The EDRi papers

EDITION 03

COME FUNZIONA INTERNET

Una guida per "policy-makers"

EUROPEAN DIGITAL RIGHTS

Per la versione italiana:

Centro Nexa su Internet & Società
Politecnico di Torino

Questo documento mette a disposizione
di tutti una guida introduttiva ad alcune
delle tecnologie che costituiscono il cuore di Internet.
Speriamo che questo testo rappresenti un utile
strumento di riferimento in grado di illustrare in
maniera accessibile il funzionamento di Internet,
la Rete globale la cui apertura è alla base di così
tanti diritti civili e di così tante attività economiche.

CONTENUTI:

Documento scritto da:
Joe McNamee, Advocacy Coordinator
Kirsten Fiedler & Marie Humeau, Advocacy Managers e Sophie Maisuradze, Intern

Design: CtrlSPATIE

La European Digital Rights (EDRi) è un gruppo di 32 associazioni sulla privacy e sui diritti civili digitali attive in 20 paesi

European Digital Rights
39 Rue Montoyer
B-1000 Brussels
tel: + 32 (0)2 550 4112
brussels@edri.org

Traduzione italiana a cura del:

Centro Nexa su Internet & Società
Politecnico di Torino

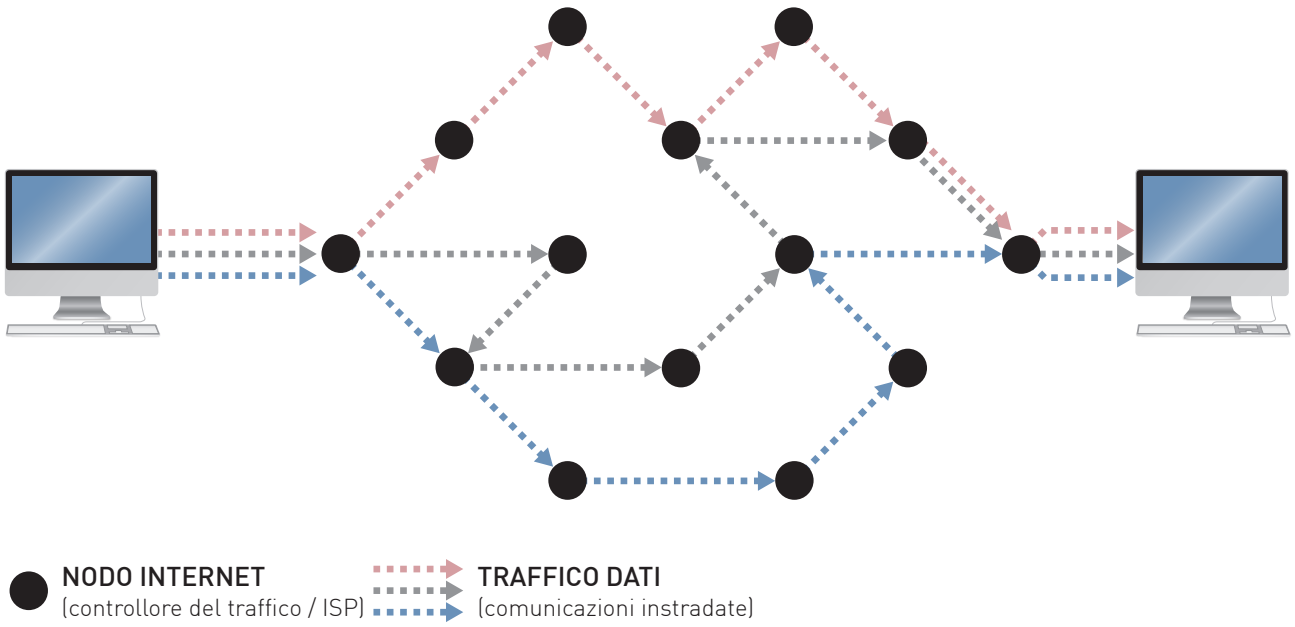
http://nexa.polito.it/ - Dip. Automatica e Informatica

Coordinatore: Juan Carlos De Martin.
Hanno contribuito: Elena Atzeni, Alberto Bellan, Fabio Chiusi, Arturo Di Corinto, Giuseppe Futia, Giovanni Battista Gallus, Raimondo Iemma, Luca Leschiutta, Luca Nicotra, Margherita Salvadori, Claudia Sarrocco, Valentin Vitkov. Si ringraziano i Fellow del Centro Nexa.

Versione 1.0 (9 maggio 2012)

INTERNET

UNA RETE DI RETI DI COMPUTER



Internet è un sistema globale di reti di computer interconnesse.

Quando due o più dispositivi elettronici vengono connessi per permettere la comunicazione reciproca, essi formano una rete. Internet è costituita dall'interconnessione su scala mondiale di reti di questo tipo, ciascuna appartenente ad aziende, governi o individui, col risultato di permettere a tutti i dispositivi connessi a tale rete di reti di comunicare tra di loro.

Per comunicare i computer devono essere in grado di comprendersi a vicenda. Su Internet la comunicazione è possibile perché tutti i dispositivi parlano la stessa "lingua" o protocollo, ovvero, il Protocollo Internet (in inglese, Internet Protocol, in sigla IP), un "mercato unico" senza barriere fisiche, tecniche o nazionali. Il protocollo IP costituisce la base di tutti gli altri sistemi di comunicazione su Internet.

Trasmettere una qualsiasi comunicazione su Internet usando il protocollo IP è come inviare le pagine di un libro per posta usando moltissime buste differenti. Tutte le buste usano lo stesso indirizzo mittente e lo stesso indirizzo di destinazione.

Anche se alcune buste viaggiano via nave e altre via aereo, alla fine tutte arrivano a destinazione ed il libro può essere ricomposto.

Su Internet il contenuto della busta (chiamata tecnicamente "pacchetto") dipende da protocolli, ossia, da convenzioni che definiscono il formato dei dati e le procedure di connessione per i diversi tipi di comunicazione. Esempi di queste convenzioni costruite sopra il protocollo IP sono:

- SMTP per spedire la posta elettronica;
- HTTP per accedere a siti web;
- BitTorrent per la condivisione di file in modalità peer-to-peer (P2P), ovvero tra pari (una modalità per condividere file di dati all'interno di gruppi di persone anche molto ampi).

Chiunque è libero di creare il proprio protocollo e usarlo su Internet, a patto che si basi sul protocollo IP.

In altre parole, il solo limite è l'immaginazione, la sola regola è che l'indirizzo sulla busta sia nel formato standard richiesto dal protocollo IP. L'apertura del sistema è ciò che ha reso Internet un fenomeno globale.

Qualsiasi restrizione dell'apertura di Internet riduce il suo potenziale di sviluppo futuro.

L'uso universale di un singolo protocollo di base per tutte le forme di comunicazione ha importanti vantaggi.

I dispositivi che sono responsabili per il trasporto dei dati su Internet (chiamati "routers", che in italiano potremmo tradurre come "instradatori") non hanno bisogno di essere programmati differentemente per trattare diversi tipi di dati. Anzi, non hanno alcun bisogno di sapere nulla dei dati che smistano, a patto che tali dati usino il protocollo IP.

Come il postino che consegna la posta tradizionale, i "router" devono solo guardare all'esterno della busta per essere in grado di consegnare il messaggio. Non importa se la busta contiene una bolletta o una lettera d'amore (tranne che per il ricevente, naturalmente).

Ciò implica:

- Possibilità di innovazione illimitata in termini di nuovi protocolli e nuove applicazioni, purchè costruite sopra il protocollo IP;
- Non c'è alcun bisogno di sapere nulla in merito al contenuto di qualsiasi comunicazione: "privacy by design"; [1]
- Flussi dati flessibili e veloci.

Essenzialmente, Internet offre un solo, flessibile servizio: trasportare dati da un dispositivo ad un altro a prescindere dalla natura dei dispositivi usati, da come e dove essi sono connessi a Internet e dalla natura o dal contenuto dei dati stessi.

"L'apertura e la flessibilità di Internet sono le ragioni primarie dei successi economici, d'innovazione e democratici resi possibili dalla Rete"

1 Con tale espressione si fa riferimento alla concezione secondo cui le tecnologie devono essere strutturate in maniera da assicurare una protezione dei dati intrinseca, di tipo tecnico-procedurale.

L'INDIRIZZO IP

UN INDIRIZZO DIGITALE

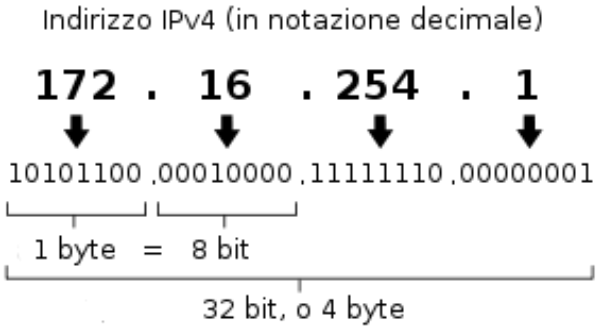
Un indirizzo IP è un indirizzo numerico che viene assegnato ad ogni dispositivo collegato ad Internet.[2]

In molti casi gli indirizzi IP possono essere utilizzati per identificare un'organizzazione o un individuo che usino un Internet Service Provider per collegare ad Internet uno o più apparecchi.

In altri casi, in particolare nelle reti aziendali, nelle connessioni wireless pubbliche o non protette e nelle connessioni mobili ad Internet, l'indirizzo IP non sempre identifica la persona che ha compiuto un atto tracciabile digitalmente.

Poiché un router casalingo o aziendale spesso mostrerà solo un indirizzo IP per tutte le persone connesse ad esso, l'indirizzo IP identificherà un gruppo di persone piuttosto che un singolo individuo. Di conseguenza, spesso è difficile, se non impossibile, essere sicuri di chi ha fatto cosa sulla base del solo indirizzo IP.

D'altra parte, gli indirizzi IP sono molto spesso associabili a specifiche persone, e perciò devono essere trattati come "dato personale" tranne nel caso in cui venga inequivocabilmente stabilito che non lo siano.



"L'indirizzo IP non sempre identifica la persona che ha compiuto un atto tracciabile digitalmente"

2 A causa della scarsità nell'attuale generazione di indirizzi IP, è sempre più comune, particolarmente nelle reti aziendali, che gli indirizzi IP vengano condivisi da tutti i computer, per esempio, di un ufficio. Questa scarsità è in via di soluzione con l'adozione dell'indirizzamento IPv6.

CRITTOGRAFIA

RISERVATEZZA IN UNA RETE PUBBLICA

Una lettera può essere aperta, letta e chiusa senza lasciare traccia. Una telefonata può essere intercettata. Come può un utente inviare un messaggio sensibile in modo che rimanga al riparo da occhi indiscreti?

Grazie alle tecnologie informatiche nel ventesimo secolo abbiamo assistito a una rapida evoluzione della crittografia. I computer hanno reso possibile non solo la cifratura rapida dei messaggi elettronici, ma anche la violazione molto più rapida delle chiavi di cifratura usate finora.

Va detto che la crittografia non è una soluzione infallibile e non garantisce una completa riservatezza. Una tecnica frequente per aggirare la crittografia è catturare il messaggio prima ancora che venga cifrato – per esempio, ad opera di un programma installato di nascosto sul computer (o sul telefono cellulare) dell'utente, programma che registra quali tasti vengano premuti sulla tastiera (i cosiddetti programmi "cavallo di Troia registra-tasti", o in inglese "Trojan keylogger").

Un altro elemento al quale bisogna porre attenzione cifrando un messaggio è la sua integrità (cioè la completezza del file), altrimenti il messaggio può essere manipolato anche senza conoscere la chiave di cifratura. I migliori strumenti crittografici verificano automaticamente l'integrità dei file cifrati.

L'immagine qui sopra mostra le fasi di una importante tecnica di crittografia chiamata crittografia a chiave pubblica ('public key encryption'), che funziona sulla base di una coppia di chiavi, una pubblica e una privata:

1. Il mittente richiede una copia della chiave pubblica del destinatario;

2. Usando un software appropriato, il mittente cifra il messaggio usando la chiave pubblica del destinatario;

3. Il messaggio viene inviato;

4. Il destinatario decifra il messaggio usando sia la sua chiave pubblica sia quella privata.

IL DOMAIN NAME SYSTEM (DNS)

L'ELENCO TELEFONICO DI INTERNET

Un sito web su Internet è raggiungibile tramite l'indirizzo IP numerico del server che lo ospita (nel momento in cui scriviamo, per esempio, l'indirizzo di EDRi.org è 217.72.179.7). Gli indirizzi IP non sono facili da ricordare per gli esseri umani. Usarli per identificare risorse online, inoltre, non è pratico, dato che i servizi su Internet devono di tanto in tanto migrare su un nuovo indirizzo IP (se cambiano Internet Service Provider, per esempio).

Dato che l'uso di indirizzi IP per siti web non è né pratico né 'user friendly', sono stati creati i 'domain names' (cioè i nomi a dominio, come edri.org). Il Domain Name System globale funziona un po' come una rubrica telefonica per Internet.

Se conoscete il nome a dominio del sito web che volete visitare, il Domain Name System è utilizzato – in modo invisibile e automatico – per reperire l'indirizzo IP corrispondente al web server presso cui si trova il sito. Perciò, quando digitate http://edri.org, il vostro computer è in grado di identificarlo come se fosse 217.72.179.7 e invia una richiesta specifica per quel sito.

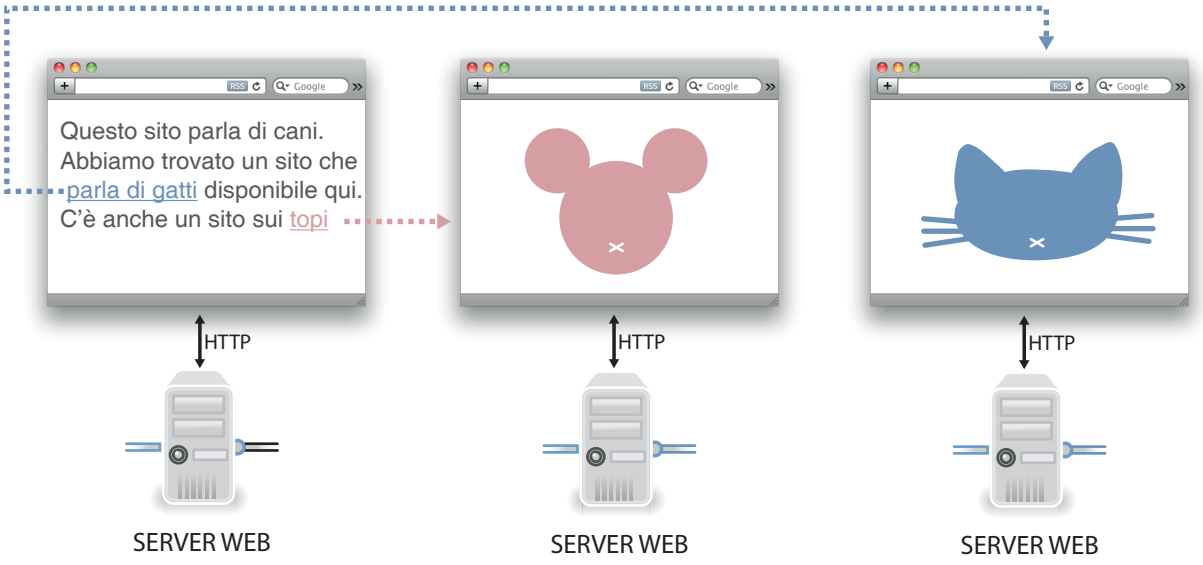
Il sistema per cercare un nome a dominio funziona in maniera gerarchica. Quando digitate http://edri.org, il vostro computer innanzitutto si connette a un server DNS per chiederne l'indirizzo.[3] Il server DNS predefinito di norma è gestito dal vostro Internet provider, ma è possibile utilizzarne uno diverso.

Se qualcuno ha effettuato l'accesso di recente a http://edri.org, il server DNS ne 'ricorderà' i dettagli e vi fornirà l'indirizzo IP corretto. In caso contrario, affiderà la richiesta a un livello più alto di autorità, dove viene seguita la stessa procedura. Al massimo livello di autorità ci sono 13 'root server' che in sostanza mettono insieme i server DNS. I 13 root server sono molto solidi e hanno un'enorme potenza di calcolo. Ne hanno talmente tanta che hanno continuato a funzionare in modo efficiente perfino quando sono stati vittima di attacchi imponenti (i cosiddetti attacchi 'distributed denial of service').

3 Se il vostro computer ha effettuato un accesso a http://edri.org di recente, allora è già a conoscenza dell'indirizzo e non ha bisogno di verificarlo con il service provider.

IL WORLD WIDE WEB

CONNETTENDO LA SOCIETÀ DELL'INFORMAZIONE



Il World Wide Web si basa sull'HTTP, un protocollo (un linguaggio di comunicazione), relativamente giovane, che a sua volta si basa sul protocollo IP. HTTP è l'acronimo del HyperText Transfert Protocol (protocollo di trasferimento dell'ipertesto), ed è stato creato per scaricare i documenti ipertestuali (cioè le pagine web) e per spedire alcune informazioni essenziali al server.

Le pagine Web possono essere create utilizzando il linguaggio HTML – HyperText Markup Language (linguaggio di marcatura dell'ipertesto). Le regole di questo linguaggio sono stabilite dal World Wide Web Consortium (W3C), e specificano marcatori speciali che indicano le proprietà tipografiche e di impaginazione del testo. Per esempio, il carattere in grassetto sarà preceduto dal segno **e sarà seguito dal segno** .

Queste specifiche tecniche hanno subito delle evoluzioni nel tempo (una delle ultime versioni è il linguaggio HTML5), perché il processo di sviluppo del linguaggio HTML è continuo nonché aperto alla partecipazione di tutti. Una volta che lo standard è stato definito, il suo uso non è soggetto ad alcuna licenza o pagamento di royalties. Il vantaggio è che tutti i computer leggono le istruzioni scritte nel linguaggio HTLM esattamente allo stesso modo, quindi chiunque può usarlo, gratis, ed essere certo che ogni apparecchio visualizzerà la pagina Web nello stesso modo. Il Web (e tutto sommato anche il mondo) sarebbe molto più povero se le persone dovessero pagare per scrivere le pagine nei linguaggi richiesti da tutti i diversi tipi di computer.

Tali caratteristiche di apertura e libertà del linguaggio HTML sono essenziali al fine di assicurare la compatibilità di tutte le pagine Web per ogni tipo di apparecchio: computer fissi, telefoni cellulari, lettori digitali, computer

portatili ed ogni altro dispositivo. La corretta applicazione delle specifiche del linguaggio HTML per il formato delle pagine Web assicura anche la libertà di accesso a tutte le persone che hanno difficoltà visive, altrimenti i sistemi di lettura dei testi non sarebbero in grado di comprendere le pagine alle quali gli utenti accedono.

Le pagine Web sono pubblicate su macchine note come Web server. Un web server è un computer che può essere individuato attraverso il suo specifico indirizzo IP (come abbiamo spiegato a pagina 5). Normalmente molti nomi a dominio (come ad esempio www.edri.org e www.bitsoffreedom.nl) possono trovarsi allo stesso indirizzo IP perché sono ospitati ("hosted") dallo stesso server. Per questo un singolo Web server con un unico indirizzo IP può ospitare numerosi siti Web. Nel caso delle società commerciali che vendono spazio ad altri siti web, sullo stesso server vengono ospitati centinaia di siti web, che non hanno alcuna relazione fra loro. Per questo eventuali tentativi di bloccare singoli siti Web in base al loro indirizzo IP hanno sempre avuto conseguenze disastrose anche per le altre pagine ospitate sullo stesso server.

Il protocollo HTTP ha una variante sicura, chiamata HTTPS. HTTP non è cifrato: di conseguenza chiunque abbia accesso ai cavi della rete o a una delle connessioni che collegano il computer dell'utente al server web può accedere a tutte le informazioni che l'utente invia al server e viceversa.

HTTPS invece cifra queste connessioni in modo che (teoricamente) solo gli utenti e il server web possono decifrare le informazioni che si scambiano. Tutto ciò è basato sulla fiducia: colui che pubblica le pagine Web chiede a un'autorità affidabile di dargli un certificato strettamente personale, una sorta di firma digitale che identifica colui che pubblica; un meccanismo simile al sigillo in ceralacca che nei secoli passati era utilizzato per chiudere i documenti.

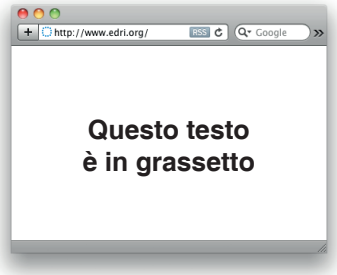
Quando un utilizzatore acquista un nuovo computer o installa un nuovo browser Web, questo contiene una lista di autorità di certificazione affidabili, cioè il riferimento ad enti il cui mestiere consiste nell'emettere certificati di sicurezza. Il computer dell'utente, collegandosi a siti web il cui certificato è stato rilasciato da uno degli enti menzionati, segnalerà all'utente che la connessione è sicura. La fragilità di questo sistema deriva da questa lista che contiene dozzine di enti. Se uno di questi enti diventa inaffidabile, i servizi che certificava diventano insicuri ma gli utenti non se ne accorgono.

LINGUAGGIO SVILUPPATO DAL WORLD WIDE WEB CONSORTIUM

W3C®

Questo testo è in grassetto

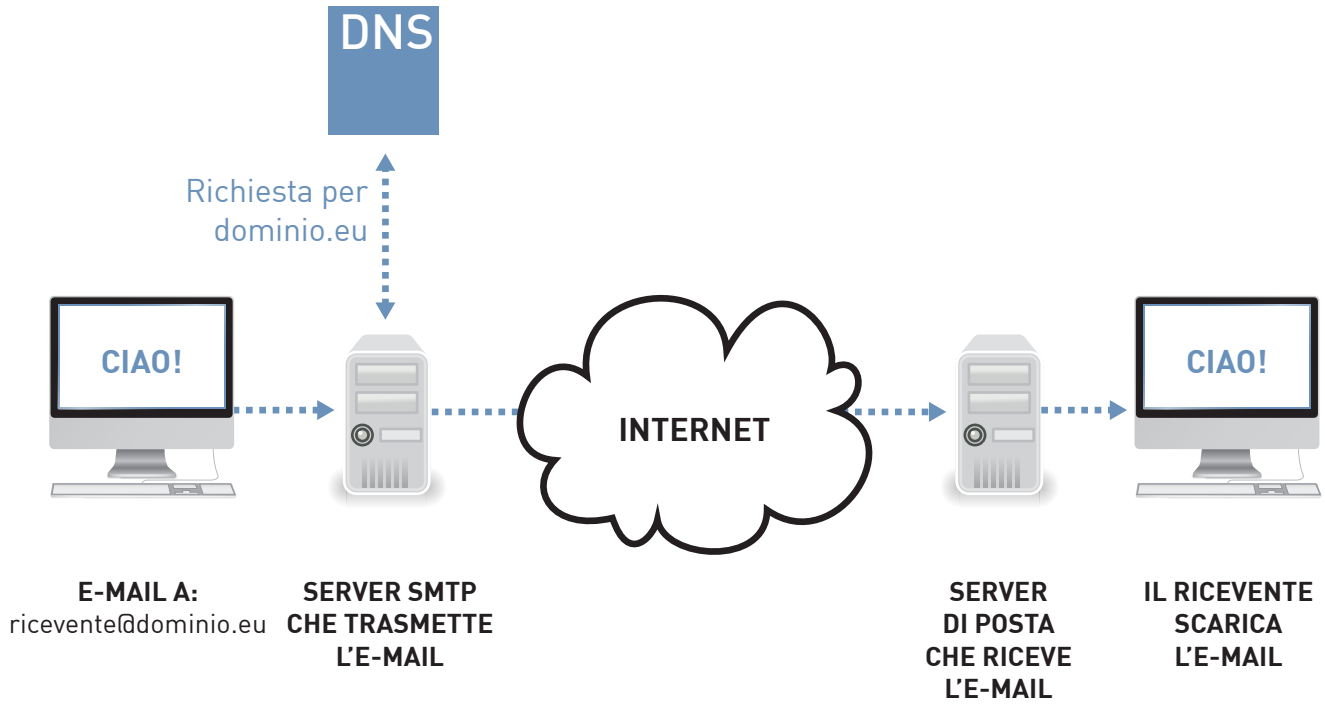
COSA SCRIVE UN PROGRAMMATORE



COSA VEDI COL BROWSER

L'E-MAIL E LA SICUREZZA

LA POSTA NEL MONDO DIGITALE



I messaggi di posta elettronica, o e-mail, sono messaggi inviati da un mittente a uno o più destinatari. L'invio di questi messaggi è gestito tramite il protocollo SMTP (Simple Mail Transfer Protocol - Protocollo semplice di invio della posta) che, come l'HTTP, è anch'esso basato sul protocollo IP.

Dopo aver composto una e-mail mediante un sito webmail o un programma di posta elettronica, essa è trasferita a un server SMTP in uscita. Viene poi trasferita da un server e-mail all'altro, sempre usando SMTP, fino a che non raggiunge il server di destinazione finale.

I server e-mail ricavano le informazioni necessarie all'invio interrogando le informazioni del Domain Name System (DNS) descritto a pagina 7. Il DNS contiene anche le informazioni relative a quali server siano deputati alla gestione delle e-mail per ogni dominio. Il dominio può essere ricavato dalla porzione dell'e-mail del destinatario successiva al segno @.

Dopo che il messaggio arriva al server e-mail che gestisce tutte le e-mail del destinatario, vi rimane fintanto che quest'ultimo non lo cancelli. Alcuni programmi di posta eseguono quest'operazione automaticamente una volta che l'utente ha scaricato la posta dal proprio PC o smartphone.

Sicurezza delle e-mail Le e-mail possono essere intercettate da terzi quando transitano da un server all'altro. Ci sono due modi per evitare che

ciò succeda: rendere sicura la comunicazione tra i server e-mail, oppure cifrare il contenuto delle stesse e-mail. La comunicazione tra server e-mail può essere resa sicura nello stesso modo in cui il protocollo HTTPS rende sicure le comunicazioni HTTP (nel modo descritto in precedenza).

Nel caso dell'e-mail vi è però una debolezza, in quanto il nostro computer non comunica direttamente con il server di destinazione finale. Ciò comporta il fatto che se anche uno solo dei server e-mail intermedi non usa la cifratura per inoltrare il messaggio, esso può essere intercettato durante questa fase del transito.

A causa di questa vulnerabilità, è preferibile cifrare il messaggio stesso. Per cifrare le e-mail si può usare un sistema diffuso e liberamente disponibile, quale ad esempio PGP (Pretty Good Privacy), anche disponibile come OpenPGP e GPG.

DEEP PACKET INSPECTION

SBIRCIANDO NEL VOSTRO TRAFFICO INTERNET

I dati su Internet sono trasmessi in "pacchetti", ovvero piccoli blocchi di dati. Ogni pacchetto ha un'intestazione che descrive la sua origine e la sua destinazione (è come una busta su cui siano scritti gli indirizzi del mittente e del destinatario). Tali informazioni permettono alle apparecchiature di rete di determinare il miglior percorso per trasmettere un pacchetto in un dato momento.

Storicamente le apparecchiature di rete si limitavano a esaminare solamente le informazioni di origine e destinazione. Tuttavia, con il rapido incremento di attività malevole, i gestori delle reti hanno deciso di dover esaminare un maggior numero di dettagli di ogni pacchetto per distinguere i pacchetti "sicuri" da pacchetti generati da intrusioni informatiche o da attacchi finalizzati a bloccare un servizio (noti in inglese come "denial of service attacks").

Ad esempio, i programmi per la sicurezza di rete [firewalls] inizialmente bloccavano solamente pacchetti che partivano da un'origine specifica ed erano indirizzati verso una destinazione specifica e uno specifico servizio. Usando tali criteri si possono bloccare tutte le richieste di servizi verso la rete di un'azienda provenienti dall'esterno, perché, per esempio, non si vogliono rendere disponibili al pubblico i propri servizi (per esempio, non si vuole che un estraneo stampi sulle nostre stampanti di rete). Allo stesso tempo, non bloccando le richieste di servizi originati dalla rete della propria azienda, si possono tranquillamente fruire di tutti i servizi disponibili su Internet.

Ad un certo punto si potrebbe decidere di attivare un server web sulla propria rete per pubblicare dei documenti. In tal caso sarebbe necessario modificare le impostazioni del proprio firewall per permettere l'accesso a richieste provenienti dall'esterno e dirette al servizio web. Tuttavia, ci sono numerosi attacchi contro server web che appaiono inoffensivi dal punto di vista degli algoritmi usati dal firewall. In altre parole, è impossibile distinguere tra pacchetti legittimi e pacchetti dannosi basandosi unicamente sui dettagli di origine e destinazione.

Gli ingegneri di rete hanno compreso velocemente che sarebbe stato più semplice individuare gli attacchi se le apparecchiature di rete avessero esaminato un po' più in profondità i pacchetti. In teoria, tale operazione è tecnicamente semplice - le intestazioni di un pacchetto non sono separate dal pacchetto se non in base a una definizione logica dei confini delle intestazioni. Si tratta solamente di analizzare pochi altri bytes rispetto a quelli che vengono normalmente analizzati, ad esempio per effettuare l'instradamento. Oppure andare ancora più in fondo e guardare l'intero contenuto

del pacchetto.

I dispositivi predisposti a fare ciò sono stati inizialmente chiamati "Sistemi di prevenzione delle intrusioni" (Intrusion Prevention Systems, IPS). Successivamente tali caratteristiche sono state introdotte nella maggior parte dei dispositivi di rete. Quando questi dispositivi venivano usati solo per bloccare attacchi informatici, ciò non causava controversie.

Tuttavia, nel corso del tempo, i governi, i fornitori di contenuti e gli operatori di rete hanno iniziato a rendersi conto che la tecnica - generalmente denominata Deep Packet Inspection (DPI) - offre loro un controllo ben maggiore sui contenuti trasmessi tramite Internet rispetto a prima.

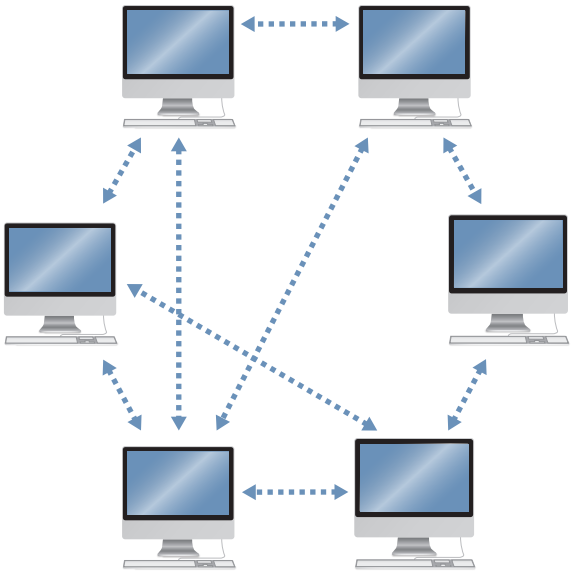
Le tecniche di Deep Packet Inspection sono già in uso per fini di giustizia (sorveglianza, blocco, ecc.), profilazione per fini di marketing, pubblicità mirata, per far rispettare livelli contrattuali di servizio, e vengono proposte come mezzo per la tutela dei diritti d'autore.

Tuttavia la DPI può costituire una pratica lesiva di diritti fondamentali quali la riservatezza ed inviolabilità delle comunicazioni e la protezione dei dati personali.

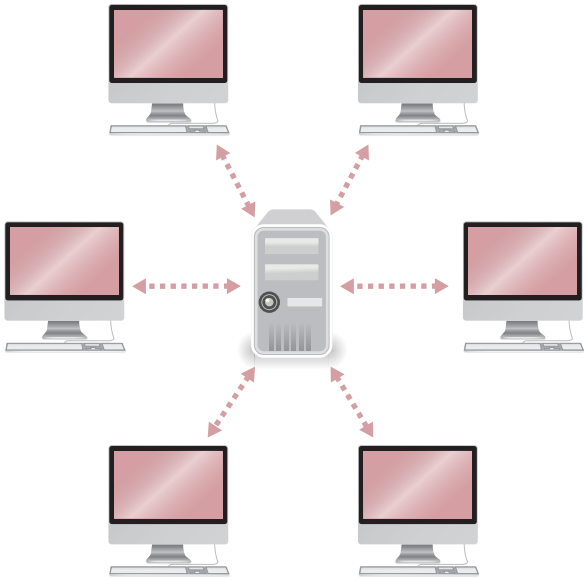
Dal punto di vista dell'utente, le tecniche di Deep Packet Inspection possono essere contrastate usando la crittografia: il contenuto "profondo" di un pacchetto crittografato, infatti, è totalmente opaco per l'operatore.

PEER-TO-PEER

DA ME A TE, CON NESSUNO IN MEZZO



PEER-TO-PEER
SISTEMA DI NODI SENZA INFRASTRUTTURA CENTRALIZZATA



CENTRALIZZATO
MODELLO DI SERVIZIO BASATO SU SERVER (NON PEER-TO-PEER)

Le reti peer-to-peer sono costituite da dispositivi (web server o computer di utenti) che comunicano su un piano paritario. Ogni "peer" (ossia ogni dispositivo) può comunicare con gli altri "peer" e non sussiste alcuna distinzione tra produttori e consumatori, client e server, ecc. Si tratta semplicemente di un certo numero di dispositivi che comunicano con altri dispositivi.

Questo modello si differenzia da quello client-server o uno-a-molti, nel quale un computer soddisfa le richieste di numerosi client - per esempio un sito web che fornisca contenuti rivolti a molti utenti (un dispositivo che comunica con molti dispositivi).

Su Internet le applicazioni peer-to-peer utilizzano protocolli peer-to-peer che sono basati - come è inevitabile che sia - sul protocollo IP.

Le reti peer-to-peer presentano un numero di particolari vantaggi:

- Non presentano i problemi legati al "punto singolo di fallimento" ("single point of failure") in quanto non ci sono entità centralizzate. In una rete uno-a-molti, se il dispositivo "uno" presenta

un problema, viene di conseguenza influenzato l'intero sistema. In una rete basata sul modello molti-a-molti, anche nel caso in cui si verificasse un guasto ad uno dei dispositivi, ciò produrrebbe un danno minimo da un punto di vista generale;

- Queste reti possono crescere agevolmente, in quanto ogni partecipante che si aggiunge porta anche risorse addizionali (capacità di traffico, memoria, potenza di calcolo) alla rete stessa;

- Non c'è nessun amministratore perché non c'è un'autorità centrale;

- I guasti hanno un impatto minimo perché non ci sono risorse centralizzate e c'è un livello di duplicazione delle risorse intrinsecamente elevato;

- Garantiscono libertà agli utenti. Non solo i dispositivi che partecipano si trovano su un piano di uguaglianza, ma anche gli stessi utenti lo sono.

Uno dei compiti importanti di un'applicazione peer-to-peer è di organizzare e individuare le risorse nella rete.

In una certa misura, i server di posta elettronica rappresentano un primo esempio di applicazioni peer-to-peer. Usando il protocollo SMTP, qualsiasi server può inviare un'email a qualsiasi altro server. Il DNS (Domain Name System) può anche elencare molteplici server in grado di gestire e-mail in entrata per un determinato dominio, aumentando l'affidabilità del sistema.

Gli utenti "peer" in una rete di condivisione file non conoscono immediatamente l'indirizzo IP degli altri utenti che partecipano alla rete e non sanno quali utenti hanno quali file (o parte di essi). Questo è tipicamente gestito mediante un processo nel quale gli utenti condividono informazioni riguardo ai contenuti di cui dispongono altri utenti. I file sono identificati usando chiavi "hash", le quali sono fondamentalmente impronte digitali che permettono a singoli file di essere identificati in modo inequivocabile. I DHT (Distributed Hash Tables) permetto ai "peer" di scoprire quali utenti mettono a disposizione una parte o la totalità di un determinato file.

Gli utenti di reti peer-to-peer hanno bisogno di un modo per ottenere le impronte "hash" dei file desiderati. Alcune di esse sono pubblicate su siti web, ad esempio quelli per scaricare versioni del sistema operativo Ubuntu. Ci sono dizionari che mappano le descrizioni leggibili dei file in impronte "hash", in modo da rendere possibile la ricerca di file in reti peer-to-peer.

Siti web quali thepiratebay.org e mininova.org mettono a disposizione questi dizionari. In ogni caso, queste impronte digitali possono essere distribuite anche tramite e-mail, chat e per mezzo dei social network - ovvero, non esiste alcun sistema centralizzato.

Esistono anche reti peer-to-peer che garantiscono l'anonimato degli utenti che vi partecipano.

PUBBLICITÀ COMPORTAMENTALE

PERSONALIZZANDO

La pubblicità comportamentale (in gergo anche "targeting comportamentale") è una tecnica che si basa sul tracciamento delle attività degli utenti su Internet. È utilizzata per costruire profili di utenti Internet, in modo da veicolare messaggi pubblicitari che, se il profilo è corretto, saranno per loro più rilevanti, e quindi più efficaci.

La pubblicità comportamentale sfrutta un principio facile da comprendere: se un utente visita un sito web dedicato, ad esempio, al calcio, il browser web (ad esempio Internet Explorer, Firefox o Chrome) memorizzerà sul computer un piccolo file, detto "cookie". Un sito web è di regola composto da contenuti provenienti da diverse fonti. Ad esempio, il testo e le immagini possono provenire dal sito digitato nel browser dell'utente, mentre altri contenuti aggiuntivi, come i messaggi pubblicitari, sono scaricati da altri indirizzi (addirittura da sorgenti senza legami con il sito web). Ogni volta che un contenuto viene scaricato, la richiesta al server può includere anche dati contenuti nei cookie ospitati sul computer dell'utente.

Per i fini della pubblicità comportamentale, i cookie contengono, di regola, un numero di identificazione. Se successivamente l'utente legge un articolo sulle automobili, le società pubblicitarie saranno in grado di fare ipotesi su chi legge articoli sia sulle auto sia sul calcio. Nel nostro esempio, un'ipotesi semplice potrebbe essere che l'utente sia qualcuno di potenzialmente sensibile a pubblicità di birra.

La pubblicità comportamentale è concepita anche per non presentare ad un utente pubblicità non pertinenti rispetto al suo profilo di consumatore.

Quanti più siti che fanno parte della stessa rete di tracciamento utilizzata per i servizi di pubblicità comportamentale (come gran parte dei siti web dei giornali e molti altri) vengono visitati dall'utente, tanti più dati relativi al suo profilo vengono raccolti. Analizzando in un arco di tempo relativamente breve le abitudini online di una persona è possibile sviluppare un profilo molto dettagliato – e "l'identificabilità" dei dati aumenta, anche se in teoria sono dati anonimi.

Grandi quantità di dati relativi ai comportamenti online possono ridurre la dimensione del gruppo di persone al quale appartiene un utente, fino ad arrivare a un numero molto esiguo di individui che possano corrispondere al suo profilo. Molti anni fa, il gestore di un motore di ricerca ha pubblicato una grande quantità di dati "anonimi" relativi alle ricerche fatte. Il risultato è stato che, analizzando tali dati "anonimi", alcuni giornalisti sono stati in grado di identificare persone specifiche, dimostrando che i dati "anonimi" alla

fine anonimi non sono.

Non è dato sapere se dati aggiuntivi, provenienti da altre fonti, vengano utilizzati per i servizi di pubblicità comportamentale. Molte società attive nel settore del targeting comportamentale, come Google e Yahoo!, forniscono anche altri servizi online, oltre alla ricerca. L'aggregazione di dati provenienti da diverse fonti rende possibile l'identificazione di singoli individui.

Si sostiene che la pubblicità comportamentale sia uno dei fattori di sviluppo dei successi economici dell'industria della pubblicità online negli ultimi anni. La tecnica è utilizzata su base sperimentale anche per fornire altri contenuti agli utenti di Internet, come ad esempio le notizie.

I fornitori di servizi ed i pubblicitari argomentano che tale tipo di tracciamento è essenziale ed è di fatto svolto nell'interesse dell'utente, in quanto permette loro di offrire molti servizi gratuiti e di proporre unicamente messaggi pubblicitari rilevanti e mirati. Coscienti peraltro dei problemi di privacy, sostengono l'adozione di procedure di "opt-out" secondo cui la profilazione per scopi legittimi è ammissibile a meno che l'utente non abbia esplicitamente dichiarato di opporsi.

A seguito della modifica apportata alle direttive europee sulla privacy nel 2009, che richiede per alcuni tipi di cookies il consenso preventivo dell'utente, tale procedura di opt-out rischia però di esser insufficiente e non aderente al dettato normativo.

Per contro, la richiesta di un consenso informato preventivo per l'utilizzo di ogni singolo cookies potrebbe rendere pressoché impossibile la navigazione.

Una soluzione di compromesso, adottata in alcuni stati membri dell'Unione e ipotizzata nella stessa normativa europea, potrebbe essere individuata nel consenso dell'utente espresso preventivamente mediante le impostazioni del browser o l'uso di applicativi quali "do not track".

Rimane il problema che molti utenti di Internet non sanno dell'esistenza dei cookies, né tantomeno cambiano le impostazioni dei loro browser.

Inoltre, i browser moderni e le loro estensioni (i cosiddetti plug-in, come ad esempio Flash) offrono molti altri modi per salvare e richiamare dati, che si aggiungono ai cookies tradizionali. Tali dati aggiuntivi sono difficilmente gestibili dall'utente medio e non sono sempre contemplati nelle preferenze dei browser relative ai cookies.

A tutt'oggi, il rapporto tra la tutela dei dati personali e la pubblicità comportamentale deve in rete ancora trovare il giusto equilibrio: molti stati membri dell'Unione Europea, tra cui l'Italia, non hanno infatti ancora recepito le modifiche alle direttive privacy in tema di cookies e molte sono le questioni tecniche e giuridiche aperte.

I MOTORI DI RICERCA

UN INDICE DI INTERNET

La navigazione sul World Wide Web avviene attraverso hyperlink (testi o immagini che, una volta cliccati, aprono altri siti o risorse).

Ogni utente può creare dei link che indirizzano a contenuti presenti sul web. Attraverso l'attività di linking, gli utenti di Internet contribuiscono all'organizzazione delle informazioni on-line in una rete di risorse connesse tra loro.

È importante sottolineare che non esiste un indice ufficiale di tutti i contenuti disponibili in rete: i motori di ricerca forniscono quindi un servizio essenziale, permettendo agli utenti di navigare su Internet in modo più efficiente.

Ci sono diversi tipi di motori di ricerca. Il più importante è il motore di ricerca detto "crawler based".

Questo tipo di motore di ricerca utilizza dei programmi (detti "crawlers" o "spiders" ovvero "striscianti" o "ragni") per ricercare le informazioni disponibili in rete, indicizzandole in modo sistematico. La complessità e l'efficienza del crawler influenza le dimensioni e il livello di aggiornamento dell'indice, entrambi elementi fondamentali per la qualità del servizio offerto da un motore di ricerca. Semplificando, lo spider/crawler segue i link contenuti su una pagina e indicizza le pagine alle quali questi link fanno riferimento, poi segue i link contenuti su queste ultime e nuovamente indicizza le pagine "linkate", e avanti così di seguito.

L'attività più importante dei motori di ricerca è stabilire una relazione tra la ricerca dell'utente e le informazioni contenute nell'indice. Il risultato di questa attività è solitamente una lista di riferimenti presentati sotto forma di classifica. In particolare, i riferimenti sono costituiti da titoli, informazioni sommarie e hyperlink che il motore di ricerca ritiene rilevanti.

A fianco dei 'risultati naturali' (i risultati selezionati dal motore di ricerca) i motori di ricerca commerciali fanno comparire risultati sponsorizzati scelti in base a un'asta di parole chiave precedentemente svolta tra diversi soggetti interessati a promuovere la propria attività. Il procedimento attraverso cui vengono individuati i risultati naturali è particolarmente complesso e i motori di ricerca commerciali proteggono come segreti industriali gli algoritmi in base ai quali individuano tali risultati. L'algoritmo PageRank di Google è uno degli algoritmi di ricerca Web più famosi. PageRank stabilisce la rilevanza dei siti presenti all'interno dell'indice sulla base della struttura dei link che vi fanno riferimento (per esempio, tenendo conto della tipologia di siti che puntano a una determinata pagina).

Altre tecniche importanti per combinare in modo efficiente le richieste degli utenti con le informazioni contenute nell'indice sono l'analisi del contenuto dei siti e l'analisi dei dati dell'utente. A quest'ultimo riguardo, i motori

di ricerca commerciali utilizzano i cookies per memorizzare le ricerche degli utenti, i link di preferenza e altri tipi di informazioni in sezioni personalizzate dei loro database, anche per un lungo periodo di tempo.

Un motore di ricerca "verticale" o "specializzato" è un servizio dedicato alla ricerca di informazioni su un determinato argomento come i viaggi, lo shopping, gli articoli accademici, le notizie o la musica. I grandi motori di ricerca "crawler based" possono fornire, come servizio aggiuntivo, anche motori di ricerca specializzati. Un "meta motore di ricerca" è un motore di ricerca che non dispone di un proprio indice e non fornisce risultati propri, ma usa i risultati di uno o più motori di ricerca diversi. Una "directory" è un insieme di link classificati in diverse categorie. Esempi celebri sono la directory Yahoo! e l'Open Directory Project.

CLOUD COMPUTING

INTERNET DIVENTA IL TUO COMPUTER

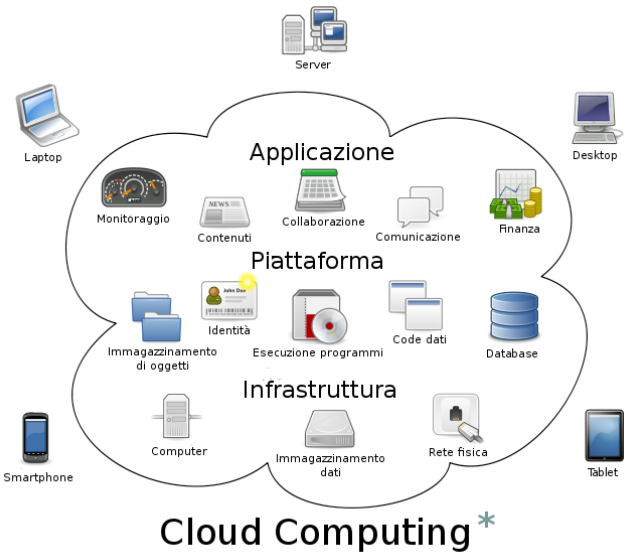
"Cloud computing" è diventato di recente un neologismo alla moda ad uso del marketing. Il concetto, di per sé, è tutt'altro che nuovo, sebbene il numero di applicazioni realizzate secondo questo paradigma stia crescendo enormemente.

Nei diagrammi che descrivono le reti di comunicazione, la nuvola ("cloud") viene usata per rappresentare la rete che sta al di fuori della rete di un utente. Cloud computing si riferisce dunque a qualsiasi servizio di elaborazione realizzato nella rete invece che all'interno del computer dell'utilizzatore.

Uno dei primi esempi di cloud computing è la versione web dell'e-mail ("webmail"). Gli utenti di una webmail possono accedere alla loro casella di posta elettronica da qualsiasi dispositivo collegato a Internet, piuttosto che da un apparecchio soltanto. I servizi di webmail più noti e usati includono Yahoo! Mail, Hotmail e Gmail.

Con il costante aumento della velocità delle connessioni a Internet, la gamma di servizi che possono essere offerti via cloud computing è molto cresciuta negli ultimi anni. Oggi, ad esempio, è possibile conservare considerevoli quantità di dati nella "cloud" facendo uso di hard disk virtuali, come quello fornito da Microsoft Live o Dropbox.

Analogamente suite per ufficio, ad esempio per la videoscrittura, e strumenti per la gestione dei database, vengono offerti online in misura



crescente.

Il progetto di sistema operativo sviluppato da Google (Google Chrome OS) è un ulteriore passo nella transizione verso sistemi di elaborazione in cloud (o "cloud-based computing"). Utilizzando il browser Google Chrome come base, il sistema operativo mira a incorporare in via automatica tecnologie cloud, così da rendere minima la quantità di software utilizzata nel proprio computer, con un forte affidamento ai servizi disponibili online - secondo un approccio che, per molti versi, è opposto a quello dell'elaborazione tradizionale, che prevede un utilizzo di software installati nel proprio computer preponderante rispetto al ridotto (o assente) utilizzo di software nella cloud.

* Immagine originale di Sam Johnston: http://commons.wikimedia.org/wiki/File:Cloud_computing.svg?uselang=de

SOCIAL MEDIA

DOVE CI INCONTRIAMO

I social media sono applicazioni Internet che consentono la creazione e lo scambio di contenuti generati dagli utenti (c.d. "User Generated Content").

I social media si differenziano dai mezzi di comunicazione ordinari poiché non si limitano a trasmettere informazioni, ma consentono all'utente di interagire con le informazioni stesse. L'interazione può consistere semplicemente nella possibilità di lasciare commenti, di votare per un articolo o di esprimere il proprio apprezzamento per una qualsiasi azione di altri utenti (il ben noto "mi piace"). Ogni utente dunque non è più un semplice spettatore, ma diventa parte del mezzo, poiché altri utenti possono leggere i suoi commenti o recensioni.

Oggi gli utenti si stanno abituando ad avere la possibilità di reagire alle informazioni ricevute e a esprimere il proprio punto di vista. Ciò incrementa la partecipazione collettiva ai dibattiti in corso. Il numero di utenti dei social media è in continua crescita, e di conseguenza cresce anche la loro forza e influenza.

Qualsiasi sito web che invita i visitatori ad interagire col sito e con gli altri visitatori può essere considerato un "social media". Questi ultimi possono essere divisi a grandi linee in sei diverse tipologie:

1. Progetti collaborativi (es. Wikipedia), nei quali gli utenti interagiscono aggiungendo articoli e modificando testi esistenti;

2. Blog e microblog (es. Twitter);

3. Aggregatori di contenuti (es. YouTube, Flickr), nei quali gli utilizzatori interagiscono condividendo e commentando foto e video;

4. Reti sociali (es. Facebook, Myspace, Hi5, Google+), nelle quali gli utenti interagiscono aggiungendo amici, commentando i profili di altri utenti, aggregandosi in gruppi e partecipando a discussioni e scambi di opinioni;

5. Giochi di ruolo virtuali (es. World of Warcraft);

6. Realtà sociali virtuali (es. Second Life).

Un aspetto importante e delicato quando si parla di social media è la tutela degli utenti - in particolare la tutela del diritto alla riservatezza (privacy). Difatti, sebbene gli utenti possano in genere scegliere quali informazioni personali condividere o nascondere, le impostazioni di base e le ulteriori forme di tutela per i bambini sono oggetto di molte controversie. In aggiunta, alcuni siti, come Facebook, hanno già in più occasioni modificato unilateralmente le impostazioni sulla privacy dei propri utenti.

INTERNET GOVERNANCE

DEMOCRAZIA DIGITALE

Il primo tentativo di definire l'espressione "Internet Governance" ebbe luogo durante le riunioni preparatorie del Summit mondiale sulla società dell'informazione (World Summit on Information Society, WSIS) delle Nazioni Unite.

Una prima definizione comunemente accettata fu prodotta dal "Working Group on Internet Governance", un gruppo multi-stakeholder (governi, privati, società civile e comunità tecniche) creato dal Segretario Generale delle Nazioni Unite, definizione poi inclusa nell'Agenda di Tunisi per la società dell'informazione:

"Lo sviluppo e l'applicazione da parte dei governi, del settore privato e della società civile, nei loro rispettivi ruoli, di principi condivisi, standard, regole, procedure decisionali e programmi sui quali si basi l'evoluzione e l'uso di Internet."

Questa definizione sottolinea l'approccio multilaterale alla discussione di politiche relative a Internet che, stante l'impatto sulla comunità, deve avvenire in modo aperto, trasparente e responsabile.

È proprio per raggiungere tale obiettivo che è stato creato l'Internet Governance Forum, un forum multi-stakeholder per la discussione e il confronto di politiche pubbliche relative ad elementi chiave dell'Internet governance. Il forum, che è già arrivato alla sua sesta edizione (dal 2006 al 2011), ha ispirato l'organizzazione di incontri simili a livello nazionale e regionale (e.g. EuroDIG – il dialogo pan-europeo sulla governance di Internet). Nonostante l'importanza di tali luoghi di discussione e confronto, è necessario evidenziare che questi meeting non funzionano come organismi decisionali, ma cionondimeno influenzano le politiche relative a Internet.

Quali temi include l'Internet Governance?

- Infrastrutture e standardizzazione;
- Questioni tecniche relative al funzionamento di Internet: infrastrutture di telecomunicazioni, standard e servizi (es. Internet Protocol, Domain Name System), standard relativi a contenuti e applicazioni (es. HyperText Markup Language);
- Questioni relative alla salvaguardia del funzionamento sicuro e stabile di Internet: cybersecurity, crittografia, spam;
- Questioni giuridiche: legislazione e regolamentazione a livello nazionale e internazionale applicabile a Internet (es. diritto d'autore, privacy e protezione dei dati, criminalità informatica);
- Questioni economiche: e-commerce, tassazione, firma elettronica, pagamenti elettronici;
- Sviluppo: digital divide, accesso universale a

Internet;

- Questioni socio-culturali: diritti umani (libertà di espressione, diritto di cercare, ricevere e fornire informazioni), politiche per l'utilizzo dei contenuti, privacy e protezione dati, multilinguismo e diversità culturali, educazione, tutela dei minori online.

Chi partecipa all'Internet Governance?

- Governi: elaborano e danno attuazione alle politiche e norme pubbliche relative a Internet;

- Settore privato: i fornitori di servizi Internet (ISPs), i gestori della rete, le società che gestiscono l'anagrafe dei nomi a dominio ("registries" in inglese), quelle che forniscono i nomi a dominio a cittadini e imprese ("registrars" in inglese), società di software, società che producono contenuti;

- Società civile: organizzazioni non governative che rappresentano gli utenti Internet;

- Organizzazioni internazionali: l'Unione Internazionale delle Telecomunicazioni (ITU), UNESCO, United Nations Development Programme (UNDP);

- Comunità tecnica: Internet Society (ISOC), Internet Engineering Task Force (IETF), Internet Architecture Board (IAB), Internet Corporation for Assigned Names and Numbers (ICANN).

Per maggiori informazioni:

Jovan Kurbalija, An Introduction to Internet Governance, Diplo Foundation, 2010

EDRI.ORG/PAPERS

Per la versione italiana:

Centro Nexa su Internet & Società

Politecnico di Torino

nexa.polito.it/publications

Il Centro Nexa è un centro di ricerca del Dipartimento di Automatica e Informatica del Politecnico di Torino.

Con il supporto finanziario del Programma per i Diritti Fondamentali e la Cittadinanza dell'Unione Europea (solo per la versione inglese).