Ethernet: principi di base

Un esame dei punti chiave per comprendere le caratteristiche di base di un protocollo, Ethernet, che insieme a USB è il mezzo di comunicazione per eccellenza dal mondo embedded verso il PC

Daniele Mastrilli

In origine, alla base dell'interfaccia di protocollo Ethernet, c'è CSMA/CD, acronimo di Carrier Sense Multiple Access with Collision Detect. In parole semplici il termine significa "Ascoltare prima di parlare": ascoltare se qualcuno nel momento in cui un utente vuole comunicare stia già occupando il relativo canale, e in caso affermativo prenderne nota.
La dimensione del payload va da 46 a 1500 Byte (a dire il vero, un byte in terminologia Ethernet è definito Ottetto) mentre il datarate va da decine di Mbit/s fino al Gbit/s.
Ethernet presenta caratteristiche fondamentali che lo rendono il mezzo di comunicazione ideale per molte applicazioni. Un motivo per tutti è la connettività Internet sviluppata su Ethernet, che ha una caratteristica unica, l'ubiquità.
Ethernet è un protocollo definito nei livelli fisico e del link secondo specifiche IEEE 802.3.
La definizione è versatile e aperta, infatti il massimo rate può essere 10Mbit/s, oppure 100Mbit/s, o ancora 1000Mbit/s e oltre (GbEthernet). Il modo di trasmissione può essere Broadband oppure Baseband. Il mezzo fisico di trasmissione può essere un cavo coassiale, una fibra ottica, e così via.
L'approccio di Ethernet è basato, come mostra la figura 1, sullo sviluppo verso l'applicazione organizzato in Layer. La figura mostra parte del protocollo TCP/IP come layer superiori ai livelli di base di IEEE 802.3.

Incapsulamento

Un concetto molto importante è quello dell'incapsulamento. Ogni livello dello stack è responsabile per una particolare funzionalità del protocollo: quello fisico è responsabile della connessione elettrica per la trasmissione del dato attraverso un mezzo fisico (filo, fibra, …). Ogni layer superiore implementa delle funzioni specifiche ed è a sua volta suddiviso in più blocchi funzionali senza sovrapposizione di ruoli o funzioni. Il dato nasce in un certo modo e nel passaggio attraverso i layer successivi viene incapsulato in un dato ancora più lungo e completo di informazione. In figura 2 si mostra come ogni livello associato a una sessione di un browser web è pianificato sul modello dello stack. Al livello di Ethernet (livelli del data link e fisico), il datagramma IP è trasportato attraverso una rete usando il protocollo IEEE 802.3. Anche a questo livello un frame MAC (Media Access Control) consiste di un header e di una sezione dati (payload). Il MAC header contiene informazioni relative al frame MAC, come il MAC Address sorgente (l'indirizzo fisico della sorgente del messaggio), il MAC address di destinazione e la lunghezza del frame. Il payload contiene invece il datagramma IP completo da trasportare.
Si noti come gli indirizzi incapsulati entro ogni protocollo sono

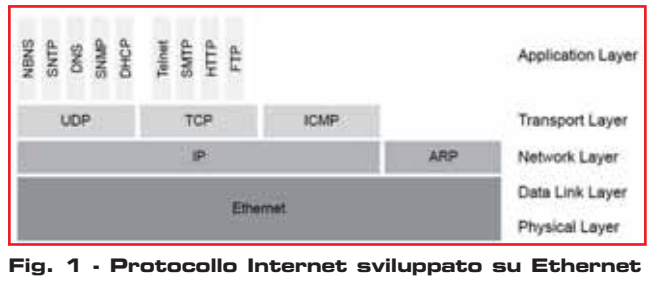


Fig. 1 - Protocollo Internet sviluppato su Ethernet

diversi e tipicamente non hanno una relazione fissa tra loro. Nell'esempio il pacchetto TCP usa un numero di porta che di solito per http è la 80. Il datagramma IP usa un indirizzo IP che viene assegnato staticamente o dinamicamente in funzione degli indirizzi internet disponibili. Il frame MAC usa l'indirizzo fisico assegnato a quel particolare dispositivo hardware.

Application Layer

È il livello che fornisce l'interfaccia verso l'utente. Quando l'application layer è associato a un protocollo inferiore, come nell'esempio di cui sopra relativo a TCP, a esso è associata una porta, che per server http è di solito come detto la numero 80.

Ci sono diversi protocolli a livello application associati a Internet, inoltre è possibile aggiornare lo stack introducendone dei nuovi, sfruttando la caratteristica del protocollo aperto. HTTP sta per Hyper Text Transfer Protocol ed è usato di solito per creare interfacce di navigazione nel World Wide Web (WWW). SMTP è un l'altro protocollo a livello di Application, significa Simple Mail Transfer Protocol, usato per il trasporto di e-mail attraverso la rete Internet. FTP (File Transfer Protocol) è il protocollo ottimizzato per il trasporto di file. DNS (Domain Name System) è usato per tradurre in un indirizzo alfabetico "NomeSitoWeb.com" l'indirizzo IP numerico. DHCP è il Dynamic Host Configuration Protocol, utilizzato per assegnare indirizzi IP dinamicamente a un particolare nodo. Telnet è usato per stabilire una connessione TCP interattiva con un nodo. SNTP (Simple Network Time Protocol) permette il sincronismo dei clock di nodo con un clock di riferimento. SNMP (Simple Network Management Protocol) è utilizzato per monitorare i dispositivi collegati alla rete per eventuali malfunzionamenti e altro.

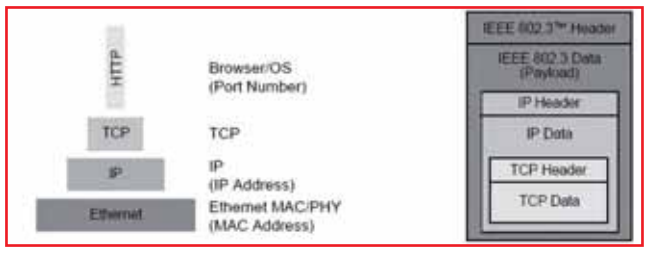


Fig. 2 - Esempio di incapsulamento del dato

Transport Layer

Racchiude dettagli che sono dipendenti dalla rete e li nasconde ai layer superiori che non si devono preoccupare del mezzo di comunicazione. Esso include anche gli indirizzi di trasporto, l'individuazione e il recupero dell'errore, e così via. Se usati al di sopra di un protocollo IP a questi layer di trasporto viene assegnato un numero IP.
Livelli di trasporto molto comuni sono TCP, che gestisce comunicazioni sicure; UDP, per comunicazioni efficienti ma non sicure; ICMP (Internet Control Message Protocol) usati per inviare errori di nodo o messaggi di stato.

Network layer

Determina come il messaggio deve essere instradato nella rete, tenendo presente l'informazione di QoS (Quality of Service) e provvede a fornire un network address al livello del trasporto.
Alcuni tra i più comuni network layer sono:
- ARP, Address Resolution Protocol, usato per traslare indirizzi di protocollo a indirizzi di interfaccia hardware, come indirizzi IP che devono essere traslati in indirizzi MAC.
- RARP, Reverse Address Resolution Protocol, usato inversamente rispetto ad ARP, per esempio per traslare un indirizzo MAC in un indirizzo IP.
- IP, Internet Protocol, un network layer privo di connessione usato da TCP, UDP, e così via.
Chiaramente tutti questi protocolli permettono di specializzare dei blocchi funzione e una piccola modifica a un blocco non deve toccare gli altri blocchi.

Link layer e fisico

Il layer fisico (PHY) fornisce una trasmissione trasparente di sequenze di bit attraverso la connessione fisica, includendo funzioni di codifica, di multiplexing, di sincronizzazione, di recupero del clock, di serializzazione e così via. Il Data Link layer è relativo invece alla trasmissione di frame in modo tale da non avere errori, gestendo la sequenza di frame, il controllo di flusso, e così via.

Il mezzo fisico

Ethernet è in parte definito dal mezzo fisico sul quale i frame di dato sono trasmessi. I mezzi fisici più comuni sono i seguenti:

- **1Mb/s**
  - 1Base5: doppino telefonico
- **10Mb/s**
  - 10Broad36: un cavo broadband
  - 10Base2: cavo coassiale RG58
  - 10Base5: un cavo coassiale
  - 10Base-F: una fibra ottica
  - 10Base-T: due paia di UTP (Unshielded Twisted Pair wire: doppino twistato non schermato) CAT3 o superiori in Full Duplex
- **100Mb/s**
  - 100Base-FX: due fibre ottiche, Full Duplex
  - 100Base-T2: due paia di UTP CAT3 o migliori, Full Duplex
  - 100Base-T4: quattro paia di UTP CAT3 o migliori, Half Duplex
  - 100Base.TX: due paia di UTP CAT5 o migliori, Full Duplex
- **1Gb/s**
  - 1000Base-CX: cavo di rame
  - 1000Base-LX: fibra ottica multimodo o monomodale per onda lunga
  - 1000Base-SX: fibra ottica multimodo per onde corte
  - 1000Base-T: 4 cavi CAT5e, CAT6 o migliori.

Il frame Ethernet 10/100

SSD (Start-of-Stream Delimiter): è il delimitatore del flusso di dati, un preambolo di sette ottetti di valore 55 esadecimale. Se funzionante su base 100Mb/s i primi quattro ottetti

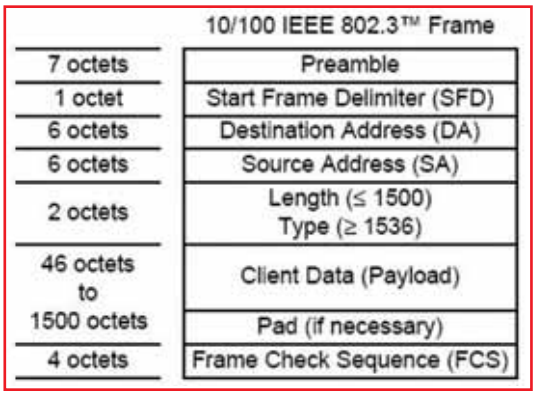


Fig. 3a - Formato base del frame Ethernet

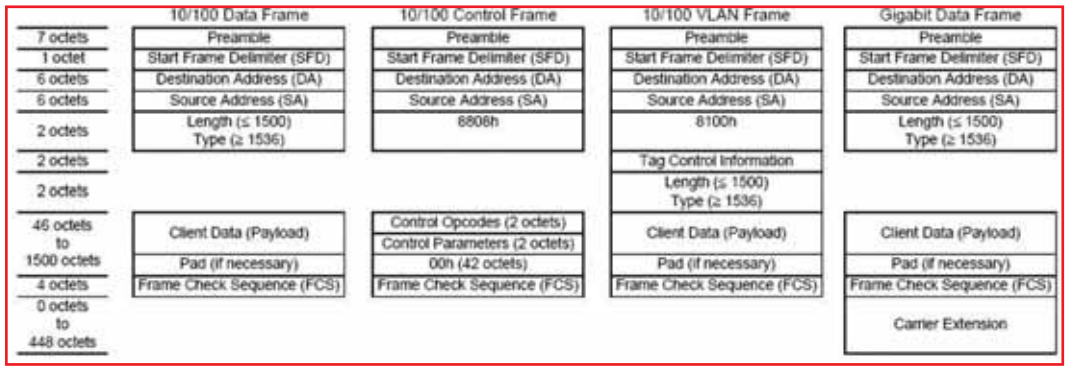
Fig. 3b - Differenze di formato per i diversi bit rate

La dimensione del Frame

IEEE 802.3 non include il preambolo nel calcolo della dimensione, pertanto la dimensione minima e massima di un frame di base o di controllo è di 64 oppure 1518 ottetti.
Di conseguenza la dimensione massima per una VLAN è di 1522 ottetti. La dimensione massima del frame di gigabit Ethernet è di 9000 ottetti. La dimensione del frame non va confusa con quella del payload.

sono codificati logicamente 4B/5B. Il preambolo è importante affinché il ricevitore possa agganciarsi al flusso di dati prima che il frame con il contenuto significativo arrivi.

SFD (Start-of-Frame Delimiter): è a volte parte del preambolo che viene così indicato essere di 8 ottetti. Il suo valore è '10101011'.



DA (Destination Address): è costituito da 6 ottetti di MAC Address dell'indirizzo di destinazione.

SA (Source Address): è costituito da 6 ottetti di MAC Address dell'indirizzo sorgente.

Length/Type: è formato da due ottetti e se ha valore minore o uguale a 1500 allora rappresenta il numero di ottetti nel payload, mentre se il suo valore supera o eguaglia 1536 allora rappresenta il tipo di Payload.

Payload (dato client): può per esempio essere un datagramma IP o altro. La sua lunghezza minima è di 46 ottetti mentre la massima di 1500.

Pad: dal momento che un datagramma reale potrebbe essere di lunghezza inferiore ai 46 minimi richiesti dalle specifiche, esistono degli ottetti detti pad che possono essere inseriti per completare il payload minimo.

FCS (Frame Check Sequence): costituito da 4 ottetti è calcolato sull'indirizzo di destinazione usando un CRC a 32bit.

ESD (End-of-Stream Delimiter): se operante in modalità 100 Mb/s il PHY, durante il ESD, trasmette paia di simboli per denotare la fine del frame. Nel caso di comunicazione 10 Mb/s al posto dello ESD è inviato uno speciale segnale detto TP-IDL che non è considerato parte del frame. Nella figura 3a viene schematizzato il formato base del frame Ethernet, mentre nella figura 3b sono riportate le differenze di formato per i diversi bit rate.

Control Frame

Frame Ethernet con un EtherType di 8808h sono specificati come MAC Control Frame e sono usati per controllare il flusso di frame attraverso una connessione. I primi due ottetti di un frame di controllo contengono lo opcode. Al momento l'unico opcode di controllo standard è quello di pausa, il quale ha un opcode e un destination address come indicato di seguito:

- Opcode: 001h
- Address: 01-80-c2-00-00-01 (multi cast).

Questo comando è inviato per richiedere alla stazione dall'altra parte del link di smettere di trasmettere per un certo periodo di tempo.

Indirizzi MAC

Un indirizzo MAC è lungo 6 ottetti e contiene un numero unico per ogni dispositivo hardware Ethernet. Esso consiste di 24-bit di OUI (Organizationally Unique Identifier – indirizzo attribuito da IEEE alla società che acquista il MAC) e 24 bit di identificatore dell'hardware (attribuito dal proprietario dello OUI ai propri dispositivi). Ci sono indirizzi il cui bit meno significativo dell'ottetto 1 è impostato in modalità multi cast che sono da interpretare come attribuiti a pacchetti multi cast. Un indirizzo MAC FF-FF-FF-FF-FF-FF è invece un indirizzo broadcast, che ha lo scopo di raggiungere tutti i nodi.

Costruzione dello stream

La funzione del fisico Ethernet e del MAC e le loro interfacce sono definite dalle specifiche IEEE 802.3.
L'interfaccia fisica verso il mezzo trasmissivo è chiamata MDI e cambia a seconda di quale sia il mezzo fisico (doppino, fibra e così via).
L'interfaccia tra PHY e MAC è chiamata MII ed è composta da un percorso di ricezione, da un percorso di trasmissione e da un percorso di gestione, il quale è usato per leggere e scrivere i registri del fisico.
La larghezza del percorso di ricezione e trasmissione è la stessa ed è determinata dalla velocità implementata da PHY e MAC. Per 10Mb/s si tratta di 4 bit a 2.5MHz, per 100Mb/s sono sempre 4bit ma a 25MHz. Ci sono poi delle particolari MII definite come Ridotta (RMII) e Seriale (SMII), da 2 bit e 1 bit rispettivamente.
In riferimento alla figura 4 (relativa alla definizione di MAC e PHY) i termini usati hanno il significato di seguito riportato:
Reconciliaton layer: mappa lo stato fisico (perdita di portante, collisione, e così via) al MAC layer.
Media independent Interface (MII) (Opzionale): è un'interfaccia a n-bit per la trasmissione e ricezione con il PHY.
Physical Coding Sublayer (PCS): codifica, Multiplexing e sincronizzazione di stream di simboli in uscita.
Physical Medium Attachment (PMA): trasmissione e ricezione di segnale (serializzazione/de serializzazione di stream di simboli, recupero del clock …).
Auto-Negotiation (Optional): negoziazione del modo più alto supportato da entrambe gli host
Medium Dependent Interface/Physical Media Dependent (MDI/PMD): RJ45 …
Medium: UTP, Fibra …

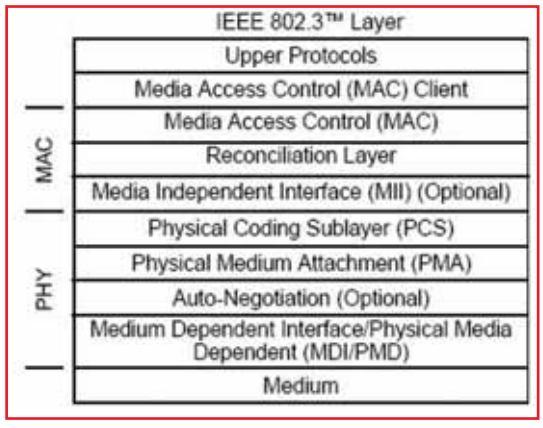


Fig. 4 - Definizione di MAC e PHY secondo IEEE 802.3 100Mbps

Temporizzazione del flusso

Prima di spiegare la temporizzazione di frame IEEE 802.3 è importante capire il motivo per cui si fa tale temporizzazione.
Originariamente Ethernet fu progettato affinché ogni nodo potesse accedere al mezzo condiviso, però solo un nodo per volta deve poter prender possesso del mezzo, e ogni nodo trasmette in half duplex. Senza un meccanismo di gestione delle collisioni, quando più nodi simultaneamente provano ad accedere al bus l'informazione viene corrotta con conseguente perdita di dati.
I requisiti di base di un protocollo di networking sono: nodi multipli possono trasmettere su un mezzo condiviso (accesso multiplo); ogni nodo deve essere in grado di sentire quando un altro nodo sta trasmettendo (Sensing della portante); ogni nodo trasmittente deve poter sentire se sul mezzo avviene già una trasmissione (individuazione di una collisione). Quando viene percepita la collisione ogni nodo deve prevedere un meccanismo per ritrasmettere senza conflitti.
Il tempo di attesa in cui un nodo aspetta prima di ritrasmettere, dopo avere sentito che il mezzo è già attivo perché una comunicazione è in atto, si chiama Inter-Packet Gap (IPG), e dipende dalla velocità del bus. Il meccanismo appena esposto è CSMA/CD. Il caso peggiore si ha quando un nodo molto lontano sente che il bus è libero e inizia a trasmettere, mentre subito dopo, un nodo più prossimo che non ha ancora percepito che il bus è stato attivato inizia anch'esso a trasmettere. Nel momento in cui il nodo vicino si accorge del conflitto invia un segnale di Jam che viene percepito dall'altro nodo e il conflitto è gestito. Il tempo necessario ad accorgersi della collisione si chiama "slot time" oppure "finestra di collisione". Quando la collisione è percepita entro lo slot time essa è definita "in-window". Collisioni fuori dalla finestra "in-window" devono invece essere gestite dal software perché non sono previste dal layer fisico e data link. Questo tempo è diverso a seconda del mezzo. Per 10Base-T la finestra in-window è definita su un diametro di connessione di 2500 con ripetitori, mentre i metri diventano 200 per 100Base-T. GbEthernet è sostanzialmente uguale a Ethernet invece.

Full Duplex

In origine Ethernet implementava CSMA/CD, ma reti Ethernet moderne sono configurate per topologie star o point-to-point. In questa tipologia un nodo è connesso al più a un altro nodo e può operare in full duplex. In questo modo il conflitto non è possibile e CSMA/CD non è più usato. In più, il throughput del mezzo, in una comunicazione full duplex, raddoppia, passando da 10Mbit/s a 20Mbit/s e da 100Mbit/s a 200Mbit/s.

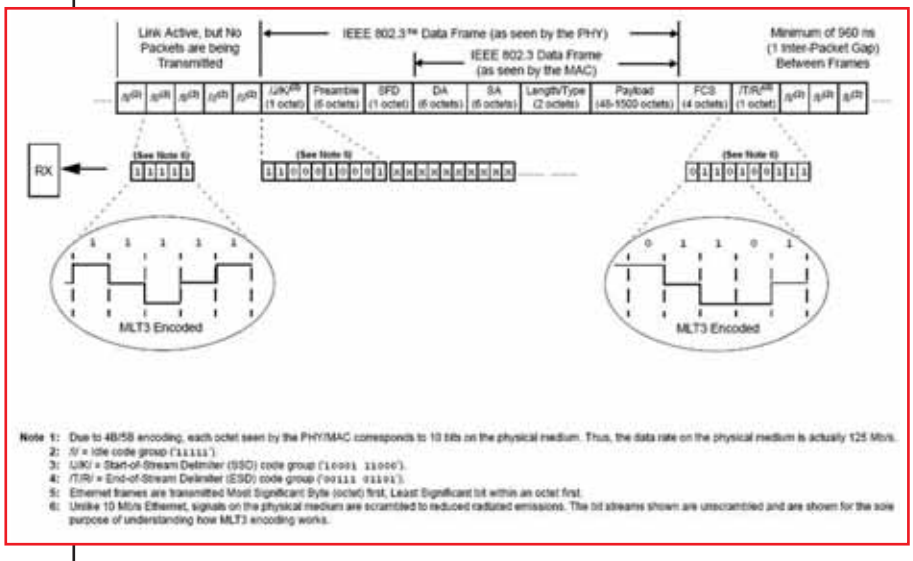


Fig. 5 - Codifica MLT3

Ci sono delle differenze tra lo stream di 100Mb/s e quello di 10Mbit/s.
Il primo passo nella trasmissione di uno stream a 10Mbit/s è quello di fare una codifica di tipo Manchester, che nel caso di Ethernet significa che lo 0 indica un passaggio di stato da alto a basso, mentre uno 1 indica il passaggio opposto, da basso ad alto. Questa codifica oltre a essere affidabile consente il recupero del clock, però richiede il doppio della banda.
Il secondo passo è quello di modificare la forma del segnale (wave shaping) al fine di ottenere il profilo definito dalle specifiche IEEE 802.3. Questi profili hanno lo scopo di garantire un adeguato livello di propagazione nel mezzo e minimizzare le emissioni elettromagnetiche indesiderate.
In fine il segnale è trasmesso attraverso il cavo usando driver di tensione o di corrente (questo dipende dal particolare PHY che si adotta).
Quando lo stream è un 100Mbit/s qualcosa cambia. I cavi intrecciati non schermati sono per natura dei filtri passa basso, così quando si aumenta la velocità lo schema di codifica Manchester non funziona più. Un altro problema riguarda la regolamentazione relativa alla potenza trasmessa, la quale deve stare al di sotto dei 30MHz.
Il segnale 100Mbit/s ha una frequenza data dalla codifica 4B/5B pari a 125MHz. Nel 100Mbit/s si hanno due codifiche delle quali una è logica ed è chiamata 4B/5B. Questa risolve un paio di problemi. Il primo è relativo al recupero del clock su stream di dati ricchi di zero. Senza transizioni non si avrebbe un clock da recuperare, perciò si perderebbe presto la sincronizzazione. In secondo luogo lo schema 4B/5B aiuta nella trasmissione di segnalazioni oltre che bit di dato. Per fare questo si inviano 4 bit di dato in 5 bit trasmessi.
Le traduzioni di sequenze di 4 bit in 5 bit sono fissate in tabella e risolvono il problema aumentando la ridondanza. Ogni codifica contiene sempre un uno pertanto si ha sempre almeno una transizione.
L'altro schema di codifica usato nel 100Mbit/s è noto come Multi -Level Transition 3 (MLT3), ed è mostrato in figura 5. Ogni livello logico 0 oppure 1 è codificato come una transizione a uno dei tre livelli. La transizione è sempre verso il livello più prossimo, e sempre nello stesso ordine (-1, 0, +1, 0, -1 …). Uno zero logico è noto come "nessuna transizione", mentre un livello logico 1 equivale a una transizione.
Allora sequenze prolungate di soli 1, dal momento che un 1 indica una transizione, equivalgono a una transizione a ogni bit. Dal momento che per completare un intero ciclo sono necessarie 4 transizioni, la massima frequenza fondamentale è ridotta di 4 volte, da 125Mhz a 31,25MHz. Questo permette di avere la potenza di trasmissione al di sotto dei 30MHz come richiesto.

Auto-negoziazione e Auto-crossing

L'auto-negoziazione è il processo secondo il quale due nodi comunicano e danno informazione dei loro requisiti, come velocità, duplex, e così via, in modo da poter ottenere la combinazione migliore ai due capi del collegamento. L'auto-negoziazione è eseguita in fase di inizializzazione ma non è sempre necessaria se non in GbEthernet.
È sempre vero che una porta di Tx sarà sempre connessa a una porta di Rx di un altro nodo e viceversa. Questa connessione è detta di crossover. Nella topologia UTP questo crossover è tipicamente fatto nello switch o hub o router. Un cavo Ethernet per questa tipologia di rete avrà il pin numero x collegato al pin numero x dall'altro capo. Esistono poi cavi "crossati" che presentano una connessione tra pin di capi estremi già pronti per un collegamento diretto tra due computer, collegamento nel quale uno switch o router o hub non c'è, e l'incrocio dei pin di Tx e Rx deve essere fatto manualmente nel cavo (anche tra router e router serve un cavo crossato). In alcuni casi l'apparato implementa un auto-crossover che modifica lo stato del pin da Tx a Rx e viceversa dopo la connessione, evitando all'utente la scelta del cavo.