

Internet: L'infrastruttura per il World Wide Web



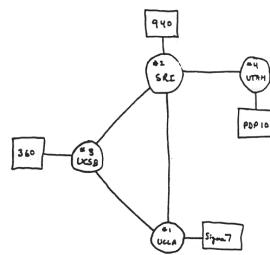
Moreno Marzolla
INFN Sezione di Padova
moreno.marzolla@pd.infn.it
<http://www.dsi.unive.it/~marzolla>

Ringraziamenti

- Parte del materiale presentato è tratto dal Corso di Tecnologie Web, prof. Fabio Vitali, Università di Bologna

Breve evoluzione di Internet / 1

Fonte: <http://www.let.leidenuniv.nl/history/ivh/chap2.htm>



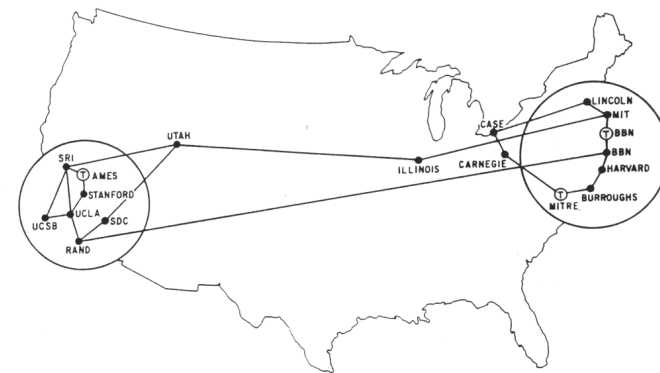
THE ARPA NETWORK

DEC 1969

4 nodes

FIGURE 6.2 Drawing of 4 Node Network (Courtesy of Alex McKenzie)

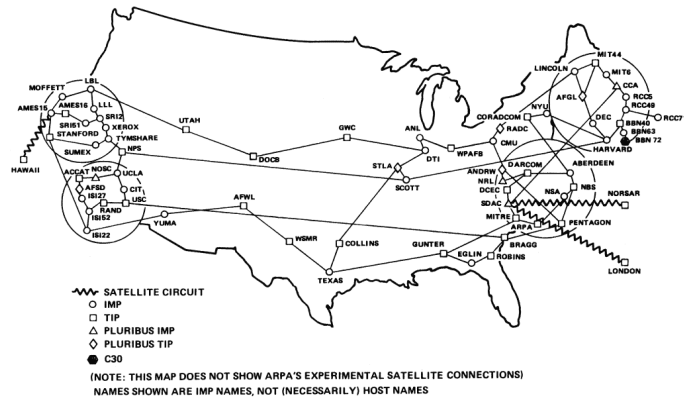
Breve evoluzione di Internet / 2



MAP 4 September 1971

Breve evoluzione di Internet / 3

ARPANET GEOGRAPHIC MAP, OCTOBER 1980



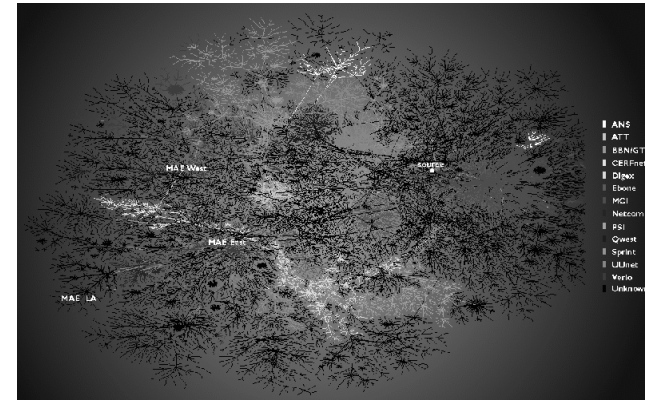
Moreno Marzolla

Tecnologie Web

5

Internet oggi

Fonte: <http://www.caida.org/outreach/papers/1999/webmatters99/>

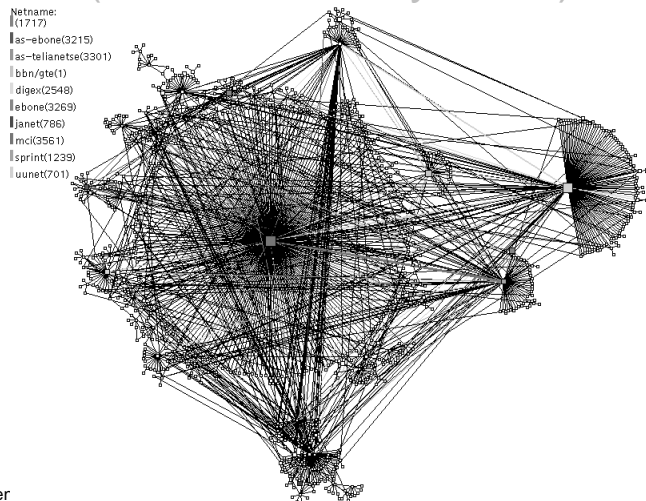


Moreno Marzolla

Tecnologie Web

6

Internet oggi (Autonomous Systems)



Morer

7

I protocolli TCP/IP

- Questi protocolli sono alla base del funzionamento di Internet
- Stabiliscono le regole fondamentali relative alle modalità di trasmissione dei dati tra le reti e tra singoli host
- Si tratta di due protocolli:
 - TCP: Transmission Control Protocol
 - IP: Internet Protocol

Moreno Marzolla

Tecnologie Web

8

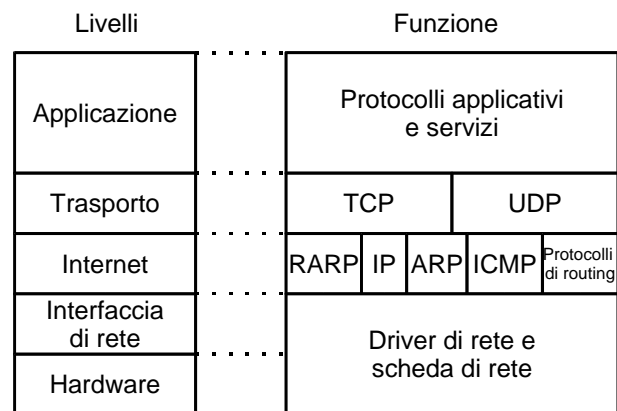
TCP

- Il protocollo TCP definisce delle regole per stabilire una comunicazione affidabile ed efficiente tra due processi in esecuzione su due macchine diverse collegate a Internet
 - Affidabile: pacchetti persi o danneggiati vengono ripristinati
 - Efficiente: Si cerca di ottimizzare la comunicazione il più possibile
- TCP fornisce l'illusione di un canale di comunicazione (un "tubo") che collega direttamente mittente e destinatario

IP

- Il protocollo IP definisce le regole di base per la comunicazione tra due macchine connesse alla rete
 - Comunicazione orientata a pacchetti
 - Comunicazione non affidabile
- IP definisce le regole per l'instradamento dei pacchetti nella rete
- Si occupa anche della gestione degli indirizzi di ciascun pacchetto, assicurando che vengano etichettati correttamente

Architettura TCP/IP



Livello Hardware

- A livello più basso, la comunicazione avviene tramite impulsi elettrici su un filo di rame
 - Oppure tramite onde radio che viaggiano nell'etere



Livello Interfaccia di Rete



- Comunicazione diretta attraverso un mezzo fisico tra due macchine direttamente collegate
 - Gestione degli errori dovuti a interferenze
 - Gestione di collisioni che avvengono se entrambi tentano di comunicare contemporaneamente

Livello Internet



- Il protocollo IP lavora a questo livello
- Comunicazione tra due macchine non necessariamente direttamente collegate
 - Protocolli di instradamento (routing)
 - Identificazione univoca delle macchine (gestione dei nomi)

Livello Trasporto



- Il protocollo TCP lavora a questo livello
- Implementa una comunicazione affidabile ed efficiente tra processi in esecuzione su macchine diverse

Livello Applicazione



- Applicazioni generiche
 - Posta Elettronica
 - HTTP
 - FTP
 - Newsgroup
 - ...

Gli indirizzi IPv4

- Gli indirizzi Internet sono rappresentati, a livello più basso, come numeri di 32 bit, suddivisi in una serie di 4 numeri di 8 bit
 - Es. 157.138.20.15
- Quando il protocollo IP invia un pacchetto, lo etichetta con l'indirizzo del mittente e del destinatario
- Tutti i nodi connessi alla rete devono avere indirizzo univoco

Gli indirizzi IPv6

- Stiamo rimanendo a corto di indirizzi IP
- Per tale ragione è stato sviluppato un nuovo protocollo (IPv6) con un nuovo schema di numerazione
 - Indirizzi IPv6 di 16 byte
 - Notazione esadecimale, 8 blocchi da 4 cifre
 - 0012:FF87:10A0:016B:3572:0000:AA7B:F07C
 - Ci sono in tutto circa $3 \cdot 10^{38}$ possibili indirizzi IPv6
 - Vuol dire $7 \cdot 10^{23}$ indirizzi IP per metro quadro terrestre, oceani inclusi

I nomi di dominio / 1

- Lavorare con i numeri può andare bene per i calcolatori, meno per gli utenti umani
- Per questa ragione, al posto dei numeri si usano i nomi di dominio
 - ganimede.dsi.unive.it
- Un nome di dominio è costituito da un numero arbitrario di parole separate da punti
- I nomi vengono tradotti in numeri da server chiamati Domain Name Server (DNS)

I nomi di dominio / 2

- Voglio accedere al sito www.dsi.unive.it
- Il computer contatta il DNS
 - Domanda: "A chi corrisponde l'indirizzo www.dsi.unive.it?"
 - Il DNS risponde: "Appartiene a 157.138.20.3"
- Ora è possibile contattare il nodo internet di indirizzo 157.138.20.3

I nomi di dominio / 3

- La parte finale (più a sinistra) dei nomi rappresenta i domini di primo livello
 - .it Italia
 - .fr Francia
 - .uk Regno Unito
 - .edu Università e scuole
 - .com Aziende
 - .gov Governo Stati Uniti
 - .net Servizi di rete
 - .org Organizzazioni no profit
 - ...

DNS

- Il DNS funziona con una struttura ad albero
 - Alle radici dell'albero ci sono i DNS "autorevoli"
- Questi passano le query ai DNS di livello più basso
 - Il DNS che "risolve" i domini che terminano con .it contiene la lista di tutti i DNS relativi a domini X.it
 - Quindi conosce l'indirizzo del DNS che serve unive.it
 - Il DNS di unive.it conosce l'indirizzo del DNS di dsi.unive.it
 - Il DNS di dsi.unive.it conosce l'indirizzo di www.dsi.unive.it

I protocolli a livello Applicazione

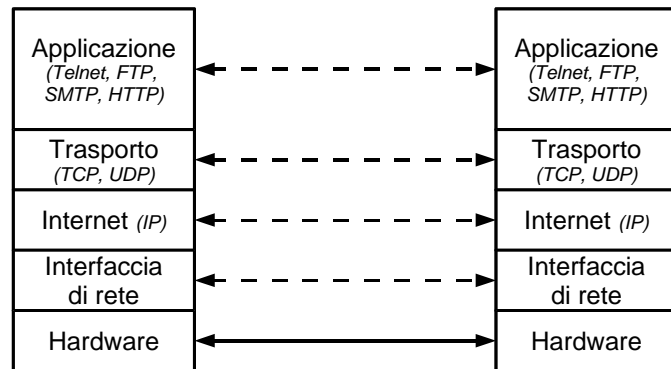
- Distinguiamo tra:
 - protocolli di applicazione vera e propria: forniscono il servizio agli utenti finali (SMTP, NNTP, HTTP, telnet, FTP, ecc.)
 - protocolli di servizio: forniscono servizi non direttamente agli utenti, ma alle applicazioni utente (SNMP, DNS, ecc.)
- Ci occupiamo di SMTP e di alcuni protocolli connessi, per la loro importanza rispetto ad HTTP

End-to-End argument / 1

- L'End-to-end argument è uno dei principi progettuali più importanti di Internet, responsabile della scalabilità e flessibilità di Internet negli ultimi vent'anni. Venne espresso per la prima volta nel 1981 da Saltzer, Reed e Clark.
- Quando si implementano protocolli a livelli, una particolare attenzione va rivolta alla decisione di quali funzionalità mettere in quale livello, e come possono i livelli inferiori aiutare i livelli superiori nello svolgimento dei loro compiti.
- Il principio dell'end-to-end argument è che i livelli inferiori non possono e non debbono fornire funzionalità di livello applicazione, perché dovrebbero avere conoscenza di quali siano queste applicazioni e di come funzionino.
- Questo ha degli effetti su quello che viene chiamata la sottorete di comunicazione.

End-to-End argument / 2

- Lo schema di comunicazione tra livelli è spesso disegnato così:



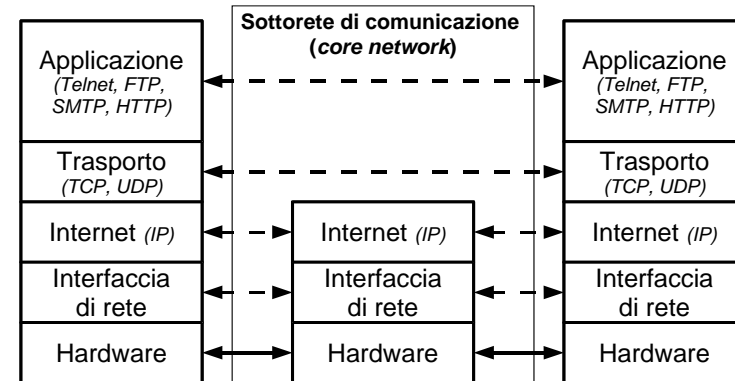
Moreno Marzolla

Tecnologie Web

25

End-to-End argument / 3

- Ma in realtà lo schema di comunicazione tra livelli è così:



Moreno Marzolla

Tecnologie Web

26

End-to-End argument / 4

- Un altro modo per esprimere l'end-to-end argument è di dire che la sottorete di comunicazione (ovvero tutti i protocolli e macchinari compresi tra i due host finali) non possono e non debbono occuparsi dei dettagli dell'applicazione, ma soltanto del completamento con successo (ed efficienza) della trasmissione.
- Tra i vantaggi:
 - La complessità del core network è drasticamente ridotta
 - La generalità della rete aumenta la possibilità che un nuovo servizio venga creato senza modifiche al core network
 - Le applicazioni non debbono dipendere da cattive implementazioni o limiti realizzativi di servizi intermedi su cui non hanno controllo, con ciò migliorando la loro affidabilità.
- Questa semplicità e generalità ha grandemente giustificato l'evoluzione e il successo di Internet negli anni passati.

Moreno Marzolla

Tecnologie Web

27

Problemi / 1

- Applicazioni più esigenti:
 - l'end-to-end necessariamente implica il best-effort service, in cui la qualità di servizio di ogni applicazione dipende dal traffico esistente, indipendentemente dalla sua importanza. Scaricare MP3 può rallentare la trasmissione video di un'operazione chirurgica in remoto.
- Differenziazione nei servizi degli ISP
 - Gli ISP attualmente competono su aspetti di commodity nella fornitura di rete: prezzo e velocità; questi sono ambiti di competizione infami e senza gloria, in cui sopravvivono i più spietati e potenti.
 - Piacerebbe invece agli ISP poter fornire abbonamenti collegati all'elenco di servizi disponibili, e non di prezzo o velocità. Ma questo implica inibire a livello di host certi servizi per cui uno non è abbonato.
 - Il caso di America On-line (e Microsoft Network)
 - Il caso dei cellulari GSM e GPRS

Moreno Marzolla

Tecnologie Web

28

Problemi / 2

- Lavorare in un mondo non fidato
 - Caratteristica dell'end-to-end è che i due host finali collaborano volontariamente al buon fine della conversazione, senza che il core network faccia niente per filtro o controllo
 - Questo ha portato a spam, virus, attacchi di hacker, siti web pirata, ecc., che assumono la neutralità del core network e approfittano dell'inadeguatezza dell'altro estremo per proprio vantaggio
- Utenti non esperti
 - L'end-to-end argument lascia completamente all'utente finale il compito di configurare correttamente i servizi. Con la crescita del numero di utenti e l'abbassamento del loro livello di conoscenze tecniche, questo può portare a problemi di inutilizzo dei servizi stessi.
 - Software e servizi configurati centralmente invece permetterebbero al fornitore di servizi di provvedere sempre indipendentemente dalle capacità tecniche dell'utente.

Problemi / 3

- C'è una crescente esigenza da parte di terze parti di intervenire nello scambio di dati tra i due estremi:
 - **Censura:** Organizzazioni con i loro dipendenti, ISP con i loro abbonati, genitori con i loro figli, stati dittatoriali con i loro cittadini, vogliono poter controllare che tipo di dati vengono scambiati e poter bloccare in anticipo gli scambi sgraditi.
 - **Indagini di polizia:** anche gli stati democratici possono voler controllare (magari a posteriori) certe comunicazioni identificate e specificate da un giudice, se questo serve a portare a termine indagini di polizia. Non parliamo poi dei servizi segreti.
 - **Tassazione:** i governi o i proprietari di diritti tipicamente vogliono una fetta di guadagno da ogni transazione commerciale, in particolare se internazionale. Sono dunque interessati a scambi di dati che comportano trasferimenti di valore e a stabilire esattamente quali valori sono stati scambiati e come ottenerne una parte in tassazione.
- Ovviamente tutte queste situazioni richiedono che lo scambio di dati venga monitorato e, sulla base del contenuto dello stesso, bloccato o registrato, possibilmente a livello di infrastruttura (core network)

Possibili risposte / 1

- Il ruolo delle leggi
 - E' possibile intervenire a livello di leggi, imponendo tasse e divieti e controlli.
 - Però la legge interviene solo a posteriori, e in molti casi solo se ne vale individualmente la pena.
 - Inoltre ha un effetto limitato, soprattutto sulle relazioni internazionali (il cattivo di turno sposta semplicemente il suo end-point in uno stato in cui questi controlli non ci sono).
- Modifica degli end-point
 - Si impone ai costruttori di hardware e software (ad esempio ai fabbricanti di browser) di creare nuovi prodotti in cui i meccanismi di filtro e controllo sono built-in
 - Problema: non si applica ai vecchi prodotti (bisogna fornire un fattore motivante per passare al nuovo), e fa nascere un mercato parallelo di prodotti che non ne tengono conto.
 - Altro problema: complica la realizzazione di nuovi software, il che riduce il numero e la flessibilità dei nuovi servizi.

Possibili risposte / 2

- Aggiungere funzioni al core network in sottoreti ad hoc
 - Di fatto, oggi firewall, filtri sul traffico e NAT sono modificatori del core network che permettono di effettuare controlli e blocchi a priori. Creano delle sottoreti (ad esempio, Intranet aziendali) all'interno delle quali l'end-to-end argument non vale.
 - Nuovi servizi vengono esaminati caso per caso e approvati solo se rispondono ai criteri globali di sicurezza/controllo/censura stabiliti. Ovviamente questo comporta una riduzione della qualità e della flessibilità della sottorete, ma non incide sulla rete globale.
- Realizzare domini di servizio
 - Terze parti fidate per entrambi gli estremi vengono utilizzate come intermediari per lo scambio di messaggi, fornendo quei meccanismi di controllo che si reputano necessari.
 - La comunicazione non avviene più in maniera diretta, ma mediata dal nodo intermedio
 - L'adesione può essere volontaria (etichette PICS per i contenuti Web, anonimizzatori, ecc.), costretta dall'ISP (Server SMTP) o dalle leggi, ma non richiede modifiche ai meccanismi della rete

SMTP, POP, IMAP

- La posta elettronica viene gestita da programmi e calcolatori chiamati mail server (o server di posta)
- I server di posta devono essere attivi 24 ore su 24, per consentire la ricezione e l'invio di messaggi
- Con i server di posta occorre comunicare tramite protocolli standard, che sono proprio SMTP, POP e IMAP

SMTP / 1

- Simple Mail Transport Protocol
- Viene utilizzato per *spedire* un nuovo messaggio
- Il protocollo è molto semplice

SMTP / 2

```
HELO abc.com
250 multivac.localdomain Hello localhost.localdomain [127.0.0.1], pleased to
meet you
MAIL FROM <pippo@abc.com>
501 5.5.2 Syntax error in parameters scanning "FROM"
MAIL FROM: <pippo@abc.com>
250 2.1.0 <pippo@abc.com>... Sender ok
RCPT TO: <pluto@xyz.com>
250 2.1.5 <pluto@xyz.com>... Recipient ok (will queue)
DATA
354 Enter mail, end with "." on a line by itself
From: <pippo@abc.com>
To: <pluto@xyz.com>
Subject: Questa e' una prova

Questo e' il testo del messaggio.
Non c'e' molto da dire, in verita'...
A presto.
.
250 2.0.0 i0UHCm501849 Message accepted for delivery
QUIT
221 2.0.0 multivac.localdomain closing connection
```

POP / 1

- Post Office Protocol
- E' il protocollo usato per *recuperare* i nuovi messaggi di posta elettronica
 - I messaggi vengono trasferiti dal server di posta al vostro PC usando il protocollo POP
 - Una volta sul vostro PC, vengono letti da un apposito programma (pine, Eudora, Outlook...)

POP / 2

- Tramite il protocollo POP è possibile chiedere al server di posta di:
 - Recuperare la posta e in seguito cancellarla dal server
 - Recuperare la posta senza cancellarla
 - Vedere se ci sono nuovi messaggi

IMAP

- Internet Message Access Protocol
- E' un nuovo protocollo che ha rimpiazzato POP per la lettura dei messaggi
 - Con IMAP è possibile chiedere al server di scaricare solo alcuni messaggi, anziché tutti
 - E' possibile scaricare solo l'intestazione e il mittente del messaggio, e il contenuto in seguito
 - IMAP consente di creare *cartelle* sul server di posta per tenere ordinati i messaggi

HTML

- HyperText Markup Language
- Linguaggio utilizzato per la definizione delle pagine WEB
- E' composto da una serie di *tag* utilizzati per definire la struttura e lo stile del documento
- Esistono dei *tag* che consentono di definire collegamenti ad altre pagine WEB (*hyperlinks*)

Intranet ed Extranet

- Intranet
 - Rete privata, basata sulle tecnologie WEB, che ospita le applicazioni Internet su una rete locale
 - E' sostanzialmente una versione "chiusa" di Internet, messa in piedi per agevolare lo scambio di informazioni all'interno della propria azienda
- Extranet
 - Estende il concetto di Intranet, prevedendo il collegamento della rete aziendale a quella dei suoi partner, clienti o fornitori

Intranet / 1

- Le *Intranet* non sono molto diverse da Internet, salvo che solo certe persone hanno il permesso di accedervi
- Basandosi su una architettura client/server, la gestione delle richieste di risorse funziona come su Internet
 - Richieste file
 - Posta elettronica
 - Scambio documenti
 - Servizi di stampa

Intranet / 2

- Le *Intranet* sono un sistema diffuso ed economico per distribuire le informazioni aziendali
- Una Intranet utilizza i protocolli di Internet
 - TCP/IP
 - FTP
 - Telnet
 - HTTP/HTML

Intranet / 3

- L'uso di protocolli standard consente a diversi reparti aziendali di condividere informazioni
- Le Intranet hanno prezzi abbordabili perché l'infrastruttura necessaria è spesso già presente se i calcolatori aziendali sono su una LAN

Intranet / 4

- Applicazioni delle Intranet
 - Invio ai dipendenti di manuali, regolamenti interni, normativa
 - Comunicazione dei compiti a ciascun dipendente
 - Comunicazione delle specifiche tecniche dei prodotti
 - Invio elenco telefonico interno aggiornato
 - Addestramento del personale tramite corsi online

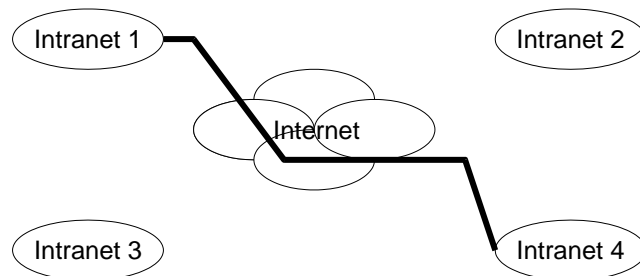
Intranet / 5

- Una considerazione di tipo economico
 - Prima di considerare l'adozione di una Intranet in una azienda, è necessario valutarne l'impatto economico in termini di costi/benefici
 - In altre parole, occorre che il costo di installazione della Intranet sia ripagato dai risparmi derivanti dall'uso della rete
 - Tali benefici sono solitamente difficili da stimare

Extranet / 1

- Le Extranet permettono di collegare tra di loro diverse Intranet per consentire alle aziende di condividere informazioni con i propri fornitori
- Le informazioni vengono trasmesse sulla Extranet in modo sicuro
- La Extranet sfrutta i protocolli di Internet
 - Ha lo scopo di creare una sottorete "sicura" che unisca più Intranet

Extranet / 2



Extranet / 3

- Caso di studio: FedEx
 - Fino a qualche tempo fa, i clienti dovevano telefonare e comunicare un codice personale per sapere lo stato delle proprie consegne
 - Successivamente, FedEx ha distribuito un software per consentire agli utenti l'accesso al proprio database interno tramite un collegamento via modem
 - A seguito dell'avvento di Internet, questo software proprietario è stato eliminato e sostituito da una interfaccia WEB

Rete privata

- E' una connessione dedicata che collega le Intranet di due aziende
 - E' una linea telefonica, che a differenza delle normali collega permanentemente i due soggetti
- Vantaggi
 - Sicurezza: a parte le due aziende collegate, nessuno ha accesso al collegamento telefonico
- Svantaggi
 - Costo
 - Ogni coppia di aziende collegate necessita una connessione dedicata

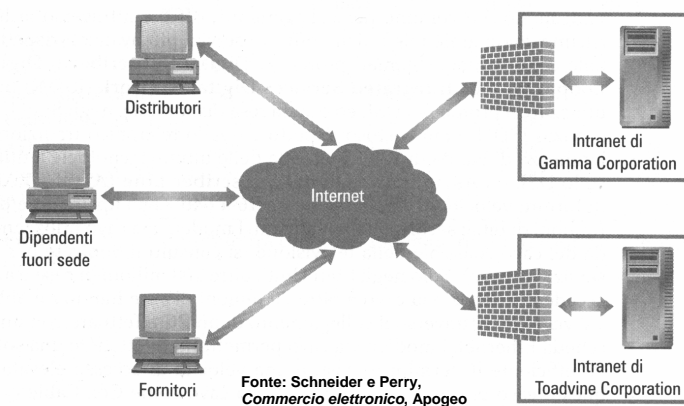
Virtual Private Network (VPN)

- Le VPN sono state create per superare le limitazioni delle connessioni dedicate
 - In pratica, anziché usare la linea telefonica, si usa Internet e i suoi protocolli per trasferire dati sensibili
 - L'idea è simile alla Extranet
 - Una VPN è un tipo di Extranet
 - In generale, una Extranet può *NON* essere implementata tramite una VPN

Vantaggi delle VPN

- Si può considerare una "corsia separata" dell'autostrada Internet
- Le VPN sono un modo relativamente semplice e sicuro per unire diverse Intranet
- La soluzione basata sulle VPN scala molto bene all'aumentare del numero di partner che vogliono condividere informazioni

VPN

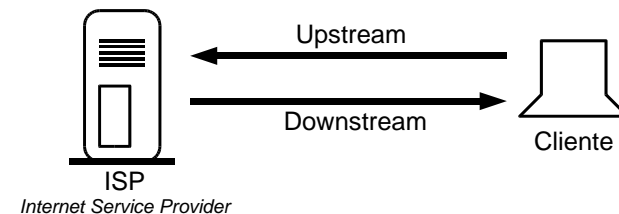


La connessione in rete / 1

- Esistono fornitori di connettività a Internet detti *Internet Service Providers* (ISP)
- Gli ISP offrono diverse modalità di connessione a Internet
 - Caratterizzate da tecnologie diverse...
 - ...velocità di trasferimento diverse...
 - ...e ovviamente costi diversi!
- Larghezza di banda=velocità effettiva di trasmissione dei dati

La connessione in rete / 2

- Upstream
 - Collegamento da voi al ISP
- Downstream
 - Collegamento dal ISP a voi



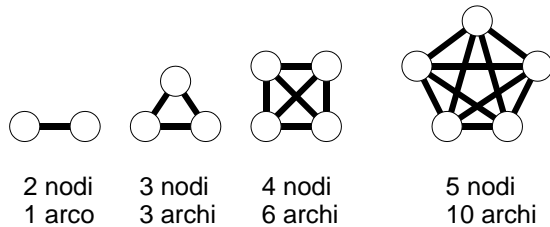
Alcune considerazioni sul “valore della rete”

- Sia n il numero degli utenti di una rete
- Sistema di distribuzione: rete broadcast (ad esempio Radio/TV)
 - Il valore della rete è $O(n)$, cioè proporzionale al numero dei suoi utenti
- Sistema di comunicazione: rete telefonica o e-mail
 - Il valore della rete è $O(n^2)$, ossia proporzionale al quadrato del numero degli utenti
- Sistemi peer-to-peer
 - $O(2^n)$

Legge di Sarnoff: Sistemi di distribuzione

- Supponiamo di possedere una fattoria, con 10 clienti che acquistano una bottiglia di latte al giorno, spendendo 1 euro ciascuno
- Se ampliamo il parco clienti del 10% acquisendo un nuovo cliente, i nostri ricavi aumenteranno anch'essi del 10%
- Legge di Sarnoff
“Il valore di una rete di distribuzione cresce in proporzione al numero di utenti”

Numero di collegamenti in una rete



In generale: n nodi, $n(n-1)/2$ archi

Legge di Metcalfe: reti di comunicazione

- Supponiamo di possedere una rete telefonica con 10 utenti che si telefonano l'un l'altro una volta al giorno, pagando 1 cent per telefonata
 - Incasso giornaliero: $(10*9)/2 = 45$
- Se acquisiamo un nuovo utente, il numero dei clienti aumenta del 10%, ma i ricavi aumenteranno del 22% ! Infatti:
 - Incasso giornaliero: $(11*10)/2 = 55 = 45 * 1,22$

Cosa significa tutto ciò?

- Supponiamo di dover scegliere di entrare a far parte di una rete. Abbiamo due alternative
 - Una rete di grandi dimensioni con tanti utenti
 - Una rete di piccole dimensioni con pochi utenti
- Sicuramente sceglieremo la rete affollata
 - E più la rete diventa affollata, più diviene conveniente entrare a farvi parte
- Questo spiega perché oggi abbiamo un unico sistema di posta elettronica, quando fino a pochi anni fa erano dozzine

Ancora oltre...

- Il numero di gruppi di persone che si possono formare su Internet non è né n né n^2
- Ci sono circa 2^n sottoinsiemi in un insieme di n elementi
 - Quindi, il valore di Internet cresce in proporzione a 2^n . Questo numero diventa spaventosamente grande al crescere di n

Potenze di 2

- Ricordate la storia araba del sultano che vuole ringraziare un bravo ministro offrendogli quello che questo chieda? Il ministro risponde che vuole soltanto due chicchi di riso nella prima casella di una scacchiera, quattro nella seconda, otto nella terza e via così – una progressione basata su 2^n . Il re dopo aver deposto 8192 chicchi sulla 13a casella realizzò di colpo che non avrebbe avuto abbastanza riso per soddisfare la richiesta — 2^{64} : è più di 10^{18}

Su Internet / 1

- Tutte e tre le leggi si applicano a Internet
 - Servizi come i siti di notizie destinati ai singoli individui beneficiano dei nuovi utenti in modo lineare
 - Servizi commerciali che facilitano le transazioni beneficiano dei nuovi utenti in modo quadratico
 - Servizi che costruiscono *comunità*, come AOL, beneficiano di nuovi utenti in modo esponenziale

Su Internet / 2

- La cosa importante è che il valore dominante di una rete tipica tende a variare dapprima come la legge di Sarnoff, poi come quella di Metcalfe e infine in modo esponenziale, man mano che la dimensione della rete aumenta