Ethernet

Ethernet is a development of the Aloha radio network developed in Hawaii about 1970, for communication with remote nodes in difficult terrain.

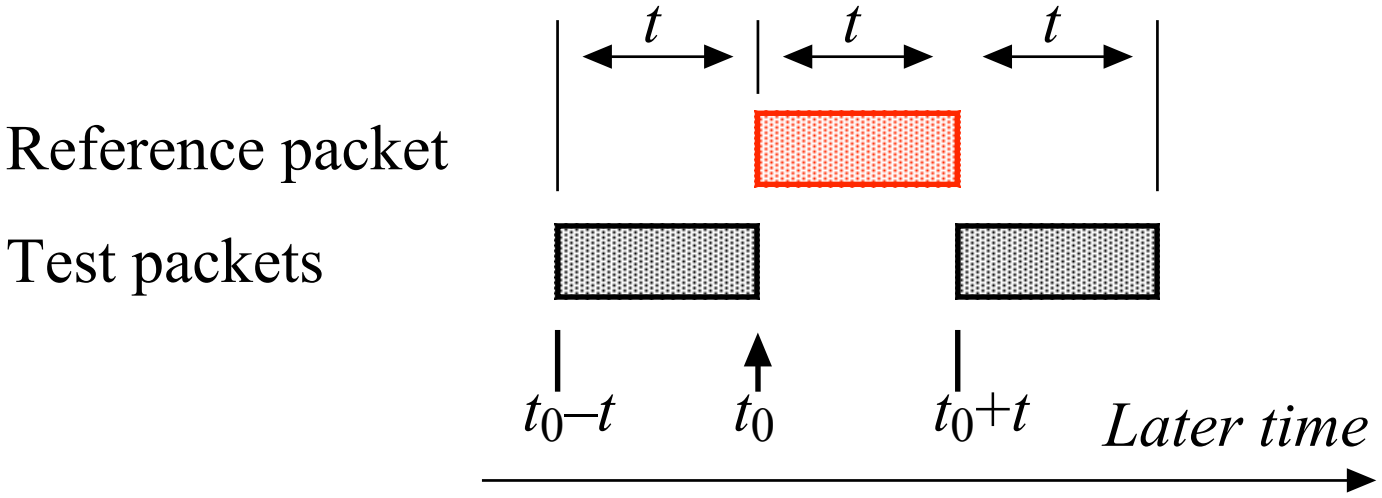
The central node can hear all remotes, and all remotes can hear the central, but remotes cannot necessarily hear each other.

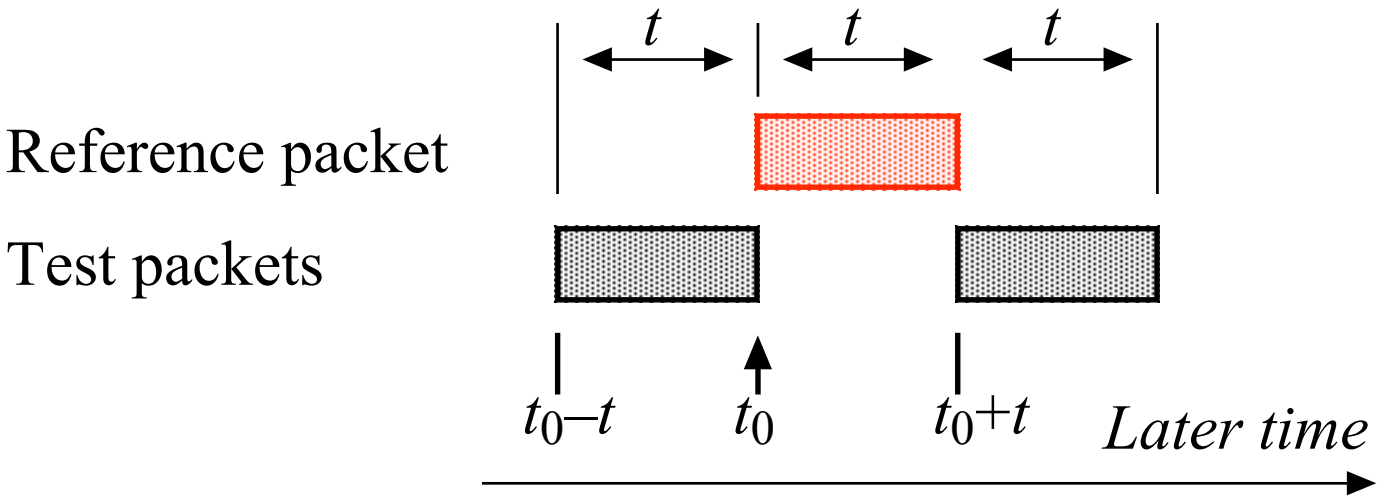
The Aloha network has —

- A single central node transmits packets on one radio channel to all remotes, whenever necessary.
- All remote nodes share a second radio channel on which they transmit back to the central site.
- When a remote node has a message, it just sends it. If it overlaps or "collides" with another transmission, both are garbled and neither is acknowledged. Both remotes wait a random time and retry.

Analysis of the Aloha protocol

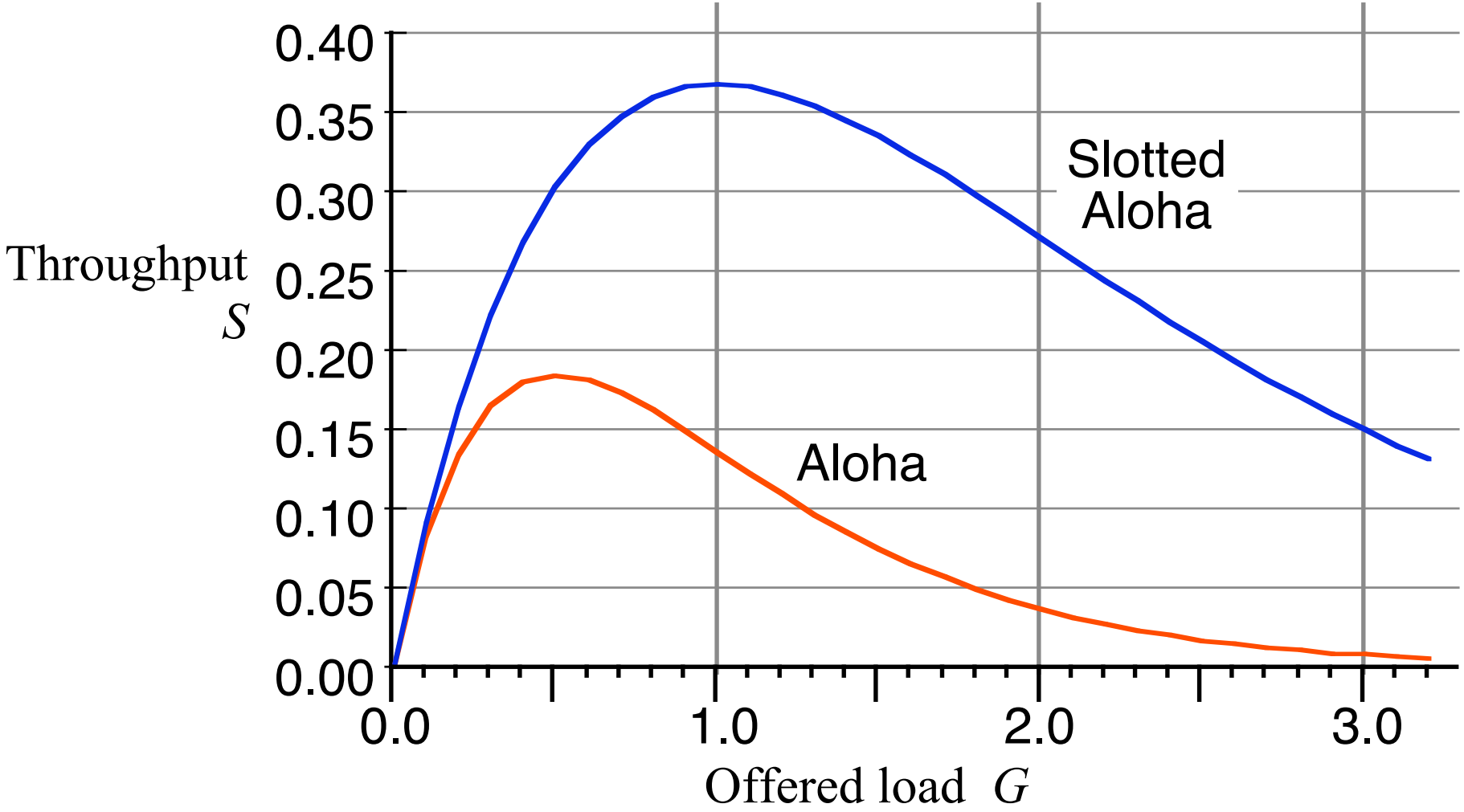
- The offered load G is the total traffic which all remote nodes try to inject into the network, as a fraction of the total channel capacity. ($G > 1$ if many stations attempt to transmit simultaneously).
- The throughput S is the traffic which is successfully transmitted ($S \leq 1$).
- Assume all packets are of length t, occurring randomly with Poisson statistics. Assuming that a reference packet starts at time t_0, then any other packet starting between times $t_0–t$ and t_0+t will collide.





- The *collision window* is then $2t$.
- The total traffic presented to the network, G, is the throughput S, plus the number of retransmissions from earlier collisions.
- The retransmission rate is $G \times prob[packet\ collision]$.
  If a Poisson process transmits at rate r, the probability of at least one transmission within period t is $1-e^{-rt}$.
  The probability of a transmission in the collision window is $1- e^{-2G}$, giving

$$G = S + G(1- e^{-2G}), \text{ or } S = G\, e^{-2G}.$$



- Aloha has a maximum throughput of $S = 1/2e = 0.184$ at $G = 0.5$.
- At $G > 0.5$ the network suffers *congestion collapse* as increasing the traffic to send (G), decreases that successfully sent (S).

Slotted Aloha

- In slotted Aloha, remote stations can send only at well-defined time "slots" (perhaps the centre sends regular ticks).
- This halves the collision window, and gives $S = Ge^{-G}$, with a maximum of $S=1/e$ =0.368 at G=1.0.

Persistent Aloha protocols retransmit with a probability $p < 1$ if a collision is detected; they increase the throughput at the cost of increasing the average delay in transmitting. (These are not important.)

Reservation Aloha protocols may be used in satellite–ground communication, with stations contending for slots within a transmission frame.

- The only way of avoiding congestion collapse is to reduce the offered traffic G, to force G below the peak of the throughput curve. (Perhaps as congestion is detected, wait progressively longer before retrying, or otherwise send less traffic.)

From Aloha to Ethernet

Ethernet, in its original form, makes several changes from Aloha —

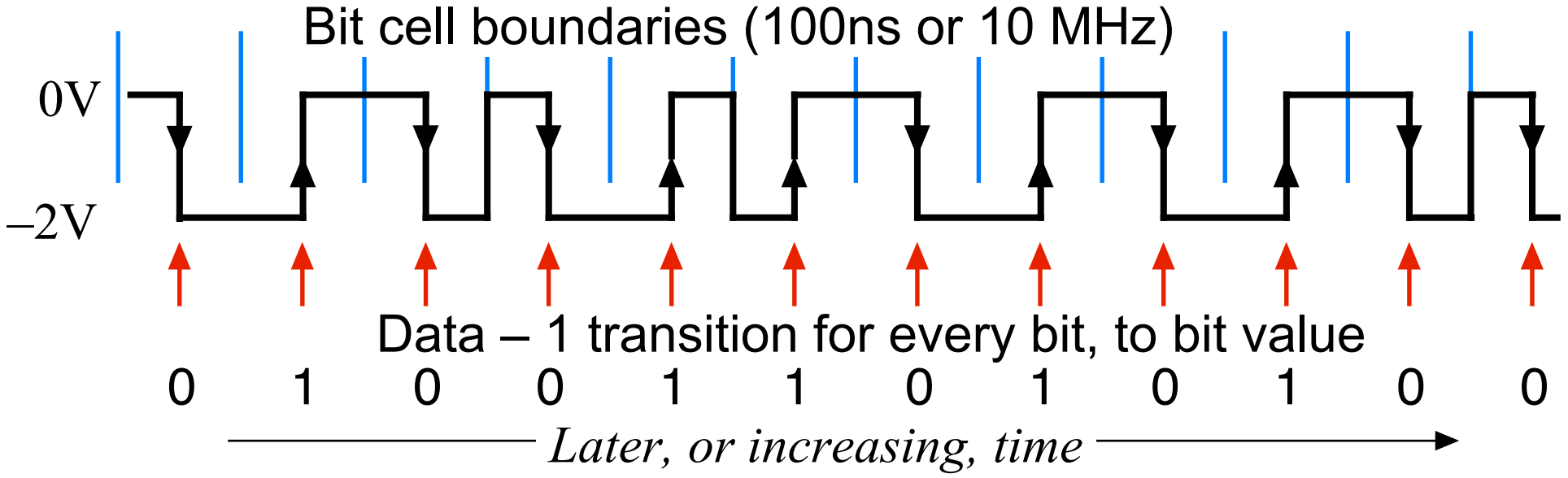
- Transmission is on a multi-dropped cable (this gives "Multiple Access" for many stations) – all stations can hear every transmission
- When a station wants to send it first monitors the cable (the "ether") and waits if any other station is sending (this is called "Carrier Sense")
- When a station does transmit, it monitors the cable (ether) for another station which sends at about the same time ("Collision Detection")
- A station which detects a collision stops transmission immediately, and "*defers*" a random time before retrying.
- If it collides again, the deferral time is doubled for each collision to reduce the network traffic and ease congestion (*binary exponential backoff*).

Ethernet is described as a CSMA/CD protocol

(Carrier Sense Multiple Access, with Collision Detection).

Original Ethernet – technical detail

- Signals are encoded with "Manchester coding" at 10 Mbit/s (100 ns clock interval), with voltages of about 0V and –2V on the 50Ω cable.
- The 100ns "bit cell" is divided into 2 halves, with a possible data transition in the middle of each cell.
- A data transition –ve → +ve encodes a "1"; and +ve → –ve a "0" bit.
- Data transitions are optionally provided at cell boundaries as needed.



Start and End of Frame

- Start of frame is signalled by the bits …10101011*xxx*...
- End of frame is signalled by the end of data transitions. (There may be a single final –2V → 0V change to restore the rest state.)

Important parameters

slotTime	512 bit times
interFrame Gap	9.6 µs
attemptLimit	16
backoffLimit	10
jamSize	32 bits
maxFrameSize	1518 octets
minFrameSize	512 bits (64 octets)
addressSize	48 bits

The Ethernet protocol

1. Wait until the medium has been idle for InterFrameGap (9.6μs)
2. Send preamble, start delimiter and frame, including FCS
3. If a collision is detected, send jamSize random bits, then stop sending and wait for a random number of slotTimes before retrying.

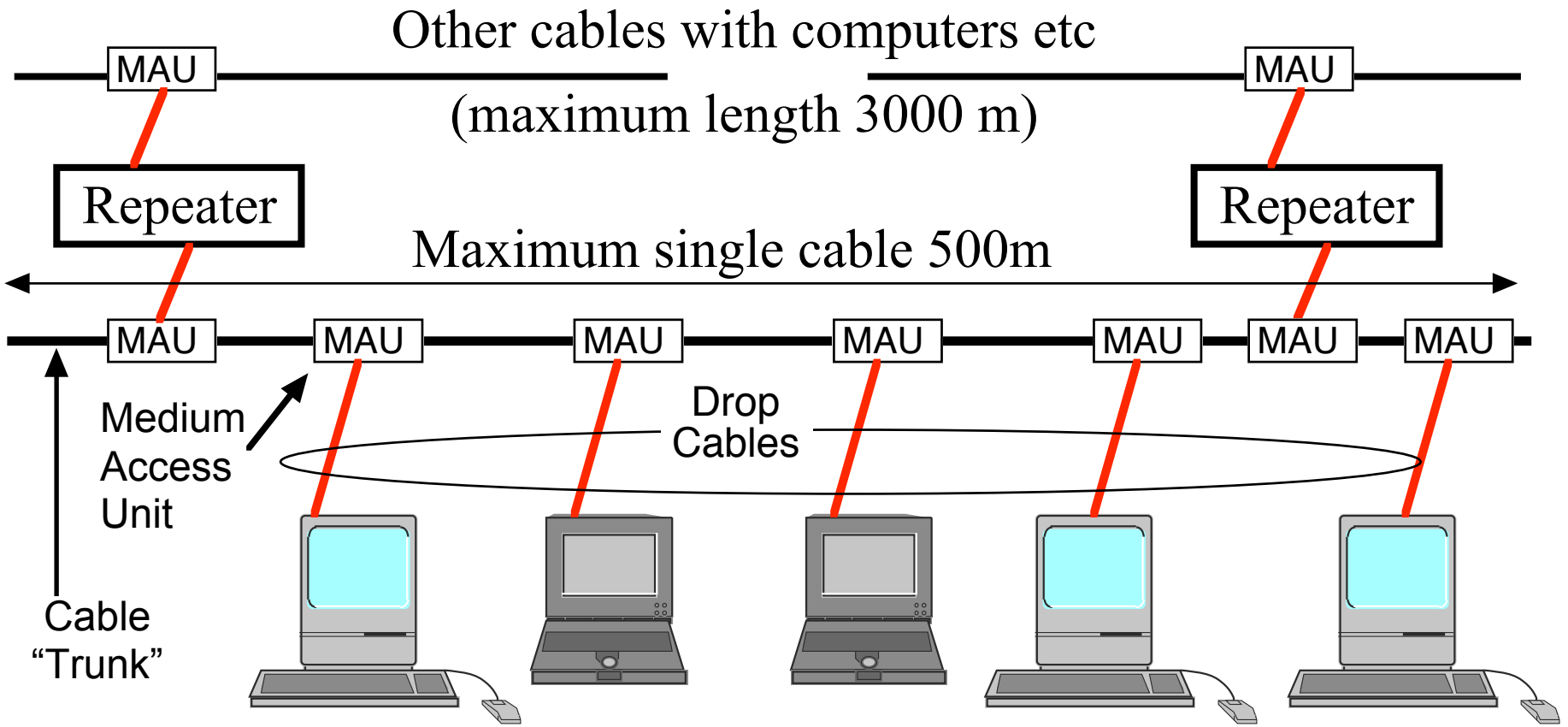
- If there have been n attempts at this transmission, the wait in step 3 is a random number k slotTimes, where $0 < k < 2^n$, for $n \leq 10$.
- If ($n > 10$) over 10 retries, the maximum delay stays at
  $2^{10} = 1024$ slotTimes $\approx$ 52ms
- If ($n \geq 16$) an irrecoverable error is signalled back to the software (maximum of 16 retries).

The successive doubling is called a "binary exponential backoff".
It assumes that any collision is due to an overloaded network and therefore reduces the offered load onto the medium to relieve the overload.

Later Ethernet developments.

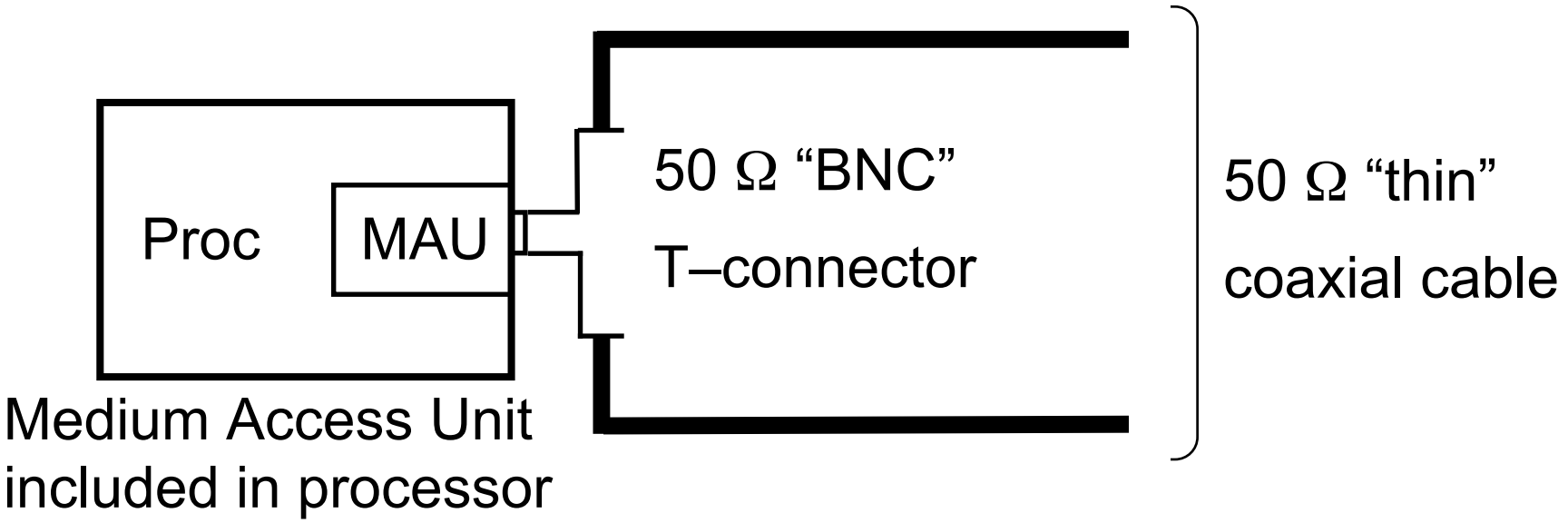
10BASE-5 (Baseband, or DC signalling, 500 m segments, “thickwire”)

This is the original Ethernet, with separate Medium Access Units to connect to a heavy “ether” cable, a maximum single cable length of 500 m, and a maximum segment of 3000m using repeaters to connect cables.



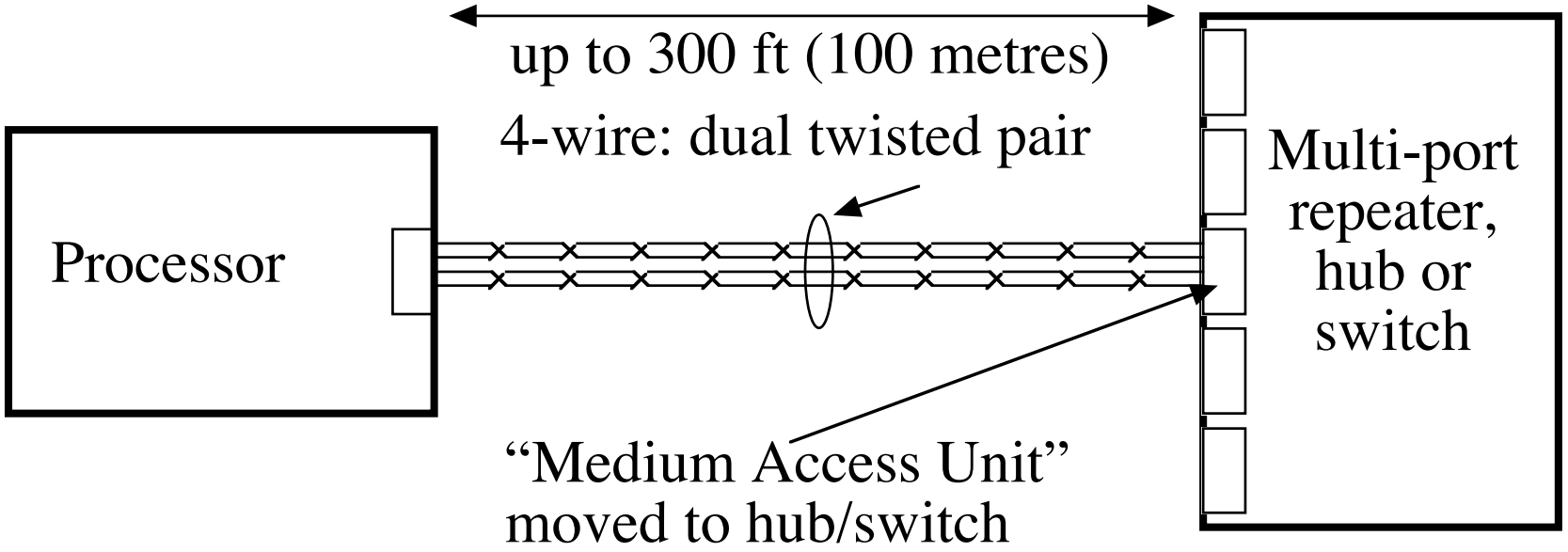
10BASE-2 Ethernet ("thinwire")

- This moves the MAU (Medium Access Unit) right into the computer, and uses a lightweight coaxial cable as the Ethernet "backbone".
- The cable is much lighter, more flexible and less expensive than "thickwire" cable, but is limited to 185 metres in a single run.
- It is electrically compatible with 10BASE-5 systems and the two can be intermixed with repeaters.



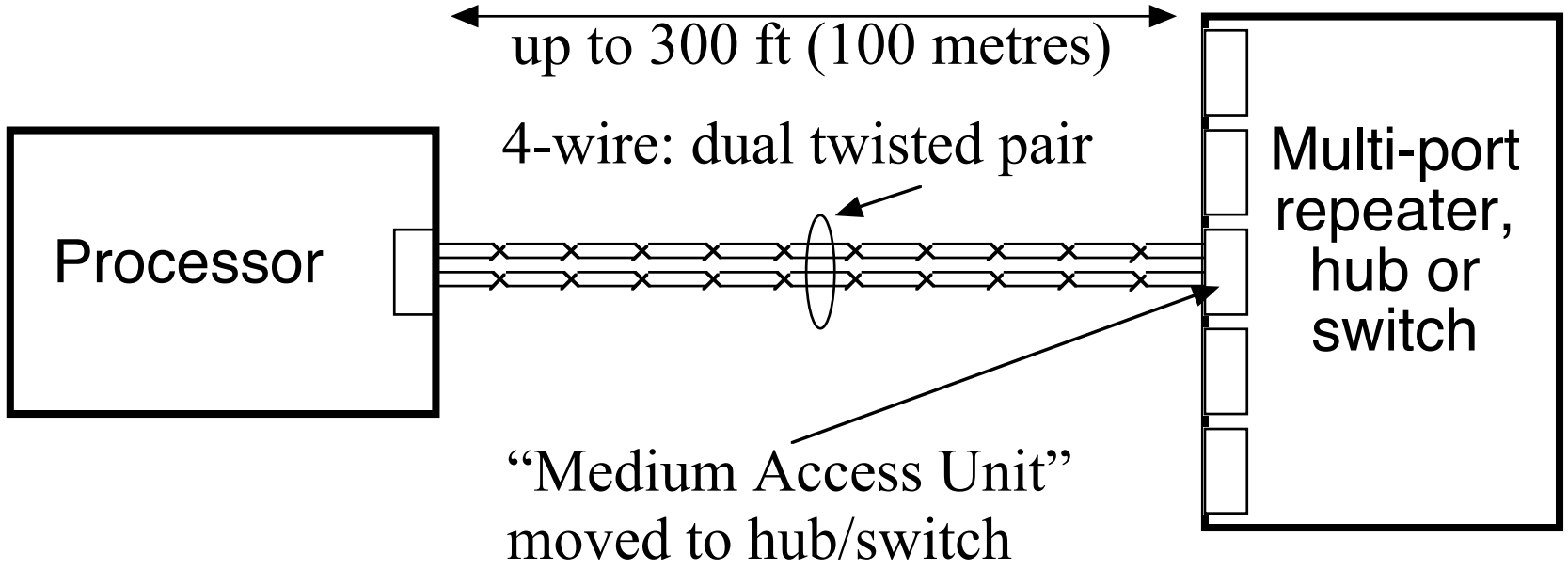
10BASE-T Ethernet (twisted pair)

- This version moves the cable itself and the MAU right into a multi-port repeater, hub or switch.
- The 10BASE-5 MAU "drop cables" are replaced by much simpler 4-wire cables (dual twisted pairs)



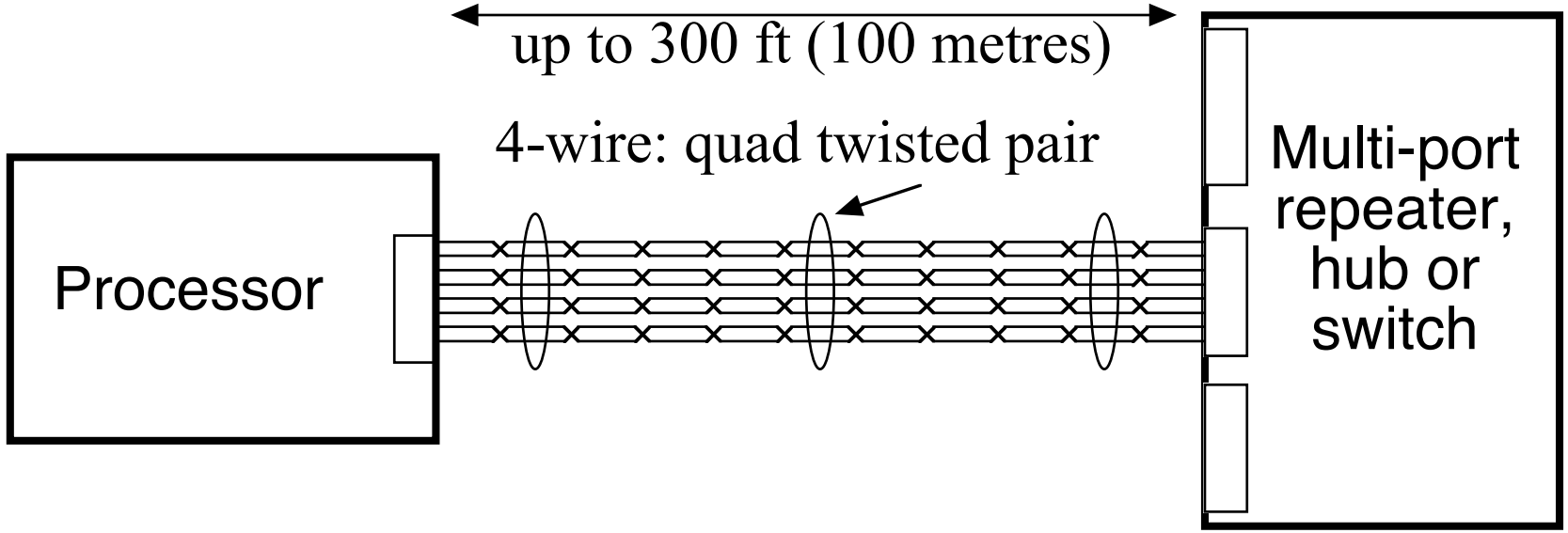
100BASE-T Ethernet ("Fast Ethernet")

- Physically identical to 10BASE-T Ethernet, but with higher quality cable (CAT-5 UTP – Unshielded Twisted Pair) (or optical fibre – 100BASE-F)
- Can "fall-back" to 10BASE-T operation if one port is 10MHz only
- Two ends "auto-negotiate" speed, full/half duplex, etc
- The user sees identical protocols to 10BASE-*x* ethernets.



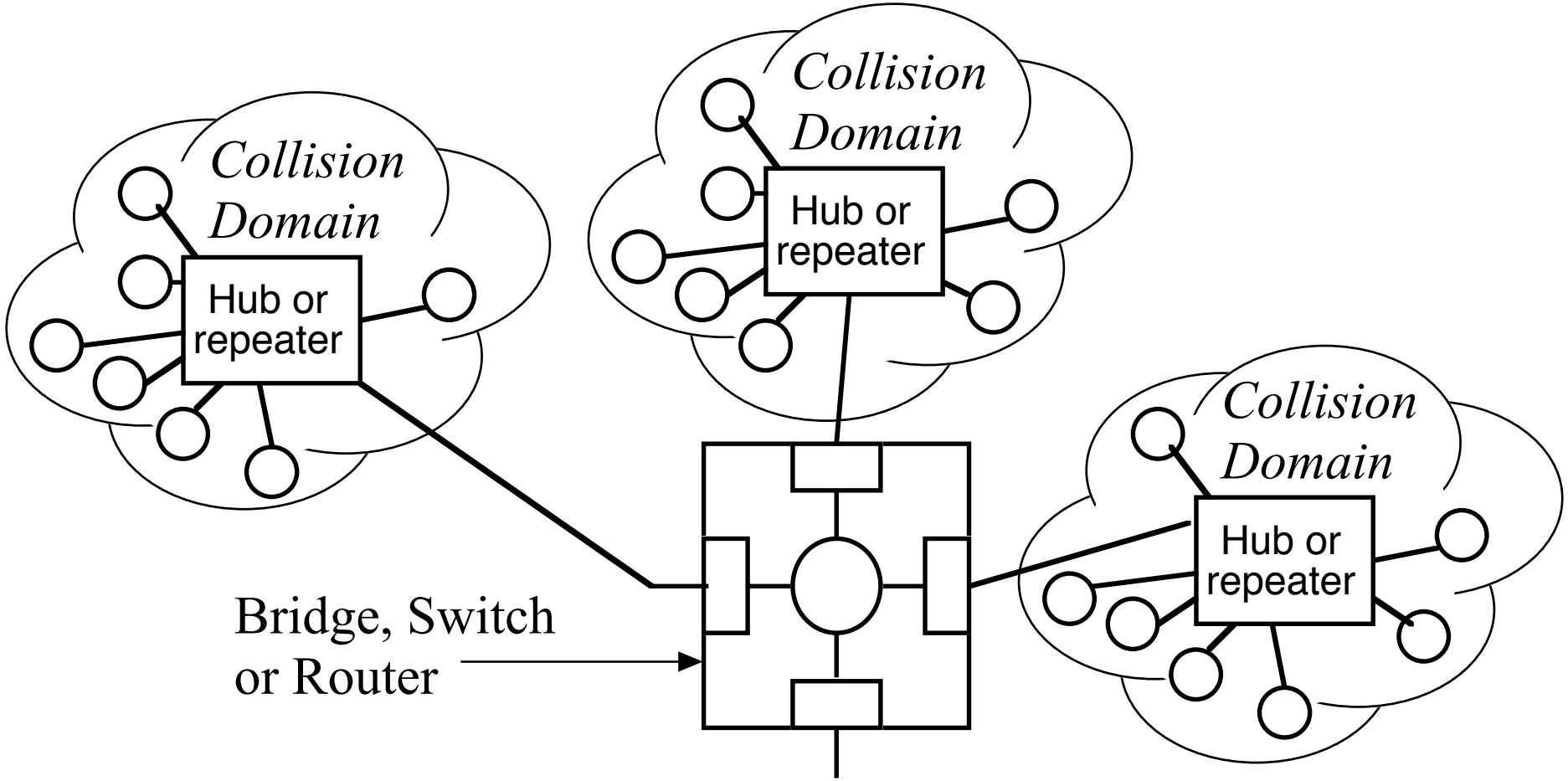
1000BASE-T Ethernet ("Gigabit Ethernet")

- Uses 4 × twisted-pair cable (1000BASE-T) or fibre (1000BASE-F)
- Two ends "auto-negotiate" communication parameters
- Identical user protocols to 10BASE-*x* ethernets.
- Can send a burst of packets, up to 8,192 bytes



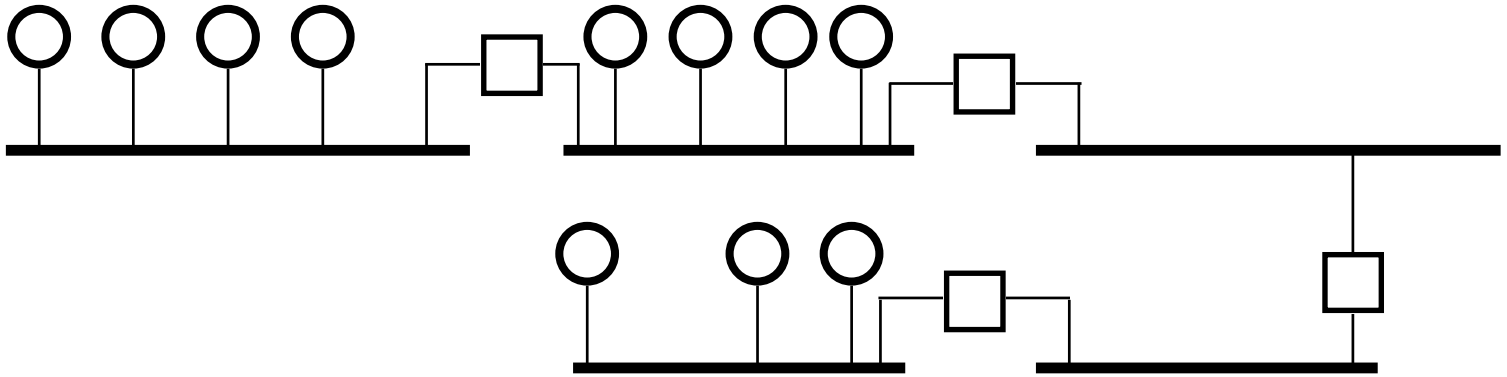
Collision Domains

- Repeaters and hubs handle bits (switches and routers recognise frames); distributing any bits received on one port to be sent on all other ports and form the equivalent of an Ethernet segment .
- The resulting "*collision domains*" are separated by switches, etc.



10 Mbps Ethernet repeater Rules

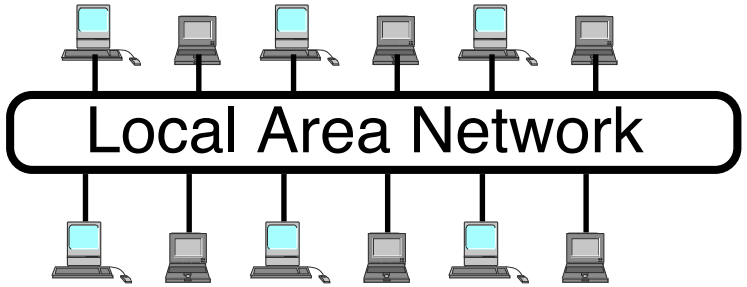
- The network may have up to *five* Segments (of up to 500 m diameter each)
- The five segments imply at most *four* repeater hops
- No more than *three* of the segments may have nodes
- This means that *two* segments must be just links
- The whole forms *one* Collision domain, diameter up to 2500 m, with no more than 1024 stations.



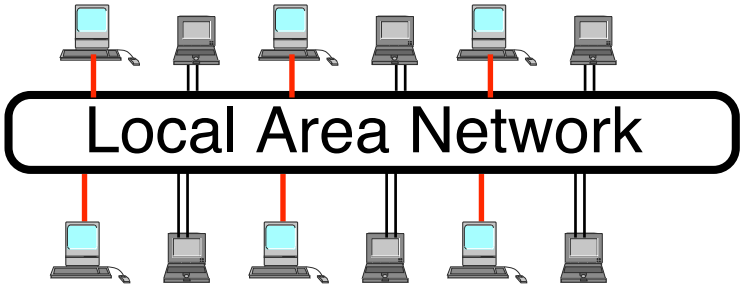
Virtual LANs (VLANs) (moved here from previous handout)

A traditional LAN consists of all the stations and bridges to one side of a router or gateway. All of the stations on the LAN can see all of the traffic. (Bridge routing may affect the detailed message distribution, but certainly all broadcast messages go to all stations.)

A Virtual LAN uses intelligent switches or bridges to selectively forward traffic to some stations but not others, even if they are on the same physical LAN. Accessing is controlled to individual nodes or stations.



Standard LAN –
all nodes see each other



Virtual LAN –
nodes divide into mutually invisible sets

- Note that a Virtual LAN depends on stations connecting individually into a switch and will not work at all with a traditional bus or loop topology.

VLANs can give –

- Flexibility. Reconfiguration can be done by commands, rather then by recabling.
- Security. Many VLANs encrypt their data and tunnel it through other protocols. Even if traffic is intercepted it may be unreadable.
- Cost. A building may have 3 or 4 VLANs all sharing a single physical network. Thus only one set of cables, switches etc needs to be provided.
- Terminals on the same *physical LAN* but different *virtual LANs* can communicate only through some mutually visible router.
- The special feature of a *switch* is that it connects *workstations* (possibly through hubs) and may aggregate them into one or more *LANs*, whereas a *bridge* connects *LANs*. In simulating a LAN, a switch includes many of the features of a bridge, such as address learning and selective forwarding (and perhaps router functions as well, such as monitoring protocols).
- Finally, DO NOT confuse Virtual LANs with Virtual Circuits.

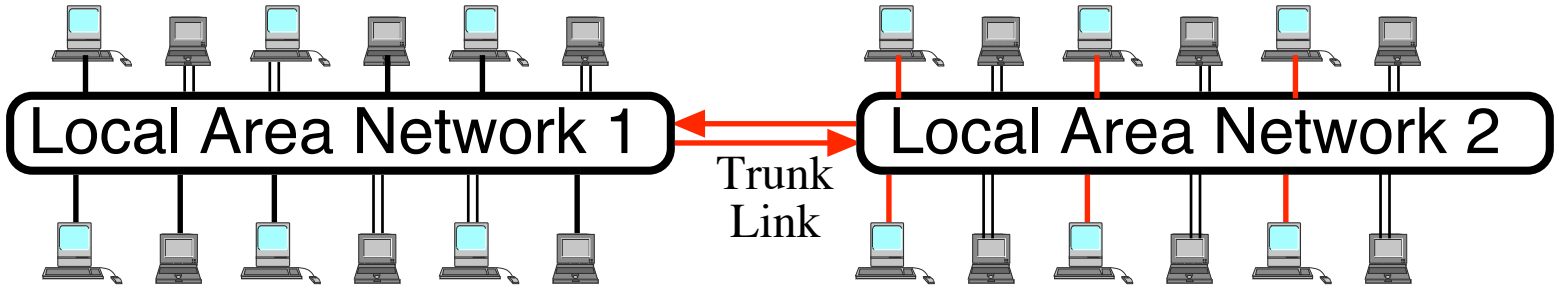
Defining a VLAN

At the simplest level a VLAN is simply the list of switch ports associated with the VLAN. For example ports 1, 2, 4, 5, 6 may be on one VLAN and ports 3, 7, 8 may be on another VLAN.

But VLANs may be distributed across several switches (bridges), connected by "trunk links" with a trunk link shared by several VLANs.

Messages between switches therefore include a *tag* sub-header (indicated by an Ethertype code), which includes a *VLAN identifier* and a priority.

Equipment that is "VLAN aware" can examine the tag and direct the message appropriately.



The GARP VLAN Registration Protocol.

The switches in a virtual LAN exchange GARP messages (*GARP = Generic Attribute Registration Protocol*) and GVRP (*GVRP = GARP VLAN Registration Protocol*) to propagate VLAN information through the physical network and establish VLAN connectivity, much as the Spanning Tree protocol establishes physical connectivity.

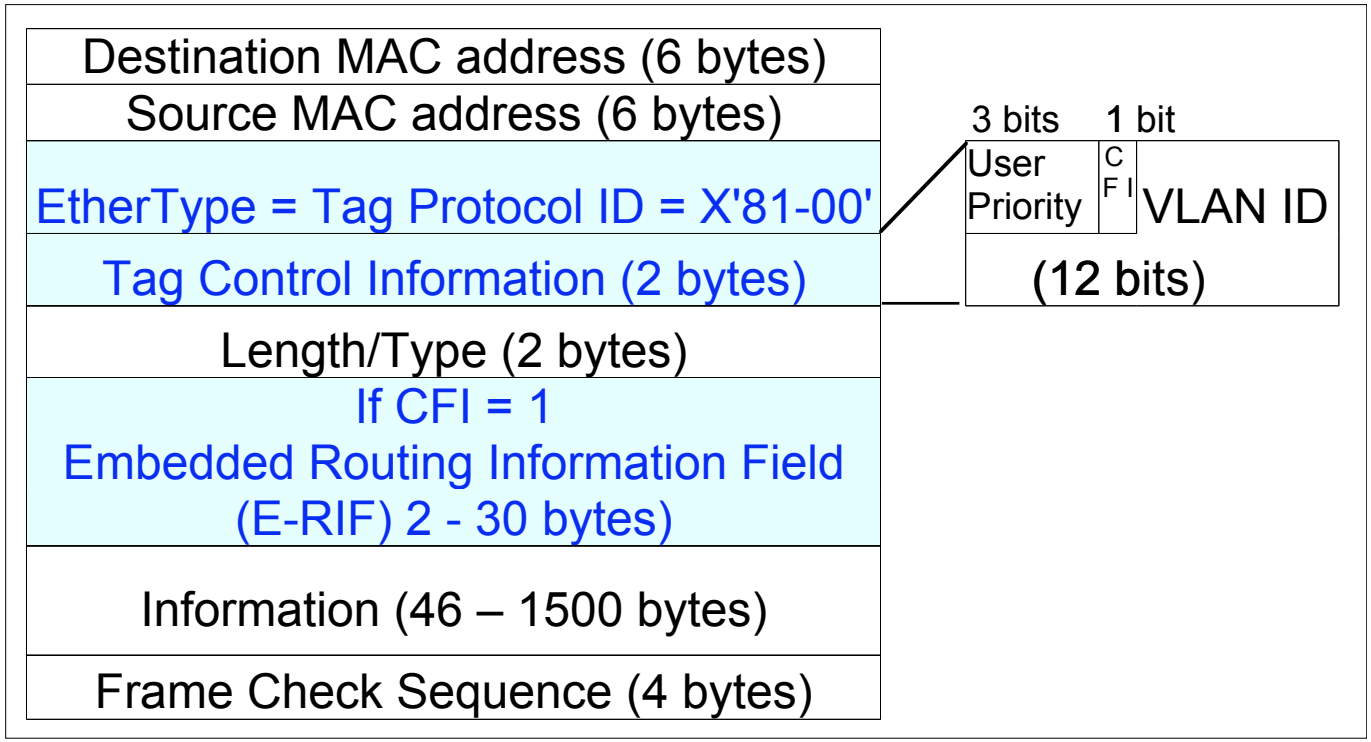
VLAN membership may be defined by —

- Default – Initially all ports belong to the same default VLAN
- Port-based – at each switch list the ports belonging to the VLAN
- Protocol based – select VLAN according to protocol type
- IP address based – select VLAN according to IP addresses
- Port and GVRP based – define VLAN through ports and GVRP

VLAN Message Tags

The Virtual LAN message format expands the Ethernet/802.3 frame, using the fields shaded in the diagram. (*This expansion may be beyond the usual 1500 octet limit but is private to the VLAN switches and used only for communication between VLAN switches*.)

The E-RIF field is used for source routing, or when combining different LANs in a single network.



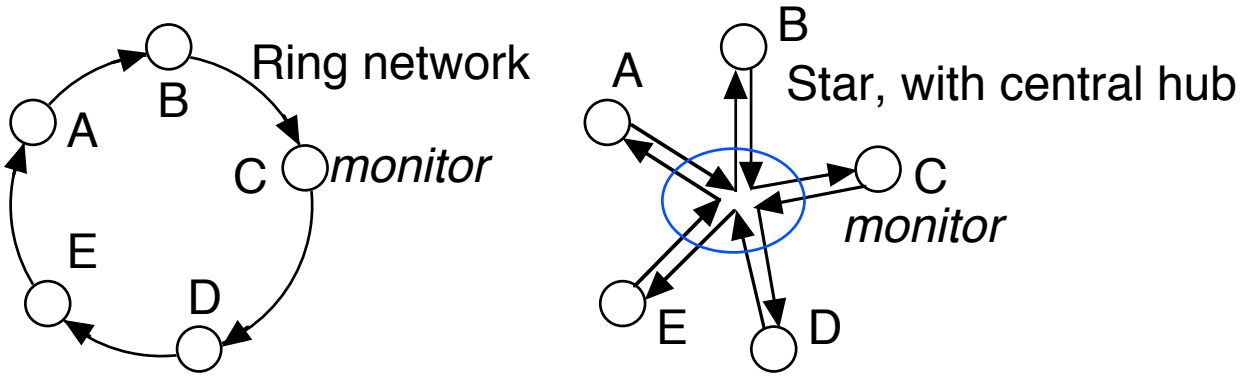
Token Networks

- All networks have user stations *competing* or *contending* for access to the network; only one can use the network at a time.
- In Ethernet it really is a battle; the stations fight it out and eventually one wins.
- Another approach is to pass a *token* or right to transmit around the network in an orderly fashion. Only the station with the token can actually send data. Others must wait until they receive the token.
- There are two approaches —

1. A *token ring* has all the stations or nodes in a physical ring, so that each node is connected to only two neighbours. All communication requires intervening nodes to relay or forward traffic.
2. A *token bus* physically resembles an Ethernet. Nodes form a logical ring based on their physical addresses. Most receive the token from the immediately higher address and send it to the immediately lower address.

Token Ring (IEEE 802.5)

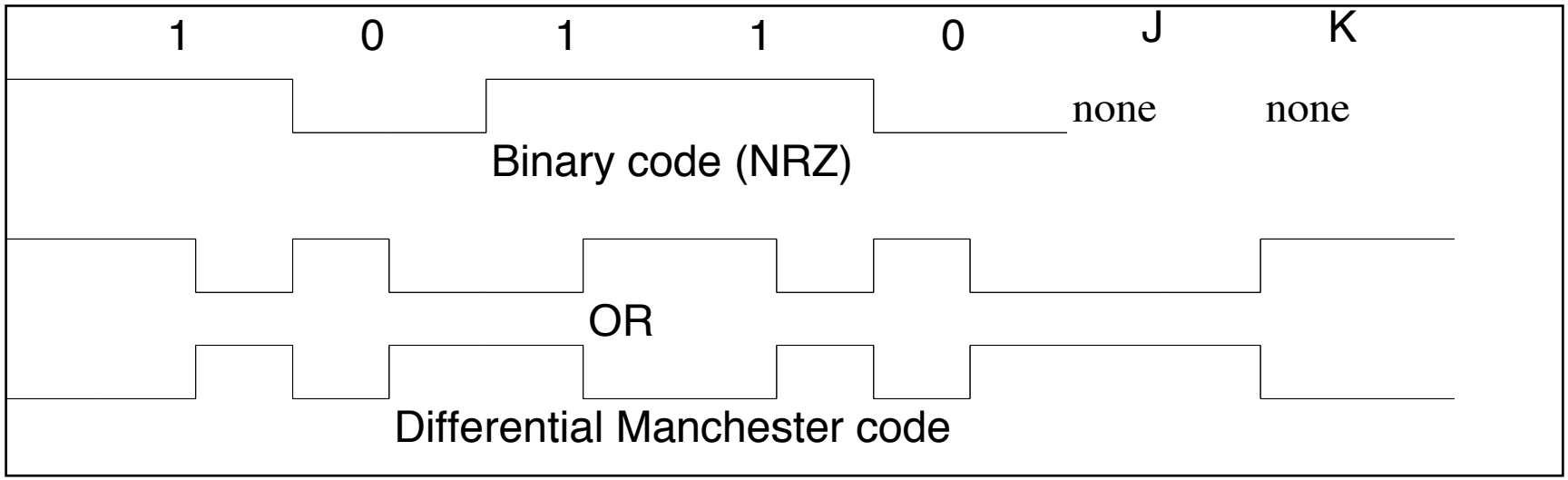
- There are two physical configurations, the obvious physical ring, and a star with central hub, which makes for easier control and maintenance.



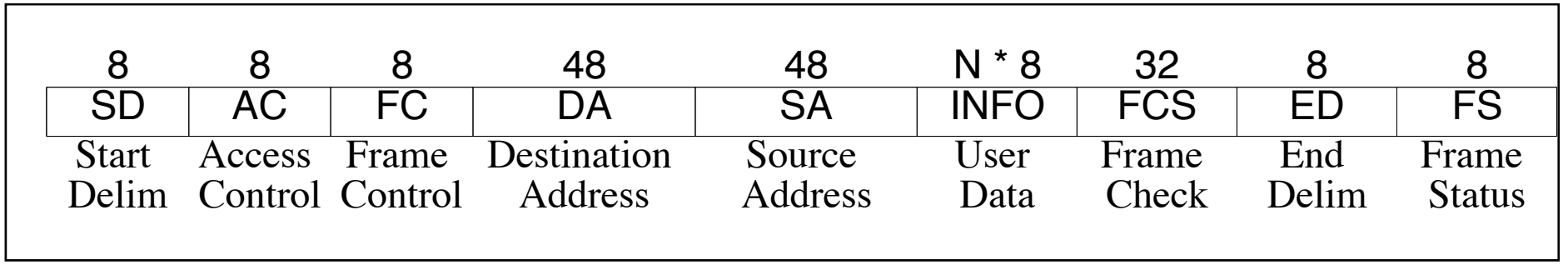
- One node becomes the *active monitor* which, apart from its normal communication, checks that the ring is operating correctly and supplies the bit-clock for the ring. All other nodes are *standby monitors*.
- As each node receives from only its up-stream neighbour there is no need for a preamble to synchronise clocks.

Token ring physical

- The ring uses differential Manchester coding, at 4 Mbps or 16 Mbps, with special non-data signals to mark frame boundaries.
- Differential Manchester coding has a transition at the middle of every bit cell, with a transition at the start of every 0 data bit and no transition at the start for a 1 data bit.
- A non-data J maintains the preceding level for 1 bit time, while a non-data K has an initial transition; neither has the mid-cell clock.
  J and K "bits" always occur in pairs.



Token Ring Frame format



- Start Delimiter is `JK0JK000`
- End Delimiter is `JK1JK1IE` I = intermediate frame, E = Error detected
- The sequence SD ED is an "abort sequence" to stop current activity
- Destination and Source Addresses are as in 802.3 (usually 48 bits, 16 is allowed but seldom used).
- Frame Check Sequence (FCS) is similar to IEEE 802.3, covers FC to FCS
- Frame Status 8 bits `ACrrACrr` A= Address recognised; C = frame copied

Access Control Octet

- This has the 8 bits — 

- PPP and RRR are used in priority control (see later).
- T is the Token Bit and is 1 for normal data and 0 for a token
- M is the Monitor bit. A token is always generated with M=0; the Monitor station will set M=1 whenever it sees a high-priority token of frame and will abort a frame if it receives M=1.
- A minimal frame **SD AC ED** is known as a *token* and circulates in the absence of other traffic.
  Any station wanting to transmit will wait for a token, seize it (priorities permitting) and transform it into an information message.
  When the message returns to the sender, the sending station will create a new token and send it on to the downstream successor.

Frame Control FC

- The Frame Control octet is — 

- The FF bits may be
  - 00 MAC (Medium Access Control) frames
  - 01 LLC (User, or other LLC data)
  - 1x reserved

The LLC frames are generated by the LLC layer, or passed on from a user.

The MAC frames coordinate operation of the ring itself, particularly protecting against failure and restarting if necessary.

The Active Monitor station

1. Generates bit timing for the ring
2. Marks frames which pass, removing them if they circulate unread
3. Puts a 24-bit buffer in the ring so a token doesn't "eat its tail"
4. Regularly generates AMP frames to check ring integrity

Don't learn Token Ring past here.

The MAC frames include

- CL_TK **Claim Token** When the ring appears to be idle (no *x*MP frames for a while), each station sends CL_TK frames with its own address, stopping if it receives a CL_TK from another station. When it receives its own CL_TK it becomes the Active Monitor
- DAT **Duplicate Address Test** Each station starting up will send a DAT with its own address. If it receives the frame with A=1 in the FS octet it shuts down because there is another station with the same address.
- AMP **Active Monitor Present** About every 3 seconds the Active Monitor sends an AMP message
- SMP **Standby Monitor Present** About 10 ms after receiving an AMP or SMP message a Standby Monitor will send an SMP message to its downstream neighbour.
- BCN **Beacon** This is used in fault recovery

Token Ring Priorities

(Not important, as Token Ring is nearly obsolete — Shay pp 379–392).

- The P field of the AC octet contains the current priority of the token (a frame is an "active token"!). The **R** field contains a request for operation at priority **R**.
- The station has a queued PDU with priority **Pm**, waiting for transmission. If **Pm**>0, it sets **R**=**Pm** on the next token if **R**<**Pm**. This will request operation at the higher priority.
- When a token is received the P and R fields from its AC are copied into the station's **Pr** and **Rr** registers.
- If when new token is generated, **Rr** > **Pr** or **Pm** > **Pr**, its priority is set to **Pn**=max(**Rr**, **Pm**). At the same time, **Pr** is saved in a stack as **Sr**, and **Pn** as **Sx**.
- Having raised the priority, a station can send PDUs at or above the ring priority **R**. When all have been sent, it can restore priorities from the stack.

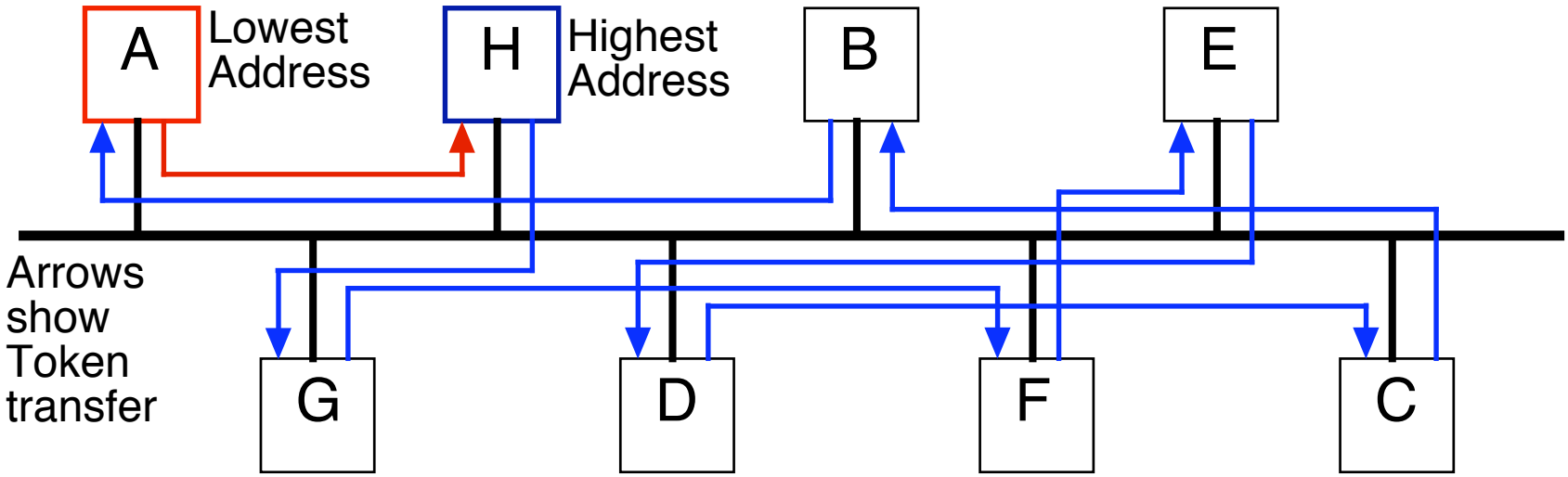
Token Ring Timers

A Token Ring is controlled by several timers, the more obvious of which are

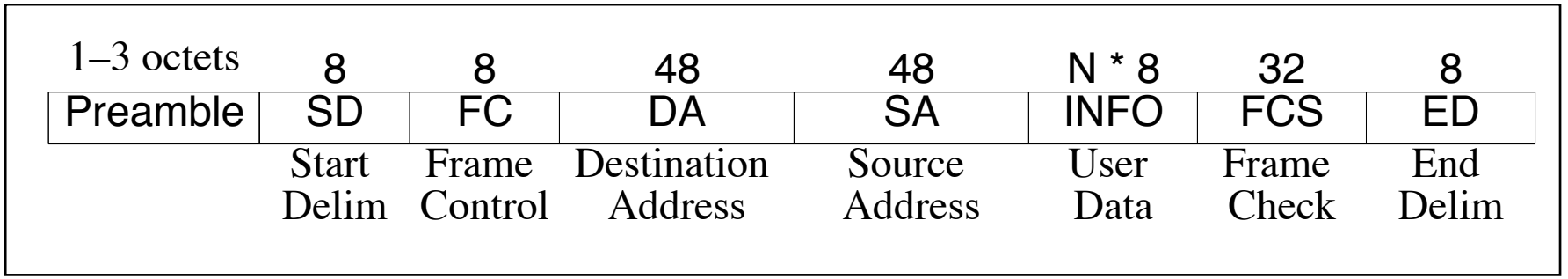
- Timer, Holding Token (**THT 10 ms**) limits the time that a station can transmit before it must release the token
- Timer, Queue PDU (**TQP, 10 ms**) is the time before forwarding an SMP
- Timer, No Token (**TNT 1 s**) The ring is assumed in trouble if no token is seen for TNT.
- Timer, Active Monitor (**TAM, 3 s**) the time before sending an AMP frame
- Timer, Standby Monitor (**TSM 7 s**) used by Standby Monitors to detect an idle ring

IEEE 802.4 Token Bus

- This was designed for manufacturing and production lines, and offers guaranteed response times.
- Physically the network resembles classic Ethernet, with stations tapped onto a single multi-dropped cable.
- The stations are connected into a *logical ring*; all except the lowest address station pass the token to the next lowest address.
- The token is a short message passed between stations, with source and destination addresses.



Token Bus Frame



- The frame resembles 802.5 Token Ring, including delimiters with "non-data" codes. The delimiters are similar to those of 802.5, but are described differently.
- Because all stations are independent, the transmissions have no agreed bit timing and a preamble is necessary (but shorter than 802.3)
- Addresses may as before be 16 or 48 bits, but 16 bits are seldom used and may be ignored.

FC (Frame Control Field)

- This defines management frames (much as Token Ring), and user data

1	2	3	4	5	6	7	8	Bits, transmission order
0	0	0	0	0	0	0	0	Claim Token
0	0	0	0	0	0	0	1	Solicit_Successor_1
0	0	0	0	0	0	1	0	Solicit_Successor_2
0	0	0	0	0	0	1	1	who_follows
0	0	0	0	0	1	0	0	Resolve Contention
0	0	0	0	1	0	0	0	Token
0	0	0	0	1	1	0	0	Set Successor
0	1	M	M	M	P	P	P	LLC Data Frame
1	0	M	M	M	P	P	P	Station Management

Token

This is the token which is passed between stations; only the station with a token has the right to transmit.

Solicit_Successor_1

Each station knows its *predecessor station* (from which it *receives* a token) and its *successor station* (to which it *forwards* the token). Periodically each station transmits a Solicit Successor frame; if any "new" station has an address in the range

`sender_address > stn_address > successor_address`.

it replies to the sender with a Set_Successor, naming itself as the successor.

Solicit_Successor_2

Like Solicit_Successor1, but sent by the station with the lowest address (`successor > self`); stations reply only if *outside* the address range.

Don't learn Token Bus past here.

Who_follows

This used in ring maintenance if the token seems lost. With the ring order C→ B → A, if C sends the token to B and sees no response, it will try again and then send a broadcast "Who-Follows B". A (whose predecessor is B) then replies to C "Set_Successor to A", removing station B from the logical ring.

Set Successor

See under Who-Follows above.

Resolve Contention

Used to resolve the confusion if more than one station responds to a Solicit_Successor.

Claim Token

When the ring starts up all stations emit Claim Token frames, including variable delays. Any station which sees another Claim Token drops out, until only one survives.

Times

Priorities (0, 2, 4, 6) are controlled by a system of *Token Rotation Times*.

- TRT_i Token Rotation Time for priority *i*. This timer is started when a station sends a token at priority *i*, and noted when the token is received by the originating station
- THT Token Holding Time. The maximum time for which a station may send Class 6 frames.
- TTRT Target TRT The maximum permitted Token Rotation Time.
- A station after receiving the token may transmit (class 6 priority messages) for max(THT, TTRT–TRT).
- If all class 6 frames are sent, it may then send Class 4, then 2 and 0 frames until THT expires.
- A station should avoid sending a frame which cannot be completed, OR must abort the frame and send SD ED as soon as the time is exceeded.