

1 Lezione del 23-09-25

1.1 Introduzione

Il corso si pone di presentare le nozioni di base sulle reti informatiche, le tecnologie di rete più diffuse, i protocolli Internet e lo sviluppo di applicazioni distribuite *client-server* e *peer-to-peer* (**P2P**).

In particolare il programma del corso comprende:

- Sviluppo di **applicazioni** in rete:
 - Client-server;
 - Peer-to-peer.
- Reti a **connessione diretta**:
 - Collegamenti punto-punto;
 - Reti locali.
- Reti a **commutazione di pacchetto**;
- **Interconnessione** di reti di tipo diverso;
- **Trasporto** end-to-end e protocolli;
- **Sicurezza**;
- Reti **wireless** e **mobili**, intese come caso particolare delle normali reti **cablate** (*wired*).

1.1.1 Applicazioni in rete

Nel dettaglio delle *applicazioni in rete*, vedremo come già detto i paradigmi *client-server* e *peer-to-peer*, di cui possiamo già fare alcuni esempi:

- Applicazioni client-server:
 - Web;
 - Trasferimento file;
 - Posta elettronica;
 - DNS;
 - Ecc...
- Applicazioni peer-to-peer:
 - Ricerca di contenuti;
 - Torrent;
 - Telefonia online;
 - Ecc...

In questo ci avvarremo del concetto di **socket** come primitiva per la gestione della rete dal lato S/O.

1.1.2 Reti dirette, a commutazione e wireless

Inizieremo con lo studio di *collegamenti punto-punto*, e quindi di trasferimento affidabile di dati fra 2 punti. Vedremo poi le reti locali, ad accesso multiplo, e i casi particolari come *Ethernet*.

Vedremo quindi le reti a *commutazione di pacchetto* per la copertura di grandi regioni. Anche qui approfondiremo tecnologie come gli *switch*, ancora *Ethernet*, ecc...

Per quanto riguarda l'*interconnessione di reti* vedremo il protocollo Internet **IPv4**, il **routing** (cioè l'*instradamento*) e i protocolli di trasporto (**UDP** e **TCP**).

Parleremo anche di reti *wireless* e *mobili*, e quindi di tecnologie come **WiFi**, le **reti cellulari**, e reti senza infrastruttura come **Bluetooth**.

1.1.3 Sicurezza

Vedremo poi le minacce alla *sicurezza* e alcune soluzioni che abbiamo a disposizione per mitigarle. In particolare, tratteremo di **crittografia** e **integrità** dei messaggi.

Nello specifico parleremo di tecnologie a livello applicazione (**PGP**), a livello trasporto (**TLS** (usata in *HTTPS*)), a livello Internet (**IP-Sec**) e difese di sicurezza come **firewall** e **IDS**.

1.2 Terminologia

Iniziamo quindi a definire la terminologia di base usata nel corso, usando Internet come esempio.

1.2.1 Internet

La prima domanda che ci poniamo è "*Che cos'è Internet?*".

Visione ingegneristica

Iniziamo col vedere la definizione di Internet agli occhi di un ingegnere che si occupa di reti:

- Si tratterà di una rete che connette miliardi di *dispositivi*, detti **host** ("*ospiti*"), che eseguono *applicazioni in rete* al cosiddetto **edge** ("*bordo*") della rete.
- Una visione **interna** della rete ci dirà invece che è un insieme di **pacchetti** che viaggiano attraverso infrastruttura (*router*, *switch*), per raggiungere il loro destinatario.
- A livello **fisico** potremmo considerare le connessioni fisiche fra dispositivi, date da cavi, segnali radio, ecc...
- Infine, potremo organizzare le **reti** come collezioni di dispositivi, router e connessioni gestite da determinate organizzazioni.

Non è esattamente corretto parlare di "*reti di calcolatori*" in quanto oggi ad essere connessi a Internet sono tutta una gamma di dispositivi non necessariamente orientati al puro *calcolo*: è questo il caso del cosiddetto *Internet of Things* (**IoT**).

Possiamo quindi intendere Internet come una "rete di reti", cioè più **ISP** (*Internet Service Providers*) connessi fra di loro, che a loro volta connettono una gamma dispositivi (host, router, switch, ecc...).

Per governare l'operazione di tali rete si necessita di **protocolli**, che definiscono il modo in cui si inviano e ricevono messaggi in rete.

In particolare, per quanto riguarda Internet notiamo l'**IETF** (*Internet Engineering Task Force*), organizzazione che gestisce diversi standard del settore (anche detti **RFC**, da *Request For Comments memoranda*).

Visione utente

Per l'utente, internet sarà un insieme di **infrastrutture** che forniscono **servizi** finali, fra cui il Web, telecomunicazioni, streaming, ecc... Dal punto di vista delle **applicazioni** in esecuzione sui dispositivi, Internet rappresenterà un'interfaccia di programmazione per consentire la comunicazione fra processi su una o più macchine. In questo parleremo di **hook** che permettono alle applicazioni di **connettersi** a Internet, cioè accedere ad un qualche protocollo di trasporto dei dati.

1.2.2 Protocolli

Un **protocollo** è una precisa *specifica* del formato secondo il quale due dispositivi in rete si scambiano informazioni. Solitamente i protocolli si sviluppano in più fasi, successive nel tempo, dove si portano avanti diverse operazioni necessarie alla comunicazione.

Esempi di protocollo sono il *protocollo Internet* **IPv4**, e il *protocollo di trasporto* **TCP** usato nel Web e visto nel corso di progettazione web.

1.2.3 Infrastruttura di Internet

Vediamo più nel dettaglio la struttura di Internet:

- Abbiamo detto che all'*edge* di internet ci sono i cosiddetti *host*, cioè i **client** e i **server**. Notiamo che non vogliamo riferirci alle macchine fisiche client o server, ma ai **processi** che si comportano come tali per l'implementazione di un'applicazione distribuita.
- I dispositivi *terminali* che abbiamo appena nominato accedono ad Internet attraverso le cosiddette **reti di accesso**, cablate o wireless e basate sulle tecnologie utilizzate (router).

Per collegare i sistemi terminali ai router si usano *reti residenziali*, *reti di accesso istituzionali* (scuola, lavoro, ecc...), nonché *reti wireless e mobili* (Wifi, o reti come 4G solitamente fornite da privati). In questo caso può interessarci la frequenza di trasmissione, in bit al secondo, di una rete di accesso, o se quella rete è ad accesso *condiviso* (pensa WiFi) o *dedicato* (pensa Ethernet).

Uno standard storico per le reti di accesso è quello della trasmissione sulla linea telefonica su **DSL** (*Digital Subscriber Line*). Negli Stati Uniti si è invece diffuso l'uso della linea televisiva cablata. Oggi, sfruttiamo invece tecnologie come **ADSL** (*Asymmetric Digital Subscriber Line*) e **FTTC** (*Fiber To The Cabinet*). La differenza principale fra queste è che la linea in ADSL è interamente in rame, sia dalla centrale all'armadio di ripartizione che dall'armadio ripartilinea agli utenti finali, mentre nella linea FTTC si porta il segnale all'armadio attraverso cavi in fibra ottica. Lo standard di ultima generazione è **FTTH** (*Fiber To The Home*), che prevede una linea in fibra ottica anche dall'armadio agli utenti finali.

Possiamo quindi vedere la rete locale (**LAN** (*Local Area Network*)) di una comune abitazione come composta da un router, connesso a un **modem** DSL (*modem* deriva da modulatore/demodulatore sulla linea telefonica dei messaggi Internet) o direttamente via cavo ad un altro centro di ripartizione, e ad eventuali dispositivi come

access point WiFi che offrono la connessione via rete mobile ai dispositivi finali (una cosiddetta **WLAN**, *Wireless Local Area Network*).

Altre soluzioni per le comunicazioni wireless sono rappresentati da reti **cellulari** su larga scala, che sono quelle usate dagli operatori telefonici (tecnologie come **4G**, ecc...).

- Dalle reti di accesso si arriva a Internet attraverso reti interconnesse di router, arrivando quindi alle *reti di reti* di cui stavamo parlando.

2 Lezione del 24-09-25

2.1 Comunicazione dati su Internet

Abbiamo visto la struttura a livello fisico della rete Internet. Vediamo adesso i meccanismi secondo cui la trasmissione di dati avviene. La rete Internet è una rete a **commutazione di pacchetto**, il compito degli host è di:

- Ottenere messaggi dalle applicazioni;
- Dividere quei messaggi in frammenti più piccoli, detti **pacchetti**, di dimensione L bit;
- Trasmettere quei pacchetti nella rete di accesso ad una *frequenza di trasmissione* (o **bit-rate**) R .

Il tempo T_{packet} necessario a trasmettere un pacchetto da L bit su una linea da R bit al secondo di bitrate sarà quindi semplicemente calcolato come:

$$T_{\text{packet}} = \frac{L}{R}$$

Il bitrate, detto anche frequenza di *link*, dipende appunto dal **link** (o *mezzo*) della trasmissione, cioè l'infrastruttura fisica che sta fra trasmettitore e ricevitore. Possiamo classificare 2 tipi di link:

- Mezzi **guidati**: segnali che si propagano in mezzi solidi (rame, fibra, cavi coassiali, ecc...).
 - Un celebre esempio di mezzo trasmissivo guidato è il classico **doppino telefonico**, o trasmettitore a *Twisted Pair* (**TP**). Questo è formato da due fili di rame isolati e avvolti l'uno sull'altro, che permettono la trasmissione differenziale e quindi la riduzione dei rumori in *common-mode*;
 - Un altro tipo di mezzo trasmissivo guidato è il **cavo coassiale**, formato da due conduttori concentrici in rame separati da un dielettrico. Il segnale è trasferito come campo magnetico fra i due conduttori: questo permette bitrate più alti rispetto al normale doppino telefonico e una migliore schermatura dalle interferenze;
 - Infine, possiamo parlare della **fibra ottica**, formata da fibra di vetro che porta impulsi luminosi ad altissima velocità. Questa tecnologia presenta velocità di trasmissione estremamente alte, e vista la natura luminosa del segnale, non è suscettibile ad interferenze (alta affidabilità, cioè piccola frequenza di errore sui bit).

- Mezzi **non guidati**: segnali che si propagano nell'etere (segnali radio, ecc...). Questi sono solitamente meno sicuri ma significativamente più comodi per l'utente finale (possibilità di spostarsi, mancanza di cavi, ecc...).

La trasmissione wireless è suscettibile a fenomeni fisici come *riflessi*, *interferenze* e *ostruzione* da parte di oggetti fisici.

- Il **WiFi** è un esempio di mezzo di trasmissione non guidato che può raggiungere centinaia di Mbps su regioni locali;
- Reti wireless più ampie possono essere quelle **cellulari**, usate nella telefonia mobile;
- Infine si può parlare delle **reti satellitari**, usate per l'interconnessione di regioni geografiche fra di loro anche molto distanti.

2.1.1 Commutazione di circuito

Prima della commutazione di pacchetto si usava la tecnica della **commutazione di circuito** (ad esempio sulle linee telefoniche). Questo prevede di dedicare completamente una certa linea di trasmissione alla comunicazione fra due host, invalidandone quindi l'uso da parte di altri host.

Chiaramente, la commutazione di pacchetto permette un carico migliore della linea, dove più pacchetti provenienti da diverse fonti possono viaggiare a istanti temporali molto vicini fra di loro.

In particolare, la commutazione di pacchetto è utile per dati trasmessi in *burst*, mentre la rete a commutazione di circuito assicura minima congestione possibile a costo di occupazione completa della linea.

2.1.2 Commutazione di pacchetto

La tecnica della **commutazione di pacchetto** o *packet-switching* permette ad una rete di **router** interconnessi di ricevere ed instradare (letteralmente, "*routing*") pacchetti provenienti da più fonti in modo che raggiungano la loro destinazione.

Questo inserisce chiaramente un ritardo nel sistema, in quanto il router deve:

- Ricevere il pacchetto *completamente* ed memorizzarlo: questo richiede L/R secondi;
- Leggere l'header del pacchetto per capire il prossimo passo dell'instradamento;
- Trasmettere il pacchetto verso la sua nuova destinazione (un altro router o l'host finale), impegnando ancora L/R secondi.

Abbiamo quindi che il ritardo end-end immesso dal router è necessariamente di almeno $2L/R$ secondi, tralasciando il tempo necessario all'instradamento stesso.

Nel caso generale si abbiano N router (quindi $N + 1$ link fra i router) e P pacchetti da inviare, dovremo considerare che il primo pacchetto arriva in $(N + 1) \frac{L}{R}$ (deve attraversare tutti i link) e i successivi $P - 1$ pacchetti arrivano in $(P - 1) \frac{L}{R}$, per cui il tempo complessivo è:

$$T_{end-to-end} = (N + P) \frac{L}{R}$$

Se troppi pacchetti arrivano in un breve lasso di tempo, cioè se la frequenza di arrivo supera quella di trasmissione:

- I pacchetti verranno messi in coda finché non sarà possibile trasmetterli;
- I pacchetti possono essere persi se il buffer di memoria dedicato alla loro memorizzazione nel router si riempie.

2.2 Prestazioni della commutazione di pacchetto

Facciamo qualche considerazione ulteriore sulle prestazioni delle linee a commutazione di pacchetto. Abbiamo ottenuto il valore $T_{\text{packet}} = \frac{L}{R}$ per la trasmissione di un singolo pacchetto da L bit su una linea con bitrate R , e $T_{\text{end-to-end}} = (N + P) \frac{L}{R}$ per più pacchetti su un numero arbitrario di router.

Da quanto abbiamo detto nella scorsa sezione, un modello più sofisticato del packet-switching terrà conto di 4 sorgenti di ritardo:

- Ritardo di **trasmissione** T_{trans} , dato dalle caratteristiche del link. Come abbiamo già detto, questo vale:

$$T_{\text{trans}} = \frac{L}{R}$$

con L lunghezza del pacchetto in bit e R bitrate del link;

- Ritardo di **propagazione** T_{prop} , dato dalle proprietà fisiche del mezzo di trasmissione. In particolare, questo è il tempo fisico di trasmissione del segnale su un link, dato da:

$$T_{\text{prop}} = \frac{d}{s}$$

con d distanza del link e s velocità del mezzo di trasmissione. Chiaramente, per i nostri scopi s sarà una frazione significativa della velocità della luce $c \approx 3 \cdot 10^8$ m/s;

- Ritardo di **laborazione** (instradamento) T_{proc} , dipende dalle caratteristiche del router ed è perlopiù costante;
- Ritardo di **accodamento** dato dalla presenza di code T_{queue} . Questo è il più complicato da trattare, in quanto dipende dal numero di pacchetti presenti nel buffer del router. Come vedremo fra poco, una buona euristica per la valutazione di questo ritardo (che è comunque trattabile solo in maniera statistica) è l'*intensità di traffico* sulla linea di trasmissione.

Sommando queste sorgenti di ritardo potremo ottenere una stima del ritardo complessivo su un router (nodo) T_{node} :

$$T_{\text{node}} = T_{\text{trans}} + T_{\text{prop}} + T_{\text{proc}} + T_{\text{queue}}$$

Dati N nodi e P pacchetti, il ritardo end-to-end potrà quindi essere calcolato semplicemente come:

$$T_{\text{end-to-end}} = (N + P)T_{\text{node}}$$

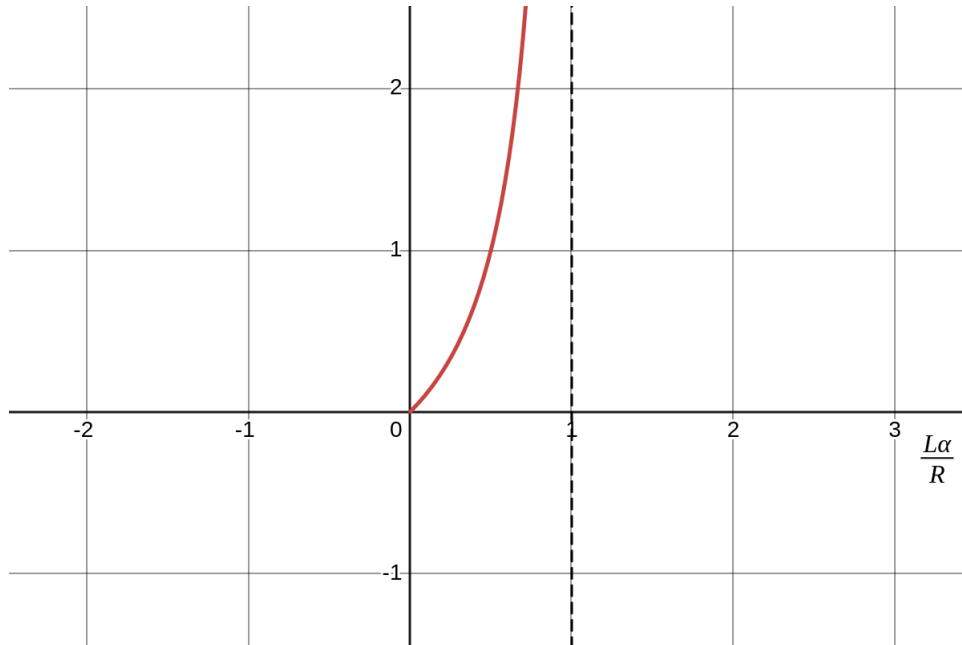
2.2.1 Ritardo di accodamento

Come anticipato, possiamo valutare statisticamente il ritardo di accodamento T_{queue} calcolando l'*intensità di traffico* su una linea:

$$I_{\text{traffic}} = \frac{L\alpha}{R}$$

presa α come la frequenza media di trasmissione di pacchetti sull'istante temporale.

Semplicemente tracciando la funzione si nota che se $I_{\text{traffic}} \geq 1$ il tempo di di accodamento tende a infinito (si ha necessariamente perdita di pacchetti), per cui vorremo mantenere $I_{\text{traffic}} < 1$, e idealmente $I_{\text{traffic}} \ll 1$.



2.2.2 Traceroute

Per fare diagnostica in situazioni reali, si possono usare software di **traceroute**, che inviano pacchetti con lo scopo di tracciare gli host incontrati e misurare il tempo impiegato nella trasmissione *round-trip*, o almeno capire se i pacchetti sono stati inviati o persi.

Più nello specifico, ipotizziamo di avere N router fra sorgente e destinazione. Traceroute invierà $N + 1$ pacchetti speciali sulla linea fra sorgente e destinazione, e ogni router reinvierà il primo di questo che riceve (l' n -esimo) indietro alla sorgente, assieme a informazioni temporali e il suo indirizzo. In questo modo, l' $N + 1$ pacchetto arriverà effettivamente alla destinazione, mentre gli N pacchetti precedenti scandaglieranno ognuno dei router sulla via per la destinazione.

In verità, questo procedimento viene ripetuto 3 volte, per avere risultati più accurati.

Vediamo ad esempio il traceroute che dal portatile su cui sto scrivendo gli appunti arriva ai server dell'Università di Pisa:

```
$ traceroute www.unipi.it
traceroute to www.unipi.it (131.114.104.6), 30 hops max, 60 byte packets
 1 _gateway (10.178.37.4)  3.745 ms  3.552 ms  3.494 ms
 2 172.16.133.129 (172.16.133.129)  300.183 ms  300.156 ms  300.133 ms
 3 172.16.133.38 (172.16.133.38)  60.081 ms  59.673 ms  60.235 ms
 4 172.16.17.50 (172.16.17.50)  52.910 ms  60.413 ms  59.354 ms
 5 151.7.56.70 (151.7.56.70)  61.571 ms  62.259 ms  60.645 ms
 6 151.7.56.66 (151.7.56.66)  89.876 ms  55.029 ms  86.501 ms
 7 151.6.3.199 (151.6.3.199)  52.064 ms  151.6.1.27 (151.6.1.27)  51.777
   ms 151.6.3.199 (151.6.3.199)  44.325 ms
 8 151.6.5.139 (151.6.5.139)  52.076 ms  151.6.1.213 (151.6.1.213)  52.842
   ms 151.6.5.139 (151.6.5.139)  52.399 ms
 9 garr.rom.namex.it (193.201.28.15)  52.662 ms  51.836 ms  52.516 ms
```

```

10  re1-rm02-rs1-rm02.rm02.garr.net (185.191.181.70)  52.317 ms  52.470 ms
    52.278 ms
11  rs1-rm02-rs1-pi01.pi01.garr.net (185.191.181.34)  56.893 ms  56.872 ms
    56.526 ms
12  rs1-pi01-rl1-pi01.pi01.garr.net (185.191.181.39)  56.495 ms  49.338 ms
    50.169 ms
13  rl1-pi01-ru-unipi.pi01.garr.net (193.206.136.90)  49.919 ms  49.808 ms
    50.003 ms
14  jser-jspg.unipi.it (131.114.191.90)  51.377 ms  50.215 ms  51.243 ms
15  * * *
16  * * *
17  * * *
18  * * *
19  * * *
20  * * *
21  * * *
22  * * *
23  * * *
24  * * *
25  * * *
26  * * *
27  * * *
28  * * *
29  * * *
30  * * *

```

I numeri a sinistra indicano l'indice del router raggiunto, seguiti dall'indirizzo e le misurazioni temporali nei 3 tentativi effettuati dal traceroute. In particolare, notiamo come i router da 2 a 8 sono anonimi, al 9 siamo arrivati nella rete del GARR, e dal router 15 in poi non abbiamo più ottenuto risposte ai pacchetti traceroute (probabilmente per via della configurazione dei router, il traffico passa comunque).

2.2.3 Throughput

Introduciamo il concetto di **throughput** come la frequenza (bit al secondo) con cui i bit vengono spediti da trasmettitore a ricevitore. Il throughput può essere:

- **Istantaneo**: calcolato ad un certo istante temporale (per quanto possibile dalla natura discreta della trasmissione);
- **Medio**: calcolato su un periodo temporale più lungo.

Ipotizziamo che un server debba inviare un file di F bit ad un client. La linea di comunicazione fra server e router ha capacità di link di R_S bit/secondo, mentre la linea fra router e client ha capacità di R_C bit/secondo. Chiaramente la capacità totale sarà:

$$R_F = \min(R_S, R_C)$$

cioè link con basse capacità di trasmissione rappresentano *bottleneck* per il throughput di tutto il collegamento fra dispositivi.

Facciamo adesso l'esempio di una rete condivisa fra più coppie client/server. In questo caso, date N coppie, la capacità della rete di interconnessione agli occhi di una singola coppia sarà, preso R come la capacità complessiva:

$$R_i = \frac{R}{N}$$

per cui ogni coppia vedrà una linea di comunicazione con capacità:

$$R_F = \min(R_S, R_C, \frac{R}{N})$$

2.3 Struttura di Internet

Abbiamo detto che Internet è una rete di reti.

- Gli host si collegano a internet attraverso le *reti di accesso* (reti LAN, istituzionali, reti mobili, ecc...);
- Le reti di accesso si collegano agli **ISP** (*Internet Server Provider*);
- Gli ISP devono essere collegati ad altri ISP per permettere la comunicazione fra host su diversi ISP (all'interno di **IXP** (*Internet eXchange Point*, ecc...)).

La struttura di reti interconnesse che si viene a formare è molto complessa, e la sua evoluzione è stata guidata da fattori economici e politici.

Abbiamo quindi una struttura gerarchica:

1. Al livello più alto troviamo i cosiddetti **tier 1 ISP** e i **content provider** (Google, Facebook, ecc...) che si occupano di copertura nazionale e internazionale. In particolare, i content provider preferiscono collegarsi direttamente agli IXP per risparmiare sugli ISP;
2. Seguono gli **IXP**, che collegano più ISP fra di loro, e gli ISP locali (regionali, ecc...);
3. Infine troviamo le reti di **accesso** locale (reti LAN, WLAN, ecc...).

2.4 Sicurezza Internet

La struttura di Internet espone i suoi utenti a diversi pericoli, fra cui:

- **Virus**: programmi maligni che si replicano modificando altri programmi;
- **Worm**: programmi maligni che si replicano con lo scopo di diffondersi in altri computer;
- **Spyware**: programmi maligni che cercano di ottenere informazioni che violano la privacy;

ed altre svariate categorie di *malware*.

2.4.1 Attacchi DoS

Un caso tipico di attacchi informatici in rete è quello degli attacchi **DoS** (*Denial of Service*), dove l'attaccante cerca di sfruttare le risorse di un server (memoria, larghezza di banda) al punto di renderlo inutilizzabile ad altri utente. Questo può essere fatto inviando traffico maligno al server (pacchetti molto grandi o corrotti).

La tecnica può essere espansa a più attaccanti (*botnet*) dando vita a tecniche più sofisticate (come il **DDoS**, *Distributed Denial of Service*).

2.4.2 Intercezione di pacchetti

Una problematica delle reti a commutazione di pacchetto è l'**intercezione** dei pacchetti instradati, o in inglese *packet sniffing*. Questo è più facile su reti wireless piuttosto che cablate, ed espone pericoli ovvi per la sicurezza (password, dati sensibili, ecc...).

Oggi, come vedremo, si usano protocolli che implementano forme di crittografia per mitigare questo tipo di problematiche.

2.4.3 Spoofing IP

Il problema di identità fasulle si presenta in Internet attraverso pacchetti malformati, con indirizzi sorgente sbagliati. Queste tecniche, seppure meno pericolose, possono comunque essere usate per confondere o comunque complicare il traffico sulle reti.