Fast Chain Recovery with Traffic Engineering

Bob Sultan (bsultan@huawei.com)

Deng Zhusheng (dengzhusheng@huawei.com)

***Abstract*—A proposal was made at the March 2008 IEEE 802.1 Plenary Meeting to initiate a project whose purpose is to reduce spanning tree convergence time in VLAN Bridged networks by using a Fast Chain Recovery (FCR) protocol in regions of the network that can be decomposed into chains of Bridges. The chain topology is provisioned by the network operator. This paper is intended to show that once such chains are provisioned, they provide a natural means of supporting Fast Re-Route (FRR) of Traffic Engineered (TE) traffic such that Ethernet Switched Path (ESP) connectivity persists in the presence of a single failure in each of the chains through which the ESP passes. This scheme, which we call FCR-TE, provides more robust protection than TE Service Instance (TESI) end-to-end protection currently described by IEEE 802.1Qay. These two types of protection are independent and can be supported concurrently on a set of Bridges. FCR-TE uses the topology and notification functions supported by FCR and adds additional protocols and algorithms appropriate for the TE environment, as described in this paper. a subset of the algorithms and protocols supported by FCR. We suggest that once chains are provisioned within a Bridged LAN, the scope of the use of those chains should not be limited to VLAN traffic, but should include TE traffic.**

I. Introduction

Virtual Local Area Network (VLAN) Bridging (IEEE Std. 802.1Q) requires the blocking of Bridge ports such that network connectivity is limited to a tree topology as shown in Figure 1. That is, for a given VLAN, there is only a single active path that may be taken from one bridge to another, where both bridges participate in the VLAN. This restriction prevents the looping of frames within the network.

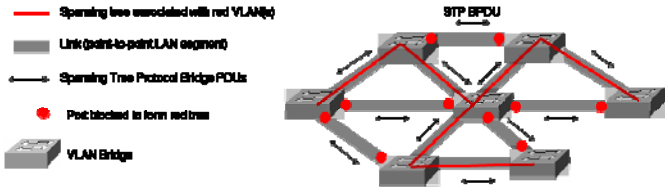

Figure 1: STP prevents looping of VLAN frames

The establishment of the spanning tree is accomplished through the use of one of several variations of a Spanning Tree Protocol (STP). When a Bridge or link fails, connectivity may be disrupted during the time required to perform the distributed tree computation. In general, time required for the STP to converge is related to the size of the network.

A proposal [1] was made at the March 2008 IEEE 802.1 Plenary Meeting to initiate a project whose purpose is to reduce spanning tree convergence time in VLAN Bridged networks by using a Fast Chain Recovery (FCR) protocol in regions of the network that can be decomposed into chains of Bridges. The chain topology is provisioned by the network operator.

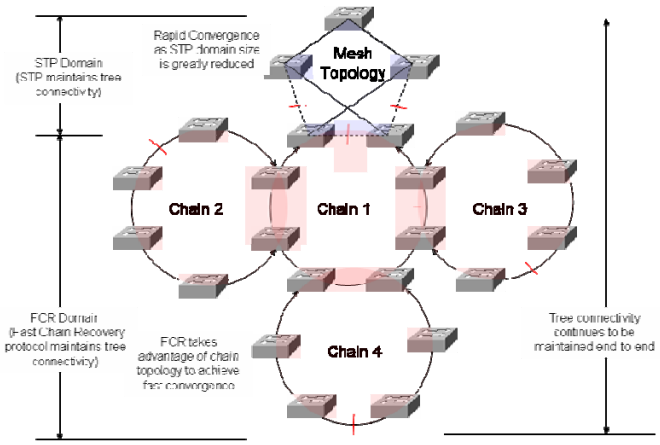

Figure 2: Reducing convergence time with FCR

Figure 2 shows a network containing a chained domain comprising three chains and a meshed domain. Proper operation of VLAN Bridging requires the establishment of the establishment of a tree spanning all bridges associated with a VLAN. FCR maintains tree connectivity within the chain domain while STP maintains tree connectivity within the mesh domain. FCR converges rapidly. Assignment of some portions of the network to chain domains reduces the size of the remaining portions of the network in which STP is applied. Hence, convergence time is reduced.

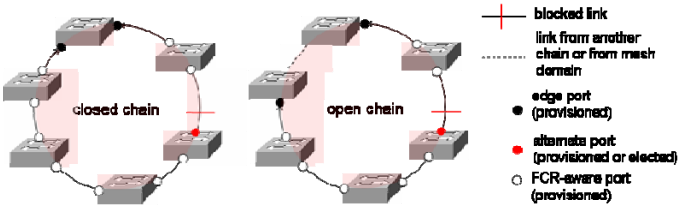

Figure 3: FCR Chain and Port types

As shown in Figure 3, chains may be *closed* or *open*. Both ends of a closed chain terminate on the same bridge. The port that is the endpoint of a chain is an edge port. Other ports on

the chain are FCR-aware ports.

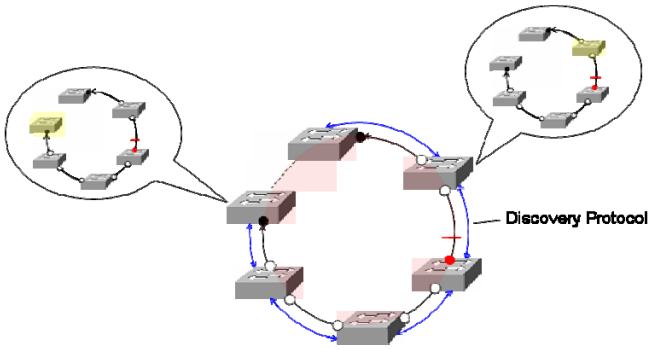

Figure 4: Replicated view of Chain connectivity

FCR requires provisioning of the topology of each chain. During operation, bridges learn the status of links and bridges on the chain by exchanging information with neighbor bridges. Combining this information with the provisioned topology information, each bridge can construct a map of connectivity within the chain as shown in Figure 4.

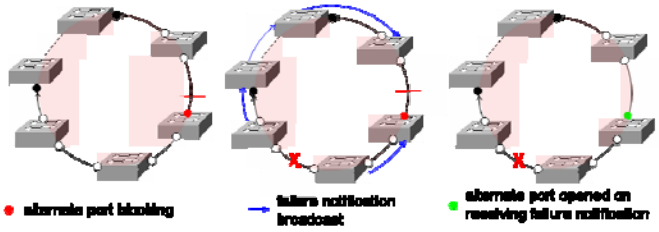

Figure 5: Opening the Alternate Port on failure detection

FCR further requires that one port within the chain, known as the alternate port, be selected, by provisioning or by election, as the port that will be placed in blocking state in the absence of faults on the chain. A Bridge discovering a local fault reports that information by broadcast notification to other bridges on the chain. On learning of such a fault, the alternate port is opened, as shown in Figure 5, as its closure is no longer necessary for loop prevention. Topology Change Notification (TCN) is propagated in order to allow Bridges to flush cached Forwarding Database (FDB) entries to force learning of MAC addresses whose associated paths have changed.

II. Traffic Engineering

IEEE Std. 802.1Qay (PBB-TE) describes the method by which a Bridged network can support Traffic Engineered (TE) services as well as VLAN services. PBB-TE specifies that traffic is carried via a unidirectional Ethernet Switched Path (ESP). The ESP may be point-to-point (P2P) or point-to-multipoint (P2MP). A P2P ESP carries traffic from a *source* to a *destination*. A P2MP ESP carries traffic from a *root* to some number of *leaves*. Forwarding along a P2P ESP is performed exactly as the forwarding of a VLAN frame carrying an Individual MAC address *except* that Forwarding Database (FDB) entries are *provisioned* rather than *learned* and failure to locate an FDB entry matching the <DA, VID> of the frame results in frame discard rather than broadcast. Forwarding along a P2MP ESP is performed exactly as the forwarding of a VLAN frame carrying a Group MAC address *except* that Forwarding Database (FDB) entries are always *provisioned* and are not established dynamically via the MRP Multiple Registration Protocol (MMRP). Thus, in all cases, the path of traffic associated with an ESP is determined by provisioned FDB entries of the form '<DA, VID> → outbound port(s)'. P2P ESPs having different SA values can share the same <DA,VID> (ie., share the same FDB entry). Such an ESP is distinguished from other ESPs having the same <DA, VID> value by the SA value. Thus a P2P ESP is uniquely identified by the 3-tuple <ESP_DA, ESP_SA, ESP_VID>, where the ESP_DA is an Individual MAC address. A P2MP ESP is uniquely identified by the 3-tuple <ESP_DA, ESP_SA, ESP_VID>, where the ESP_DA is a Group MAC address but ESPs with different values of ESP_SA cannot share the same value of <ESP_DA, ESP_VID>.

A key characteristic of 802.1Qay is that TE traffic can coexist in the same network with VLAN traffic. The roles are partitioned by VID value. VIDs are provisioned as being used to identify a VLAN or as being used as part of the ESP identifier.

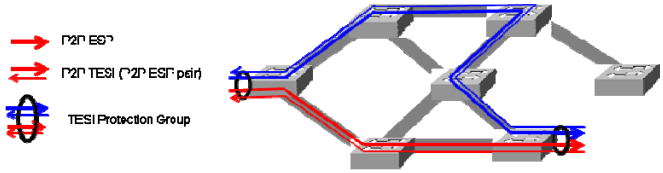

Figure 6: TESI Protection Group

TE service is supported through the use of a bidirectional TE Service Instance (TESI). A P2P TESI consists of a pair of P2P ESPs having the same endpoints but different directions of traffic flow. A P2MP TESI consists of a P2MP ESP carrying traffic from root to leaves, and a P2P ESP carrying traffic from each leaf to the root. A P2P TESI can be protected end-to-end by pairing TESIs having the same endpoints but disjoint paths. The pair of TESIs is called a TESI Protection Group (PG) as illustrated in Figure 6.

III. FCR-TE Provisioning

As previously noted, a key objective of FCR is to reduce spanning tree convergence time in VLAN-based networks. TE services do not require the establishment of a spanning tree for loop prevention as the path of each ESP is provisioned. However, an extension to FCR, called FCR-TE, can protect TE traffic from a single failure in each chain through which an ESP passes. This method of protection has advantages as compared to the end-to-end TESI protection currently specified by IEEE 802.1Qay.

802.1Qay specifies that a P2P TESI can be protected from the failure of a single element in its path. This is done by forming a TESI PG from two TESIs having endpoints in the same two Bridges. A failure along the path of one TESI in the PG will cause traffic associated with that TESI to be switched to the other TESI in the PG, as illustrated in Figure 7. Such protection switching is controlled at the endpoints of the PG.

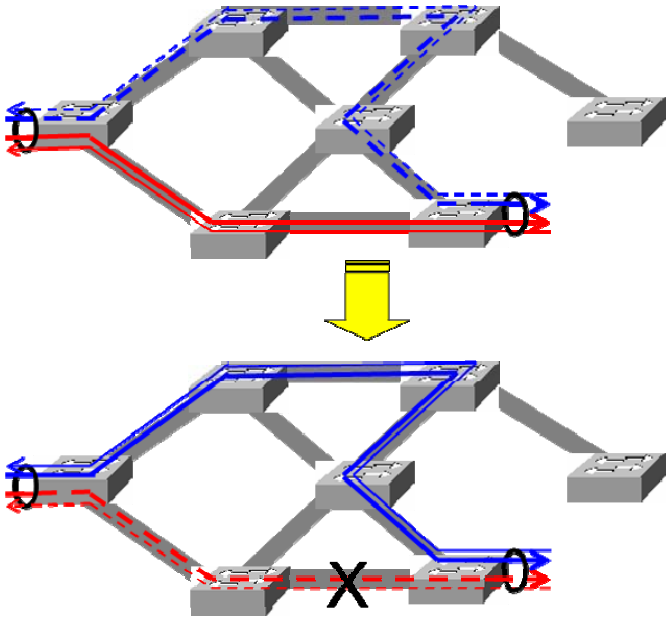

Figure 7: TESI end-to-end Protection

End-to-end TESI protection has the limitations that (1) connectivity will not survive a single fault occurring in both TESIs of the PG, (2) failure notification must propagate to the PG endpoint before protection action can be taken, and (3) protection actions will cause traffic load to change along the entire path of the TESIs.

FCR-TE describes modifications to FCR making it a suitable method to protect an ESP from a single failure in *each* chain through it passes. FCR-TE does not perform *protection switching*, as protection switching is an activity performed at the endpoints of a connection entity. Instead, FCR-TE uses a method of Fast Re-Route (FRR) to bypass faults in a manner that is not visible to the endpoints of the connection entity. Thus, this method of protection is entirely independent of TESI protection switching described in IEEE 802.1Qay. Either one, or both, can be deployed in a TE-capable network.

We introduce several terms useful in describing FCR-TE. An *Ingress Bridge* is the Bridge by which traffic associated with a specific ESP enters a chain. An *Egress Bridge* is a Bridge by which traffic associated with a specific ESP exits a chain. A *Transit Bridge* is a Bridge that is neither the Ingress Bridge nor the Egress Bridge of a chain with respect to a given ESP.

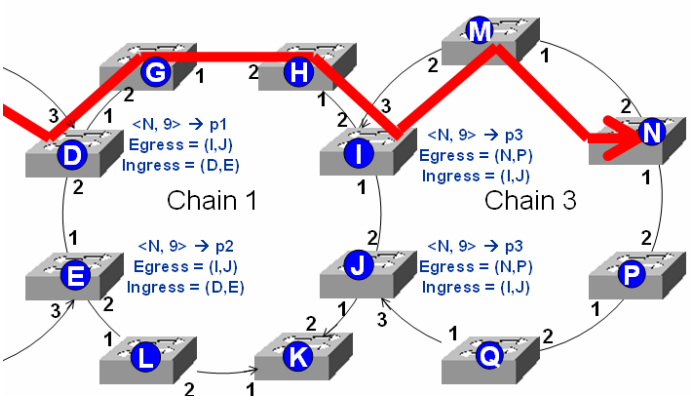

Figure 7: Provisioning an ESP across a Chain

In Figure 7, Bridge D is the Working Ingress Bridge for ESP <N, A, 9> in Chain 1, Bridge E is the Protection Ingress Bridge, Bridge I is the Working Egress Bridge, and Bridge J is the Protection Egress Bridge. Numbering denotes the local port identity.

No explicit per-ESP provisioning is required on Bridges that are transit Bridges with respect to an ESP. Instead, each Bridge on the chain, except for a Chain Edge Bridge on an Open Chain, is provisioned with a 'wildcard' FDB entry <*, VID> → (p1,p2) for each VID value that has been provisioned for TE usage. The asterisk denotes 'any MAC Address for which a more specific FDB entry does not exist' and p1 and p2 are the two Bridge ports lying along the chain. A frame arriving at a Bridge provisioned with such a wildcard entry, and no FDB entry exactly matching the DA and VID, will be forwarded from one chain port to the other.

Thus, the use of chains can reduce the provisioning required to establish the route of an ESP. The chain is defined once, and individual ESPs associated with that Chain need not be provisioned in the FDB of Bridges along the Chain through which they pass.

Per-ESP provisioning of FDB entries is required at the Ingress Bridge and Egress Bridge of an ESP with respect to a chain. The entries provisioned are exactly those that would be provisioned in the absence of chains. In xxx, the Ingress Bridge D is provisioned with the entry <N, 9> → p1, where P1 is a Chain port of Chain 1, and the Egress Bridge I is provisioned with the entry <N, 9> → p3, where P3 is *not* a Chain port of Chain 1.

As described thus far, provisioning of the ESP in Ingress and Egress bridges is identical to provisioning that would normally be performed at bridges along the path of an ESP. Per-ESP provisioning is eliminated at transit bridges.

In addition to provisioning the FDB entry at the Ingress Bridge, it is necessary to provision the identity of the Egress Bridge associated with the ESP and Chain. Knowing the identity of the Egress Bridge allows the Ingress Bridge to

determine whether a fault reported on the Chain lies on the provisioned path of the ESP. If so, the Ingress Bridge can initiate FRR by changing the identity of the outbound port in the FDB entry to forward ESP traffic in the opposing direction on the Chain.

Additionally, the ESP must be protected from failure of the Ingress or Egress Bridge. This is done by identifying both a Working and a Protection Ingress Bridge and Egress Bridge. The Working Ingress Bridge performs the Ingress Bridge functions when the Working Ingress Bridge is operational. Otherwise, the Ingress Bridge functions are performed by the Protection Ingress Bridge. Only one of the two Bridges is Active at any time. Working and Protection Egress Bridges operate similarly.

Each Ingress Bridge is provisioned with the identities of the Egress Bridges associated with the Chain and ESP. In each case, the Working Egress Bridge is listed before the Protection Egress Bridge. Each Ingress Bridge is also provisioned with the identities of the Ingress Bridges associated with the Chain and ESP. In this way, the Working Ingress Bridge knows the identity of the Protection Ingress Bridge and vice versa. Finally, each Ingress Bridge is provisioned with the identity of the local port on which frames associated with the ESP are forwarded from the Ingress Bridge. This value is installed in the FDB for the Active Ingress Bridge.

IV. FCR-TE Operation

In the absence of faults on the chain, traffic associated with the ESP is forwarded as shown in Figure 8.

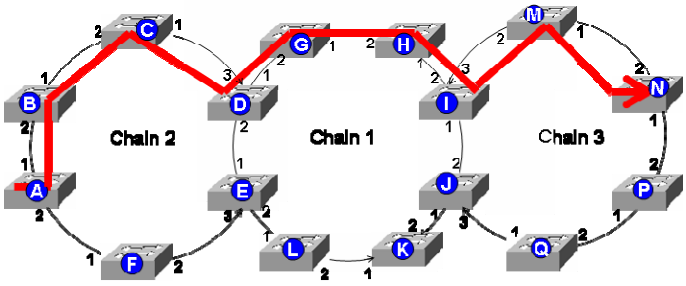

Figure 8: ESP Forwarding in the Absence of Faults

When a fault occurs along a Chain, notification is broadcast to Bridges on the Chain using FCR protocols. The Active Ingress Bridge knows the identity of the Working and Protection Egress Bridges. It can infer the identity of the Active Egress Bridge as it knows the status of the Working Egress Bridge. If the Working Egress Bridge is operational, it is the Active Egress Bridge; otherwise the Protection Egress Bridge is the Active Egress Bridge.

The Active Ingress Bridge, having a map of the chain topology, can determine whether the fault lies between the Active Ingress Bridge and the Active Egress Bridge in the direction in which ESP traffic is being forwarded by the Active Ingress Bridge. The direction is inferred from examination of the outbound port value in the FDB entry for the ESP in the Active Ingress Bridge.

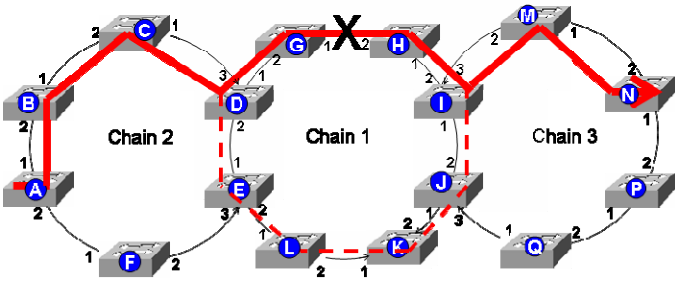

Figure 9: Fault on Provisioned Path of the ESP

If the fault does lie between the Active Ingress Bridge and the Active Egress Bridge in the direction in which ESP traffic is being forwarded by the Active Ingress Bridge, then the outbound port field of the FDB entry is changed to the opposing Chain port, effectively steering traffic in the opposing direction on the Chain as shown in Figure 9. When the Active Ingress Bridge learns that the fault has cleared, and no fault exists on the chain, the FDB entry reverts to its originally provisioned value, possibly after the expiration of a Wait-to-Restore timer.

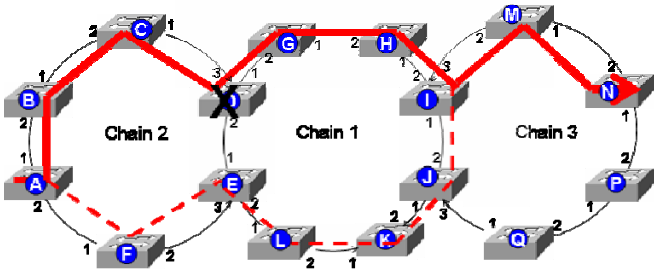

Figure 10: Failure of the Working Ingress Bridge

When the Working Ingress Bridge is the Active Ingress Bridge, a failure of the Working Ingress Bridge will be observed by the Protection Ingress Bridge. The Protection Ingress Bridge will then assume the role of Working Ingress Bridge. The Protection Ingress Bridge installs an ESP-specific FDB entry with an outbound port value that will carry traffic from Active Ingress Bridge (the Protection Ingress Bridge) to Active Egress Bridge in the direction that will bypass the failed Working Ingress Bridge, as shown in Figure 10. When the Protection Ingress Bridge learns that the Working Ingress Bridge is operational, the Working Ingress Bridge reverts to its role as the Active Ingress Bridge, possibly after the expiration of a Wait-to-Restore timer.

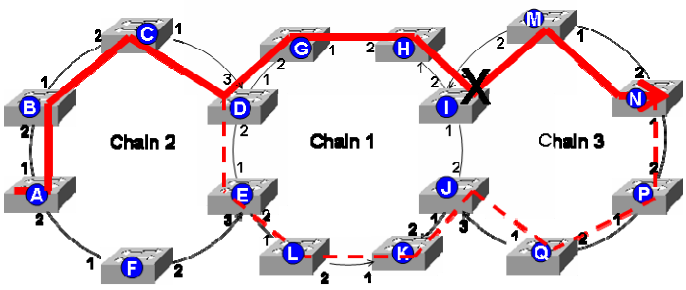

Figure 11: Failure of the Working Egress Bridge

Failure and restoration of Egress Bridges operates in a similar manner, as shown in Figure 11. If the path of the ESP takes it to an adjacent chain, then the Egress Bridge is an Ingress Bridge with respect to that adjacent Chain. In this case, the value of the FDB entry is determined by the Ingress Bridge function associated with the adjacent chain.

The Ingress Bridge associated with an ESP entering a chained Region from an unchained Region is not protected. That is, there is a Working Ingress Bridge, but no Protection Ingress Bridge. Thus, a failure of this Ingress Bridge results in a loss of ESP connectivity, just as would occur as the result of the failure of an unchained bridge on the path of the ESP. A TESI PG can be deployed to protect the TESI from such a failure. The same is true of an Egress Bridge by which an ESP exits a chained region and enters an unchained region.

V. CONCLUSIONS

FCR decomposes a Bridged LAN into Regions containing Chains in order to reduce spanning tree convergence time. The same chains can be used to support Fast Re-Route to bypass faults on a Chain. This method provides protection against a single failure in each chain through which an ESP passes.

REFERENCES

[1] Fast Recovery for Chains and Rings, Norman Finn http://www.ieee802.org/1/files/public/docs2008/new-nfinn-fast-chains-rings-par5c-0308-v1.ppt.