Running with scissors

Cut-through in bridged networks

Mick Seaman

This note describes some of the potential problems associated with cut-through in bridged networks.[1] The negative impacts of cut-through might be addressed, or at least reduced or contained to the point they are tolerable, by explicit restrictions on both network topology and how existing[2] and future protocols are used. A detailed specification of cut-through capable devices and the way they are deployed would be required. The restrictions described are compatible with the use of cut-through in rings of many bridgelets[3], which might be a suitable specification target.

1. Basics

A forwarded frame is ‘cut-through’, by a relaying network node such as a bridge, if outbound transmission begins before the inbound frame has been fully received. The aim is to reduce store and forward delay, which would otherwise be at least the time taken to receive the entire frame. Store and forward delays can dominate the total end to end delay when they exceed node to node propagation times, when there are many nodes in the end to end path, and when the design of the network topology and bandwidth allocation schemes minimize the queuing delay due to interfering traffic. Conversely the benefit of cut-through is low if the transmission speeds are so high as to reduce the effective wire-length of a propagating frames, where the network is richly connected (thus reducing end-to-end hop counts), and where best effort over-provisioning is used. The demand for cut-through is therefore highest in networks where there are good reasons for retaining existing cabling or cabling topologies. It is not the purpose of this note to argue the need and demand for cut-through, but to note that it exists, discuss the associated difficulties, and suggest solutions.

The essential cut-through problem is that onward transmission is begun before knowing the results of checks that can only be made when the entire frame has been received. Unknowns include, but are not limited to, whether the frame has been corrupted (to the extent detectable by checking the FCS) and whether frame is too long. It may be convenient to specify or at least allow the reception of a minimum number of octets before cut-through begins. The specified minimum might include just the MAC destination address, both the destination and source MAC addresses, any possibly present VLAN tag, the fields that can contribute to an IP-based ECMP flow hash, or (for simplicity) a minimum frame sized number of bytes. Any protocol that uses a trailer field has to be studied to see whether it can be used with cut-through, or whether strict conditions have to be imposed (including prohibition).[4]

If any protocol error condition,[5] indicating that the frame should not be forwarded, is encountered after cut-through has started, the frame could still be forwarded in its entirety or it could be truncated.[6] Clearly the intended recipients of a truncated frame should either simply not receive the frame or should be aware that truncation has occurred and not process the frame. Truncated or not, it is desirable that each relaying system understand whether the error occurred on the immediately attached LAN, or was detected by its neighbor or some prior system on the frame’s path. The difference can be signaled by using a reserved or algorithmically determined check sequence[7] or integrity check value.

[1] The discussion may appear long-winded, even tendentious. This is because I am attempting to work through the issues methodically and discover both issues and solutions I might have missed. The reader has my sympathy but not my apologies.

[2] Including, for example, LLDP.

[3] Similar to a TPMR, a bridging device providing cut-through between two ‘ring ports’ (each of which may aggregate a number of links, with all the ‘ring ports’ operating at the same speed) and one or more ‘station ports’, each providing connectivity to an end station. Traffic passing between a ‘station port’ and a ‘ring port’ may be cut-through, but if so their links are constrained to operate at the same speed. If the attached end stations are local this may be unnecessary and undesirable, a principal argument for cut through being the need to operate using existing cabling only capable of operating at lower speeds. This constraint should not apply to locally attached end stations. The advantage of requiring a store-and-forward step before finally delivering the packet to the end station is that the cabling plant can be fully managed, independent of the operation of each end station.

[4] Security protocols are particularly interesting, for reasons described later.

[5] Including but not limited to corruption detected by FCS checking.

[6] I use ‘truncated’ rather than ‘aborted’ to reflect the fact that next recipient of the frame might also cut-through the frame, and thus process the received fragment before realizing that it has been truncated.

Whether frames are being cut-through or not, a frame received by an intended destination (independent of whether that is a separate end station or a logical end station) is received completely before being processed by the apparently intended client.

2. Problems and answers

The issues raised by cut-through and the potential remedies don't necessarily fit into distinct non-interacting categories.[8] For a first cut we consider what should be done to address the potential consequences of the following:

—Forwarding[9] (part of) a corrupted frame.

—Forwarding (part of) an overlong frame.

—Forwarding (part of) a frame that purports to come from an authorized source, but has been modified or transmitted by another party.

—Inability to cut through (or usefully cut-through) frames of certain protocols.

At a minimum the occurrence of corrupted and over-long frames has to be detected and made visible to network administration. Detecting errors when frames are processed by end stations might not meet that requirement. Frame corruption should be localized to particular network segments,[10] so network administrators can locate and then address the root causes of the corruption. The remainder of this note assumes that the error rate on any particular segment is very low and uncorrelated with errors on other segments, so the probability of any given frame suffering repeated corruption is so low as to be ignorable. This assumption relies on prompt action from network administrators if problems are detected, including taking the network out of service if necessary. Given this proviso, and depending on the severity of the impact of each of the above on the operation of any particular protocol we might decide on one or a combination of the following:

a) Live with it. Acceptable if the effect on the forwarding of other frames (if any) is strictly time limited, and confined to only one of any route diverse paths used for interoperability. Applicable or possibly applicable to the following:

   1) Management counters incremented by the apparently correct cut-through frame, before corruption or truncation is detected. Typical bridging implementations do most of the intelligent receive processing of a frame in a single shot. Keeping track of each frame to correct counts that might vary on a frame-by-frame basis (such as unicast, broadcast, or multicast), when frames from other ports (or preempting frames) might intervene between processing the initial octets of the frame and completing reception, is an unnecessary implementation constraint. It is sufficient[11] to maintain (separate) counts of received corrupted and truncated frames.

   2) Bandwidth including bandwidth intended for use by frames of another flow, and consumed by a frame that has been corrupted in a way that makes it resemble a frame for that flow. Any attempt to recover any change made to reservation state is potentially complex, as it requires an assumption that there is surplus bandwidth to permit such a recovery without impacting other flows. On that assumption it would better to allow for additional bandwidth for use by flows pushed into non-conformance through bandwidth loss to corrupted frames, though such an over-flow decision would also add complexity by requiring cut-through specifications to constrain otherwise applicable flow policing techniques. A better approach is simply to accept the possibility of bandwidth loss and to enhance reliability through the use of completely disjoint diverse paths.

b) Mitigate it, by requiring additional or different protocol procedures in cut-through nodes. Applicable to the following:

   1) To help network management determine where corruption (possibly sporadic) is occurring, cut-through corrupted frame and truncated over-long frames should be marked so the following nodes on the path do not conclude that the corruption occurred on the immediately connected physical media. One possibility more such marking would be to modify the FCS to be both invalid but algorithmically dependent on the forwarded fragment in a way that is most unlikely to have occurred as a result of corruption. The physical media MAC technology standardization might define other marking method and, if FCS modification is used, should

[7] It might be the case that there is no obviously 'reserved value' but there is a 'in this case so obviously wrong that it must have been deliberately set wrong'.
[8] Not a problem limited to cut-through.
[9] 'Forwarding' includes all the side-effects (such as incrementing counters) associated with relaying a frame.
[10] i.e. instances of real physical media.
[11] As should be permitted by any standard

define the modification algorithm so it generates a truly unlikely value when the details of physical encoding and signaling are taken into account.

2) Learning incorrect station location information from a corrupted source address or VID can disrupt connectivity to a station for an indefinite period. Direct source address learning could be replaced by an alternate mechanism, or be restricted to frames addressed to the learning node and thus received in their entirety before processing.

c) Require store-and-forward for all frames of particular protocols, but permit cut-through for other protocols. This is unacceptable, as the subsequent transmission of the stored-and-forwarded frame can interfere with cut-through of following frames. That effect might be mitigated by limiting the store-and-forward to preemptable frames, but making class of service adjustments has potentially complex effects on network systems unaware of the use of cut-through in part of the network.

d) Require selected relaying nodes in a network to store-and-forward all protocols. This can mitigate some of the worst possibilities arising from a permissive approach [a) above] when a frame is corrupted as follows:

1) A unicast frame becomes multicast, and thus steals bandwidth from a large number of links.

2) A VID is changed, and the corrupted frame steals bandwidth from regions of the network from which the original frame was excluded. In a richly connected network with shortest path or engineered connectivity, the subsequent path of such a modified frame might traverse and steal bandwidth from segments of both (or all) of the disjoint paths used by other pairs of stations to provide resilient redundant connectivity. Note that a change of VID can result in the destination address being treated as unknown, leading to frame flooding.

3) A source address is changed, and loop-free operation of part of the network depended on source address recognition and subsequent frame discard. Although such operation is not specified by existing 802.1 standards we need to be aware of the possibility of future protocol standardization by ourselves and by others, the need to have an approach that will not give rise to unpleasant surprises in the future, and the difficulty of standardizing and enforcing comprehensive sets of prohibitions.

4) A hop count, of a protocol that relies on that count for temporary loop mitigation is changed, with consequent bandwidth loss in part of the network.

At a minimum there should be a least one store-and-forward node in each loop in the physical topology. That provision deals with d.3) and d.4) above, though not with all cases of multiple disjoint path bandwidth theft [d.2)] or with the combination of high port count bridges and d.1). These can be addressed by requiring store-and-forward in nodes with more than two ports attached to other forwarding nodes (i.e. excluding ports providing connectivity to end stations). While requiring store-and-forward only on certain paths (i.e. certain pairs of ingress and egress ports) through a bridge would seem possible, that complicates the issues described in e) below and their impact on the logical network topology and address scopes.

In some networks forwarding of any particular frame may be constrained to a single path, with forwarding only to specifically registered addresses. In those networks the topological restrictions in the immediately prior paragraph are unnecessary.

e) Prohibit protocol processing for certain protocols in cut-through nodes, i.e. forward frames those protocols transparently (or simply discard them) as if the node was operating at a lower sub-layer (e.g. as if the node was a TPMR, with the protocol peers in Customer Bridges; or, equivalently, as if one or a succession of cut-through nodes behaved as shared media). Considerations include the following:

1) It may not be possible to both process and cut-through MACsec protected frames when confidentiality is being provided (see 3. below for discussion). CFM is another candidate protocol requiring further study.[12] Both MACsec and CFM protect the operation of other protocols, and thus provide a further reason [beyond the discussion of c) above] for using store-and-forward for all protocols processed in nodes that store-and-forward some frames (as opposed to being the destination of those frames, and taking some further protocol action).

[12] I have left detailed analysis of CFM to others; it may be argued that important CFM frames are so short as to be readily accommodated by requiring reception of a minimum number of octets prior to initiating cut-through, with no difference in CFM frame handling when cut-through is enabled and when it is not.

2) Transparent forwarding of one or more protocols, as described in e) above, does introduce a new sub-layer and an associated addressing scope. Frames sent to the Nearest Bridge group address (used by LLDP to support Power over Ethernet (PoE) will no longer reach the "Nearest Store-and-forward Bridge".

We need at least one additional IEEE Std 802.1Q Reserved Address, with a scope that lies between the existing "Individual LAN Scope/Nearest Bridge group address" and the "Nearest non-TPMR Bridge".[13] Use of that address by protocols whose immediate peers need to provide store-and-forward necessarily involves some implementation change for those protocols.

In principle additional sub-layers could be introduced, with associated address scopes e.g. "Nearest Store-and-forward Customer Bridge"[14]. This poses the problem of running out of 802.1Q Reserved Addresses.[15] However cut-through is not necessarily desirable in all environments so we may be able to avoid defining a Provider Bridge cut-through sublayer and an associated address.

If the use of cut-through is restricted to nodes with no more than two ports attached to other relaying nodes [see discussion of d.4) above] we may be able to consider cut-through as only operating at a sub-layer logically positioned between a TPMR and physical media. A potential difficulty is providing suitable addresses and protocol procedures to share of VLAN information with cut-through bridges that need it to restrict frame propagation.

f) Prohibit the use of particular protocols in networks or network regions with cut-through. Applicable considerations include the following:

1) The first question is are there any protocols that are candidates for such a drastic measure other than those with trailer fields that have to be processed before forwarding [see c above for the associated ill-effect], or whether potential candidates protocols can be accommodated by their transparent forwarding through cut-through regions with appropriately scoped addressing for their control planes such that participating protocol peers are positioned in store-and-forward nodes, as described in e) above.

g) Specify a new type of tag that adds a check sequence or an cryptographic integrity check value to protect the initial octets of each frame, including those that affect forwarding actions. Each potentially cut through frame would include the tag value included within those initial octets, and cut through would not begin until they were received and the tag checked. Considerations include the following:

1) This is unlikely to be an attractive solution as it would require additional support at the store-and-forward nodes neighbouring cut-through nodes.

3. Security considerations

Frame integrity, and data origin authenticity are provided by MACsec and similar protocols by a cryptographic computation over the frame data that adds a trailing tag [called the Integrity Check Vector (ICV) in IEEE Std 802.1AE. Successful validation of the tag requires processing of the entire frame,[16] and confirms that the frame has not modified since the tag was added and that the system adding the tag possessed the cryptographic key used to generate the tag—and was thus an authenticated and authorized party (or at least a party trusted by the party that participated in the initial authentication and subsequent key agreement procedures used to generate or distribute that key). Until the entire frame has been received there is no guarantee that the frame has not been generated or modified as part of an attack. Where the ill-effects of frame corruption described in 2 above might be dismissed as occurring rarely, an adversary with access to the physical media might cause them at will as cut-through frames are necessarily forwarded before the tag/ICV is checked.

Selecting confidentiality as part of the cryptographic protection in an attempt to make it hard for an adversary to perform a targeted attack is a poor defence. The GCM based Cipher Suites that are most commonly used (and specified for MACsec) are not proof against nonce reuse. If a nonce is reused for two different frames the XOR of the confidentiality

[13] For a detailed description

[14] This is just by way of example. A better approach would probably be to define the current "Nearest Customer Bridge" as applying to Customer Bridges that use store-and-forward exclusively (and thus meet the current Customer Bridge specification) and introduce a new address (if necessary) for any intervening Customer Bridge like device using cut-through.

[15] See http://www.ieee802.org/1/files/public/docs2015/ae-seaman-ede-address-scopes-1115-v02.pdf for an extended discussion of Reserved Address scopes.

[16] This would be the case even if the tag were computed when the frame was first constructed and the tag added to a reserved field in the initial octets of the frame.

protected ciphertexts will be the same as the XOR of the corresponding plaintexts. It has to be assumed that an adversary can make well informed guesses as to the plain text of at least some frames, and will therefore (by reusing the nonces of those frames) be able to construct a targeted attack.

Given the attack surface offered by cut-through there seems little point in supplementing the MACsec protection with additional procedures to be followed by a node that has begun to cut-through a frame and then finds it carries an invalid tag/ICV. While that would help locate attacks it comes at the cost of requiring hop-by-hop security processing at each cut-through node without the usual defensive benefits of that approach. It would be possible to use the approach described in 2.g) above in which the initial octets (the first 64, perhaps, which include the fields most likely to cause network disruption) of a frame are protected by an ICV inserted into the frame and the frame not cut through until those octets are integrity checked, but deployment of such an approach is likely to be difficult, particularly in the lower bandwidth cost sensitive environments where cut-through is most desired. The most feasible approach is to forward MACsec protected frames without any security processing, truncating and marking any FCS errors just as if they were unprotected frames.

Since the scope of MACsec protection would be between store-and-forward devices, intervening cut-through bridges that wished to limit the propagation of frames based on their content would need the cryptographic protection to be limited to protecting integrity (rather than including confidentiality) and would need to know how to skip the SecTAG when interpreting frame fields.

4. Conclusion

The use of cut-through carries a risk of network disruption when frames are corrupted as the FCS is not checked before forwarding begins. The consequences of forwarding corrupted frames can be mitigated by topological restrictions, requiring at least one store-and-forward node in any physical node and at branch points. Fortunately these restrictions are not onerous in topologies where cut-through is most desirable—long chains or rings of forwarding nodes with attached end stations. Using cut-through does reduce the options available when securing a network and identify where an attacker has gained access to the network. Restrictions on protocol processing in cut-through nodes does impact the use of 802.1Q Reserved Address as they define address scopes based on the location of neighbouring protocol participants.

In general network protocols need to be examined on a case-by-case basis to see how, and if, they should be supported in a bridge providing cut-through. Some protocol functionality (e.g. learning station locations) may need to be supported in a different way. The need to state and restate restrictions and exceptions when cut-through is being used may be onerous in the more general 802.1Q specification, with a likelihood of omissions, oversights, and over optimism as well as missing opportunities for simple solutions tailored to cut-through topologies. A separate specification focused on the specific areas where cut-through is believed to be required would seem a better approach.