Frame Duplication and Misordering in RSTP Bridges

Tony Jeffree
Wednesday, July 5, 2000

This is a discussion paper for 802.1, addressing the potential issues of frame misordering and duplication in Bridged LANs containing Rapid Spanning Tree Bridges. It has been revised following discussion at the Ottawa 802.1 meeting in May 2000.

1. Background

In 802.1D:1998 (and earlier) Bridges using STP, the inherent delays imposed by STP before a Port is made Forwarding ensure that any frames of a conversation that were "in transit" prior to a reconfiguration event will have been delivered or discarded before any frames are transmitted on a Port that is made Forwarding as a result of the reconfiguration event. Hence, the risk of frame duplication and/or misordering on reconfiguration in LANs comprising legacy Bridges is small.

RSTP by its nature and intent can reduce the delay before a Port is made Forwarding to very small values; in the case of Root Port transitions, the delay can be as short as the hardware can manage it, or in the case of Designated Port transitions, as short as two frame transmission times on a single segment. Hence, it is conceivable that a reconfiguration event can take place, and a Port be made Forwarding as a result, while prior frames of a conversation are still in flight.

This paper attempts to examine these issues and evaluate the circumstances in which a problem might occur.

2. Frame Duplication

For a unicast frame, where the destination address of the frame has been learnt by the Bridges that may be required to forward it, that frame can only be buffered for transmission on one Port of one Bridge at any one time. Hence, it would appear to be impossible for frame duplication to occur in this situation.

If the unicast address has not been learnt, then the frame will be flooded (and therefore buffered) on all outbound Ports; there is therefore a possibility that a reconfiguration event could cause duplication of a flooded frame. Figure 1 illustrates this scenario. The figure shows a fragment of a Bridged LAN, consisting of three three-Port Bridges. The Root Bridge is assumed to be somewhere beyond Port 1 of Bridge A; the path costs associated with the various Ports of the three Bridges result in the configuration shown, with Bridge B Port 3 in Blocking and the remaining Ports all in Forwarding.

If the configuration is stable, then a unicast frame arriving at Port A1 and destined for an unlearnt unicast destination reachable through Port B2 will be flooded by Bridge A to its Ports 2 and 3, and by Bridge C to its Ports 2 and 3. However, as Port B3 is blocked, only the frame reaching Port B1 will be transmitted onwards through B2. Hence, a station reachable through B2 sees only a single copy of the frame.

If the configuration is not stable, then it is possible for the following sequence of events to occur:

a) Bridge A receives the frame through A1 and queues the frame on A2 and A3.
b) Bridge A transmits the frame through A2 and A3.
c) Bridges B and C receive the frame through B1 and C1, and queue the frame for transmission on B2, C2, and C3.
d) Bridge B transmits the frame through B2.
e) Bridge B detects that link A2-B1 has failed, and immediately places B3 in Forwarding as its new Root Port.
f) Bridge C transmits the frame through C2.
g) Bridge B receives a second copy of the frame through B3, and queues it for transmission through B2.
h) The second copy of the frame is transmitted through B2.

As can be seen from the above, any station downstream of B2 has seen the same frame twice. However, as any subsequent response from the destination station would cause the address to be learned, the worst case for "normal" LAN

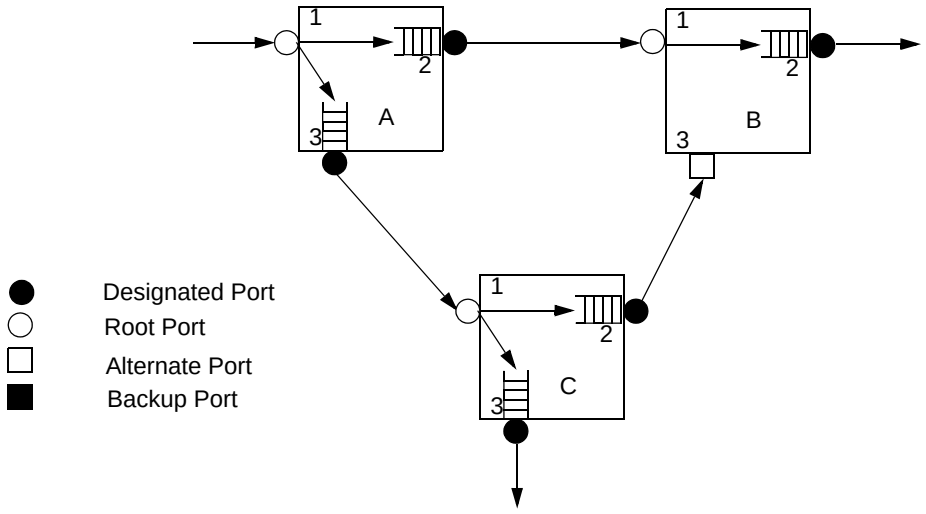

Figure 1—Frame duplication scenario

protocols arising from this scenario would be two connect requests being seen by the destination machine, as all subsequent components of the conversation would be based on learned addresses.

There may be some increase in the probability that this effect will cause duplication of unicast frames, due to the fact that the reconfiguration event will itself cause addresses to be flushed from the filtering databases of Bridges in the LAN, and the flushing will be repeated over the duration of the reconfiguration event. However, the effect of repeated flushing would be most marked if the frequency of TCN transmission is comparable to the typical time period for which a frame is buffered in a Bridge.

Multicast frames, as these will in general be buffered at multiple outbound Ports, are all potentially subject to duplication on reconfiguration events. However, of the multicast-based protocols that are in common use in LANs, the only one that would be sensitive to frame duplication is GARP, and GARP frames are not subject to frame forwarding in Bridges. Hence, this scenario seems not to be an issue.

3. Duplication – Conclusions

There is a small risk of duplication of unlearned unicast frames, and of multicast frames, on reconfiguration events. The risk will depend upon the configuration of the LAN, and the opportunity that the configuration gives for frames to be “in transit” in buffer storage on alternate paths to the destination(s) when a reconfiguration event takes place. The frequency of reconfiguration events should itself be very low in a Bridged LAN. Hence, a small risk of duplication occurring as a result of an event that itself is of very low frequency adds up to an extremely small risk of frame duplication actually occurring. For practical purposes, the contribution of this source of frame duplication to the behaviour of the MAC service appears to be negligible, and if it occurs, it does not appear to result in any damage to known LAN protocols.

4. Frame Misordering

It is possible to conceive of a situation where, prior to a reconfiguration event, an end-to-end conversation was required to transit the maximum diameter of the Bridged LAN, and following reconfiguration, the same conversation only transits a single Bridge. This could happen, for example, with a Bridge that has a Root Port connected to one end of a LAN “diameter” and a Alternate Port connected to the other end of the “diameter”, as in the stable configuration shown in Figure 2. In this configuration, a conversation is taking place between a station connected to A1 and a station connected to B2; as A2 is a blocked Alternate Port, all traffic between these two stations is relayed via Bridge C. Block the Root Port and make the Alternate Port Forwarding, and an almost instantaneous switch of location occurs for end stations downstream of that Bridge, from one “end” of the LAN to the other. As there could be frames in transit before the reconfiguration, frames transmitted immediately following the reconfiguration could arrive at their des-

tination before the ones in transit, resulting in frame misordering as seen by the recipient. The following sequence of events illustrates the point:

a) Frame 1 of a conversation is received through A1, buffered and transmitted through A3, received through C1 and buffered awaiting transmission through C2.
b) Link A3-C1 fails; Bridge A immediately places A2 in Forwarding as its new Root Port.
c) Frame 2 of the same conversation is received through A1, buffered and transmitted through A2, received through B1 and buffered awaiting transmission through B2.
d) Bridge C transmits Frame 1 through C2 and Bridge B transmits Frame 2 through B2.
e) Frame 1 is received through B3, buffered for transmission through B2, and transmitted through B2.

Clearly, the receiving station connected to B2 sees the frames in the reverse of the order in which the originating station transmitted them.

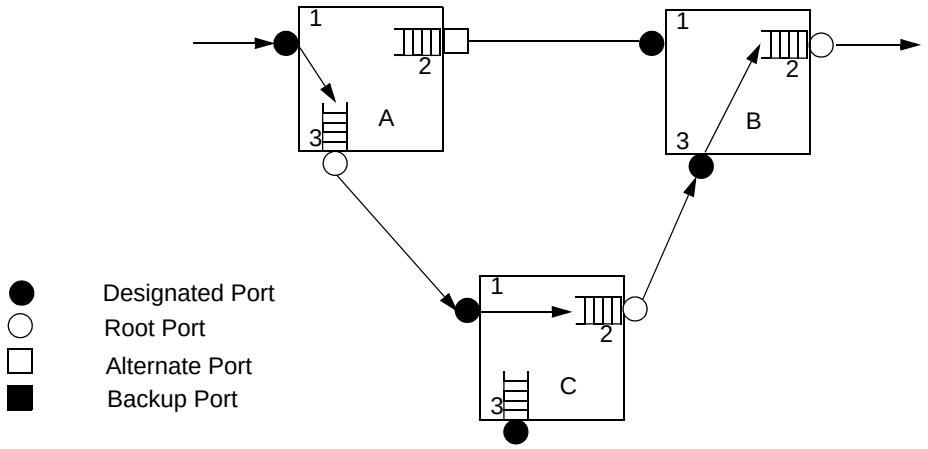

Figure 2—Frame duplication scenario

5. Misordering – Conclusions

As with frame duplication, there is a small risk of misordering of unicast and multicast frames on reconfiguration events. The risk will depend upon the configuration of the LAN, and the opportunity that the configuration gives for frames to be “in transit” in buffer storage on alternate paths to the destination(s) when a reconfiguration event takes place. The frequency of reconfiguration events should itself be very low in a Bridged LAN. Hence, a small risk of misordering occurring as a result of an event that itself is of very low frequency adds up to an extremely small risk of frame misordering actually occurring. For practical purposes, the contribution of this source of frame misordering to the behaviour of the MAC service appears to be negligible.

Some LAN protocols, for example LAT, LLC2, and NETBEUI, are sensitive to misordering, so even a low incidence of misordering could result in perceived problems in networks that support these protocols. There is no obvious solution within the operation of Rapid Spanning Tree that would remove this residual level of misordering. However, it is possible that by using protocol sensitive VLAN classification supported in a multiple Spanning Tree environment there might be the possibility of segregating any sensitive protocols on one or more VLANs that are carried over one or more legacy Spanning Trees. Whether this solution is practical will require further study.

6. Other considerations

The possibility that frames may be stored “in transit” on old routes after a reconfiguration has completed means that there is also a possibility that addresses will be mis-learnt, leading to a risk of denial of service for the addresses concerned. Hence, it is necessary to run short ageing timers for a period after a reconfiguration to allow such frames to be delivered or discarded.

The Bridge model described in 802.1D/P802.1w does not include the concept of queueing on input to a Port; however, practical Bridge designs generally include some input queueing. While this is not a solution to the effects descrived above, a Bridge should flush any input queue associated with a Port that becomes Disabled. This behaviour is consistent with the Bridge model; if a frame is in an input queue, it has not been received as far as the Bridge Port is concerned, and therefore should not be received after the Port becomes Disabled.