Migliardi, M., Schiappacasse, S., *Towards Energy-Aware Intrusion Detection Systems on Mobile Devices*, Proc. of the 8th International Workshop on Security and High Performance Computing Systems, 1-5 July 2013, Helsinki. Finland

12. Dagon, D., Martin, T., Starner, T., *Mobile phones as computing devices: the viruses are coming!* Pervasive Computing, IEEE, 2004, 3(4), pp. 11-15, doi: 10.1109/MPRV.2004.21

13. Lei, Z., Shui, Y., Di, W., Watters, P., *A Survey on Latest Botnet Attack and Defense*, Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on , pp. 53-60, doi: 10.1109/TrustCom.2011.11

14. Traynor, P., Lin, M., Ongtang, M., Rao, V., Jaeger, T., McDaniel, P., La Porta, T., *On cellular botnets: measuring the impact of malicious devices on a cellular network core*. In: Proceedings of the 16th ACM conference on Computer and communications security, 2009. pp. 223-234

15. Gobbo, N., Merlo, A., Migliardi, M. *A Denial of Service Attack to GSM Networks via Attach Procedure*, Proc. of ARES 2013 Workshops, IFIP International Federation for Information Processing (2013), LNCS 8128, pp. 361-376

Biografie

Mauro Migliardi si e' laureato in Ingegneria Elettronica nel 1991 e ha conseguito il dottorato in Ingegneria Informatica nel 1995. E' stato ricercatore presso l'Universita' degli Studi di Genova e ricercatore associato presso la Emory University di Atlanta (USA) come COPI del progetto HARNESS. Attualmente e' Professore Associato presso l'Universita' degli Studi di Padova e nel 2013 ha vinto il CanadaItaly Innovation Award. In generale, i suoi interessi di ricerca riguardano l'ingegnerizzazione dei sistemi distribuiti, ma recentemente si concentra sull'uso di dispositivi mobili per i servizi di supporto alla memoria umana, le problematiche energetiche e la sicurezza informatica.

E-mail: mauro.migliardi@unipd.it

Alessio Merlo ha ricevuto il dottorato in Informatica nel 2010, presso l'Universita' degli Studi di Genova, dove ha lavorato su problematiche di performance e di controllo degli accessi in ambito Grid Computing. Attualmente e' ricercatore presso l'Universita' E-Campus e un ricercatore associato presso Artificial Intelligence Laboratory (AILab) presso il DIBRIS, Universita' degli Studi di Genova. I suoi interessi di ricerca riguardano la sicurezza di dispositivi mobili, dei sistemi distribuiti e del Web.

E-mail: alessio.merlo@unige.it

L’evoluzione normativa sul testing del software

Antonio Piva - Attilio Rampazzo

Dopo lo sviluppo di varie metodologie su questo argomento molto controverso nel mondo dello sviluppo o della manutenzione del software, ISO ha cercato di colmare le mancanze con una nuova norma la ISO 29119. La norma è suddivisa in più parti e fornisce un unico ed integrato standard che affronta il tema in tutte le sue parti: vocabolario, processo, documentazione, tecniche. Si sta inoltre discutendo la predisposizione di un modello di valutazione del processo di testing.
Rappresenta uno standard (best practice) che si possa applicare a tutte le tipologie di sviluppo software e ai diversi cicli di vita.

“Le operazioni di testing possono individuare la presenza di errori nel software ma non possono dimostrarne la correttezza” (Dijkstra, 1972)

Introduzione

L’introduzione nella pratica dello sviluppo del software dei concetti di *ciclo di vita* e di *processo software,* coincidenti con la nascita dell' ingegneria del software, rappresenta un passaggio storico dallo sviluppo del software, inteso come attività "individuale" (ovvero affidata alla libera creatività dei singoli individui), ad un approccio più industriale, in cui la creazione di programmi e sistemi software viene considerata come un processo complesso. Questo insieme di attività e di processi complessi richiedono pianificazione, controllo, e documentazione appropriati (così come avviene abitualmente nei settori tradizionali dell’ingegneria).

Il collaudo del software (detto anche *testing* o *software testing* secondo le denominazioni anglosassoni) è un procedimento, che fa parte del ciclo di vita del software, utilizzato per individuare le carenze di correttezza, completezza e affidabilità delle componenti software in corso di sviluppo. Consiste nell'eseguire il software da collaudare, da solo o in combinazione ad altri programmi di servizio, e nel valutare se il comportamento del software rispetta i requisiti: fa parte delle procedure di assicurazione di qualità, ma non è l'unica.

L'ingegneria del software focalizza molto bene l'importanza rivestita dall'attività di testing all'interno del ciclo di vita del software.

Al testing vengono attribuiti due obiettivi importanti quanto imprescindibili per la qualità del prodotto finale:

- assicurare che tutti i requisiti siano stati correttamente implementati e
- rimuovere il maggior numero di errori presenti nel software.

Purtroppo, non sempre tali obiettivi sono raggiunti dato il ruolo “tampone” che l’attività di testing assume nel ciclo di vita del software, essendo tipicamente un attività posta al termine dello sviluppo e prima della consegna.

La posizione nel ciclo di vita, infatti, fa si che la fase di testing sia molto spesso compromessa per sopperire al ritardo accumulato e recuperare i costi eccessivi prodotti alla data. Purtroppo la decisione che più di frequente si vede prendere è quella di ridurre il test per recuperare il ritardo e i costi.

Conseguenza inevitabile è il parziale test effettuato sul software (magari tralasciando anche parti importanti o critiche per il business). A tale mancanza non può rimediare alcun standard, ovviamente. E' un fatto di cultura, di professionalità, di conoscenza, di studio, di applicazione.

Il collaudo a volte viene confuso con il *debugging*, con il *profiling*, o con il *benchmarking*.

A districare questa matassa sono intervenute alcune normative al fine di dare una linea guida su come effettuare il collaudo del software tra queste l’*IEEE Standard for Software Test Documentation.*

Su queste specifiche sono poi nate delle particolari certificazioni professionali che hanno qualificato moltissimi professionisti in tutto il mondo.

ISO/IEC IEEE 29119, uno standard internazionale per gestire i test del software.

La nuova norma ISO/IEC IEEE 29119 nasce dall'esigenza di mettere ordine ai molti di standard e/o norme esistenti sul testing. Fino a poco tempo fa, infatti, l'attività di test era documentata da diversi standard, alcuni lacunosi e altri contradditori di cui si citano i più conosciuti:

- IEEE 829 standard sulla documentazione del testing ,
- IEEE 1008 standard limitato allo Unit Testing,
- BS 7925-1 e BS 7925-2, standard dedicati pure questi allo Unit testing.

Si è resa manifesta, quindi, la necessità di sopperire a mancanze evidenti e definire:

- in maniera completa e strutturata anche ulteriori tipologie di test quali: di integrazione, di sistema e di accettazione;
- un modello esplicito del processo di test;
- strategie e politiche per il testing;
- un processo particolarmente indirizzato alla gestione del testing nei progetti software;
- tecniche specifiche di testing;
- test non-funzionali, che coprano l'intero ciclo di vita del software.

Rispondendo a queste esigenze, già dal 2007, ISO si è prodigata per lo sviluppo dello standard - ISO/IEC IEEE 29119 Software Testing - che indirizza tutti gli aspetti mancanti in maniera completa ed adeguata. Uno standard "scritto dai professionisti del testing per i professionisti del testing", infatti nel gruppo di lavoro ISO JTC1/SC7 WG26 sono state coinvolti una ottantina di specialisti provenienti da 27 paesi, riuniti due volte l'anno, al fine di rappresentare le esperienze e le norme dei singoli paesi di origine.

Struttura dello standard ISO/IEC 29119

Lo standard ISO/IEC 29119 è composto da quattro parti così suddivise:

- Parte 1: Concetti e Vocabolario
- Parte 2: Processi
- Parte 3: Documentazione
- Parte 4: Tecniche di Testing

La figura 1 mostra le quattro parti della norma e le relazioni con gli altri standard da cui la norma ha preso spunto.

Ad oggi, sono disponibili solo alcune parti della ISO/IEC 29119, e come tutti gli standard ISO sono a pagamento.

Le quattro parti della norma sono incentrate su un modello di processo di *risk-based* a tre livelli per il test del software che fornisce le linee guida per lo sviluppo di strategie organizzative di test e politiche, la gestione di progetti di test tra cui la progettazione del progetto, livelli di strategie e piani di test, di monitoraggio e controllo test, ed un processo di prova dinamica che fornisce le linee guida per l'analisi e la progettazione dei test, dei test di ambiente e di set-up, della manutenzione ivi inclusi l'esecuzione di test e del reporting.

Parte 1: Concetti e Vocabolario

La prima parte della norma prevede di fornire una panoramica dello standard, i concetti generali del test del software e la descrizione delle sue relazioni con le altre attività di sviluppo e di manutenzione. Fornisce inoltre:

- una panoramica sulle implicazioni delle attività di testing nei vari modelli di ciclo di vita del software, inclusi i modelli "agili"
- una descrizione dei vari approcci al testing.

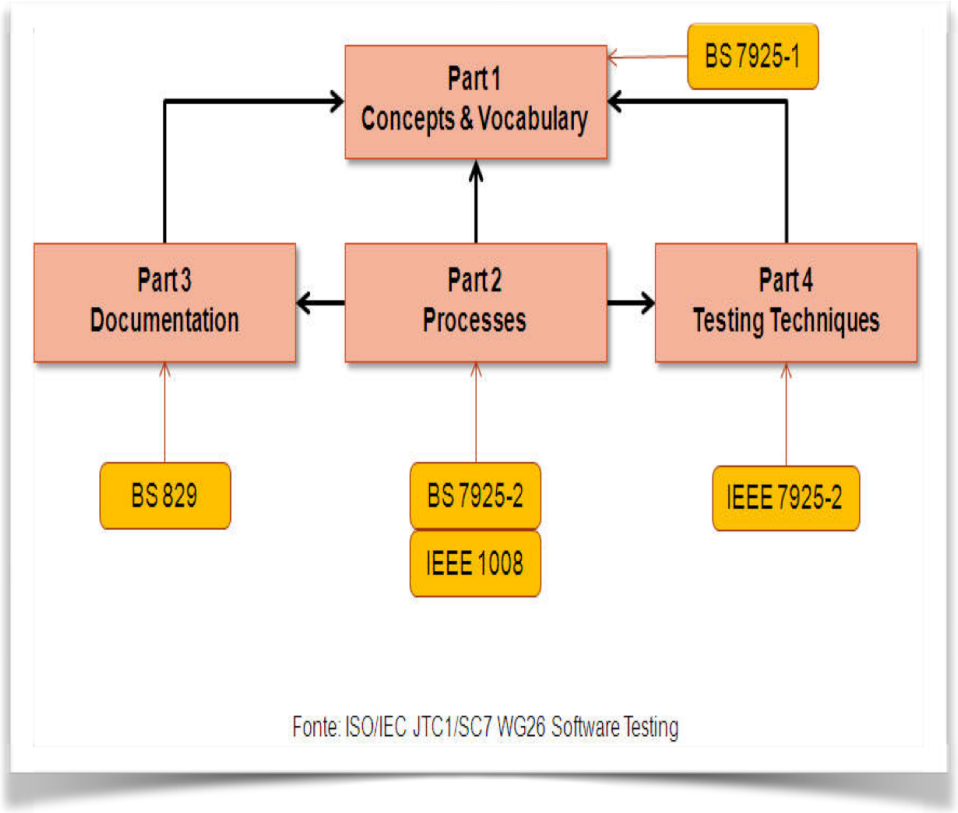


Figura 1
Le parti della norma ISO/IEC IEEE 29119 e relazioni con gli standard da cui ha perso spunto

E' incluso anche un utile vocabolario dei termini utilizzati nel testing al fine di creare una base comune di comprensione tra i vari addetti al settore.

Parte 2: Processi di Testing

Il modello dei processi che contiene le attività di testing le quali danno rilievo:

- al processo organizzativo (strategia e politiche di testing),
- al processo gestionale (gestione dei test nel progetto),
- al processo di esecuzione vera e propria (test statico e test dinamico).

La figura 2 mostra il modello dei processi definiti da questa parte dello standard.

Ciascuna area di processo prevede più processi definiti nei loro componenti di base e negli input e output che li relazionano gli uni agli altri.

In questa parte vi sono alcune figure chiarificatrici con i flussi dei processi (*Workflow*) e viene fornita una descrizione molto ampia di ciascun processo in termini di Scopo, Artefatti prodotti, Attività e compiti, Informazioni di dettaglio.

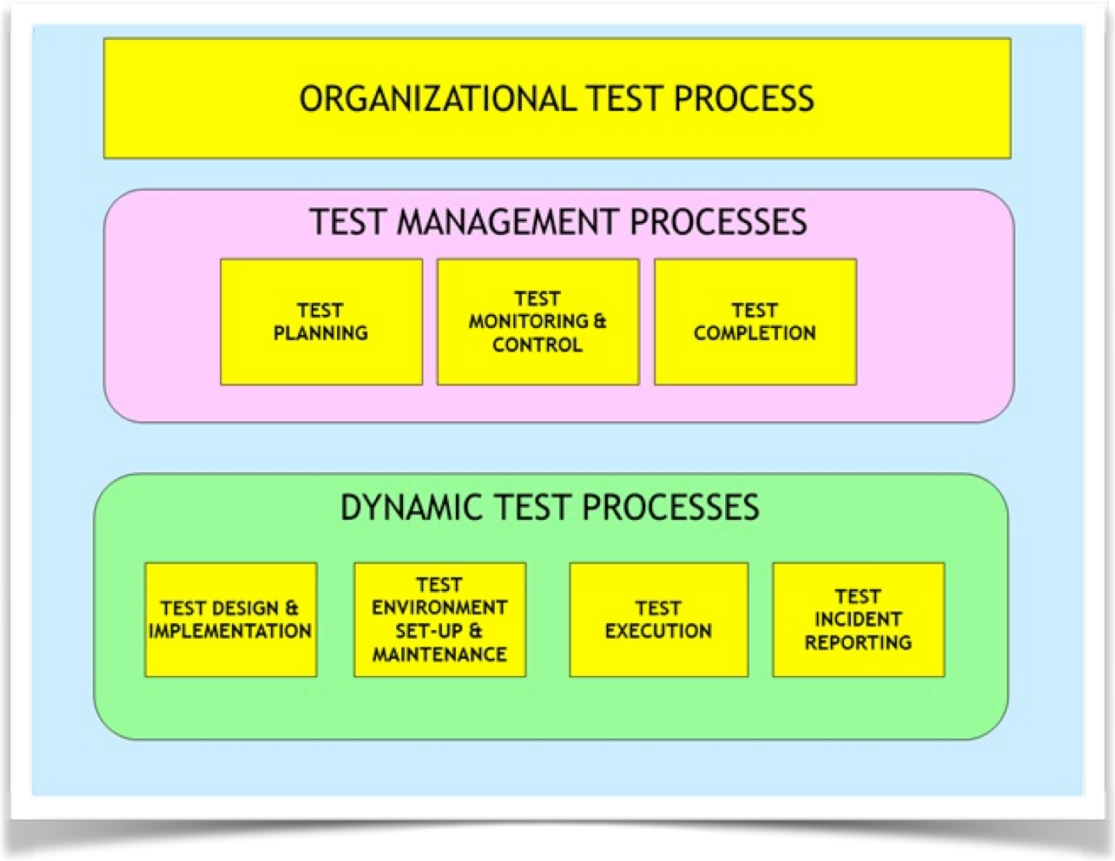


Figura 2
Modello dei processi di testing

Parte 3: Documentazione

La parte relativa alla documentazione è molto ricca e fornisce nelle appendici esempi pratici di come costruire i vari documenti previsti dallo standard.

IEEE ha autorizzato ISO ad utilizzare il conosciuto standard per la documentazione di test IEEE 829 come base per questa parte di standard.

La descrizione della documentazione è organizzata secondo la seguente tipologia:

- **Organizzativa**: descrive la documentazione relativa alle politiche e alle strategie per il testing.
- **Progetto**: descrive la documentazione richiesta a livello di progetto.
- **Testing**: descrive la documentazione tecnica più specifica del testing.
- **Appendici**: vengono proposti vari esempi di documenti da produrre nell'attività di testing.

Parte 4: Tecniche di Testing

La parte relativa alle tecniche di testing è molto ricca e vengono descritte raggruppandole secondo categorie specifiche:

- Specification-Based Testing Techniques: ovvero le tecniche di testing guidate dalle specifiche del prodotto.
- Structure-Based Testing Techniques: ovvero le tecniche di testing basate sulla struttura del codice.
- Quality-Related Types of Testing: ovvero le tecniche di testing relative alle caratteristiche del prodotto e al suo utilizzo e alla sua gestione.

Ulteriori implementazioni: modello di processo di valutazione

Durante lo sviluppo dello standard è stato deciso di aggiungere una ulteriore parte dedicata alla valutazione del processo di testing.

Con tutta probabilità questa parte non sarà inclusa nelle ISO/IEC IEEE 29119 ma dovrebbe far parte di una nuova versione dello standard ISO/IEC 15504 Information Technology - Process Assessment (conosciuto come SPICE-Software Process Improvement and Capability Determination) nella serie 33000 e dovrebbe essere: **ISO/IEC 33063 Process Assessment Model for Software testing Processes.**

La valutazione "è un processo attraverso il quale osserviamo, interpretiamo, formuliamo un giudizio di valore su un fenomeno o su un soggetto in rapporto a parametri". La valutazione implica quindi l'osservazione di un determinato fenomeno (come è applicato il test del software nel rispetto della norma ISO/IEC IEEE 29119), la misurazione di questo fenomeno attraverso parametri oggettivi precedentemente condivisi e la formulazione di un giudizio di valore, che non è mai sulla persona ma è sempre riferito alle azioni osservate ed al livello di competenza espressa attraverso tali azioni.

Anche per questa valutazione i livelli di competenza saranno quelli già proposti da ISO/IEC 15504 e che sono evidenziai in figura 3.

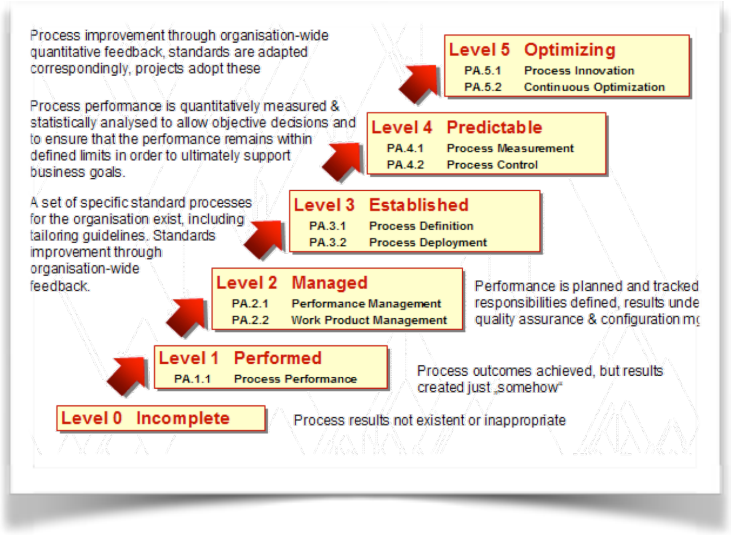


Figura 3
Livelli di competenza del Process Assessment

Conclusioni

Questo nuovo standard internazionale fornisce ai professionisti del testing delle linee guida che coprono tutti gli aspetti del ciclo di vita del software: non c'è nulla di obbligatorio.

Lo standard ISO/IEC 29119 fornisce un insieme consistente di definizioni, processi, procedure, tecniche e modelli per la documentazione del testing. Al momento non si hanno significativi esempi del suo utilizzo, anche se certamente dovrebbe avere un importante impatto sul mondo del test del software.

Al giorno d'oggi si sente sempre più necessita di modelli che ci guidino nella attività quotidiana anche se questa è relativa alla produzione del software dove è sempre difficile controllare l'assenza di errori: un metodo di conduzione dei test dovrebbe sopperire a ciò. Sarà'comunque importante che lo standard sia sempre aggiornato con la continua evoluzione del mondo del software, dei sistemi informativi e relativo hardware.

Bibliografia

- ISO/IEC/IEEE 29119-1 Software and systems engineering -- Software testing -- Part 1: Concepts and definitions
- ISO/IEC/IEEE 29119-2 Software and systems engineering -- Software testing -- Part 2: Test process
- ISO/IEC/IEEE 29119-3 Software and systems engineering -- Software testing -- Part 3: Test documentation
- ISO/IEC/IEEE 29119-4 Software and systems engineering -- Software testing -- Part 4: Test techniques
- A.Avellone, M.Cislaghi, E.Colonese Qualità e collaudo del software Ed. UNI Service

Sitografia

- www.softwaretestingstandard.org

Biografia

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, qualità e comunicazione all'Università di Udine. Consulente sistemi informatici e Governo Elettronico nella PA locale, valutatore di sistemi di qualità ISO9000 ed ispettore AICA. Presidente della Sezione Territoriale AICA del Nord Est.

E-mail: antonio@piva.mobi

Attilio Rampazzo *CISA CRISC*, *C|CISO* consulente di Sistemi Informativi e Sicurezza delle Informazioni in primaria azienda di Servizi Informatici italiana. Ha maturato un'esperienza più che trentacinquennale nello sviluppo e conduzione di progetti informatici in ambito bancario e finanziario, nei quali la qualità e la sicurezza hanno ricoperto un ruolo determinante.

E' Vice Presidente di AICA sez. Nord Est e CISA Coordinator e Research Director in ISACA Venice chapter.

Svolge attività come Valutatore di Sistemi di Sicurezza delle Informazioni e di Sistemi di Gestione dei Servizi (cert. AICQ Sicev) presso CSQA Certificazioni.

Socio AICA, AICQ, ISACA Venice chapter, itSMF Italia, ASSOVAL, ANIP.

e-mail: attilio.rampazzo@gmail.com

e-leadership: guidare l'innovazione nell'era digitale

Roberto Bellini

La diffusione e penetrazione delle tecnologie digitali e dell'ICT è in costante crescita, ma il livello di competenza per il loro utilizzo concreto anche a favore della innovazione di business non lo è altrettanto. . Varie analisi e teorie americane ed europee indicano che queste Tecnologie svolgono un ruolo sistematico di abilitazione e governo dei processi digitali e che costituiscono inoltre un componente fondamentale per la innovazione di prodotti/servizi intelligenti. Emerge un nuovo spazio di competenze per l'innovazione che l'Europa indica con il termine di e-leadership e su cui si appoggia, insieme alle competenze digitali di cittadinanza e a quelle specialistiche, il programma della Grand Coalition for Digital Jobs .
Questo articolo vuole contribuire alla definizione dello spazio per quelle che indicheremo come competenze per l'innovazione in ambiente digitale e propone una metodologia descrittiva per la qualificazione e la popolazione dei relativi contenuti per le varie aree disciplinari riconoscibili nello sviluppo delle strategie competitive delle imprese e degli enti della pubblica amministrazione.

Introduzione

Il tema della innovazione è costantemente presente nel lavoro di AICA: data la focalizzazione sul tema delle competenze digitali, l'innovazione vista da AICA riguarda prima di tutto le "competenze per l'innovazione digitale" a cui segue il tema del come queste competenze si possono riconoscere, accrescere, insegnare, verificare, ecc.

Il contributo di AICA al tema si è espresso fino ad oggi attraverso una serie di ricerche e approfondimenti che vanno dal sistematico aggiornamento del costo dell'ignoranza informatica ai vari contributi di ricerca che vengono dagli allenamenti delle squadre Olimpiche di informatica, ai più approfonditi contributi sul come fare didattica nel web e alle caratteristiche dell'Innovatore pubblicate in precedenti articoli.

A tutt'oggi la filiera cognitiva delle competenze digitali che il CEPIS e AICA promuovono è costituita da tre componenti fondamentali:

- le competenze degli Specialisti ICT, di coloro cioè che progettano, realizzano e mantengono in esercizio i sistemi informativi e digitali con la loro crescente espansione in tutti gli angoli delle organizzazioni pubbliche e private; la numerosità di questi specialisti è dell'ordine di qualche % rispetto alla forza lavoro di un paese e di una organizzazione;
- le competenze d'uso delle tecnologie informatiche, che iniziata dal CEPIS e AICA a metà degli anni 90 ha ormai raggiunto milioni di utenti, anche se potrebbero raggiungere come potenziale la totalità delle persone che lavorano;
- le competenze minime di navigazione per il cittadino/consumatore, compresi tutti coloro che o ancora non lavorano o ormai non lavorano più.

Secondo questa classificazione la filiera delle competenze digitali risulta articolata come indicato in Fig. 1.

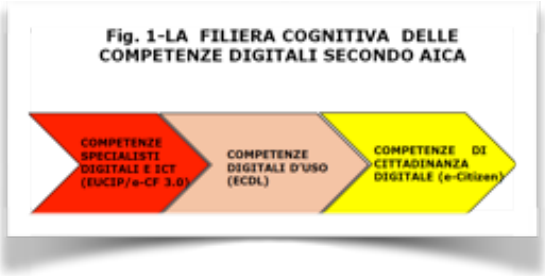


Fig. 1-LA FILIERA COGNITIVA DELLE COMPETENZE DIGITALI SECONDO AICA

Oggi è maturo un nuovo passo di approfondimento, in coerenza con quanto sta facendo l'Europa: è stato individuato un grandissimo spazio per lo sviluppo delle competenze per la innovazione digitale, chiamato *e-Leader* : gli vengono dedicati dalla Commissione Europea convegni e programmi di lavoro, rispettivamente gli *e-skills Forum* degli ultimi due anni, e la *Grand Coalition for Digital Jobs* a partire da marzo del 2013; l'obiettivo è quello di identificare, configurare e realizzare programmi educativi e di apprendimento che ottengano come risultato quello di adeguare e arricchire le competenze digitali di uso di chi lavora in aree disciplinari tradizionali e diverse da quelle digitali. In realtà sulla base della constatazione che le tecnologie digitali e dell'ICT si ritrovano in qualunque tipo di innovazione sta maturando una vera e propria spinta ad accrescere prima di tutto la consapevolezza di questa caratteristica abilitante delle TIC ma poi soprattutto a disegnare percorsi di innovazione anche nel mondo dei sistemi dell'industria manifatturiera in cui sfruttare questa leva abilitante.

Il bisogno ormai emerso viene descritto come segue: "**Gli *e-leadership skills* sono quelli che abilitano le persone a guidare gruppi di lavoro/team di progetto costituiti da professional e manager di impresa di aree disciplinari diverse a identificare e progettare modelli di business e capacità di realizzazione di opportunità di business facendo il miglior uso di tecnologie digitali e nella generazione di maggior valore per l'organizzazione in cui e per cui lavorano**". Questa definizione mette in evidenza il valore prioritario

assegnato agli e-skills per l'innovazione di business, che costituisce una priorità europea e nazionale: che non vuole dire ovviamente diminuire l'attenzione ad una "normale" attività e attenzione a complementare con competenze digitali tutti coloro (operatori di processo, professional e manager) che lavorano utilizzando competenze di altre aree disciplinari: si tratta di ridefinire adeguatamente il campo su cui lavorare per le competenze d'uso, scegliendo come focalizzazione ulteriore quella della innovazione per le imprese sia nel settore dei servizi che nei settori manifatturieri di prodotti materiali.

Questo contributo di AICA riguarda quindi la progettazione di profili di **e-Leadership** per organizzazioni che non operano nel settore informatico, in cui sono di gran lunga prevalenti competenze di aree disciplinari diverse da quelle ICT e che possono arrivare a comprendere fino al 95-99% delle risorse.

Il contributo si basa su due approfondimenti:

¥ quali fattori spingono verso i bisogni di innovazione e come classifichiamo le innovazioni: è essenziale infatti approfondire per quanto possibile le analisi per poter disegnare le competenze più specifiche richieste per fare innovazione, , di cui quelle digitali sono un componente;

¥ un modello descrittivo generale della combinazione di competenze e professionalità NON-ICT (altre aree disciplinari) con quelle ICT: essendo le competenze digitali per la innovazione aggiuntive a quelle disciplinari che si trovano normalmente in una impresa o ente pubblico, il modello descrittivo delle competenze è la base su cui poter identificare di quali nuove combinazioni di competenza si stia trattando; il modello descrittivo proposto viene popolato in particolare per il cluster della Digital Innovation Leadership e per quello della Marketing Innovation e-leadership.

Fattori che spingono verso l'innovazione e classificazione delle innovazioni

L'innovazione di business costituisce una richiesta generica delle organizzazioni pubbliche e delle imprese private per mantenere o migliorare la probabilità di assicurare maggior valore ai propri clienti e proprietari; sulla spinta di una turbolente pervasività della automazione e della digitalizzazione in ogni angolo delle organizzazioni, è opportuno definire un nuovo tipo di competenza per catturare i nuovi bisogni di soluzioni digitali per i vari componenti delle filiere di business.

Semplificando al massimo, i 4 principali fattori che spingono l'innovazione possono essere considerati i seguenti:

- una "società sempre più flessibile", in cui le reti sociali e professionali facilitano le comunicazioni interpersonali e creano nuovi bisogni;
- una "evoluzione delle tecnologie" che facilitano la cooperazione digitale (tutto si connette) e accompagnata dalla riduzione dei costi;
- i cambiamenti delle regole istituzionali (leggi, normative, regolamenti attuativi) che ad esempio in Italia vengono buttati fuori a getto continuo;
- i cambiamenti dell'economia reale, che impongono maggiore competitività su platee di mercato "glocali" (globali e locali insieme) e una crescente egemonia di player anch'essi glocali: grandi banche, grandi compagnie

assicurative, imprese del settore energia, dell'automotive, delle costruzioni, dell'elettronica, dei servizi professionali, del turismo, del commercio, ecc.

Questi cambiamenti spingono a far emergere, nelle singole persone che lavorano, le competenze di cui sono portatori, ottenute attraverso lo studio e l'esperienza su progetti innovativi, con l'ulteriore vincolo/opportunità di sviluppare il loro percorso di carriera professionale attraverso i delicati equilibri che si instaurano per ciascuno fra la qualificazione e i titoli di studio ottenuti al termine dei rispettivi curricula di istruzione, la retribuzione, il tipo di inquadramento da contratto, il ruolo svolto nella realtà operativa in cui è inserito, le opportunità di lavoro fra cui scegliere, consuntivate dalle eventuali certificazioni di competenze perseguite.

Le competenze acquisite e messe a disposizione da ciascun lavoratore infine si classificano fra:

- replicabili, che favoriscono l'automazione dei processi migliorando la produttività (costo per unità di prodotto) e la occupazione, se i volumi di produzione si incrementano
- innovative, che nelle organizzazioni spingono verso la creazione di valore per clienti e utenti e che per definizione portano a prodotti e servizi personalizzati o personalizzabili in funzione delle esigenze del singolo cliente o di segmenti omogenei di clienti.

La combinazione di competenze replicabili e competenze innovative si configura nella esperienza di ciascun lavoratore e di ciascuna organizzazione in funzione di una molteplicità di fattori che vanno analizzati in altra sede, ma che comunque devono considerare le conoscenze, le abilità e le capacità come un presupposto su cui potersi basare nella nuova organizzazione del lavoro centrata sulle competenze.

La nuova organizzazione che nasce dalla spinta di questi fattori dovrebbe essere maggiormente capace di

- reagire più rapidamente agli stimoli del mercato e alle sollecitazioni dei clienti
- forzare l'adozione di processi ottimizzati e di reti in grado di sostenere le catene del valore nella loro interezza dall'origine al termine.

Entriamo nel merito dell'analisi dell'innovazione adottando la classificazione proposta nella Teoria della Innovazione Dirompente, di C. Christensen [1 e 2] . Vengono riconosciuti due tipi di innovazione:

- la innovazione dirompente: è quella che permette, utilizzando una nuova combinazione di tecnologie, materiali e risorse, di realizzare un nuovo prodotto/servizio (precedentemente non presente) per soddisfare i bisogni di un nuovo mercato;

- la innovazione di sostegno: è quella conseguente agli interventi di miglioramento dei processi operativi di vendita, distribuzione, produzione e fornitura.

Per entrambi i tipi di innovazione assumono un ruolo importante oltre ai nuovi materiali, tutte le tecnologie di processo e informative, la modellazione del business che razionalizza magari nuove fonti di ricavo, e molti altri fattori che emergono per ogni settore di industria in cui l'innovazione si esprime.

L'innovazione dirompente

- consideriamo tale quella di un prodotto/servizio le cui funzioni d'uso e il cui senso è percepito e acquistato con soddisfazione da nuovi clienti disponibili a

riconoscere un più alto valore rispetto a prodotti tradizionali; migliorano il livello di competitività dell'impresa o dell'unità di business che li propone;

- esempi di innovazione dirompente sono:
  - l'iPhone e l'iPad della Apple e il negozio virtuale delle apps che è nato contestualmente; su entrambi i tipi di prodotto-servizio si è scatenata una guerra competitiva furibonda che in pochi anni ha travolto la "vecchia" informatica pesante e presente solo nelle imprese e ha dato luogo ad un mercato di consumo di prodotti/servizi per il consumatore che si inerpica in crescite esponenziali;
  - ancora possiamo citare le innovazioni introdotte con la "economia della condivisione": ad esempio con biciclette e auto utilizzabili da parte di chiunque entri a far parte della comunità di condivisione dei mezzi per la mobilità urbana, in cui i mezzi di trasporto non sono più di proprietà del singolo ma ognuno li può usare sulla base del tempo necessario e pagare di conseguenza;
  - un cambiamento simile si è verificato nel mercato della musica che è diventata tutta digitale in termini sia di produzione che di distribuzione e consumo e la stessa cosa sta avvenendo, anche se con più lentezza, per l'editoria con gli e-book.

L'innovazione di sostegno

- Per innovazione di sostegno intendiamo quella del piccolo imprenditore commerciale che apre un secondo punto vendita dopo il successo del primo e così in avanti; questo tipo di innovazione si focalizza sulla ottimizzazione delle funzioni d'uso secondarie del nuovo prodotto/servizio o ancora di più sulle modalità di promozione, commercializzazione, produzione e gestione introducendo tecnologia, procedure e organizzazione con l'obiettivo di migliorare la produttività dei processi a parità di prodotto/servizio ormai affermato sul mercato.

- Nella innovazione di sostegno verrà ottimizzata l'intera filiera che va dalla definizione e acquisizione dei componenti all'assemblaggio del nuovo prodotto/servizio, e successivamente alla distribuzione fino alla assistenza tecnica del prodotto installato presso il suo utente finale; nella filiera di fornitura che parte dal cliente finale sono coinvolti diversi partner con specializzazioni diverse: l'ottimizzazione della produttività della filiera passa attraverso l'analisi delle modalità con cui sia possibile aumentare i volumi di produzione del prodotto di successo a fronte di costi unitari decrescenti, spinti dall'incremento dei livelli di automazione in cui aumenta il capitale investito e diminuisce la quantità di lavoro umano.

La diffusione del web ha introdotto, in generale, una modalità completamente diversa del fare innovazione: prima del web le innovazioni erano appannaggio dei grandi sistemi industriali, della ricerca istituzionale e delle multinazionali e seguivano un approccio sostanzialmente top down; con il web le applicazioni promosse e commercializzate sul web hanno cominciato a svilupparsi sempre più in fretta e con crescente intensità, affermandosi come un vero e proprio motore per l'innovazione. Dopo il web, con la possibilità di produrre "contenuti" basati sulla conoscenza, chiunque può diventare un potenziale innovatore; una delle ragioni è che il costo per provare a realizzare un nuovo servizio (prodotto immateriale) si è ridotto fino ad avvicinarsi a zero; l'innovatore investe su se stesso e sperimenta fino a che vince con un risultato positivo oppure cambia

strada o rinuncia; per un sito, per una app, per sviluppare un programma puoi spendere poche migliaia di euro.

Una situazione del tutto simile sta prendendo forma anche per la realizzazione di prodotti materiali attraverso le stampanti 3D: questa nuova tecnologia favorisce la crescita di una nuova ondata di innovazioni nel mondo manifatturiero. Come la introduzione del web ha permesso che chiunque possa diventare produttore di "contenuti", anche le stampanti 3D permettono a chiunque di diventare produttore di artefatti materiali: la difficoltà da superare è costituita più che dal funzionamento della macchina di stampa, dalla conoscenza dei nuovi materiali utilizzabili per queste produzione; le tecnologie di produzione diventano accessibili e con una spesa per macchina di 2-3 mila € e si può cominciare a sperimentare una produzione impegnando il proprio tempo volontariamente; **i** costi unitari di produzione sono bassi e indifferenti ai volumi di produzione e alla complessità del prodotto; nascono i **fablab** come centri di servizio in cui promuovere la sperimentazione e i **kit per la costruzione di una impresa start-up** su cui inoltre si possono ottenere finanziamenti incentivanti. Con la stampante 3D, l'innovazione di prodotto (per i prodotti consumer più semplici - fino a 30 componenti) diventa molto più alla portata (vedi la possibilità di prototipare e verificare n funzioni d'uso aggiuntive), ma ci possiamo portare dietro anche le innovazioni di processo simulando prima e realizzando poi, con opportune integrazioni, i miglioramenti operativi della realizzazione del nuovo prodotto; per l'innovazione di prodotti industriali (B2B) complessi la stampante 3D può venire utilizzata prevalentemente per l'approvvigionamento da fornitori di componenti relativamente semplici; diventa praticabile la produzione a distanza di prodotti, ciascuno dei quali può anche essere personalizzato in funzione delle esigenze del compratore: i disegni digitalizzati dei vari prodotti, corredati dalle indicazioni sui materiali di produzione (che costituiscono il materiale stampato con la tecnica additiva), possono essere trasmessi ad una macchina remota che "stampa" il prodotto corrispondente al disegno; basta con costi di produzione centralizzati e focalizzati sulla riduzione dei costi unitari di produzione per alti volumi di prodotti assolutamente identici.

Infine una ultima osservazione: mentre nella innovazione di prodotto/servizio sono più importanti e diffuse le competenze innovative, nella innovazione di sostegno assumono maggiore importanza le competenze replicabili. Per poter analizzare in modo più preciso quali siano le competenze innovative e quelle replicabili abbiamo però bisogno di un ulteriore delimitazione di campo: quali sono le tipologie di processo che vengono attivate dalla realizzazione di un prodotto/servizio dirompente. L'ipotesi di lavoro è quella di considerare a questo scopo l'articolazione dei processi nella catena del valore [3].

Il valore generato dalla azione di vendita del prodotto/servizio costituisce il miglior indicatore di successo per tutti gli operatori della filiera che contribuiscono alla realizzazione del prodotto/servizio e alla sua "vita" una volta installato/utilizzato da un consumatore soddisfatto. Identifichiamo i seguenti 8 processi che corrispondono ad altrettante aree disciplinari diverse dal punto di vista delle competenze:

- Marketing & Vendite (M&V)
- Assistenza Tecnica al Cliente (ATC)
- Infrastrutture, Processi Operativi e Logistica (IPOL)
- Catena di Approvvigionamento (CA)
- Risorse Umane (RU)
- Ricerca e Sviluppo (R&S)

- Tecnologie Digitali e dell'Informazione (TDI)
- Amministrazione, Finanza e Controllo (AF&C)

L'innovazione organizzativa vede sempre più i 7 processi fondamentali dell'impresa concentrati nel servire il cliente finale a cui viene proposto un prodotto/servizio: tale concentrazione si sviluppa comunque, ma diventa essenziale per una organizzazione che spinge un nuovo prodotto intelligente sostenuto da una soluzione tecnologica dirompente. La figura 2 rappresenta schematicamente la centralità del prodotto acquistato e utilizzato dal cliente finale (P); viene messo in evidenza che la fondamentale leva del marketing mix è costituita dal prodotto/servizio.

Fig. 2-L'organizzazione innovativa centrata sul cliente di un prodotto

Può essere utile, tenendo conto delle indicazioni precedenti, introdurre una nuova competenza di e-business: possiamo definirla come l'uso dell'ICT nelle attività dell'impresa che generano business. La vendita è basata sullo scambio di prodotti e servizi fra persone, gruppi e imprese ed è visto come una attività essenziale per ogni "affare". I metodi di e-business abilitano le imprese e le organizzazioni ad integrare i loro sistemi di elaborazione dei dati interni ed esterni per ottenere prestazioni più flessibili, per lavorare in modo più efficace fra fornitori e partner, e per meglio soddisfare i bisogni e le aspettative dei loro clienti e infine per un efficace ed efficiente gestione delle funzioni interne.

Un'ultima osservazione sul nuovo prodotto: il nuovo prodotto ha tanto maggiore successo quanto più il suo posizionamento è diverso da quelli concorrenti, fornendo in termini di funzioni d'uso una diversa e più efficace soluzione per soddisfare bisogni latenti di clienti e utenti, non serviti in precedenza. Il nuovo prodotto intelligente d'altra parte, oltre alle caratteristiche distintive che lo fanno apprezzare ai clienti precursori (quelli che capiscono in anticipo le nuove funzioni d'uso e le adottano acquistando il nuovo prodotto), è caratterizzato dall'arricchimento delle sue funzioni con l'aggiunta di sensori e di una piattaforma di comunicazione incorporata, che permette a ciascun prodotto, una volta venduto e installato, di :

- ¥ conoscere in quale contesto viene utilizzato e da chi, associando il modello del prodotto all'utente e al cliente;
- ¥ comunicare il proprio stato di funzionamento e la prestazione relativa alla centrale di assistenza tecnica a cui viene connesso al momento della installazione;
- ¥ monitorare il proprio livello di prestazione, diagnosticare, la dove ne emerga l'esigenza, eventuali guasti che giustifichino la riduzione della prestazione, intervenire nella eventuale sostituzione dei componenti guasti seguendo o il programma incorporato preventivamente nel prodotto oppure eseguendo le operazioni di ripristino ricevute dalla centrale.

Ricordiamo che le altre leve del mix sono il prezzo, la promozione e il punto vendita e che la missione del Marketing è quella di posizionare il prodotto e di

pianificarne l'uscita in modo che la combinazione delle 4 leve ricordate permettano di ottimizzare il 1° margine di contribuzione a seguito di ciascuna vendita. In figura 3 vengono indicati in rosso i componenti che aggiungono intelligenza ad un prodotto materiale tradizionale. Nel caso di nuovi prodotti materiali realizzati con la tecnologia delle stampanti 3D è possibile inoltre aggiungere intelligenza alle modalità con cui il prodotto viene disegnato su misura per ogni specifico cliente.

Con le considerazioni fatte, siamo in grado di proporre una delimitazione di campo per la rilevazione e la messa a fuoco delle competenze per l'innovazione, tenendo conto anche di un primo risultato derivante dalla analisi di centinaia di casi di innovazione di successo:

- da una parte, per l'innovazione di business (dirompente o di sostegno), l'organizzazione ha bisogno di una combinazione di competenze innovative per ognuna delle 8 aree disciplinari identificate;
- dall'altra, emerge con chiarezza che l'innovazione dirompente è sempre legata ad un nuovo prodotto/servizio portato sul mercato attraverso un altrettanto nuovo e adeguato modello di business.

In sintesi: l'acquisizione di nuovi clienti e mercati è fortemente legata a nuova offerta di prodotti materiali e immateriali (servizi) e il successo di business è nella grande parte dei casi dovuto ad un cluster di 4 tipologie di competenze: Marketing & Vendite (M&V), Ricerca &Sviluppo (R&S), Tecnologie Digitali e dell'Informazione (STI), Amministrazione Finanza & Controllo (AF&C).

Viceversa il successo della innovazione di sostegno è nella grande maggioranza dei casi dovuto ad un cluster di 3 tipologie di competenze: Infrastrutture, Processi Operativi e Logistica (IPOL) nonché Tecnologie Digitali e dell'Informazione (TDI) e Amministrazione Finanza & Controllo (AF&C).

In conclusione il contributo delle Tecnologie al successo della innovazione di business può essere specificato come segue:

- ¥ l'ICT contribuisce, insieme alle Tecnologie Operative, sia all'innovazione dirompente che a quella di sostegno; la modularizzazione del lavoro in varie fasi di trasformazione specializzata è costituita da una combinazione specifica per impresa di Tecnologie ICT e TO;
- ¥ l' ICT è particolarmente importante nella innovazione di sostegno dove aiuta ad ottimizzare le prestazioni di tutti gli operatori coinvolti nella catena del valore. Se poi il prodotto innovativo è immateriale (si colloca cioè fra i servizi) il contributo delle Tecnologie ICT può essere ancora più rilevante perché la tecnologia in questo caso assume anche il ruolo di tecnologia di produzione dei servizi vendibili, per i prodotti materiali svolto dalle Tecnologie Operative.
- ¥ In sintesi, l'ICT contribuisce ad entrambi I due tipi di innovazione (quella dirompente e quella di sostegno) per ottimizzare tutti i processi operativi di fornitura, produzione e distribuzione. L'implicazione sul piano delle competenze IT è duplice:

- ♣ gli specialisti ICT devono imparare ad analizzare, modellare e condividere con i colleghi professional di altre aree disciplinari le caratteristiche dei prodotti offerti e completare le relative competenze d'uso professionale delle tecnologie e dei processi operativi di sviluppo, gestione e manutenzione;
- ♣ gli specialisti di altre aree disciplinari (professional competenti nelle aree del Marketing &Vendite, delle Tecnologie Operative, degli Approvvigionamenti, della Assistenza Tecnica, della Amministrazione, ecc.) devono acquisire competenze d'uso professionale (e quindi di alto livello) delle tecnologie digitali e dell'ICT .

Come dice E. Moretti nel suo saggio "La nuova geografia del lavoro" [4], è ormai evidente "la necessità di recuperare il tempo perduto, soprattutto in Italia, nell'accettare la svolta postindustriale per far fronte al nuovo mostro sacro della innovazione continua basata sulla scienza e sulla creatività; ogni Paese è oggi obbligato a prendere atto che una omissione gigantesca è stata compiuta, e che due gruppi umani fondamentali ora devono essere messi in prima linea: i professionisti e i creativi. Diventa più che evidente che nella geografia del lavoro è il fattore tecnologico-scientifico che fa la differenza tra gli operatori che sopravvivono e quelli che soccombono".

Con le considerazioni svolte possiamo ora proporre una modifica alla filiera cognitiva delle competenze digitali secondo il punto di vista di AICA: la modifica consiste nella introduzione di una ulteriore articolazione della filiera in particolare per quanto riguarda la componente delle competenze d'uso a livello professionale che riguardano i professionisti di impresa di aree disciplinari diverse dal quelle ICT e che comunque riguardano almeno i 4/5 dell'insieme delle risorse professionali di ogni organizzazione NON-ICT (stima riportata in "Makers" di Chris Anderson [5]).

Per raccogliere la sfida delineata nel saggio di Moretti proponiamo infine di sviluppare le competenze digitali dei cluster professionali individuati focalizzandoli soprattutto sui temi della innovazione (sia dirompente che di sostegno): il CEN/ISSS ha denominato queste competenze come di e-leadership e nei paragrafi seguenti cercheremo di tenere conto di questa indicazione (fig. 4) e del rilievo che assumono le competenze di Marketing & Vendite nel caso di innovazione di prodotto.

Fig.4-LA NUOVA FILIERA COGNITIVA DELLE COMPETENZE DIGITALI

L'approccio proposto permette di sviluppare una linea di approfondimento che va oltre quella già sviluppata nel progetto sulla e-leadership di Empirica, INSEAD e IDC i cui risultati sono stati presentati a metà del 2013: in questa ricerca il focus è stato prevalentemente sui cosidetti CIO-Chief Information Officer e meno sui contributi dei **professionisti** delle altre aree disciplinari.

Il modello che segue propone un approfondimento per la descrizione delle competenze digitali di uso professionale per profili di aree disciplinari diverse da quelle ICT e finalizzate alla innovazione.

Il modello descrittivo adottato per rappresentare combinazioni di competenze digitali per l'innovazione

Le considerazioni fatte fino a qui delineano il problema da affrontare per la definizione del mix di competenze necessarie per un innovatore: mentre il mondo dell'istruzione di qualunque livello è organizzato per competenze da costruire per i propri studenti secondo aree disciplinari riconosciute e delimitate, nel mondo del lavoro sono in corso fenomeni di scomposizione e ricomposizione delle varie professionalità in funzione delle direzioni verso cui si orientano le innovazioni di business basate su nuovi prodotti ad alto contenuto scientifico e tecnologico: nascono quindi, lentamente ma inesorabilmente, nuove configurazioni di competenze richieste dalle imprese che propongono e innovano questi prodotti. A titolo esemplificativo possiamo ricordare nuove aree disciplinari già emerse come quelle dei bioingegneri , piuttosto che degli ingegneri ambientali o dei meccatronici, degli architetti di ambienti ospedalieri umanizzati, ecc. Ma le innovazioni di business accelerano queste ri-configurazioni di competenze e tendono ad ampliarle per poter sostenere nuove competenze di tipo imprenditoriale come quelle richieste, sempre a titolo esemplificativo, dai nuovi ambienti digitali in cui si sviluppano le attività didattiche a distanza piuttosto che le attività manifatturiere legate alle tecnologie di stampa tridimensionale. I nuovi spazi di istruzione e quanto richiesto in termini anche di apprendimento permanente, definiti come Sistemi di Prodotti Materiali e Immateriali, riguardano *"configurazioni dinamiche di persone, tecnologie, organizzazioni e informazioni condivise che creano ed erogano valore a clienti, fornitori e altri portatori di interesse (stakeholders)*.

La crescente importanza dell'innovazione dei prodotti e dei servizi costituisce una sfida fondamentale per chi lavora sia nelle imprese e nelle istituzioni di governo che per gli accademici che lavorano nella educazione di base e nella ricerca. Come risposta a questa esigenza sta emergendo un campo di ricerca con proprie specifiche caratteristiche una nuova "Scienza dei Prodotti e Servizi Innovativi", mirato a scoprire le logiche sottostanti di sistemi di prodotto e servizio ad alto livello di complessità e di mettere a fuoco un linguaggio comune e un framework concettuale condiviso.

A questo scopo è necessario adottare un approccio interdisciplinare per la ricerca e la educazione sui sistemi di prodotto e servizio innovativi. Sta prendendo piede la definizione di **T-shaped professionals** (professionisti dotati di mix di competenze a forma di T): sono coloro in grado di fornire soluzioni in profondità in base alle loro capacità di esperti nella loro disciplina centrale, ma che inoltre possiedono capacità di comunicazione complessa per interagire con specialisti di una ampia gamma di altre aree disciplinari e funzionali.

Il modello del portafoglio delle competenze a T è stato proposto come risultato del progetto sulla e-Leadership di Insead, Empirica, IDC-2013: il livello a cui si colloca il portafoglio delle competenze è quello del CIO che assume, nel modello proposto, il ruolo di innovatore digitale dell'impresa in cui opera. Il risultato della

ricerca è interessante per quanto riguarda la stima degli e-leader necessari all'Europa, mentre proponiamo ulteriori approfondimenti per quanto riguarda la definizione delle competenze trasversali che l'e-leader dovrebbe possedere.

Sul modo in cui contribuire ad accelerare la costruzione intenzionale di competenze per l'innovazione di business, comunque supportata dalle competenze digitali, può essere utile articolare un percorso su due livelli di complessità.

Il **primo livello di complessità** è quello che si riscontra quando la e-leadership viene esercitata da una singola persona che sviluppa una propria idea di prodotto e business successivo mettendo in piedi una start-up; in questa ipotesi l'innovatore deve possedere un portafoglio di competenze tecniche a forma di T, dove:

- le competenze dell'asse verticale sono quelle della esperienza specialistica accumulata dall'innovatore in una delle aree disciplinari identificate nel capitolo precedente, combinate con una adeguata conoscenza del settore di business nel cui contesto si colloca l'innovazione stessa (commercio, servizi professionali, costruzioni, elettronica, energia, informatica, ecc.); per definizione, in questo caso la nuova idea di business è centrata su un nuovo prodotto/servizio che dovrà posizionarsi in uno spazio di mercato competitivo;
- le competenze dell'asse orizzontale dovrebbero comprendere oltre a quelle della visione e della comunicazione, quelle dell'area digitale e del marketing; in mancanza di queste sarebbe opportuna l'integrazione delle competenze specialistiche di riferimento con una competenza di uso professionale di metodi, strumenti e tecniche di queste due aree, digitale e marketing. Sulla base dei casi di successo fino ad oggi esaminati questo tipo di competenze vengono rilevate come sempre presenti

L'innovatore svilupperà gradualmente il suo team di progetto mano a mano che la start-up si consolida con le prime verifiche positive del nuovo prodotto e la selezione dei componenti del team insieme alla capacità di farli lavorare congiuntamente per l'obiettivo sarà uno dei principali fattori di successo del nuovo business.

Il **secondo livello di complessità** è quello che si incontra quando l'innovazione si sviluppa nell'ambito di una organizzazione già strutturata; in questo caso diventa importante valutare sia le caratteristiche delle persone che innovano che quelle della struttura organizzativa in cui l'innovazione si sviluppa: in un team di innovazione ci devono essere persone con professionalità diverse, dato che è il loro incontro che crea ricchezza. Questa ricchezza di competenze funziona bene quando si innesca un nuovo skill strutturale, come la capacità di fare planning e ri-planning dinamico sulla base dei vincoli di budget e di costo, piuttosto che la capacità di riadattare le specifiche di progetto in funzione di nuovi requisiti espressi dal cliente. I progetti di innovazione sono in continua ripianificazione. Serve una capacità di adattamento continuo, che significa modificare e adattare dinamicamente i piani di progetto, gli obiettivi, la struttura dei team avendo chiari sia gli obiettivi del cliente che i vincoli e la cultura di impresa.

Secondo quanto indicato da A.Fuggetta e G.De Michelis in "Quali sono le competenze di un innovatore digitale? - Mondo Digitale n° 1-marzo 2011", questo livello di complessità è tanto più alto quanto più l'innovazione riguarda un nuovo prodotto in un nuovo mercato. In questo caso si acquista un vantaggio di posizionamento in termini temporali che si può sfruttare con le nuove versioni del prodotto da rilasciare al momento opportuno, anche quando la concorrenza ha

cominciato a muoversi: il caso richiamato a titolo esemplificativo è quello dell'iPhone e dell'iPad visti, oltre che come terminali altamente "amichevoli" rispettivamente per la comunicazione e per l'intrattenimento e il lavoro in mobilità, anche (ormai oggi soprattutto) come terminali per l'acquisto e l'utilizzo di una molteplicità di servizi on line (Apple ha raggiunto con il suo negozio virtuale globale il milione di apps, per un mercato che va verso il miliardo di pezzi venduti).

Anche in questo caso il richiamo della configurazione a T delle competenze può essere utile, ma il risultato di questa combinazione riguarda sia le persone che l'organizzazione e quindi può assumere configurazioni diverse da una organizzazione all'altra; il team si configura come costituito da specialisti esperti ma ciascuno dotato della capacità di parlare con e capire dal tuo collega che ha un'altra specializzazione quanto va messo in comune e come per lo sviluppo della innovazione su cui tutto il team sta lavorando; tutti devono poi avere almeno l'idea che dietro al progetto c'è un costo, c'è una manutenzione, c'è una vita di progetto. Servono persone con competenze forti ma che sanno aprire lo sguardo, parlare con gli altri, interagire.

Le caratteristiche sono quelle di una struttura organizzativa che valorizza l'incontro delle persone e riesce a coniugare il caos di un progetto innovativo con il bisogno di convergere. Non è solo caratteristica della persona o delle singole persone: deve essere anche dentro le procedure interne di gestione, dei meccanismi di allocazione delle risorse, dei meccanismi di valutazione e premio, dei modi in cui il management prende le decisioni. E il progetto va continuamente ri-pianificato, e non perché ha dei ritardi, ma perché fa parte della *normalità* non innamorarsi di quello che è stato fatto e si sta facendo; è intrinseco della *natura* di un progetto di innovazione che i membri del team sappiano riconoscere che quanto fatto finora debba essere rifatto, recuperando ovviamente quanto ancora di utilizzabile può essere salvato.

Con queste premesse vediamo come si configurano, a titolo esemplificativo, i portafogli di competenze a T di due aree disciplinari: quelle relative alla Digital Innovation Leadership (Fig. 5) e alla Marketing e-leadership (Fig. 6).

Fig.5-Il portafoglio a forma di T delle competenze nell'area della Digital Innovation Ledership

La configurazione del portafoglio di competenze a T per un innovatore in ciascuna area disciplinare è stata costruita sulla base dalle seguenti assunzioni:

- di solito l'innovatore è uno specialista che ha accumulato una esperienza importante (non necessariamente perché lavora da lungo tempo ma piuttosto per l'intensità e il livello di complessità affrontati) in una specifica area disciplinare e in uno o più settori merceologici; anche se ha assunto una responsabilità gestionale, nel momento in cui si assume la responsabilità di un progetto di innovazione, soprattutto se si tratta di una innovazione di prodotto, prevale la sua competenza professionale rispetto a quella manageriale;

- nel caso dello sviluppo di una innovazione di prodotto le competenze fondamentali sono quelle del settore merceologico, quelle del marketing per il posizionamento (il nuovo prodotto deve nascere competitivo rispetto ai prodotti già sul mercato) e la scelta del modello di business e, infine, quelle digitali per poter aggiungere intelligenza al prodotto e gestirne poi la realizzazione e la distribuzione, monitorando il miglioramento della prestazione dei processi;

- si intende, infine, che le competenze aggiuntive dell'innovatore (quelle cioè su cui non ha sviluppato la sua esperienza di riferimento, che sono da collocare sull'asse orizzontale della T), siano quelle a livello di area disciplinare e sia sufficiente che vengano acquisite in termini di comprensione delle finalità e dell'utilizzo delle tecniche, dei metodi e degli strumenti da applicare al caso della innovazione, senza necessariamente diventarne uno specialista; la vera specializzazione dell'innovatore sarà infatti quella di tipo imprenditoriale, a valle sia di un insuccesso che di un eventuale successo.

Fig.6-Il portafoglio a forma di T delle competenze nell'area della Marketing e-leadership

L'innovatore digitale o quello di marketing è responsabile di un progetto complesso, che può condurre sia nell'ambito di un'impresa già consolidata che come neoimprenditore di una start-up; deve avere una profonda competenza specialistica sia di tipo tecnologico , ma anche sul settore specifico e sui principali concorrenti, avere capacità di interagire con altre culture specialistiche come quella del designer e dell'esperto di marketing, con cui verge il nuovo prodotto/servizio , spesso co-progettato con cliente/utente finale, avere ancora capacità di ascolto delle indicazioni del cliente finale a cui aggiungere una capacità interpretativa che permette di trasformare indicazioni latenti in dettagli utili per la definizione di nuove funzioni d'uso da realizzare attraverso adeguate soluzioni tecnico operative.

L'innovatore, infine, opera in un contesto in cui il risultato finale del progetto è scarsamente o per nulla definito dato che si deve confrontare con nuovi bisogni fino ad oggi non emersi (latenti); deve quindi saper prendere decisioni che correlano aspetti di tipo tecnico e aspetti di mercato e di costo e saper gestire una molteplicità di stakeholders.

Conclusioni

Abbiamo preso spunto dalla evoluzione delle tecnologie digitali per cercare di capire che cosa si potrebbe fare per sviluppare iniziative di formazione sulle competenze per l'innovazione: la molteplicità dei fattori sociologici, economici, tecnologici e istituzionali che spingono verso l'innovazione prefigurano anche un modello di lavoro flessibile, in cui competenze replicabili e competenze innovative si integrano per favorire l'innovazione per ogni tipo di organizzazione.

In particolare ci siamo concentrati sulla innovazione di organizzazioni al di fuori del settore digitale, mettendo a fuoco che ogni organizzazione pubblica o privata,piccola o grande, potrebbe fare innovazione studiando e realizzando prodotti/servizi intelligenti, accompagnandoli con adeguate e coerenti innovazioni di sostegno.

Le due domanda a cui abbiamo cercato risposte sono:

1) quanti sono gli innovatori nei settori diversi da quello digitale?

2) quali sono le competenze digitali che dovrebbero avere i professional e i manager di aree disciplinari diverse da quella digitale e dell'ICT, per poter spingere l'innovazione?

Per la domanda 1) la stima formulata a livello europeo nella ricerca Empirica, Insead e IDC è che ci siano al di fuori del settore digitale da 4 a 5 volte il numero degli innovatori che potrebbero avere bisogno di competenze digitali per migliorare la loro capacità di contribuzione alle innovazioni nei diversi settori.

Per la domanda 2) abbiamo provato a rispondere su due piani:

- la messa a fuoco di una nuova filiera cognitiva delle competenze digitali in cui dare corpo e spessore alle competenze di uso professionale, in particolare selezionate fra quelle creative che possono contribuire alla progettazione e realizzazione di nuovi prodotti intelligenti, realizzati, distribuiti e assistiti con intelligenza digitale; nulla toglie che siano comprese anche competenze digitali replicabili per la ottimizzazione della produzione, della distribuzione e della assistenza di prodotti/servizi esistenti;
- la definizione di un modello di portafoglio delle competenze a T, valido per gli innovatori che operano o in start-up o in imprese strutturate e consolidate in cui lanciare nuovi prodotti/nuovi business; la esemplificazione fatta per l'Innovatore Digitale e per l'Innovatore di Marketing in ambiente digitale (Marketing e-leader) va estesa ad altre aree.

C'è moltissimo lavoro da fare per completare le risposte abbozzate, se considerate valide.

Bibliografia

[1] Clayton Christensen, Il dilemma dell'innovatore, Franco Angeli, 2001

[2] Clayton Christensen, Michael Raynor, Il dilemma dell'innovatore: la soluzione, Etas, 2004

[3] Michael Porter, Il vantaggio competitivo, Einaudi, 2004

[4] Enrico Moretti, La nuova geografia del Lavoro, Mondadori, 2013

[5] Chris Anderson, Makers - il ritorno dei produttori, Rizzoli Etas, 2013

Biografia

Roberto Bellini, ingegnere, è past-President della Sezione AICA di Milano (Associazione di Informatici) e consigliere nel Direttivo AISM (Associazione di Marketing); per AICA è inoltre responsabile dell'area delle Competenze degli Specialisti ICT (EUCIP - European Certification of Informatics Professionals e e-Competence Services).

È membro del Comitato Scientifico di Mondo Digitale, rivista di AICA, per cui è co-responsabile delle rubriche "ICT per il successo di business" e "Professioni ICT". Contribuisce dal 2012 al Tavolo Tecnico di UNINFO che ha definito la NormaTecnica UNI 11506 sulle competenze europee ICT e dal 2005 al Workshop CEN/ISSS che lavora sulla produzione del sistema del e-competence frame work e dei profili europei. Dal 2010 coordina la Ricerca europea del CEPIS sulle Competenze e le Professionalità ICT.

Ha fatto parte, dal 2010 al 2012, della Giuria del Premio Nazionale Innovazione nei Servizi-Confcommercio, che concorreva al Premio dei Premi per l'Innovazione patrocinato dal Presidente della Repubblica.

Nel 2009-2010 ha contribuito, come esperto CNEL, allo sviluppo del Modello CNEL su "Competenze e Professionalità". Come consulente scientifico ha collaborato al progetto Regione Veneto 2010-2011 su "Validazione e certificazione delle competenze negli ambiti formali di apprendimento". Nel 2012 ha contribuito alla stesura del position paper RUIAP sul Diritto alla Competenza.

Svolge da anni attività di ricerca e docenza (prima all'Università di Bergamo e successivamente al Politecnico di Milano) sui temi della Innovazione e delle relative Competenze digitali e di marketing, in collaborazione con la Fondazione Politecnico di Milano.

Email: roberto.bellini@aicanet.it

AICA
Associazione Italiana per l'Informatica ed il Calcolo Automatico

ROTARY INTERNATIONAL
Distretti: 2031, 2032, 2041, 2042, 2050, 2080, 2100, 2110 e 2120

ETIC

2013-2014

Etica e Tecnologie dell'Informazione e della Comunicazione

10 PREMI

1 premio nazionale dell'importo di € 3.150
9 premi distrettuali dell'importo di € 2.650

messi in palio da:

AICA e ROTARY INTERNATIONAL

per tesi di laurea o di dottorato
su argomenti concernenti l'area dell' ETIC

Etica e Tecnologie dell'Informazione e della Comunicazione

Possono partecipare al concorso laureati che abbiamo conseguito il titolo presso una Università italiana o dottori di ricerca che abbiano consegnato formalmente la tesi o superato l'esame finale di dottorato nel periodo:

1 aprile 2013 - 28 febbraio 2014

LA DOMANDA DI PARTECIPAZIONE DOVRA' ESSERE PRESENTATA
ENTRO IL 7 MARZO 2014
PER VIA TELEMATICA, ACCEDENDO AL SITO DI AICA:
www.aicanet.it

con il patrocinio di
Fondazione CRUI

Segreteria del premio:
AICA - P.le R. Morandi 2 - 20121 Milano
email: segreteria@aicanet.it

Premi ETIC 2013-2014 per Tesi di laurea e di dottorato di ricerca sul tema:

Etica e Tecnologie dell'Informazione e della Comunicazione

INTRODUZIONE

Il tema dell'Etica e Tecnologie dell'Informazione e della Comunicazione (o Computer Ethics) è sempre più di attualità nella società dell'informazione e della conoscenza. L'evoluzione rapida e continua delle tecnologie dell'informazione, la loro pervasività in tutte le attività e la criticità crescente dei servizi offerti rendono sempre più importante che gli operatori del settore abbiano piena coscienza delle implicazioni etiche delle loro scelte e decisioni e che la scuola e le associazioni professionali si occupino di questi problemi.

A titolo indicativo, potranno essere considerate ai fini del presente concorso tesi di laurea e di dottorato di ricerca che trattano le implicazioni etiche e sociali dell'ICT e in generale delle tecnologie digitali in settori quali:

- la formazione (e-learning, digital divide, nativi digitali…….)
- l'economia (processi decisionali, finanza computerizzata, globalizzazione, …)
- il lavoro (rapporti personali, garanzie sociali, organizzazione del lavoro, home office, nomadismo digitale, occupazione e disoccupazione indotta….)
- la ricerca (intelligenza artificiale, rapporto uomo-macchina, robotica, potenziali effetti delle nanotecnologie…..)
- la salute (impatti sul sistema sanitario, rapporto medico-paziente, applicazioni mediche delle tecnologie digitali,…..)
- l'informazione (riservatezza, proprietà intellettuale, affidabilità dei sistemi informativi, green ICT…)

Art. 1

Rotary International, con i Distretti 2031, 2032, 2041, 2042, 2050, 2080, 2100, 2110 e 2120, e AICA (Associazione Italiana per l'Informatica ed il Calcolo Automatico) con il patrocinio della Fondazione CRUI (Conferenza dei Rettori delle Università Italiane) indicono un concorso per l'assegnazione di 10 premi, uno dell'importo di € 3.150,00 (tremilacentocinquanta) e nove dell'importo di € 2.650,00 (duemilaseicentocinquanta), al lordo di eventuali oneri di legge, da destinare a laureati specialistici, magistrali o quinquennali o dottori di ricerca delle Università Italiane che abbiano svolto una tesi di laurea o di dottorato su argomenti concernenti l'area dell'ETIC: Etica e Tecnologie dell' Informazione e della Comunicazione.

Art. 2

1. Il concorso è riservato ai laureati specialistici, magistrali e ai laureati in corsi di laurea a ciclo unico di durata quinquennale e a dottori di ricerca.

2. Possono partecipare al concorso laureati che abbiano conseguito il titolo presso un'Università Italiana con un punteggio non inferiore a 106/110 o 96/100 nel periodo 1 aprile 2013 – 28 febbraio 2014 o dottori di ricerca che abbiano consegnato formalmente la tesi o superato l'esame finale di dottorato nel periodo 1 aprile 2013 – 28 febbraio 2014.

3. Il premio di € 3.150,00 è assegnabile a candidati residenti su tutto il territorio nazionale, mentre ciascuno dei rimanenti nove premi (da € 2,650,00) è riservato ai candidati residenti nel territorio di pertinenza di ciascuno dei Distretti Rotary partecipanti a questo bando.

Art. 3

La domanda di partecipazione al concorso dovrà essere presentata entro il 7 marzo 2014 esclusivamente in formato elettronico accedendo al sito di AICA (www.aicanet.it).

Art. 4

1. Nella domanda di partecipazione il candidato deve dichiarare, sotto la propria responsabilità, pena l'esclusione dal concorso:

a. il cognome ed il nome;
b. la data e il luogo di nascita;
c. residenza, recapito telefonico, indirizzo di posta elettronica;
d. titolo della tesi di laurea o di dottorato;
e. cognome e nome del Relatore della tesi
f. Ateneo dove ha conseguito laurea/dottorato

2. Unitamente alla domanda, il candidato, pena l'esclusione, deve presentare:

a. la copia del certificato di laurea (in formato pdf) con riportate la data del conseguimento della laurea e la votazione conseguita oppure la copia (in formato pdf) del certificato attestante in conseguimento del titolo di dottore di ricerca ovvero la ricevuta (in formato pdf) della consegna della tesi di dottorato
b. una presentazione di non più di 5 pagine, in formato pdf, della tesi formulata secondo lo schema indicato nell'Allegato 1;
c. una lettera, in formato pdf, di presentazione da parte del relatore della tesi, che copra i punti indicati nell'Allegato 2; la lettera dovrà essere redatta su carta intestata e dovrà recare la firma del relatore, pena l'esclusione dal concorso;
d. l'autorizzazione a utilizzare i dati forniti dal candidato ai fini del presente bando, nonchè alla divulgazione, con i mezzi e nei modi ritenuti più opportuni, delle tesi premiate;
e. elenco di eventuali pubblicazioni redatte dal candidato sull'argomento della tesi.

Art. 5

1. La Commissione giudicatrice nominata da Rotary International, Distretti 2031, 2032, 2041, 2042, 2050, 2080, 2100, 2110 e 2120 e da AICA, è costituita: da un rappresentante AICA che ne assume la presidenza, da un rappresentante per ognuno dei nove Distretti. La Segreteria della Commissione è a cura di AICA.

2. La Commissione, a suo giudizio insindacabile, dopo aver valutato secondo i criteri indicati nell'Allegato 3 le domande regolarmente pervenute, seleziona fino a 20 candidati finalisti per i premi. Ai candidati finalisti sarà richiesto l'invio in formato pdf della tesi di laurea e di eventuali pubblicazioni scientifiche redatte dal candidato.

3. La Commissione, esaminate le tesi presentate dai candidati in base ai criteri indicati nell'Allegato 3, proclama a suo insindacabile giudizio, il vincitore del premio nazionale e i vincitori dei nove premi legati alla residenza dei candidati nel territorio di ciascun Distretto partecipante al presente bando.

Art. 6

1. I vincitori del concorso riceveranno comunicazione scritta del conferimento del premio a mezzo posta elettronica. Dopo aver ricevuto la comunicazione ciascun vincitore dovrà tempestivamente contattare la Segreteria del Premio per confermare la partecipazione alla cerimonia di premiazione.

2. La consegna dei premi avverrà nel corso di una cerimonia che si svolgerà entro il mese di giugno 2014.

3. Il premio dovrà esser ritirato personalmente da ciascun vincitore. Il mancato ritiro personale del premio, se non dovuto a cause eccezionali, comporta la perdita del premio stesso.

Art. 7

Nel caso in cui non si presentassero Candidati oppure la Commissione giudicatrice ritenesse di non assegnare tutti i premi, l'importo residuo relativo sarà utilizzato per i premi dell'anno successivo.

Allegato 1 (Articolo 4, comma 2, lettera b)

Struttura della presentazione della tesi

La presentazione della tesi deve essere organizzata nelle seguenti sezioni:

- autore

- titolo

- relatore

- inquadramento del tema trattato e del lavoro svolto

- contributo ai temi di cui al presente bando

- innovatività dei risultati ottenuti

- impatto etico e sociale dei risultati ottenuti

- rilevanza scientifica dei risultati ottenuti ed eventuali pubblicazioni

Allegato 2 (Articolo 4, comma 2, lettera c)

Struttura della lettera di presentazione del relatore

La lettera di presentazione del relatore su carta intestata e firmata deve coprire i seguenti punti:

- impegno temporale nello svolgimento della tesi

- grado di autonomia nel lavoro svolto

- innovatività dei risultati ottenuti

- impatto etico e sociale dei risultati ottenuti

- rilevanza scientifica dei risultati ottenuti

Allegato 3 (Articolo 5, commi 2 e 3)

Criteri di valutazione della Commissione giudicatrice

1. Impatto etico e sociale dei risultati ottenuti
2. Rilevanza scientifica ed eventuali pubblicazioni
3. Innovatività dei risultati ottenuti
4. Chiarezza espositiva
5. Lettera di presentazione del docente relatore della tesi

****

Premi di laurea/dottorato AICA-CINI-CNIT

Edizione 2013

All'edizione 2013 del bando per l'assegnazione dei tre premi messi a disposizione da AICA,CINI e CNIT, sono state presentate 67 domande. I laureati e dottori di ricerca partecipanti al concorso provenivano da 33 Atenei Italiani. Erano rappresentati 23 diversi Corsi di Laurea Magistrale (o Specialistica) e 7 Corsi di Dottorato di ricerca. Gli Atenei con il maggior numero di domande sono stati il Politecnico di Milano e le Università di Pisa, Roma "La Sapienza" e Roma Tre (con 5 domande ciascuno).

La Commissione preposta alla selezione delle tesi di laurea e dottorato, ha operato come previsto da ciascun bando. Dopo aver esaminato che le domande presentassero i criteri esposti nei bandi, la Commissione ha proceduto alla valutazione dei sommari delle tesi e successivamente ha valutato le tesi complete di 12 candidati finalisti.

I lavori della Commissione, composta da rappresentanti di AICA e da docenti universitari, sono iniziati nel mese di luglio 2013 e si sono conclusi nel mese di novembre 2013. A conclusione della selezione, i premi sono stati assegnati come segue:

Premio AICA

Sara Amendola - Laurea magistrale in Ingegneria Medica – Università di Roma "Tor Vergata" per la tesi su "*Sistema per la classificazione dei movimenti corporei con la Tecnologia di Identificazione a Radiofrequenza*". (Relatore: prof. Gaetano Marocco).

Premio CINI

Michele Martinelli - Laurea Specialistica in Informatica – Università di Roma "La Sapienza" per la tesi su "Integrazioni di sensori sonar per under water SLAM (Simultaneous Localization and Mapping). (Relatore: prof. Andrea Sterbini).

Premio CNIT

Francesca Carminati – Laurea Specialistica in Ingegneria delle Telecomunicazioni – Politecnico di Milano per la tesi su "Passive radio localization: channel modeling and algorithms". (Relatore: prof. Monica Nicoli).

* * *

I premi (3.000 euro ciascuno) sono stati consegnati ai vincitori il 20 dicembre 2013.

La cerimonia di premiazione si è svolta a Pisa, presso il Museo degli Strumenti di Calcolo dell'Università, dove ciascuno dei premiati ha presentato il proprio lavoro di tesi.

Una sintesi della loro tesi premiate è presentata nelle pagine seguenti.

PREMIO AICA

Vincitore: Sara Amendola

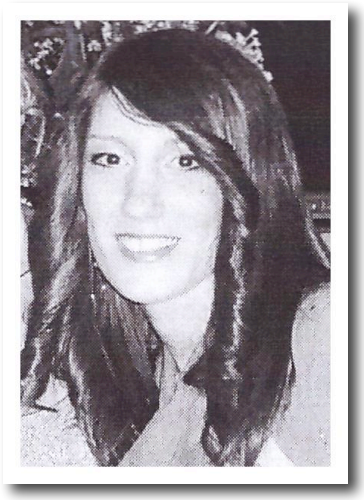
Titolo della tesi:

"Sistema per la classificazione dei movimenti corporei con la Tecnologia di Identificazione a Radiofrequenza"

Laurea Magistrale in Ingegneria Medica — Università di Roma "Tor Vergata"

Relatore: Prof. Gaetano Marrocco

Abstract della tesi



Nell'ambito dell'*Activity Recognition* confluiscono tutti quei sistemi che, avvalendosi di tecnologie eterogenee, mirano al riconoscimento automatico delle azioni e alla classificazione di pattern motori, attraverso l'integrazione e l'elaborazione dei dati registrati da sensori indossabili e/o ambientali.
In ambito clinico, la disponibilità di dispositivi per il tracciamento e la misura del movimento dei segmenti corporei potrebbe supportare la diagnosi dei disordini neurologici associati a movimenti compulsivi degli arti o, ad esempio, consentire la messa a punto protocolli di riabilitazione personalizzati per soggetti con disabilità motorie e la costruzione di reti di rilevazione remota di eventi anomali e criticità quali crisi epilettiche o cadute accidentali.
Sistemi optoelettronici e dispositivi indossabili equipaggiati con sensori inerziali costituiscono ad oggi il riferimento per le tecnologie wireless di monitoraggio del movimento. Tale approcci, pur offrendo ottima affidabilità ed accuratezza, necessitano spesso di un ambiente fortemente strutturato e del costante intervento dell'utente per la manutenzione e la sostituzione delle batterie. Peso, dimensioni e costi non contenuti rendono tale classe dei dispositivi attivi poco adatti ad applicazioni su larga scala che implicano l'integrazione "trasparente" delle unità sensoristiche all'interno di cerotti ed indumenti lavabili e, al limite, monouso.
Soluzioni alternative valide potrebbero derivare dall'applicazione dalla tecnologia passiva di identificazione a Radiofrequenza (RFID) il cui utilizzo, ormai ampiamente diffuso nella logistica, appare promettente per lo sviluppo di dispositivi *low-cost* di monitoraggio del movimento del corpo umano.
In tale contesto è maturata l'idea del presente lavoro di tesi, concepito come sinergia tra le tematiche classiche dell'ICT, in particolare nell'ambito della Biotelemetria passiva, e le moderne tecniche di riconoscimento automatico e classificazione, mutuate da diversi campi di ricerca quali la *Brain Computer Interface* e la *Human Computer Interaction*. L'obiettivo principale dello studio è stato infatti quello di valutare le potenzialità di un innovativo sistema di classificazione dei movimenti basato su tecnologia RFID, attraverso un approccio multidisciplinare che fondesse insieme le competenze elettromagnetiche di progettazione e caratterizzazione sperimentale di antenne indossabili operanti nella banda UHF e lo sviluppo di algoritmi di *Signal Processing* e *Machine Learning*.
Sono stati progettati e realizzati radio-sensori privi di batteria che, posizionati sugli arti ed interrogati da un'antenna remota, sono in grado di trasmettere, unitamente ad un codice

identificativo, segnali elettromagnetici multi-canale che presentano una modulazione in ampiezza movimento-dipendente.
I risultati ottenuti nel corso delle numerose campagne sperimentali hanno rivelato che, applicando algoritmi standard di classificazione supervisionata, è possibile discriminare otto differenti movimenti degli arti con un'accuratezza superiore al 90%, che risulta del tutto confrontabile con quella ottenibile processando un segnale elettromagnetico acquisito con ricetrasmettitori attivi.
E' stata inoltre investigata la possibilità di impiegare tale piattaforma nello sviluppo interfacce di comunicazioni reali capaci di codificare i movimenti del corpo in un segnale di controllo per periferiche che permettono l'interazione del soggetto con l'ambiente circostante (*Assistive Communication*)
Merita infine menzionare che la metodologia di acquisizione del movimento messa a punto è estremamente versatile ed è stata applicata per la realizzazione di più complessi sistemi di *Ambient*
Intelligence facilmente adattabili a qualsiasi contesto assistenziale domestico o ospedaliero.

PREMIO CINI

Vincitore: Michele Martinelli

Titolo della tesi:

"Integrazione di sensori sonar per underwater SLAM"

Laurea Specialistica in Informatica — Università di Roma "La Sapienza"

Relatore: Prof. Andrea Sterbini

Abstract della tesi

Il progetto si incentra sullo studio dell'integrazione di un sensore sonar panoramico sul Venus, un sottomarino autonomo sviluppato dal Laboratorio di Robotica dell'ENEA per monitoraggio ed esplorazione. Attraverso l'analisi delle rilevazioni del sonar si cerca di risolvere il problema dello SLAM (*Simultaneous Localization And Mapping*), ossia costruire una mappa dell'ambiente subacqueo e utilizzarla allo stesso tempo per localizzare il veicolo.

PREMIO CNIT

Vincitore: Francesca Carminati

Titolo della tesi:

"*Passive radio localization: channel modeling and algorithms"*
Laurea Specialistica in Ingegneria delle Telecomunicazioni – Politecnico di Milano
Relatore: Prof. Monica Nicoli

Abstract della tesi

Oggetto della tesi è la definizione di una metodologia originale per la localizzazione passiva ("*device-free*") di una persona in movimento in un'area coperta da una rete wireless. Viene proposto un modello analitico, basato sulla teoria della diffrazione, per legare l'attenuazione della potenza ricevuta (in media e deviazione standard) alla posizione della persona, approssimata con un'interfaccia perfettamente assorbente, nell'area del collegamento radio. Il modello consente quindi di localizzare una persona (non dotata di dispositivo radio) a partire da misure di potenza effettuate lungo i collegamenti di una rete wireless.
È stato messo a punto un sistema di tracciamento utilizzando il modello analitico proposto e la tecnica del *particle filtering*. Attraverso il modello sono state ricavate in modo automatico le mappe radio riducendo notevolmente sia la complessità sia il dispendio temporale della fase di calibrazione, necessaria per adattare il sistema alle condizioni ambientali circostanti. Le prestazioni di posizionamento sono state migliorate attraverso l'implementazione dell'algoritmo *particle filter*.
Successivamente è stata svolta un'analisi dell'accuratezza di localizzazione, sia con simulazioni che in forma analitica valutando il limite fondamentale di Cramer Rao. Il metodo è stato validato attraverso svariati test sperimentali condotti con dispositivi ZigBee IEEE 802.15.4, presso il Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), sia in ambienti chiusi che aperti, in condizioni LOS e NLOS, con disposizione uniforme o irregolare dei dispositivi (rilassando i vincoli presenti nei convenzionali sistemi) ottenendo ottimi risultati di localizzazione.
La radio localizzazione passiva è un tema oggetto di crescente interesse negli ultimi anni per il potenziale impatto industriale e le possibili applicazioni ai cosiddetti servizi location-based. È una tecnologia efficiente e flessibile che permette di localizzare una persona senza che questa debba portare con sé un dispositivo né partecipare attivamente al processo di posizionamento (completamente passiva). Inoltre può essere applicata a reti radio già esistenti nell'area di interesse (es., WiFi, cellulari, reti di sensori, etc.), non richiedendo il dispiegamento di un'infrastruttura ad-hoc per la localizzazione. Possibili applicazioni sono nell'ambito della sorveglianza e la sicurezza, la domotica, gli edifici intelligenti e automatizzati (smart spaces e ambient-assisted-living), il monitoraggio di strutture critiche (ad es., per il tracciamento di operatori in ambienti industriali), emergenza e operazioni di soccorso, monitoraggio di flussi pedonali per la rilevazione automatica di situazioni a rischio in aree pubbliche e/o critiche (ad es., stazioni o aeroporti). I primi test sperimentali condotti nell'ambito della tesi dimostrano la validità della metodologia e fanno ben sperare in un suo futuro impiego.

PREMIO “Mario Tchou”

Il premio intitolato a Mario Tchou, pioniere dell’Informatica Italiana, è stato bandito nel 2013 e dedicato a tesi di laurea o dottorato di ricerca su argomenti concernenti l’impatto economico, industriale e sociale delle Tecnologie dell’Informazione. Al concorso sono state presentate 26 domande di laureati (laurea magistrale o specialistica) e di dottori di ricerca. I candidati provenivano da 18 Atenei Italiani ed erano rappresentati 10 diversi Corsi di Laurea Magistrale o Specialistica e 7 Corsi di Dottorato di Ricerca. La grande maggioranza dei candidati documentava un voto di laurea pari a 110 e lode.

La Commissione preposta alla selezione era costituita da rappresentanti di AICA e da docenti universitari dei settori delle Tecnologie dell’Informazione della Comunicazione.

La commissione ha operato come previsto dal bando e, dopo aver verificato che le domande presentate rispettassero i criteri esposti nel bando, ha proceduto alla valutazione dei sommari delle tesi e successivamente ha valutato le tesi complete dei 4 candidati finalisti.

I lavori della Commissione sono iniziati nel mese di luglio 2013 e sono terminati nel mese di novembre 2013. A conclusione della selezione, la Commissione, riconosciuta l’elevata qualità di tutte le tesi presentate, ha proclamato vincitore:

Nicholas Caporusso, autore della tesi:

“Issues, challenges and practices in advancing pervasive human-computer interaction for people with combined hearing and vision impairments”,

Dottorato di Ricerca in Computer Science and Engineering,
Scuola IMT – Alti Studi di Lucca – Tutore: prof. Licia Sbattella.

Il premio di 3.000 euro è stato consegnato al vincitore il 20 dicembre 2013.

La cerimonia di premiazione si è svolta a Pisa, presso il Museo degli Strumenti di Calcolo dell’Università, dove il vincitore ha presentato il proprio lavoro di tesi.

Qui di seguito, una breve biografia del premiato e una sintesi della tesi.

Vincitore: Nicholas Caporusso

Titolo della Tesi: "*Issues, challenges and practices in advancing pervasive Human-Computer Interaction for people with combined hearing and vision impairments"*

Biografia

Nato nel 1981, Nicholas è un imprenditore sociale. Nel 2004 si è laureato in Informatica con una tesi su "Un sistema per la comunicazione e l'interazione per persone sordocieche" (110 e lode e plauso della commissione); nel 2007 ha conseguito la Laurea specialistica in Informatica con un lavoro su "Feedback multimodale per sistemi portatili di interfaccia cervello-computer" (110 e lode), lavoro premiato dall'Associazione Italiana per il Calcolo Automatico (AICA) come miglior tesi tecnologica del 2008. In seguito ha sviluppato ricerche nell`ambito delle tecnologie assistive con il laboratorio di Neurofisiopatologia dell'IRCCS Fondazione Santa Lucia di Roma, con il laboratorio IVULab del Dipartimento di Informatica di Bari e con il Politecnico di Milano.
Nel 2008 è stato primo classificato del bando Principi Attivi della Regione Puglia. Nel 2009 ha lavorato a Singapore, presso l'Agency for Science Technology And Research (A*STAR), dove ha svolto ricerche sulle Interfacce Cervello-Computer. Nel 2011 ha studiato Technology entrepreneurship presso la Santa Clara University grazie a una borsa Fulbright BEST. Nel 2012 ha conseguito il dottorato in Computer Science and Engineering (IMT - Institute of Advanced Studies, Lucca).
Dal 2009 è fondatore e coordinatore di QIRIS, un acceleratore di imprese innovative; dal 2012 è responsabile del programma Laboratori dal Basso dell'Agenzia Regionale per la Tecnologia e l'Innovazione della Regione Puglia. Dal 2013 è fondatore e AD di INTACT, un'azienda che sviluppa medical devices. È Presidente della sezione Giovani Imprenditori di CNA Puglia. Collabora con Tom's Hardware, per cui cura la sezione "maker".

Abstract della tesi

dbGLOVE è un dispositivo interattivo che permette alle persone cieche e sordo-cieche di comunicare con gli altri, di essere supportati nelle attività quotidiane, di interagire con il mondo esterno e di vivere in autonomia.

dbGLOVE consiste in un pad che il cieco o sordo-cieco indossa sul palmo della mano sinistra e che funziona come un'interfaccia multi-touch di input e di output basata sull'alfabeto Malossi. Il pad si connette a smartphone, tablet, PC e a qualsiasi dispositivo he supporta la connessione Bluetooth. In questo modo, dbGLOVE è allo stesso tempo una tastiera e un monitor tattile che permette alle persone cieche e sordocieche di comunicare con gli altri (in presenza o in remoto), di controllare dispositivi ed elettrodomestici, di connettersi a Internet, e di utilizzare software standard e App dedicate.