

CNT5106C Computer Networks, Fall 2009

Instructor: Prof. Ahmed Helmy

Homework #5

On Multicast at the Network Layer (IP Multicast and Membership)

[Date Assigned: Dec 6th, 2009. Due Date: Dec 15th, to the TA]

Total points: 124 points, including 24 extra points

Q1. (12 points) In IP multicast, separate mechanisms are used for member management on LANs (consider only IGMP as introduced in class) by the last-hop router and multicast routing in networks. What are the benefits of such separation? What are potential problems of such separation? (list 2 each)

A1.

Benefits:

- (1) Member management and multicast routing mechanisms can be designed and changed independent of the other.
- (2) The “request aggregation” by consider LANs instead of hosts as the entries for multicast routing reduces the load on routers.

Potential Problems:

- (1) Due to this separation, there is no way to figure out which host or how many hosts are members of a group at the sender or any router in the network.
- (2) It would be more difficult to provide member authentication and restricted access to a multicast group, since the sender does not know who the actual receiver is.

Q2. (10 points) In several multicast routing protocols we use RPF check (reversed-path forwarding check). What is the purpose of such check? How does it work? What are the underlying assumptions this check relies on?

A2. The purpose of RPF check is to avoid loops in multicast routing. We want to build a tree, which is loop-free, with the sender serves as the root. Hence a router only accepts incoming multicast traffic when it comes from the interface that is used to send packet toward the sender’s IP address on unicast routing table. By doing so a router only accepts incoming multicast traffic from each sender on one interface, hence avoids the possibility of forming routing loops.

The underlying assumptions of RPF check are:

- (1) It depends on unicast routing table, so unicast routing table must be correct and converged for RPF checks to work properly.

(2) It assumes that the path used from a sender to a router and the reversed path from the router back to the sender are symmetric. If they are not, RPF check would reject multicast traffic on the shortest path from the sender to the router. It leads to non-optimal multicast tree.

Q3. (10 points) What are the differences between the targeted environments (potential number of group members etc.) for PIM-DM and PIM-SM? How does this lead to different protocol design?

A3. In PIM-DM it assumes that most parts of the network want the multicast traffic. Therefore it starts with a flooding to build a complete multicast tree and later removes unwanted branches. Since most parts of the network want the multicast traffic, it makes sense to have a routing state on all the routers for each group, as it would be useful with high probability.

In PIM-SM, on the contrary, it assumes that most parts of the network do not want the multicast traffic. In stead of initial flooding, specific join mechanism is used in PIM-SM, and states are only created on routers that must know the state to forward the multicast traffic.

Q4. (18 points) IGMP provides membership information to the first hop router regarding the existence of receivers on a directly connected LAN.

(a) Why are group-specific query messages introduced in IGMPv2? Argue showing what a router does when it receives a 'leave' message from a host.

(b) The multicast host model does not define any interaction between the sources and IGMP. Do you see any problems with that? Explain. [Hint: Think of a case where there are no members for a group.]

(c) A researcher suggested to have the source multicast a query to the group in order to find out whether there are members in the group or not. Receivers would respond to this query by sending response to the source.

(I) Do you think that helps to solve the problem in (b) above? How?

(II) What is a main problem associated with this approach (i.e., the multicast query/response)? Suggest two alternative modifications to alleviate this problem. [Hint: You may use timers at the receivers, but how? You may also suggest simple modifications to IGMP.]

(d) Suggest a simple modification to IGMP (as an alternative to the approach in (c)) to solve the problem in (b).

A4.

(a) (4 points) The group specific query is needed to query the LAN for membership in a certain group in case of a leave (IGMPv2 attempts to reduce leave latency by allowing explicit leave messages). A router receiving a leave message would trigger/send a query message that is group specific to the LAN. If the router does not receive any membership reports in response to this query it assumes there are no longer any members on that LAN and removes the LAN from its outgoing interface list.

(b) (4 points) The problem is that of network overhead. If there are no members in the group, then the source will keep on sending packets, and the leaf router (that has created a prune state) will keep dropping the packets on the floor. So the leaf network resources will be consumed unnecessarily (in terms of bandwidth and processing overhead in the leaf router).

(c) (I) (3 points) the approach helps to solve the problem because if the source knew there are no members in the group, it would stop sending packets onto the LAN (thus pruning this last hop towards the source).

(II) The problem of multicast query, is that we may get 'response' implosion at the source. One way to alleviate this is to use randomized timers at the receivers. (the idea is that the source needs to get at least one response). A receiver getting a request would set a randomized timer, when it fires the receiver would send a response back to the source and multicast the response to the group as well. While the timer is running, however, if the receiver receives a response from another receiver, it would suppress its own response, thus reducing the number of responses sent to the source.

(4 points)

Another suggestion is to have IGMP send a 'no-member' report back to the source, if there are no members in the group (the leaf router for the source would know that from the prunes it gets) [this approach is called everse IGMP] (3 points)

Q5. (12 points) SPT bit in PIM-SM:

(a) What is the main use of the SPT (Shortest Path Tree) bit in PIM-SM? [Clarify using two simple scenarios]

(b) Someone suggested to remove this bit to reduce the state required in the routers and reduce the complexity of the protocol. What could be the effect on PIM-SM operation? [Mention a scenario to illustrate]

A5.

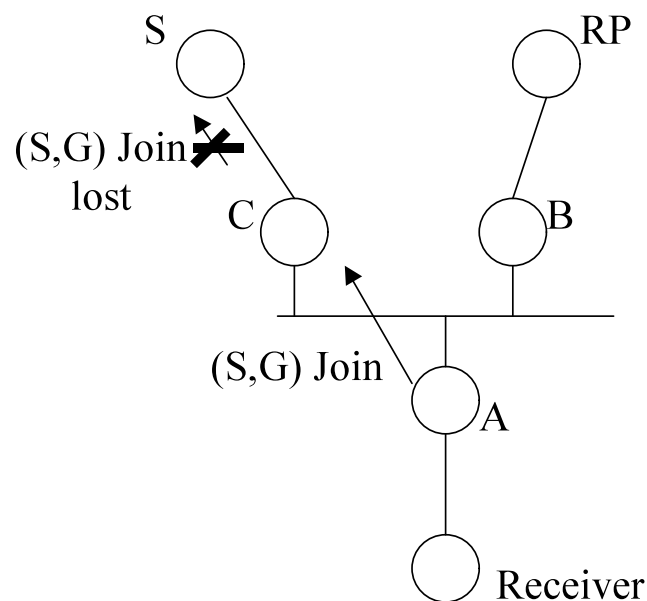
(a) (7 points, 3.5 points for each scenario)

The SPT bit is used when switching to the shortest path. It provides for 'smooth' switching to reduce the number of packets lost during the process:

- 1- First scenario: when the threshold of registers sent by the sender's first-hop router is high, and the RP sends S,G join towards the source. When the RP starts receiving 'native' data packets from the source, it sets the SPT bit to '1', which triggers 'register stop' messages to the sender's first-hop router with the next register received.
 - 2- Second scenario: when the rate of the data packets received for certain (source, group)-pair exceeds a threshold and the receiver's last-hop router decides to switch to the shortest path tree (SPT). The last-hop router sends 'hop-by-hop' (S,G) join towards the source. The 'fork' router [the router that has different incoming interfaces towards the RP and the source], sets the SPT to '0'. When the fork router starts receiving packets from the incoming interface towards the sender, it sets the SPT to '1', which in turn triggers prune messages towards the RP to stop the flow of (S,G) packets down that branch of the shared tree.
- (b) (5 points) Removing the SPT bit may affect the 'smooth' switch to the source scenarios described above. In PIM-SM, all actions are soft state, they are triggered by the state created in the router, and so the SPT bit helps in (for example) sending periodic prunes up the shared tree when switching to the SPT.
- If we do not use the SPT and have the fork router send prunes up the shared tree as soon as it sends a (S,G) join towards the source, and if the S,G join is lost, then the downstream router will get black holes and will lose packets.

Q6. (18 points) Assert mechanism in PIM-SM:

Based on understanding of the previous question, try to point out the problem with the following scenario:



Node *A* has an interface on a LAN leading to two routers *B* and *C*. Router *B* is the next-hop router on the shortest path to the rendezvous point (*RP*) and router *C* is the next-hop router on the shortest path to the source *S*. One version of the PIM-SM specification states that:

“Upon switching to the shortest path tree (SPT), the router with two different next-hop routers towards the RP and the source sets the SPT bit to ‘0’. Upon receiving the first packet from the interface leading to the source, the SPT bit is set to ‘1’ and a prune is sent towards the RP”.

(a) If *A* decides to switch to the shortest path for (S,G), but the (S,G) join from router *C* towards *S* is lost in the above scenario (see figure above), what would happen? [mention which messages will be sent by *A* and what would happen to the packets being forwarded to *A*]

(b) Someone suggested to change the wording in the specification to add: “only if the ‘interfaces’ (not next-hop routers) towards the RP and the source are different, does the router send a prune towards the RP when it receives the first packet on the interface leading to the source”. Do you think that is a good or bad idea (i.e., will it solve the previous problem or not, and will it introduce any other problems)?

[Mention a complete scenario saying what would happen in this case. Would *A* be getting duplicates from the LAN from both *B* and *C*. Hint: you need to understand the Assert mechanism.]

(c) From your understanding in (b) above, comment on the need for distinction between an inactive created state (one that was created with a Join message but not yet used to forward data packets), and an active state (one that has already been used to forward packets). Which of these states should be used in the Assert mechanism, and why?

A6. [Note: this was a real scenario we faced while designing PIM-SM and the problem was fixed in the later versions of the specification. Check RFC2117 and RFC2362].

(6 points for each a, b and c)

(a) When *A* sends the (S,G) Join towards *C* it sets the SPT bit to ‘0’, and (according to the rule given above from the PIM-SM spec) upon receiving the next packet for (S,G) *A* will set the SPT to ‘1’ and send a prune towards the RP (i.e. to *B*), and so the packets being forwarded down the shared tree from *B* will stop*. Note that the branch from *C* to *S* has not been established yet, and so *A* will be totally cut off and will suffer a ‘black hole’. [This occurs when any Join from *A* towards *S* is lost].

[* Node *A* cannot tell which node (*B* or *C*) forwarded the packet on the LAN, since the IP header has *S* in the source field and *G* in the destination field in both cases. So, the first packet from either upstream routers will trigger setting the SPT bit and the prune.]

Furthermore, when *B* sends a packet on the LAN after *C* has created S,G state, this will trigger an Assert from *C*. Accordingly *B* will get this Assert and will know that *C* is on the shortest path, so *B* will lose the Assert and will prune off *S*'s packets sent to the LAN from the shared tree.

(b) From (a) above, the suggestion is a good suggestion, although it does not solve the whole problem. The part of the problem that is solved is when the join from *A* to *C* is lost (when no Assert is triggered), since *A* in this case will not send prune messages towards *B*. This suggestion, however, does not solve the problem of Asserts by *C* as explained above.

(c) From (a) and (b) we need to distinguish between an entry that has been used to forward packets and another that was created by joins and has not yet been used to forward packets. In the above scenario router *C* has created a state when it got the S,G join from *A*, but it was not used to forward packets, because the joins from *C* towards *S* were lost.

The 'active' state is the state that should be allowed to participate in the Assert mechanism to avoid the black holes scenario explained in (a). [Note that both the ideas from (b) and (c) combined solve the whole problem. Think of scenario of join loss from *A* to *C* and from *C* towards *S* as explained above.]

Q7. (16 points) RP Bootstrap mechanism

(a) Someone suggested to use application-level mechanisms to distribute information about the RP in PIM-SM. An application wishing to join or send to a group would listen to an RP announcement group (a multicast group used to disseminate RP to group mapping) to get the RP address for the group. Do you see any problem with the above scenario?

(b) Hashing for Group-to-RP mapping:

Someone suggested a hash algorithm that takes as input the number of the RPs in the RP-list and the group address to get the RP for the group. For example, a list has 3 RPs, RP1, RP2 and RP3, the hash algorithm takes a group address *G* and applies $f(G,3)=2$ (for example) so it chooses RP2 to be the RP for the group. Do you think that is a good approach to the RP mapping? Why?

(8 points for each a and b)

A7.

- (a) The problem with the above scenario is that of ‘bootstrap’. To clarify, we need a ‘multicast group’ to disseminate the ‘RP-to-group’ mapping information. In order to support a multicast group using PIM-SM we need information about RP-to-group mapping. So we cannot possibly disseminate information about new RPs until we already have old RPs to support the dissemination group, so this idea will not work without a bootstrap solution.

[The above is sufficient as complete solution. If a student mentions the coming without the above, they get half credit]

Even if a bootstrap mechanism (say static configuration for the RPs for the well-known dissemination group) was in place, another problem with the above proposal is that it changes the ‘host model’ for multicast. Since, currently, a sender is not required to initiate any signaling before sending packets, with the proposed model, the sender needs to join this RP dissemination group and get the RP information before sending packets. This entails changing all current multicast applications.

- (b) No, this is not a good approach for RP mapping. One of the requirements for a good mapping algorithm is that it attempts to minimize group disruption during RP failures. So, a good mapping algorithm would re-map only those groups that previously mapped to the failed RP.

Keeping this in mind, the proposed algorithm does not achieve the above merit of ‘goodness’. To illustrate this, take a group G that mapped to RP_2 before RP_1 (the RP to fail) actually failed. Before failure RP_2 was chosen based on the number of RPs in the list (say 3). Now, after the failure of RP_1 , the number of RPs in the list is 2, and so $f(G,2)$ may be unequal to $f(G,3)$, and hence G (which used to map to an RP other than the failed RP) may get re-mapped due to failure of RP_1 .

Q8. (18 points) Soft state vs. hard state

PIM-SM uses a concept called ‘soft-state’ in its messaging protocol. This concept simply indicates that a join message (for example) is sent periodically by the downstream routers to the upstream router to refresh the join state. An alternative would be to use ‘hard-state’ messaging, in which an ack is sent for each message, such that a join and a join-ack (2 messages) only need to be sent between an upstream and a downstream router on a link.

- (a) Which protocol incurs less overhead on the network?
(b) Why are soft-state mechanisms sometimes preferred over hard-state mechanisms?

(c) Soft-state protocols incur more overhead on the network, especially when the number of states in the router (source-group pair state, for example) increases, as a state refresh needs to be sent upstream for every state at fixed periods. Obviously, this approach does not scale. Suggest an approach to alleviate this problem and discuss its advantages and disadvantages. [Hint: You may use timers].

A8.

- (a) (5 points) Soft state protocols incur periodic overhead to refresh the live states, while hard state only establishes the state using an ‘acked’ mechanism and it remains in place until/unless an explicit message removes it.
- (b) (5 points) Soft-state protocols are more robust to network failures, in specific, router crashes. Since the soft state protocol uses periodic timers, the state can be re-created in a crashed router, by this periodic refresh. On the other hand, hard-state protocols do not recover gracefully from crashes, since they do not send periodic refreshes.
- (c) (8 points) Similar to RTP (the real-time transport protocol) we can use the concept of ‘scalable timers’. This concept states to keep the percentage of bandwidth allocated for control traffic fixed (e.g., keep control traffic to not more than 5% of the link bandwidth). As the number of states (that need to be refreshed) increases, the ‘frequency’ of refresh is decreased, and the refresh timers are ‘scaled’ (i.e., increased in value).
- The advantage is to achieve more scalability by reducing control overhead. The disadvantage is that the refresh period will be increased, and so recovery time (after failure or crash) is increased, and join latency (incurred if the join message is lost for example) will be increased.

Q9. What is the timer-suppression mechanism, and why is it used? Mention at least two mechanisms for multicast routing (either in IGMP or multicast routing protocols) that use such a scheme.

A9. The timer suppression mechanism is used to counter the ‘implosion’ problem that may occur when multiple recipients of a multicast message attempt to respond almost simultaneously. Each node receiving the message sets a randomized timer, while the timer is running it listens for multicast messages. If the information that it wants to send has already been multicast then it suppresses its own transmission. Otherwise, it transmits the message after the expiration of the timer (more details in the lecture).

IGMP uses this mechanism to suppress excess membership reports, and PIM-SM uses it to suppress multiple joins (from multiple downstream routers) on a LAN. Similarly PIM-DM uses it to suppress multiple Join override messages on a LAN. (2 examples are sufficient).