ECSE-6600: Internet Protocols

Informal Quiz

Shivkumar Kalyanaraman: shivkuma@ecse.rpi.edu
Routing: general

T  F
☐  ☐  Forwarding works in the control plane whereas routing works in the data plane
☐  ☐  A routing protocol summarizes global information to setup a local next-hop entry in the forwarding table
☐  ☐  The distance-vector protocol involves checking neighbors’ distance vectors and updating its own distance vector.
☐  ☐  The poisoned reverse modification of DV algorithm is less effective in cases where the cost of a remote link (not the first or second) in a path increases.
☐  ☐  The link state method does not face the count-to-infinity problem because it has complete global information (a map in terms of link-states).
☐  ☐  Both the distance-vector and link-state approaches could lead to transient routing loops because the information maintained could be incomplete.
☐  ☐  Hierarchical addressing, and proper address assignment allows entire subnets to be viewed by interior routers as “virtual nodes”, leading to routing scalability
Routing II: Protocols

- RIP uses a 16-bit weight field to indicate the weight of each link.
- RIP has convergence problems because of issues like count-to-infinity, whereas the complexity in OSPF is in distributing the link states efficiently.
- A distance vector approach has a complete network map at every node.
- Diffusing computations (eg: DUAL) works because inconsistent information is not accepted while the routing tables are “frozen”.
- OSPFv2 uses the lollipop sequence number space.
- A low value of the age field and a high value of the sequence number field indicates a stable routing entry.
- On a point-to-point link, OSPFv2 performs database synchronization by exchanging its entire database between neighbors.
- The database synchronization operation is done upon discovering a new neighbor.
- On a broadcast LAN subnet, OSPFv2 prescribes the sole use of Router-LSAs due to its efficiency in encoding.
- A broadcast LAN subnet is viewed by the Dijkstra algorithm as a full mesh of links.
- A NBMA subnet is viewed by the Dijkstra algorithm as a full mesh of links.
- A pt-mpt subnet is viewed by the Dijkstra algorithm as a full mesh of links.
- The DR/BDR concept is required on pt-mpt subnets.
Routing II: Protocols

- Hellos and LSAs are multicast in broadcast LANs.
- LSA-acks are sent only to the DR and BDR, but Hello-Acks are piggybacked onto Hello multicasts on broadcast LAN subnets.
- A routing adjacency is equivalent to a separate physical link.
- The neighbor relationship is a unidirectional relationship.
- Hellos are sent periodically, whereas LSAs are sent only when a link state changes.
- The pt-mpt subnet model violates the IP subnet model assumption that nodes on the same subnet should be able to directly communicate with each other.
- A network-LSA is generated by any random router on the broadcast LAN subnet.
- An NBMA subnet allows cheap broadcast capability.
- The NBMA model requires a (costly) VC between any pair of routers on the subnet.
- Neighbor discovery on an NBMA is automatic: just multicast a Hello message to AllSPFRouters multicast address.
- The pt-mpt model allows OSPF to operate efficiently over partial meshed non-broadcast networks, even if some IP subnet assumptions are broken.
- Address abstraction is equivalent to topology abstraction in a hierarchical network like IP.
Routing II: Protocols

- OSPF supports arbitrary number of levels in its hierarchy
- An area ID can be encoded into an IP address, and hence areas can be auto-configured.
- AS-BRs operate at borders of areas and send summary information in and out of an area.
- ABRs generate external LSAs, which is summary information from other areas in the same routing domain.
- The metric field in a summary-LSA advertised by an ABR is the cost of the longest path from the ABR to any node within the area.
- Stubby areas filter all external LSAs, but may allow summary-LSAs to be optionally flooded within the area.
- The difference between an “area” and a “domain” is that different routing protocols operate beyond the boundaries of domains.
- NSSA areas allow partial filtering of external LSAs.
- Filtering of external-LSAs is a big concern because external BGP routes may number more than 100,000!
- IS-IS operates over IP whereas OSPF operates over the link layer directly
- IS-IS provides highly extensible TLV encoding, but OSPF focuses on optimization and alignment of fields.
- PNNI is a source-routed protocol and supports the QoS signaling in ATM
- The entire route in PNNI is encoded as a DTL and is processed at every hop.
- In general, signaled protocols can afford to be wasteful in terms of encoding and complexity during the signaling phase and efficient in the packet-transfer phase.
- PNNI is limited to only 2 levels of hierarchy.
Routing III: BGP

- Path-vector based distance vector algorithms have a full map of the network like Link state algorithms.
- The Bellman-Ford algorithm is used in policy-based distance-vector routing for BGP.
- EGP is restricted to a tree topology because it is incapable of comparing paths and therefore would lead to stable loops otherwise.
- Currently core routers have about 100000 routes, which suggests poor address aggregation.
- A stub AS could have traffic neither originating or terminating at the AS.
- An ORIGIN attribute of “INCOMPLETE” indicates that the routes were injected dynamically into BGP by IGPs.
- The routes in Adj-RIB-Out are likely to be different from Adj-RIB-In because BGP does policy-based route filtering.
- One of the steps of the BGP “tie-breaker” algorithm prefers the lowest ORIGIN attribute because statically injected routes are likely to be more stable than dynamically injected routes.
- The AS path length attribute cannot be used by IBGP for loop-detection because the IBGP operates within a single AS.
- Default routing works because there exists a set of “core” routers which do not use default routing.
- BGP uses a fixed tree structure to propagate reachability information from AS to the core.
- CIDR solves the router-table size explosion problem by allocating only contiguous blocks of addresses which are summarizable.
- The MED and LOCAL_PREF attributes in BGP can be used for load-balancing.
Routing III: BGP

The neighbor reachability algorithm in EGP is same as that of OSPF, i.e., send a hello and wait for a DeadInterval for a response.

Like RIP, EGP and BGP send out full routing tables to their neighbors periodically.

Today’s inter-AS topology is more complex, but it still has a roughly hierarchical structure embedded in its complexity.

An AS number can be encoded into an IP address just like a network ID.

BGP finds inter-AS routes, and then resolves it to find the physical next-hop.

All default-free routers on the Internet speak BGP.

An AS can be internally disconnected, and use an inter-AS route to reach a destination within the AS.

A public ASN assignment to an AS means that it can formulate its own routing policy.

A transit-AS differs from a peer-AS primarily in the fact that one party necessarily pays in a transit relationship.

Recursive lookup in BGP guarantees loop-free paths.

Policy routing essentially allows an arbitrary choice between available set of paths.

The CIDR part of BGP-4 allows address aggregation.

Link-state based policy routing is less preferred to vectoring protocols (like BGP) because local policies need to be announced globally, and convergence of the flooding protocol is problematic in link-state.

The route-reflector concept converts a full-mesh of iBGP sessions to a tree-structure of iBGP sessions.

BGP NEXT-HOP is the same as the IP notion of next-hop.
Routing III: BGP

- MED allows outbound load-balancing
- LOCAL-PREF allows inbound load-balancing
- AS-path Padding is used as a rough way to control inbound load, but it may not work, if the AS is providing the only path to the destination prefix
- Hot-potato routing refers to carrying traffic in the same AS as far as possible before letting it cross AS boundaries.
- Multi-homed ASes have exactly one outbound link to the external Internet.
- An AS may be multi-homed to a single transit provider, and MED is useful in this situation
- Deaggregation or punching of holes in an address prefix essentially subverts the CIDR address aggregation process and may lead to larger routing tables in the Internet
- Since the MED field is sometimes the IGP routing metric, it could lead to route-flapping and a lot of eBGP update traffic.
- Subverting the CIDR aggregation by punching a hole and advertising it to a different ISP may lead to some inbound load-balancing benefit, at the expense of the entire Internet
- A community attribute allows arbitrary coloring and processing of routes. But the community values (colors) have to be agreed upon by the set of ASes involved.
- The first 16 bits of the community attribute is just the AS number.
- The BGP decision process is a simple tie-breaker set of rules, with the recursive lookup and local-pref rules being the highest priority
- A stateful route flap dampening algorithm has been used to dramatically reduce the average number of updates sent by BGP
- BGP often takes a long time to converge after route changes.

Shivkumar Kalyanaraman

Kensselaer Polytechnic Institute
TCP

- TCP can re-assemble IP fragments
- Path-MTU refers to the procedure of finding the minimum MTU of the path to reduce the probability of fragmentation.
- The IP header checksum field is the 16-bit two’s complement of the one’s complement sum of all 16-bit words in the header.
- TCP provides reliability only at a packet-level.
- The TOS byte semantics is inconsistent with the IP model of providing only best-effort service.
- Transport protocols are minimally required because IP does not provide application multiplexing support.
- The Nagle algorithm in TCP is intended to allow the ack and echo data to be combined.
- TCP is called “self-clocking” because the source sends traffic whenever it likes.
- TCP by default uses a selective retransmission policy.
- The RFC 793 RTT estimator could only tolerate variances of upto 30%.
- The TCP congestion control algorithm is stable because it detects congestion reliably and its rate of window decrease is faster than its rate of window increase.
- TCP’s use of cumulative acks reduces the need for any timeout/retransmission of acks.
- Karn’s algorithm would be triggered often on a wireless or radio link which is very lossy.
- Delayed-acks are good for bulk traffic, but bad for interactive traffic.
- A two-way handshake is sufficient for the robust setup of a half-duplex connection, but a three-way handshake is necessary for the robust setup of a full-duplex connection.
TCP

- If timeouts are not used, in general, packet or ack-losses cannot be recovered from.
- A duplicate ack gives the same information as a NAK, but it presumes the notion of a sequence number.
- Sequence numbers allow the detection of duplicate packets, but the sequence number space must be sized sufficiently large compared to the window size depending upon the retransmission algorithm (go-back-N or selective-repeat) used.
- In a lossless network, window-based transmission can achieve full utilization.
- TCP sets its RTO to an average RTT measure + 4*mean deviation of RTT, based upon Chebyshev’s theorem.
- Retransmission ambiguity would not occur if timestamps were used on packets.
- Self-clocking of TCP can be a liability in asymmetric networks where the reverse path can artificially constrain the forward path.
- Self-clocking can also lead to burstiness if the reverse path is congested, and/or the receiver uses a delay-ack time to suppress ACKs.
- The end-to-end congestion control model is the only one that can guarantee avoidance of congestion collapse.
- The notions of efficiency and fairness define an equilibrium point to which congestion control algorithms attempt to converge.
- A stable congestion control algorithm converges to its equilibrium point.
- In the \((w, \alpha)\) notion of fairness, \(\alpha = 1\) leads to max-min fairness.
- In equilibrium, TCP attempts to conserve packets and operate at high utilization.
- TCP does not guarantee low queueing delays because it depends upon packet loss for congestion detection.
TCP/Congestion Control

- Fast retransmit refers to the procedure of using three duplicate acks to infer packet loss.
- TCP Tahoe sets its window to 1 after every loss detection.
- TCP Reno may timeout quickly in a multiple packet loss scenario.
- TCP SACK uses selective retransmit, and like NewReno, it does not reduce its window more than once per window of packets.
- With a 28kbps reverse link, 1500 byte packets are regular TCP behavior, the forward link throughput is at most around 2 Mbps.
- Header compression and link level ack suppression/regeneration could help in asymmetric bandwidth scenarios.
- Scheduling refers to the control of which packet is dropped from buffers.
- FIFO+droptail provides service isolation among the participating TCP flows.
- Synchronization occurs because DropTail leads to bursty and correlated packet losses amongst flows; and flows react to same events.
- Dropping packets early has the risk that transient burstiness may be mistaken for true overload (demand > capacity).
- Marking packets instead of dropping them avoids nonlinearities caused by loss detection and retransmission mechanisms.
- RED determines random drop probability by comparing the average queue size to a max and min thresholds.
- FRED protects fragile flows and isolates effects of mis-behaving flows, but incurs per-flow state maintenance cost.
- Random dropping/marking with a bias in RED helps break synchronization.
Routing: general (SOLNS)

T  F
☐  ✓ Forwarding works in the control plane whereas routing works in the data plane
✓  ☐ A routing protocol summarizes global information to setup a local next-hop entry in the forwarding table
✓  ☐ The distance-vector protocol involves checking neighbors’ distance vectors and updating its own distance vector.
✓  ☐ The poisoned reverse modification of DV algorithm is less effective in cases where the cost of a remote link (not the first or second) in a path increases.
✓  ☐ The link state method does not face the count-to-infinity problem because it has complete global information (a map in terms of link-states).
✓  ☐ Both the distance-vector and link-state approaches could lead to transient routing loops because the information maintained could be incomplete.
✓  ☐ Hierarchical addressing, and proper address assignment allows entire subnets to be viewed by interior routers as “virtual nodes”, leading to routing scalability

Shivkumar Kalyanaraman
Routing II: Protocols (SOLNS)

- √ RIP uses a 16-bit weight field to indicate the weight of each link
- √ RIP has convergence problems because of issues like count-to-infinity, whereas the complexity in OSPF is in distributing the link states efficiently
- √ A distance vector approach has a complete network map at every node.
- √ Diffusing computations (eg: DUAL) works because inconsistent information is not accepted while the routing tables are “frozen”.
- √ OSPFv2 uses the lollipop sequence number space
- √ A low value of the age field and a high value of the sequence number field indicates a stable routing entry
- √ On a point-to-point link, OSPFv2 performs database synchronization by exchanging its entire database between neighbors
- √ The database synchronization operation is done upon discovering a new neighbor
- √ On a broadcast LAN subnet, OSPFv2 prescribes the sole use of Router-LSAs due to its efficiency in encoding
- √ A broadcast LAN subnet is viewed by the Dijkstra algorithm as a full mesh of links
- √ A NBMA subnet is viewed by the Dijkstra algorithm as a full mesh of links
- √ A pt-mpt subnet is viewed by the Dijkstra algorithm as a full mesh of links
- √ The DR/BDR concept is required on pt-mpt subnets.

Rensselaer Polytechnic Institute
Shivkumar Kalyanaraman
### Routing II: Protocols (SOLNS)

- **√** Hellos and LSAs are multicast in broadcast LANs.
- **√** LSA-acks are sent only to the DR and BDR, but Hello-Acks are piggybacked onto Hello multicasts on broadcast LAN subnets.
- **□** A routing adjacency is equivalent to a separate physical link.
- **√** The neighbor relationship is a unidirectional relationship.
- **√** Hellos are sent periodically, whereas LSAs are sent only when a link state changes.
- **√** The pt-mpt subnet model violates the IP subnet model assumption that nodes on the same subnet should be able to directly communicate with each other.
- **□** A network-LSA is generated by any random router on the broadcast LAN subnet.
- **√** An NBMA subnet allows cheap broadcast capability.
- **□** The NBMA model requires a (costly) VC between any pair of routers on the subnet.
- **√** Neighbor discovery on an NBMA is automatic: just multicast a Hello message to AllSPFRouters multicast address.
- **□** The pt-mpt model allows OSPF to operate efficiently over partial meshed non-broadcast networks, even if some IP subnet assumptions are broken.
- **√** Address abstraction is equivalent to topology abstraction in a hierarchical network like IP.

Rensselaer Polytechnic Institute

Shivkumar Kalyanaraman
Routing II: Protocols (SOLNS)

- ✓ OSPF supports arbitrary number of levels in its hierarchy
- ✓ An area ID can be encoded into an IP address, and hence areas can be auto-configured.
- ✓ AS-BRs operate at borders of areas and send summary information in and out of an area.
- ✓ ABRs generate external LSAs, which is summary information from other areas in the same routing domain.
- ✓ The metric field in a summary-LSA advertised by an ABR is the cost of the longest path from the ABR to any node within the area.
- ✓ Stubby areas filter all external LSAs, but may allow summary-LSAs to be optionally flooded within the area.
- ✓ The difference between an “area” and a “domain” is that different routing protocols operate beyond the boundaries of domains.
- ✓ NSSA areas allow partial filtering of external LSAs.
- ✓ Filtering of external-LSAs is a big concern because external BGP routes may number more than 100,000!
- ✓ IS-IS operates over IP whereas OSPF operates over the link layer directly
- ✓ IS-IS provides highly extensible TLV encoding, but OSPF focuses on optimization and alignment of fields.
- ✓ PNNI is a source-routed protocol and supports the QoS signaling in ATM
- ✓ The entire route in PNNI is encoded as a DTL and is processed at every hop.
- ✓ In general, signaled protocols can afford to be wasteful in terms of encoding and complexity during the signaling phase and efficient in the packet-transfer phase.
- ✓ PNNI is limited to only 2 levels of hierarchy.

Shivkumar Kalyanaraman
Routing III: BGP (SOLNS)

- √ Path-vector based distance vector algorithms have a full map of the network like Link state algorithms.
- √ The Bellman-Ford algorithm is used in policy-based distance-vector routing for BGP.
- √ EGP is restricted to a tree topology because it is incapable of comparing paths and therefore would lead to stable loops otherwise.
- √ Currently core routers have about 100000 routes, which suggests poor address aggregation.
- √ A stub AS could have traffic neither originating or terminating at the AS.
- √ An ORIGIN attribute of “INCOMPLETE” indicates that the routes were injected dynamically into BGP by IGPs.
- √ The routes in Adj-RIB-Out are likely to be different from Adj-RIB-In because BGP does policy-based route filtering.
- √ One of the steps of the BGP “tie-breaker” algorithm prefers the lowest ORIGIN attribute because statically injected routes are likely to be more stable than dynamically injected routes.
- √ The AS path length attribute cannot be used by IBGP for loop-detection because the IBGP operates within a single AS.
- √ Default routing works because there exists a set of “core” routers which do not use default routing.
- √ BGP uses a fixed tree structure to propagate reachability information from AS to the core.
- √ CIDR solves the router-table size explosion problem by allocating only contiguous blocks of addresses which are summarizable.
- √ The MED and LOCAL_PREF attributes in BGP can be used for load-balancing.
Routing III: BGP (SOLNS)

- ✓ The neighbor reachability algorithm in EGP is same as that of OSPF, i.e., send a hello and wait for a DeadInterval for a response.
- ✓ Like RIP, EGP and BGP send out full routing tables to their neighbors periodically
- ✓ Today’s inter-AS topology is more complex, but it still has a roughly hierarchical structure embedded in its complexity
- ✓ An AS number can be encoded into an IP address just like a network ID
- ✗ BGP finds inter-AS routes, and then resolves it to find the physical next-hop.
- ✓ All default-free routers on the Internet speak BGP
- ✓ An AS can be internally disconnected, and use an inter-AS route to reach a destination within the AS
- ✗ A public ASN assignment to an AS means that it can formulate its own routing policy
- ✓ A transit-AS differs from a peer-AS primarily in the fact that one party necessarily pays in a transit relationship
- ✓ Recursive lookup in BGP guarantees loop-free paths
- ✗ Policy routing essentially allows an arbitrary choice between available set of paths
- ✗ The CIDR part of BGP-4 allows address aggregation
- ✓ Link-state based policy routing is less preferred to vectoring protocols (like BGP) because local policies need to be announced globally, and convergence of the flooding protocol is problematic in link-state.
- ✓ The route-reflector concept converts a full-mesh of iBGP sessions to a tree-structure of iBGP sessions.
- ✓ BGP NEXT-HOP is the same as the IP notion of next-hop

Rensselaer Polytechnic Institute

Shivkumar Kalyanaraman
<table>
<thead>
<tr>
<th></th>
<th>✓</th>
<th></th>
<th>✓</th>
</tr>
</thead>
<tbody>
<tr>
<td>□</td>
<td>✓</td>
<td>MED allows outbound load-balancing</td>
<td></td>
</tr>
<tr>
<td>□</td>
<td>✓</td>
<td>LOCAL-PREF allows inbound load-balancing</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>□</td>
<td>AS-path Padding is used as a rough way to control inbound load, but it may not work, if the AS is providing the only path to the destination prefix</td>
<td></td>
</tr>
<tr>
<td>□</td>
<td>✓</td>
<td>Hot-potato routing refers to carrying traffic in the same AS as far as possible before letting it cross AS boundaries.</td>
<td></td>
</tr>
<tr>
<td>□</td>
<td>✓</td>
<td>Multi-homed ASes have exactly one outbound link to the external Internet.</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>□</td>
<td>An AS may be multi-homed to a single transit provider, and MED is useful in this situation</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>□</td>
<td>Deaggregation or punching of holes in an address prefix essentially subverts the CIDR address aggregation process and may lead to larger routing tables in the Internet</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>□</td>
<td>Since the MED field is sometimes the IGP routing metric, it could lead to route-flapping and a lot of eBGP update traffic.</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>□</td>
<td>Subverting the CIDR aggregation by punching a hole and advertising it to a different ISP may lead to some inbound load-balancing benefit, at the expense of the entire Internet</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>□</td>
<td>A community attribute allows arbitrary coloring and processing of routes. But the community values (colors) have to be agreed upon by the set of ASes involved.</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>□</td>
<td>The first 16 bits of the community attribute is just the AS number.</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>□</td>
<td>The BGP decision process is a simple tie-breaker set of rules, with the recursive lookup and local-pref rules being the highest priority</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>□</td>
<td>A stateful route flap dampening algorithm has been used to dramatically reduce the average number of updates sent by BGP</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>□</td>
<td>BGP often takes a long time to converge after route changes.</td>
<td></td>
</tr>
</tbody>
</table>

Shivkumar Kalyanaraman

Rensselaer Polytechnic Institute
TCP (SOLNS)

- √ TCP can re-assemble IP fragments
- √ Path-MTU refers to the procedure of finding the minimum MTU of the path to reduce the probability of fragmentation.
- √ The IP header checksum field is the 16-bit two’s complement of the one’s complement sum of all 16-bit words in the header.
- √ TCP provides reliability only at a packet-level.
- √ The TOS byte semantics is inconsistent with the IP model of providing only best-effort service.
- √ Transport protocols are minimally required because IP does not provide application multiplexing support
- √ The Nagle algorithm in TCP is intended to allow the ack and echo data to be combined.
- √ TCP is called “self-clocking” because the source sends traffic whenever it likes
- √ TCP by default uses a selective retransmission policy
- √ The RFC 793 RTT estimator could only tolerate variances of upto 30%
- √ The TCP congestion control algorithm is stable because it detects congestion reliably and its rate of window decrease is faster than its rate of window increase
- √ TCP’s use of cumulative acks reduces the need for any timeout/retransmission of acks
- √ Karn’s algorithm would be triggered often on a wireless or radio link which is very lossy
- √ Delayed-acks are good for bulk traffic, but bad for interactive traffic.
- √ A two-way handshake is sufficient for the robust setup of a half-duplex connection, but a three-way handshake is necessary for the robust setup of a full-duplex connection

Rensselaer Polytechnic Institute

Shivkumar Kalyanaraman
TCP (SOLNS)

- If timeouts are not used, in general, packet or ack-losses cannot be recovered from.
- A duplicate ack gives the same information as a NAK, but it presumes the notion of a sequence number.
- Sequence numbers allow the detection of duplicate packets, but the sequence number space must be sized sufficiently large compared to the window size depending upon the retransmission algorithm (go-back-N or selective-repeat) used.
- In a lossless network, window-based transmission can achieve full utilization.
- TCP sets its RTO to an average RTT measure + 4*mean deviation of RTT, based upon Chebyshev’s theorem.
- Retransmission ambiguity would not occur if timestamps were used on packets.
- Self-clocking of TCP can be a liability in asymmetric networks where the reverse path can artificially constrain the forward path.
- Self-clocking can also lead to burstiness if the reverse path is congested, and/or the receiver uses a delay-ack time to suppress ACKs.
- The end-to-end congestion control model is the only one that can guarantee avoidance of congestion collapse.
- The notions of efficiency and fairness define an equilibrium point to which congestion control algorithms attempt to converge.
- A stable congestion control algorithm converges to its equilibrium point.
- In the (w, α) notion of fairness, α = 1 leads to max-min fairness.
- In equilibrium, TCP attempts to conserve packets and operate at high utilization.
- TCP does not guarantee low queueing delays because it depends upon packet loss for congestion detection.

Shivkumar Kalyanaraman

Rensselaer Polytechnic Institute
## TCP/Congestion Control (SOLNS)

- ✔ Fast retransmit refers to the procedure of using three duplicate acks to infer packet loss
- ✔ TCP Tahoe sets its window to 1 after every loss detection
- ✔ TCP Reno may timeout quickly in a multiple packet loss scenario
- ✔ TCP SACK uses selective retransmit, and like NewReno, it does not reduce its window more than once per window of packets
- ✔ With a 28kbps reverse link, 1500 byte packets are regular TCP behavior, the forward link throughput is at most around 2 Mbps
- ✔ Header compression and link level ack suppression/regeneration could help in asymmetric bandwidth scenarios
- ✔ Scheduling refers to the control of which packet is dropped from buffers
- ✔ FIFO+droptail provides service isolation among the participating TCP flows
- ✔ Synchronization occurs because DropTail leads to bursty and correlated packet losses amongst flows; and flows react to same events
- ✔ Dropping packets early has the risk that transient burstiness may be mistaken for true overload (demand > capacity)
- ✔ Marking packets instead of dropping them avoids nonlinearities caused by loss detection and retransmission mechanisms
- ✔ RED determines random drop probability by comparing the average queue size to a max and min thresholds
- ✔ FRED protects fragile flows and isolates effects of mis-behaving flows, but incurs per-flow state maintainence cost.
- ✔ Random dropping/marking with a bias in RED helps break synchronization