

The Internet Protocol and Packet Routing Fundamentals

Second Semester: 2007-2008

Dr. Rahul Banerjee

Associate Professor, Computer Science Group

Birla Institute of Technology and Science, Pilani - 333 031, INDIA

E-mail: rahul@bits-pilani.ac.in

Home Page: <http://discovery.bits-pilani.ac.in/rahul/>



Interaction Goals



- **Introduction to Packet Routing / Switching**
- **Connectionless versus Connection-oriented Routing / Switching**
- **Introduction to the Internet Protocol**
- **IPv4 and IPv6**
- **A comparative study of representative Routing Schemes and Protocols**
- **An Understanding of Current Industry Practices, Evolving Trends and Research Directions**

Introduction



- **Packet Handling in Datagram Networks**
- **Cell Handling in Virtual Circuit-based Networks**
- **Packet Handling in TCP/IP versus ATM**
- **Comparison**
- **Our Focus**

Network Layer: Recap



- **Network Layer is a layer of the Network Architecture that is primarily concerned with getting NLDU / Packets from the source node and delivering it to the intended destination node (through none or many intermediate nodes).**
- **Additional responsibilities of this layer include:**
 - **Providing support for connection-oriented / connectionless services as the case may be (depending upon the protocol stack and need)**
 - **Provide diagnostic support for network monitoring, configuration, management and trouble-shooting at the Network Layer or higher layer.**
- **Packet handling, packet management, Routing are its major responsibilities.**

Network Layer Goals



- In the context of packet routing, network layer structural design goals include:
 - Ensuring the shortest possible delay and thereby the highest throughput at the least possible cost
 - Ensuring acceptably reliable packet delivery (may be optional in some cases)
 - Ensuring secure packet delivery (may be optional in some cases)

Major Issues Related to the Design of the Network Layer



- Choice of the Services to be provided by the Network Layer to the Transport Layer and their nature:

The primary issue:

- Choice of any one of the Connection-oriented or Connectionless types of service to be provided by the Network Layer.

Major Issues Related to the Design of the Network Layer ..



- Services to be provided by the Network Layer to the Data Link Layer and their nature:
 - The primary issue:
 - The format and size of the Network Layer Data Unit (e.g. Packet or Cell) in which the data is to be passed to the Data Link Layer (if it does exist) with or without encapsulation and additional information.
 - It needs to be noted here that in practice, the Network Layer and the lower layers (e.g. DLL and PL) are not really totally transparent to one-another. Degree of transparency is, therefore, another major issue, which is seldom discussed in literature.

Major Issues Related to the Design of the Network Layer ..._



- Architecture and internal organization of the Network Layer so as to be able to meet the N.L. design goals:
 - In this case, the real issue is ‘how to provide the mechanism that would satisfy the primary design goals of the designated Network Layer.
 - A possible example involves making choice of PVMs or SVMs or Datagrams (again reliable ones or the default unreliable ones?) or a mix of them in a given situation.

Major Issues Related to the Design of the Network Layer ..._



- Choice of Interior and Exterior Routing and Protocol Translation schemes.
- Choice of Security to be provided at the subnet level.
- Choice of conventional / mobile / hybrid routing support.
- Choice of support for QoS, FTRT etc. for the intended class of applications.

Packet Routing Problems



- Loss of packets
- Receipt and circulation of duplicate packets
- Packet Choking / Network Congestion
- Network Cleansing
- Worst-case upper bound problem
- QoS negotiation
- Failure Handling
- Quick Recovery Requirement
- Route Tracing
- Network Management Support

Static Packet Routing Schemes



Shortest Path Routing:

- This is one of the simplest routing schemes and the primary technique involved here is the determination of the shortest available path between a source and a destination.
- The term shortest path may be interpreted in a variety of ways including:
 - path of the least geographical distance
 - path of the least congestion
 - path of the least number of Hops
 - path of the least mean queuing delay
 - path of the least propagation / transmission delay

Any weighted average based metric can be yet another choice for employing this scheme.

Dijkstra's Shortest Path Routing Algorithm



- One of the best known algorithms that may be employed for determination of the shortest path is the one suggested by Dr. E. W. Dijkstra in as early as 1959. The gist of this strategy is given below.
- 1. Each node is labeled with the name of the source node and its distance from the current node. Normally, the labeling is done in the reverse order, i.e. the label (9, A) represents distance of the current node from the source node (9) followed by the name of the source node (A).

A label may be permanent or tentative.

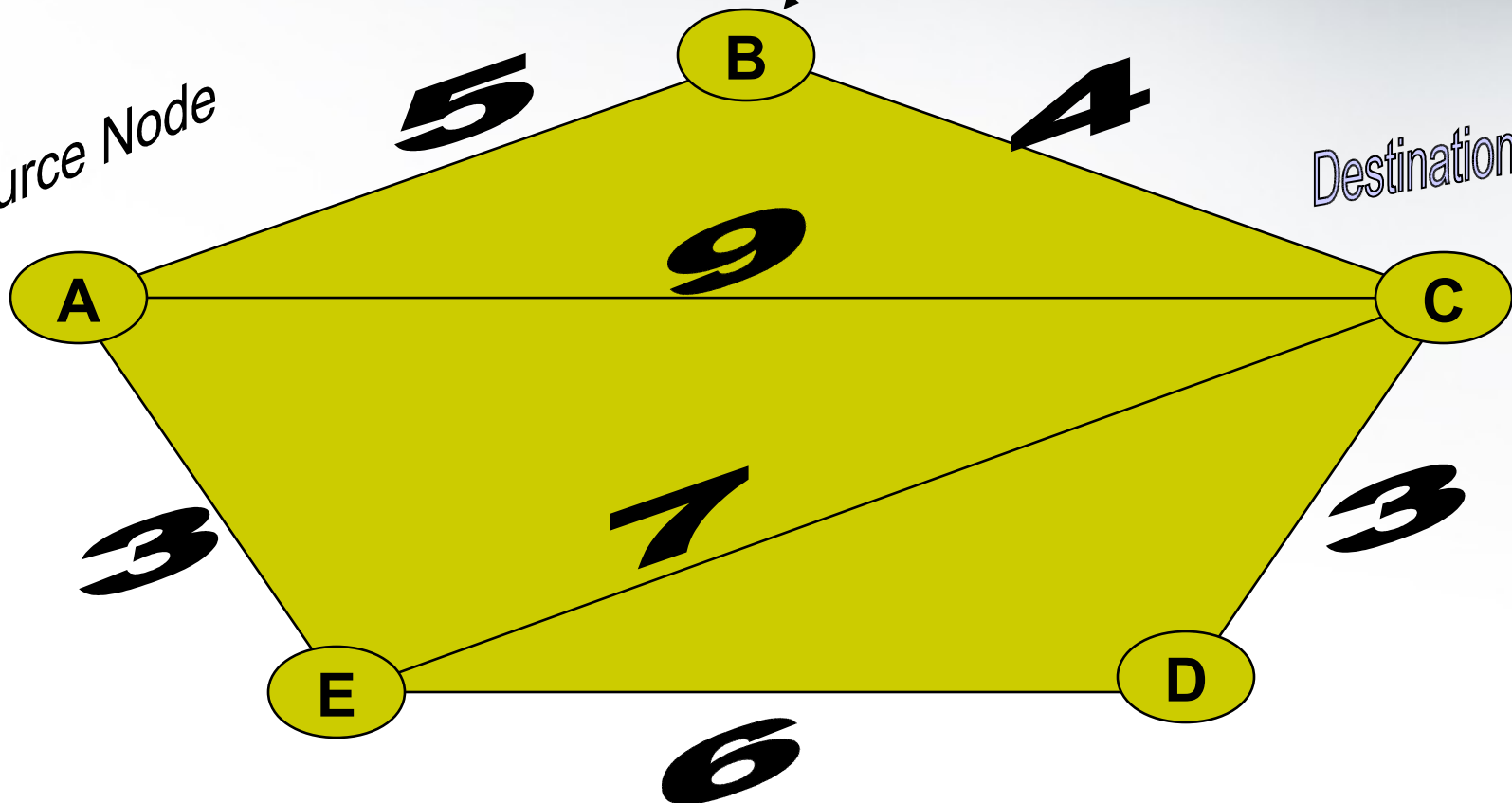
Dijkstra's Algorithm ...



Initially, all nodes, except A, shall be labelled as (Infinity, A)
Label of this node may be: (5, A)

Source Node

Destination Node



Dijkstra's Shortest Path Routing Algorithm



- 2. At the start of the algorithm all nodes are labeled tentatively.
- 3. As the algorithm progresses, the labels may change.
- 4. At any stage, when it becomes clear that the current label represents the smallest distance / shortest path between a node and the source node, former's label is marked as a permanent label.
- 5. As the algorithm progresses, more and more nodes acquire permanent labels.
- 6. The algorithm terminates when the destination node gets a permanent label.

Flooding-based Routing Schemes



- There exist three major variants:
 - Pure / Unconstrained Flooding
 - Hop-Count Based / Constrained Flooding
 - Selective / Direction-Constrained Flooding

Each of these types finds a brief description in the following slides.

Hop-Count-based Flooding Algorithm



– This algorithm may be expressed as follows:

1. At any originating node 's', structure a packet such that its header contains a 'hop count' that be initialized to length of the path (if known) or full diameter of the subnet.

Hop-Count-based Flooding Algorithm



2. At every intermediate node 'i' examine the incoming queue of packets, take the packet at the head of the queue and note the packet-id, line on which it arrived on, its hop count and destination address.
3. Decrement the hop count by one (1).
4. If the count becomes zero, discard / drop the packet and flush the corresponding entries in the local table. Otherwise, generate (n-1) replicas of the packet (where 'n' is number of arcs converging at this node) and transmit one replica on all arcs / lines except the one this packet arrived on.
5. Examine the incoming queue and if it is non-empty, repeat steps 2 to 5 else wait until a new packet arrives and then repeat steps 2 to 5.

Selective / Direction-Constrained Flooding Algorithm



- It is a variant of the basic Flooding Algorithm with the constraint of direction thrown in for the purpose of improved efficiency.
- In this scheme, packets are selectively flooded by the routers in such a way that they move approximately in the right direction (i.e. leading towards the Destination).

Flow-based Routing Algorithm



- This is yet another Static Routing Algorithm; but unlike the Shortest Path based Routing Algorithm and the Flooding based Routing Algorithm, which primarily consider the Subnet Topology alone, it considers Subnet Topology as well as Load (Traffic).
- This is particularly suitable for the subnets characterized by nearly stable average data transfer rate / mean data flow rate.
- In other words, this scheme may not prove to be effective if the mean inter-node data flow in a given subnet cannot be reliably predicted / estimated.

Flow-based Routing Algorithm



- This algorithm, unlike the other algorithms discussed so far, has several pre-requisites including the following:
 - Subnet topology must be known in advance.
 - Link / Line Capacity Matrix must be known in advance.
 - Traffic Matrix must be available a priori.
 - Mean packet-size must be known.
 - Some preliminary Routing Algorithm must be available.

Flow-based Routing Algorithm



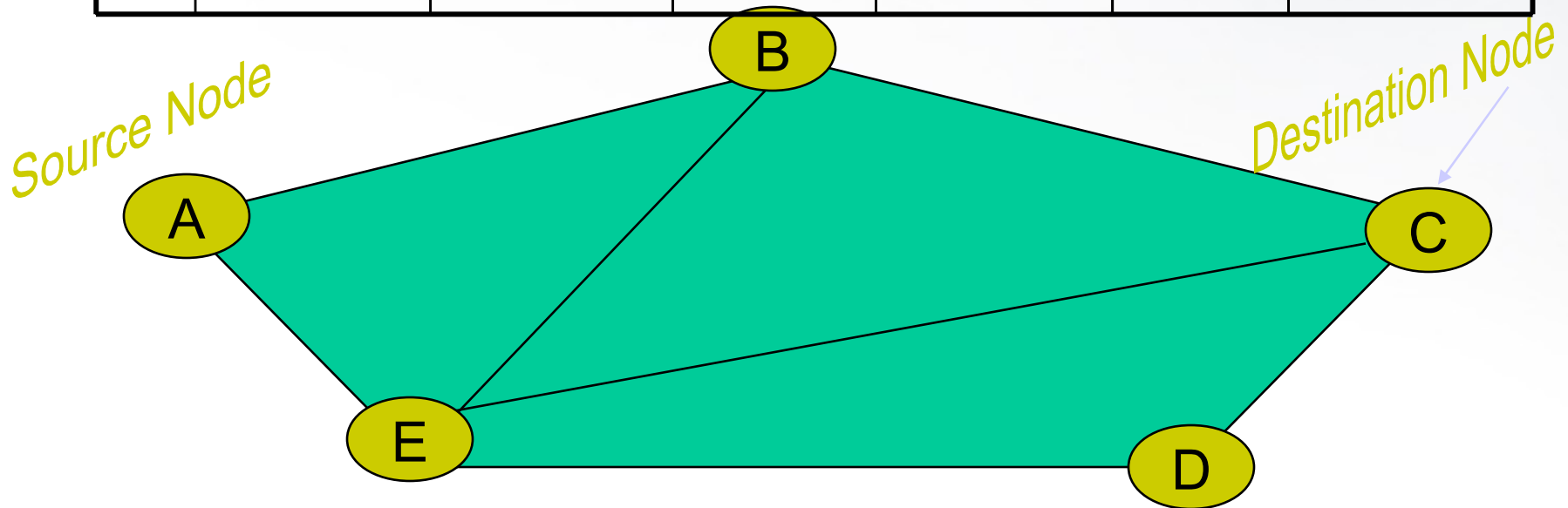
- The scheme makes use of the fact that under the above stated circumstances, for each of the links, if the link-capacity, average rate of data-flow and topology are known and if the traffic -matrix and subnet topology is available in advance, then it is possible to:
 - 1. compute the mean delay in packet-delivery per link,
 - 2. Compute the mean (overall) delay in packet-delivery over the given subnet,
 - 3. Compute the most appropriate route between any pair of Source and Destination .

Routing Table

Another example subnet



Sr No	Link-Id	Link-Traffic	Link-Speed	Link-Capacity	Mean Delay on the Link	Link-Weight
1	AB	12 pps	400 kbps	... pps	... ms	...
2
..						



Bellman-Ford Distance Vector Routing Algorithm



- This is also known as the Ford-Fulkerson Routing Algorithm.
- It is the original Dynamic Routing Algorithm used in the erstwhile ARPANET.
- For quite some time, it was popular over the Internet where a variant of it called Routing Internet Protocol (RIP) was used.
- Many Routers still use one or other variation of this algorithm.

Bellman-Ford Distance Vector Routing Algorithm ...



- **In brief, this scheme may be expressed as:**
 - Each Router knows / discovers its distance from its neighbours.
 - Each Router locally maintains a Routing Table indexed by an entry for every other Router in the subnet and identification of a preferred neighbour / link leading to that Router.
 - Metric of estimation may vary. For instance, it may be any one of Physical Distance, Hops, Delay etc.

Bellman-Ford Distance Vector Routing Algorithm ...



- Periodically, each Router sends a Vector to its neighbouring Routers. As this vector contains estimated distances, it is called a Distance Vector.
- On receipt of such Vectors from its neighbours, every Router recomputes its estimates and updates its local routing table.

Figure on the following slide presents an example subnet and a sample distance vector generated at one of its nodes.



An example subnet

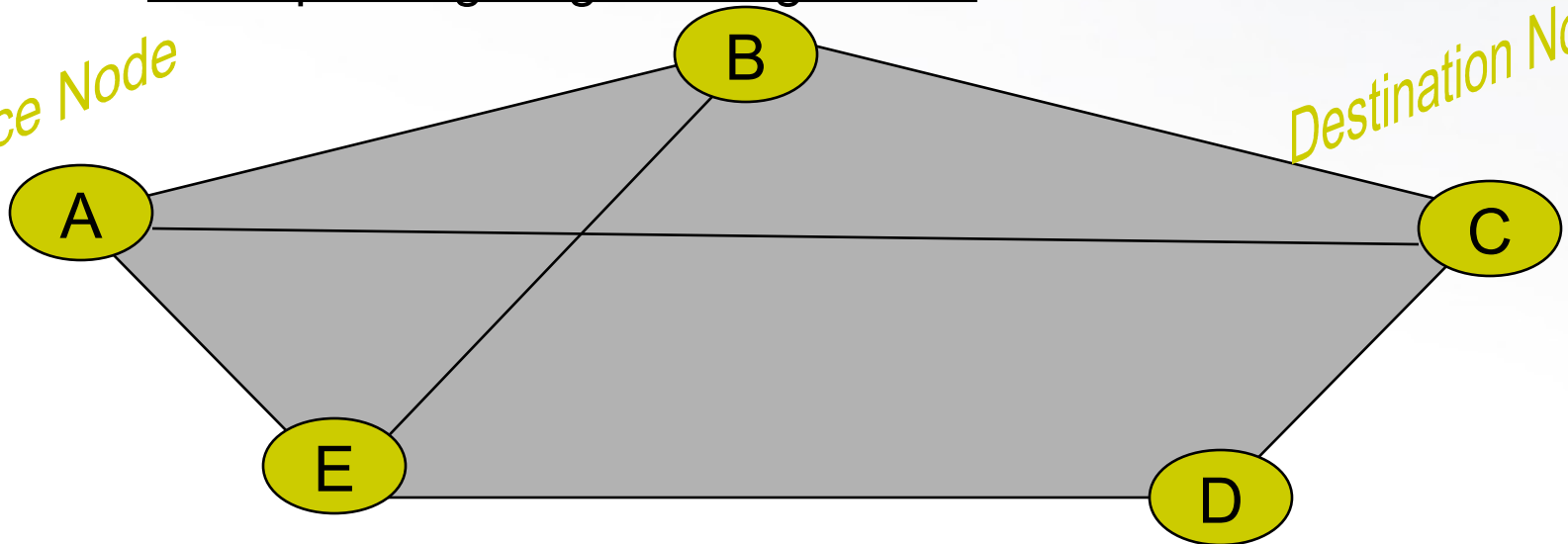
A typical Distance Vector for a Router 'A' may look like:

2	B
13	C
5	E

For the given subnet

where, the first column indicates Current Estimates and the second column refers to Identification Symbol for the corresponding neighbouring Router.

Source Node



Bellman-Ford Distance Vector Routing Algorithm ...



- For the given subnet, the Routing Table at the Router A might look like:

<i>Destination</i>	<i>Distance</i>	<i>Next Hop Via Router</i>
A	0	A
B	2	B
C	6	D
D	7	D
E	3	E

Dynamic Packet Routing Schemes ...



- Bellman-Ford Distance Vector Routing Algorithm ...
 - Novell's well-known IPX used this scheme for quite a while.
 - The primary drawback of this algorithm is its vulnerability to the 'Count-to-Infinity' problem.
 - Another drawback of this scheme is that it does not take into account Link Bandwidth.
 - Yet another problem with this algorithm is that it takes appreciably long time for convergence.

Dynamic Packet Routing Schemes ...



- Link-State Routing Algorithm:

In this algorithm, each router:

- Discovers its neighbours and their Network Addresses by sending special packets called 'Hello' packets.
- Estimates delay / cost or any other metric for reaching its neighbours by sending another special packet type called 'Echo' packets.

Dynamic Packet Routing Schemes



Link-State Routing Algorithm ..

In this algorithm, each router ...

- Immediately applies its recent knowledge to form Link-state packets which encapsulate this estimate; and, sends copies to all the discovered neighbouring routers.
- Computes the shortest path to every other router using the Shortest Path Algorithm.

Dynamic Packet Routing Schemes



Link-State Routing Algorithm ..

- In this case, fresh link-state packets are built:
 - periodically or
 - upon occurrence of an event like node-failure / link-failure / addition of a node or link / revival of a failed node or link.
- The algorithm requires that names / identifiers representing the routers be unique (globally).



Dynamic Packet Routing Schemes...

Link-State Routing Algorithm ...

A typical Link-State Packet for a Router 'A' may look like:

A	
1101..1100	
60	
B	7
E	5

where, the first row indicates the Originating Router, the second row refers to the Sequence Number (usually a 64-bit or higher number) of the link-state packet, third row shows the Age of the packet, the fourth and subsequent rows indicate estimated metrics for each of the neighbouring routers (B and E in this case).

Dynamic Packet Routing Schemes ...



Link-State Routing Algorithm ..

- **Data Structure for the Packet Buffer at the Routers has the format:**

Source	Sequence No.	Age	Send Flags	Acknowledgement Flags	Data
--------	--------------	-----	------------	-----------------------	------

- **Examples of some of the well known implementations of this scheme include:**
 - **Open Shortest Path First (OSPF) scheme**
 - **Intermediate System- Intermediate System (IS-IS) scheme**

IP: What is it?



- **IP stands for the Internet Protocol**. It is the network layer equivalent in the TCP / IP stack.
- IP has two prominent versions the IPv4 (which was designed in keeping with the technologies of the early Seventies) and the IPv6 (which is the latest version).
- The IPv6 **does away with some features of the IPv4 while adds many new features**.
- One **basic advantage of the new version is the enlarged address space** (adequate for a reasonably long time).

Major Goals of the IPv4 Design:



- Simplification of the Network Layer functionality across an internetwork.
- Reduction in the packet processing time at the routers.
- Providing support for an acceptable scheme of addresses.
- Reduction in **the size of** Routing Tables.



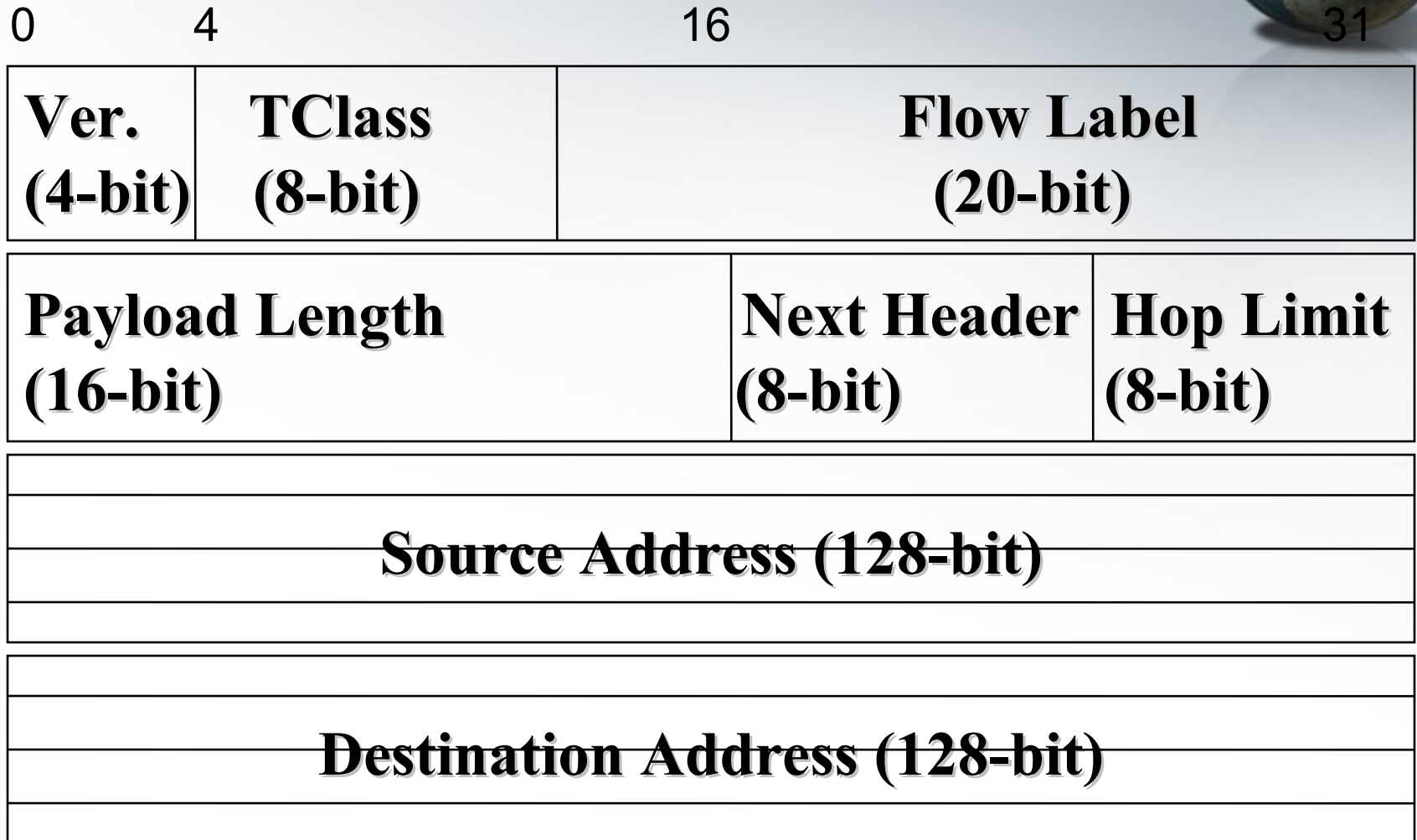
IPv4: The Header Structure

0

31

Ver.	IHL	Type of Service	Total Length
Identification		Flags	Fragment Offset
TTL	Protocol Type	Header Checksum	
Source Address (32-bit)			
Destination Address (32-bit)			
Options+Padding			

IPv6: The Header Structure



What is different in IPv4 and IPv6?



1. Options were replaced by Extension Headers.
2. IPv6 has a Flow Label field in its header primarily meant for supporting the real-time applications.
3. The Total Length field of IPv4 was replaced with the Payload Length field.
4. The Protocol Type field of the IPv4 was replaced with the Next Header Type field.
5. The Time-To-Live field of the IPv4 was replaced with the Hop Limit field .

IPv6 Versus IPv4 ...



6. Autoconfiguration capability supported for the first time.
7. Built-in provisions for security made available for the first time explicitly.
8. Support for Jumbograms has been added.
9. Both permit Fragmentation, but the IPv6 format keeps it in an extension header specifically meant for the job unlike the IPv4 format in which this information was to be maintained in a fixed field within the IP header.
10. The Service Type field has been replaced by the) Flow Label field.

IPv6 Versus IPv4 ...



11. The IPv6 header has no Header Checksum field.

12. In IPv4, there were five address classes (A to C of Network / Host combination types, D for Multicasting and E reserved for future use).

In IPv6, the IPv4 Classes have been replaced with Types. 13. Unlike the IPv4, that permits a two-level hierarchy of network and host prefixes, the IPv6 proposes to offer multi-level hierarchy or even multiple hierarchies of prefixes. In IPv6, the first byte of the address refers to the type of address.

IPv4 Options:



- **Security Option**
- **Strict Source Routing Option**
- **Loose Source Routing Option**
- **Record Route Option**
- **Timestamp Option**

IPv4 and the World of Classes:



- In the IPv4, any address is 32-bit long and is represented in four parts of one byte each separated by decimal points or dots.
- There exist two ways of looking at the IPv4 world:
 - Class-based view (A, B, C, D, E)
 - Classless view
- The 32-bit address comprises of two parts:
 - Network address / identifier
 - Host address / identifier

IPv4 and the World of Classes ...



- In the class-based / classfull version, the classes are designated based on the first few bits of the Network Address portion of the IP address.
- For instance:
 - If the first bit in this field is 0 (zero), it is referred to as a Class-A IP address.
 - If the first two bits in this field are 10 , it is referred to as a Class-B IP address.
 - If the first three bits in this field are 110 , it is referred to as a Class-C IP address.



IPv4 and the World of Classes ...

- In Class-A address, the first byte constitutes the Network Address and remaining three bytes constitute Host Addresses.
- In Class-B address, the first two bytes constitute the Network Address and the remaining two bytes constitute Host Addresses.
- In Class-C address, the first three bytes constitute the Network Address and the remaining byte represents the Host Addresses.



IPv4 and the World of Classes ...

- Example of a Class-A address:

12.0.0.3

Here, 12 is the Network Address whereas 3 is the Host Address.

(Technically, this means: Network Address is 12.0.0.0 and the Host Address is 0.0.0.3.)

- Example of a Class-B address:

180.16.0.1

- Example of a Class-C address:

192.12.7.8

IPv4 and the World of Classes ...



- Class-A Address Range:
1.0.0.0 - 127.255.255.255
- Class-B Address Range:
128.0.0.0 - 191.255.255.255
- Class-C Address Range:
192.0.0.0 - 223.255.255.255
- Class-D Address Range:
224.0.0.0 - 239.255.255.255
- Class-E Address Range:
240.0.0.0 - 247.255.255.255

Concept of Subnetting and Subnet Masks



- An IPv4 Subnetwork is often referred to mean a subset of one of the three IPv4 classes (A, B and C).
- A Subnet Mask is a sequence of bits that is used to separate Network and Host Addresses from each other. This mask divides the Address portion into another set of Network-Host Addresses.
- **Types of Subnetting:**
 - Fixed-Length Subnetting / Basic Subnetting
 - Variable-Length Subnetting
- **Types of Masks:**
 - Natural Masks
 - Extra-natural Masks

More on Masking ...



- Natural Mask for Class-A: 255.0.0.0
- Natural Mask for Class-B: 255.255.0.0
- Natural Mask for Class-C: 255.255.255.0
- Why masks?
 - They help in separating network address from the host address.
 - They make subnetting possible.
 - The subnetting in turn helps in fighting the IPv4 Address Depletion problem in some limited but effective way.
 - Every LAN segment is usually associated with at least one network number (more are possible) and if no subnetting is done, only one segment can use a given network address.

More on Masking ...



- Variable Length Subnet Masking

In this scheme, a given network can be masked with masks of different lengths thereby providing required flexibility of having network segments as required (instead of just dividing a given network into 'n' number of networks of equal sizes ---- as is the case with the fixed-length subnetting).

(All masks have a string of '1's to the left and string of '0's to the right.)

The Classless Inter-Domain Routing (CIDR) in IPv4 Subnets:



Primary Objective:

- Finding a temporary solution to the IPv4 Address space depletion

The basic idea behind the CIDR:

- Allocate the unallocated set of Class-C IPv4 network addresses in variable-sized address blocks.
- These blocks, in effect, refer to contiguous Class-C IPv4 network addresses.

The Classless Inter-Domain Routing (CIDR) ...



The RFC 1519 allocation rules for the IPv4 world:

1. The whole world was suggested to be divided into four zones each of which could use nearly 32 Million Addresses:
 - Asia-Pacific:
 - Central-Southern America
 - Europe
 - North America

The Classless Inter-Domain Routing (CIDR) ...



The RFC 1519 allocation rules for the IPv4 world ...

2. A set of nearly 320000000 addresses were suggested to set aside for future use.
3. If a router 'X' get a packet that belongs to the IPv4 addresses of one these four zones, the packet is simply forwarded to the zonal gateway.

The Classless Inter-Domain Routing (CIDR) ...



The Supernetting:

- Terms ‘Aggregation’, ‘CIDR Block allocation’, ‘Supernetting’ etc. are often used interchangeably in the IPv4-CIDR literature. (This is however, done in casual discussion alone!)
- In principle, ‘a network whose prefix-boundary has lesser number of bits than the natural mask of the network itself, is called a Supernet’.

Two ways to represent the same CIDR address are :

- 199.28.0.0/16
- 199.28.0.0 255.255.0.0

The Internet Control Message Protocol (ICMP):



- During the normal operation of the Internet, many a times, errors, crashes and some other unexpected events may occur.
- The protocol, that reports these problems to the Routers is called the Internet Control Message Protocol (ICMP).

The Internet Control Message Protocol (ICMP) ...



- Some of the common ICMP messages include:
 - Echo Request
 - Echo Reply
 - Timestamp Request
 - Timestamp Reply
 - Redirect
 - Destination Unreachable
 - Time Expired / Exceeded

The Address Resolution Protocol (ARP):



- All computers in the IP world must be associated with one IP address or the other.
- For the purpose of actual delivery of a packet, the packet has to be sent through the Host-to-Network layer which means, for actual transmission the association of a given IP address to the lower layer address (say an Ethernet Adapter address / MAC Sub-layer Address) is required.
- The protocol that permits a machine holding an IP address to enquire about this lower layer address is called the 'Address Resolution Protocol' (ARP).



The Reverse Address Resolution Protocol (RARP):

- As indicated earlier, all computers in the IP world must be associated with at least one IP address that is associated with a MAC Sub-Layer address for the purpose of communication; therefore, a machine that knows just its Mac address will need to learn / discover about the associated IP Address as well.
- The protocol that permits a machine holding its lower layer address (say its Ethernet Address) to enquire about its associated IP address is called the 'Reverse Address Resolution Protocol' (RARP).



The Interior Gateway Routing Protocol (IGRP):

- The Interior Gateway Routing Protocol (IGRP) was originally developed in the mid-1980s by Cisco Systems. This protocol did not support VLSM scheme. Its successor, EIGRP, supports VLSM.
- Basic objective of the IGRP was to provide a robust protocol for routing within what was called as an 'autonomous system' (AS).

(An AS is a collection of networks under common administration that share a common routing strategy. Every AS is normally uniquely identified by a 16-bit number.)

The Interior Gateway Routing Protocol (IGRP) ...



- The most commonly used AS-AS routing protocol prior to the advent of the IGRP was the Routing Information Protocol (RIP).
- As mentioned earlier, very small hop limit (only 16 hops) restricted the size of RIP based internetworks.
- Moreover, as pointed out in the slides related to the Dynamic Routing Algorithms, RIP proved sub-optimal and less flexible.



The Exterior Gateway Routing Protocol (EGRP):

- **The Exterior Gateway Protocol (EGP) is an inter-domain connectivity / reachability protocol.**
- **The original version of the EGP that enjoyed quite a bit of popularity is gradually giving way to other competing exterior gateway routing protocols (like the BGP and the IDRP).**

(This is because of certain weaknesses that came to light with the exponential growth of the Internet over the years.)



The Exterior Gateway Routing Protocol (EGRP) ...

- Currently, the most well known Exterior Gateway Routing Protocol is the Border Gateway Protocol Version 4.
- Incidentally, BGP4 also happens to be the first version that is capable of handling the CIDR and Supernetting.

The Border Gateway Protocol (BGP) ...



- BGP uses the TCP as its transport protocol of choice. **(Advantage of this approach is that BGP can relieve itself of reliability specific concerns.)**
- BGP is a path vector protocol. (Since, the routing information used by the BGP consists of a vector of Autonomous System ID Nos., which actually maps to a traversed path / route, it is called as a path vector protocol.)
- It is primarily used for exchange of information between autonomous systems.

Mobile IP



- The variant of the Internet Protocol which has been specifically designed for providing support for Mobile Hosts willing to communicate over the Internet is called as the Mobile IP.
- It is the result of deliberations of a special IETF workgroup and has been described in the RFC.



Mobile IP ...

- Why the Mobile IP had to be devised?
 - The basic IP has an Addressing Scheme that comprises of Class Id., Network Number and Host Number. Therefore, any packet intended for a given Mobile Host shall have no problem as long as the MH stays on the Home LAN; since Routers all over the world can continue to use the Class plus Network Address information to route any information to it. The problem of discontinuation of service would arise as soon as soon as the MH moves out of its Home Zone; since now, the Routers would still continue to send traffic meant for this MH to the Home LAN address they have in the know of!



Mobile IP ...

- Why the Mobile IP had to be devised ...
 - One solution to this problem could have been assigning a new IP address to this MH once it moved away from its Home Zone. However, this is a non-solution primarily because such an assignment would require this information to be specifically sent to a large number of Routers, Databases and of course the intending communicators / collaborators, every time such a transition takes place. Given the large amount of transactions and inconvenience involved in implementing this solution and increasing number of people using the MHs, this would translate into a huge network traffic / bandwidth requirement by itself.

Mobile IP ...



- Why the Mobile IP had to be devised ...
 - Yet another possible solution could have been requiring the Routers to take routing decisions on the complete IPv4 address, instead of the customary Class-Id. plus Network Address. This, again, is a non-solution since the this requirement would translate into the requirement of huge Routing Table space, which in turn would mean the unacceptably high cost of transmission over the Internet.
 - Clearly, any acceptable solution had to avoid these traps and at the same time should have provided the required mobility, along with the continuity of communication at an acceptably low cost and without forcing the existing software to undergo any major change. **And, thus the Mobile IP was born!**

Mobile IP ...



- **Goals of the Mobile IP Design:**
 - **Just because the Mobile Hosts are to be accommodated, the Stationary Hosts should not be required to make any change in their local software.**
 - **Routers, all over the world, should not be required to alter their Routing Software as well as Routing Table structures or entries merely for this purpose.**
 - **Databases and other collaborating entities should not be required to be explicitly informed of the changed Id. of the MH.**
 - **No extra cost should be required to be added to the transactions while an MH was in its home zone.**
 - **As a consequence of some of the above referred goals, an important goal of not assigning a new IP address to the MH was added to the IETF-WG list.**



Mobile IP ...

The Proposed Solution:

- Each of the site supporting Mobile IP should create an Agent called 'Home Agent' / 'Home Address Agent'. This HA / HAA should be in charge of keeping track of which MHs of its home network are currently visiting a foreign network zone; and providing support services to these MHs as per need.
- Each site supporting the visiting MHs should create its own Agent called 'Foreign Agent' / 'Care-of Agent'. This FA / COA should be responsible for identifying the visiting MHs, keeping their track, authenticating their credentials by communicating with the corresponding HA / HAA and providing support services as per need.



Mobile IP

The Proposed Solution ...

- Whenever anyone sends packets for a MH that is currently visiting a foreign zone, the Router at the home zone attempts to resolve the address of the intended MH in the usual way of employing the ARP. The response to this ARP broadcast then comes from the HA / HAA, which supplies its address to the enquiring Router. A technique called Gratuitous ARP (G-ARP) is used to take care of invalid cache-entry in the Router.
- Once the packet is received by the HA / HAA, it encapsulates it and passes it to the IP address of the COA / FA, who on receipt, decapsulates and sends the packet to the visiting MH.
- This is immediately followed by sending the IP address of the current COA / FA to the original sender, so that any subsequent communication could use this new address thereafter.

Running the TCP/IP over the ATM: Facts and Issues



- The IP layer in the Internet acts as a Connectionless Network Layer whereas the ATM layer provides a Connection-oriented Network Layer functionality. Therefore, when IP layer is made to function atop the ATM layer, then even before a transaction really takes place, an ATM Network Layer connection is established between the Source and the Destination Hosts.

Running the TCP/IP over the ATM ...



- Thereafter, the IP layer takes over and sends independent IP Layer Packets (called IP Packets) over the established path. Naturally, although the ATM requires just a brief Virtual Circuit Number, the IP layer atop it generates IP packets that contain full Source as well as full Destination Address. This arrangement clearly leads to unsolicited overhead. Furthermore, the routing exercise done by the IP as well as the ATM Layer proves a sort of duplication, which adds to this overhead.

Running the TCP/IP over the ATM ...



- The TCP Layer sitting on the top of the IP layer does not know about the ATM layer and therefore assumes that its underlying IP Layer might deliver the packets (at the destination) out of order and instructs its mechanism to ensure that the packets arriving at the host are reordered correctly.

Running the TCP/IP over the ATM ...



- Since this involves duplication of much of the functionality in various layers, undue overheads in terms of connection-establishment / termination, packet-generation / routing / extraction and redundant use of the packet reordering mechanism etc., this arrangement (i.e. TCP / IP over the ATM) is considered inherently inefficient.

Running the TCP/IP over the ATM ...



- If still TCP/IP over ATM is used in certain cases, it is primarily because there exists a huge software base that talks the TCP / IP way and there are a lot of people who are used to it! This must not be misconstrued to mean that acceptable combination of IP over ATM cannot exist.



Congestion Control:

- Types of Congestion Control Schemes:
 - Open Loop Congestion Control Schemes
 - Traffic Filtering Schemes (use accept / reject rules)
 - Traffic Scheduling Schemes
 - Closed Loop Congestion Control Schemes
 - Uni-Variable Feedback based schemes
 - Multi-Variable Feedback based schemes



Congestion Control ...

- Congestion Metrics:
 - Average / Mean Queue Length
 - Average number or percentage of lost / discarded packets
 - Number of retransmitted packets those had to be sent again because of Transmitter's Time-out
 - Average / Mean Delay in Packet Delivery

Congestion Control ...



Congestion Control Strategies:

- Congestion control by regulating admission of Packets / Cells
- Congestion control by regulating traffic based on traffic-type / traffic-rate (packet rate / cell rate / bit rate etc.) analysis
- Congestion control by admission-time resource reservation

Congestion Control ...



Congestion Control Strategies ...

- Congestion control by threshold monitoring and message passing
- Congestion control by preferential restraint **(in research stage)**
- Congestion control by Ostrich algorithm **(debatable)**
- Congestion control by supervised blocking / rerouting **(under investigation)**

Congestion Control ...



The Anticipatory Buffer Allocation Scheme:

- In this scheme, which is particularly suitable for Virtual Circuit Subnets, congestion can be effectively controlled / avoided by estimating the optimal buffering needs of the Switches and allocating this buffer capacity to Virtual Circuits on anticipatory / pro-active basis.
- It is a variation of pre-allocation scheme since it allocates estimated capacity in advance.
- This scheme differs from the standard VC establishment scheme in the way that in the latter no buffer-space allocation is reserved at the Switches by the call-request packet. Also, no permanent buffer allocation is done a-priori, in the latter scheme.

Congestion Control ...



The Anticipatory Buffer Allocation Scheme...

- This scheme may be implemented using many different protocols including the Sliding Window and Stop-and-Wait protocols.
- Choice of a protocol, in any case depends on the desired throughput, available buffer capacity and the associated price.
- However, for the VCs that may not, at an average, have adequate traffic so as to effectively use a sizeable chunk of such pre-allocated buffer-space, the economics may not be favourable.
- Moreover, this is, in effect, a Congestion Avoidance Scheme rather than an adaptable Congestion Control Scheme.

Congestion Control ...



The Anticipatory Buffer Allocation Scheme...

- A possible variation of this scheme could be, as suggested in the beginning, a dynamic allocation scheme that is proactive by nature and that, by using some adaptive / statistical buffering need-determination algorithm, estimates / anticipates the required buffer size and if available, allocates the VC in question.
- The primary difference here is that the call request packet need not ask for any buffer reservation. Moreover, this allocation may be done after the establishment of the VC. This scheme duals as an Avoidance as well as a Control scheme since if invoked during VC establishment, it provides avoidance whereas if triggered by anticipation of congestion, could simply reduce the chances of its building up.
- However, this solution is relatively complex to implement and has a potential of occasional misfire.
- Both of the discussed solutions, therefore, do not prove attractive.

Congestion Control ...



'Arbitrary Packet Rejection-based' / 'Reject-on-Getting-Full' Congestion Control Scheme

- This scheme is the simplest of all congestion control schemes.
- It controls further building up of congestion just by dropping any further packets reaching the node in question, entirely arbitrarily, without any learned analysis. As a result, even ACKs might get rejected and cause a series of unwarranted problems.
- This scheme requires absolutely no buffer reservation / advance allocation, in complete contrast to the earlier scheme. A variation of this scheme is called the Leaky Bucket Algorithm.
- This too is not an attractive solution because of obvious potential for creating deadlocks.

Congestion Control ...



Selective Packet Rejection based Congestion Control Scheme

- This scheme is the modified version of the previous congestion control schemes.
- It controls further building up of congestion by selective dropping of packets reaching the node in question.
- The choice of selective acceptance / rejection is governed by a set of rules.
- This scheme, like its predecessor, requires absolutely no buffer reservation / advance allocation.

Congestion Control ...



Permit-based / Token-based / Isarithmic Congestion Control Scheme:

- As the name itself suggests, this algorithm uses a Permit / Token based admission control with respect to entry to a node.
- Any sender node willing to transmit 'n' packets to a receiving node is first required to capture 'n' Tokens / 'Permit for sending 'n' packets'. If only one Token is captured, only one packet can be transmitted.
- The number of total Tokens available is usually kept constant; and as result, this scheme ensures a predictable constant traffic, without any loss of packets.
- A variation of this algorithm is known as the Token Bucket Algorithm.

Congestion Control ...



The Choke Packet Scheme of Congestion Control:

- One of the possible ways to control congestion is to cut down the incoming traffic to a node by informing the originator of the traffic that a state of congestion has occurred and the originator should cut down its packet transmission rate intended to reach / pass through this receiver.
- This scheme uses just that! It makes use of what is termed as 'Choke Packet' for indicating to the originator about the congestion and expects it to cut down its transmission rate by a pre-defined percentage.

Congestion Control ...



The Choke Packet Scheme of Congestion Control ...

- What exactly happens is this! The various Routing nodes periodically run a routine for estimating the state of utilization of their one or more output lines and compute an index that could, on crossing a certain threshold value, normally suggest that a state of congestion is about to arrive or has arrived.
- Whenever this threshold value is reached, the congestion control routine gets fired.

Congestion Control ...



The Choke Packet Scheme of Congestion Control ...

- Once this routine swings into action, any packet other than an ACK that arrives at this node intending to be forwarded on any one of the congested output lines is blocked and a special packet called the “Choke Packet” is constructed by extracting the originating node’s address from the Sender’s Address field of the packet that has been blocked.
- The original packet itself is tagged / included as payload (to the generated header with a bit set) so as to help the originator learn so that it does not generate traffic any further / more than the default cut-down rate thereafter for a stipulated period of time.

Congestion Control ...

The Choke Packet Scheme of Congestion Control ...



- Many variations of the Choke Packet-based scheme exist. However, most of them have potential to generate further network congestion due to a lot of possible choke-packet traffic.
- One such possible solution is the Hop-by-Hop Choke Packet-based scheme of congestion control. This scheme has a special feature of helping in cutting down the incoming traffic systematically and gradually by informing every intermediate Router along the way of the Choke Packet. Thus, in effect, at every hop, the scheme succeeds in immediately initiating reduction in traffic towards the congested node; rather than allowing the flow to continue until the Choke Packet reaches its destination and an action is taken.

Congestion Control ...



- Deadlock due to congestion: There exists an extreme effect of failure to timely control of congestion! That's the Transmission Deadlock / Lock-up State.
- Such deadlocks can be of several types including Direct Store-and-Forward Deadlock / Lockup and Indirect Store-and-Forward Deadlock / Lockup.
- A well-known solution to such deadlocks was suggested long back by Merlin and Schwietzer that involved use of a specially constructed directed graph showing Buffers as nodes and arcs connecting a pair of buffers in the same or adjacent router.
- Several other solutions have been proposed since then.

Recommended Readings:



- S. Keshav: An Engineering Approach to Computer Networking, AWL, 1997.
- A. S. Tanenbaum: Computer Networks, Third Edition, PHI, 1996.
- C. Huitema: IPv6, Second Edition, Prentice-Hall PTR, 1998.
- U. D. Black: Computer Networks, Second Edition, PHI, 1993.
- D. Bertsekas and R. Gallager: Computer Networks, Second Edition, PHI, 1992.
- G. R. McClain (Ed.): Handbook of Networking and Connectivity, AP Professional (Academic Press), 1994.

Recommended Readings ...



- RFC 1009 (Requirements for Internet Gateways)
- RFC 1254 (Gateway Congestion Control)
- RFC 1360 (Official Protocol Standards of the Internet Architecture Board)
- RFC 1124 (Policy Issues in Interconnecting Networks)
- RFC 1125 (Policy Requirements for Inter-Administrative Domain Routing)
- RFC 781 (IP Timestamp)
- RFC 791 (IP)
- RFC 815 (IP Datagram Reassembly)
- RFC 1042 (IP over IEEE 802.3)
- RFC 1011 (Official IP)

Recommended Readings ...



- RFC 1883 (IPv6 Specification)
- RFC 1825 (IP Security Architecture)
- RFC 1826 (IP Authentication Header)
- RFC 1827 (IP Encapsulation Security Payload)
- RFC 1828 (IP Authentication using MD5)
- RFC 1175 (FYI : A very useful reference-list on Internetworking related information)
- RFC 1208 (Glossary of Networking Terms)
- Smoot Carl-Mitchell & John S. Quarterman: Practical Internetworking with TCP / IP and UNIX, Addison-Wesley, Reading, 1993. (This book does not really discuss the IPv6. This however, helps the reader to take a look at the pre-IPv6 days and realize the wisdom of evolution of the IP.)

Recommended Readings



...

- Larry Hughes: Introduction to Data Communication: A Practical Approach, Narosa Publishers, 1997.
- Prakash C. Gupta: Data Communications, PHI, 1996.
- A. Shah: FDDI: A High Speed Network, PTR Prentice Hall, 1994.
- M. R. Tolhurst (Ed.): Open System Interconnection, Macmillan, 1988.
- William Stallings: Data and Computer Communications, Fifth Edition, PHI, 1998.

Recommended Readings ...



- D. Comer: Internetworking with TCP / IP , Vol..-1, PHI, 1995.
- D. Comer & D. L. Stevens: Internetworking with TCP /IP, Vol.. 2-3, PHI, 1994, 1993.
- W. Buchanan: Advanced Data Communication and Networks, Chapman & Hall, London, 1997.
- Uyles D. Black: TCP / IP & Related Protocols, Second Edition, McGraw-Hill, N. Y., 1995.
- RFC 1519 (CIDR)
- RFC 1997 (BGP community attribute)
- Bassam Hallabi: Internet Routing Architectures, Cisco Press, New Riders Publishing, 1997.
- RFC 904 (Exterior Gateway Protocol)