Can critical real-time services of public infrastructures run over Ethernet and MPLS networks?

Jaime Lloret[1], Francisco Javier Sanchez[2], Hugo Coll[3], Fernando Boronat[4]

[1,3,4]Department of Communications, Polytechnic University of Valencia, Camino Vera s/n,
46022, Valencia (Spain)
[2]ADIF, Molino de las Fuentes s/n, 46026, Valencia (Spain)
[1]jlloret@dcom.upv.es, [2]jsbolumar@adif.es, [3]hucolfer@posgrado.upv.es,
[4]fboronat@dcom.upv.es

Abstract. — With the adoption of Ethernet-IP as the main technology for building end-to-end real-time networks, the requirement to deliver high availability, quality and secure services over Ethernet and MPLS has become strategic. Critical real-time traffic is generally penalized and the maximum restoration time of 50 msec. sometimes is exceeded because of devices hangings or the high delay to find an alternative path. Existing standard networks are not ready to transport real-time critical information. This issue must be solved specially in critical infrastructures such as railway control systems, because passengers' safety could be committed. In this work, first we have checked in which conditions the devices of the network will have a reply in real-time. Then, we will explain why existing Ethernet and MPLS combined solutions don't offer recovery time lower than 50 msec., and we will propose a new approach that achieves these conditions. Our statements are based on real measurements in real railway environments and using real testbeds.

Keywords: real-time networks, critical systems, Ethernet, MPLS.

1 Introduction

Distribution and transport companies (gas, oil, water, railway, etc.) often have their own voice and data communication infrastructure, usually based on SDH/SONET transmission systems over fiber or copper cables. Only few years ago, industrial systems were "islands" for control and system safety, but now they are integrating network technologies, to permit easy growing, and central management. We can distinguish two main environments depending on the type of service: non critical real-time services and critical real-time services. Normally, non critical real-time services are based on Ethernet access to corporate MPLS/VPLS network. The main structure and technology for this network responds basically to multimedia traffic requirements, but the behaviour is completely different when we are connecting critical real time systems to the same network. In fact, to avoid occasional blocking problems, critical real-time services are provisioned through independent local Ethernet networks (most of them in ring topology to provide resiliency), with no interaction with corporate MPLS/VPLS network. These solutions increase complexity and budget, and present many difficulties for integrated management on critical infrastructure companies.

In the case of railway systems (where we are going to do our tests and simulations), non critical real-time services are provisioned through direct access to corporate MPLS network, such as railway station auxiliary services: automatic stairs, lifts, evacuation and emergency facilities, IP phones, video, remote alarm and supervision, etc. Station services and equipments are connected to third level nodes (usually MPLS routers) organized by a double ring topology. Both ends of the ring are connected to regional gigabit ring, through physical link between third level node and second level node (this is typically a layer 3 switch).

On the other hand, critical real-time services are provisioned through independent Ethernet rings with resiliency mechanisms activated: level 2 methods based on Ethernet common topologies (like Spanning Tree Protocol -STP-, Rapid STP, Multiple STP and link aggregation), or specific ring topologies (RPR/802.17, EAPS/RFC3619) and level 3 methods based on interior routing protocols (EIGRP, ISIS, OSPF) and default-gateway backup protocols (VRRP, HSRP). [1]

A real example of a complete railway control system network is shown in figure 1. The Circles represent routers in each station interlocking system, the rectangles represent switches using HSRP protocol and the squares represent interlocking CPUs. Dotted lines are the backup network lines. [2]

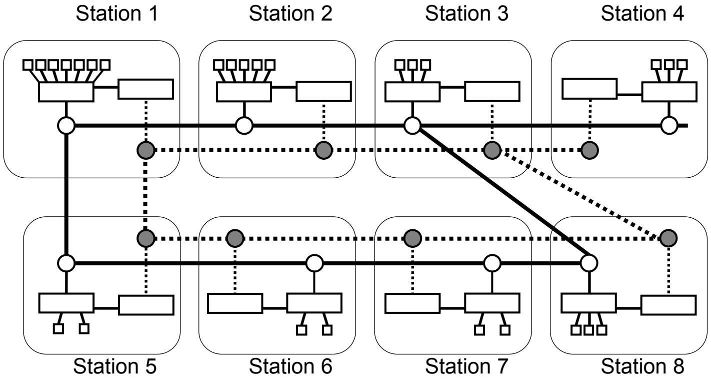


Fig. 1. Network topology of a railway line critical control system

In critical real-time networks, any type of failure must be recovered in less than 50 msec. The non achievement of this goal could suppose to lose lives, such as in critical railway control system networks [3]. It is vital to study the viability of running critical real-time services over standard networks such as Ethernet and MPLS. This goal will allow the companies to use a unique network infrastructure.

This paper is structured as follows. Section 2 describes last methods for protection and restoration in Ethernet and MPLS networks. Section 3 shows the measurements and analysis of real Ethernet and MPLS scenarios on the Spanish railway administrator (ADIF) company. Section 4 presents our proposal to improve the availability and fault tolerance for critical real-time systems using the integrated network. Finally, section 5 summarizes the conclusions and gives future works.

2 Failure detection, protection and restoration in Ethernet and MPLS networks

When there is a failure in the network, there must be a way to detect it, so the recovery operation must start. Failure detection depends on the type of failure and may be done by an adjacent node to the failed one or by a configured point of repair in the network. Failure detection involves multiple layer mechanisms such as [4]:

- Hardware: connectivity failures can be detected by line cards.
- Signaling: errors in SDH/SONET are detected using signaling mechanisms.
- Link: specific messages can be sent to poll node state or to notify an error.
- Network: when there is a failure, “hello” messages are missing.
- Application: SNMP traps can be used to notify a failure to a central manager.
- Specific protocols: failure detection could also be based on a specialized protocol like BFD (Bidirectional Forwarding Detection) [5].

The design and reconfiguration of the network, regarding with the components availability, is fundamental to deliver the required performance and reliability to real-time services. There are complex methods based on two complementary approaches to avoid the failure impact of the network components. The first one is to design high availability working paths, where the resilience is provided by selecting a working path with high connection availability. Few protocols include in their metrics availability factors, but usually data link protocols don’t include any. The second one is the use of backup paths. There are two main methods for resiliency provisioning:

- Protection: static scheme in which backup paths are pre-computed and pre-provisioned before the failure occurs. Protection can be done in two different forms: dedicated protection and shared protection. In dedicated protection, the resources used for protection are dedicated for backup traffic. There are two types of dedicated protection (1+1 and M:N protection). In shared protection, when a backup path is activated, other backup paths, that share resources with it, will have to be rerouted. Shared protection is more complex to maintain dedicated protection, but it can offer higher network utilization.
- Restoration: dynamic scheme in which backup paths are discovered dynamically due to failed connections, so they are created and routed after the failure.

In both cases have one mechanism to detect the failure and notify it to the rest of nodes and another one to re-route traffic from the working the path to a backup path.

Recovery schemes can also be categorized by the scope of the recovery. Global recovery covers link and node failure by calculating a new unique end-to-end path (this recovery action run in the ingress node), while with local recovery faults are handled locally by the neighbors. Keeping failures local cause minimum disruption in the traffic, but it is necessary to reserve several paths. The topology of the network affects on how the recovery can performs [6].

Network recovery management is difficult when the network is viewed as an unstructured collection of backup paths. These schemes are based on building a set of sub-topologies of the network, serving as a more intuitive abstraction of the recovery paths, and can be applied to both, local and global recovery. The main methods are:

- Self-healing rings. In these methods each node in the network is connected with two links to its upstream and downstream neighbours. In regular operation, the traffic is sent in one direction of the ring. When a link fails, the traffic is sent to the other link in the reverse direction, so the failed link is avoided. There are several mechanisms that rely on ring topologies, like SONET/SDH or Ethernet ring solutions (RPR/802.17 [7] and EAPS/RFC3619 [8]).

- Disjoint paths. These methods are based on a graph property described by Menger [9]. These solutions are applied to MPLS networks, because they can be adapted easier to MPLS signalling. This technique uses pre-provisioned backup paths with local recovery [10].
- Protection cycles. The goal is to provide fast recovery speed (usually offered in rings) to mesh networks. In order to combine the best properties from both ring and mesh, a new concept called p-cycles has been introduced. It is based on the formation of closed paths which are formed in the spare capacity only of a mesh restorable network. When a link fails, the traffic is locally switched or routed according to the cycle instead of to the original shortest path [11].
- Redundant trees. These methods are applied to standard Ethernet access networks. Ethernet traditionally relies on Spanning Tree Protocol (STP) defined in IEEE 802.1D, which provides loop-free communication between all nodes. The restoration after a link failure involves the reconstruction of the whole spanning tree. Typically, the recovery of the worst case requires around 50 seconds, which is not acceptable for real-time critical applications. In Rapid STP convergence, the time is reduced but is still in the order of few seconds. Multiple STP defines several spanning trees instances organized in regions and map each Virtual LAN onto a single spanning tree. MSTP restoration time is still high [12].

3 Measurements and analysis in Ethernet and MPLS networks

Our first goal is to know which delays are introduced in the network because of the data transmission through the wire and because of the devices placed inside. We will give a law that shows which conditions we should follow when we want to have delays lower than 50 msec. Then, some measurements will demonstrate that existing recovery path restoration systems for Ethernet (such as STP) don't accomplish real-time requisites. Next, we have measured the real environment to know which delays are in a real MPLS network when there is regular traffic through it. Finally, we have measured the path restoration time when there is a failure in the network.

3.1 Delays in network devices and their links.

Network delays are produced because of the transmission time T_{trans} (time to send the information from the network adapter to the transmission medium), the propagation time T_{prop} (elapsed time in the medium) and the processing time T_{proc} (time to process the received frame in the station and transmit the next one). The Round Trip Time (RTT) of a ping is the elapsed time needed to send the frame from the station to the medium, plus the elapsed time to receive the last bit at the destination, plus the elapsed time to process the last bit and send the ACK, plus the elapsed time to send from the destination station to the medium the ACK, plus the ACK propagation time, plus the processing time of the source station to process last bit of the ACK, plus the processing time of the source station. The expression for RTT is given by equation 1.

$$RTT = 2 \cdot T_{prop} + 2 \cdot T_{proc} + T_{trans\_frame} + T_{trans\_ACK} \quad \textbf{(1)}$$

We have considered the same processing time for both stations when they have the same hardware characteristics and the same operative system. The transmission time is determined by the bits of the frame sent divided by the Mbps of the interface card. The propagation time is the distance between stations divided by the propagation velocity in medium (for non guided transmissions through air or ether, it is the light speed, approximately $3 \cdot 10^8$ m/s, for guided transmissions through fiber and copper, this velocity is approximately 0.67 times the light speed).

In order to determine the mean elapsed time, we have taken the average time of 100 pings. Each ping has 548 bytes, so the total packet length is 602 bytes (548 bytes of ping data + 8 bytes of the ICMP + 20 bytes of the IP + 26 bytes of Ethernet). Figure 2 shows obtained results using UTP category 5 wires. It also shows the equation that relates the meters of wire with the RTT using the obtained results.

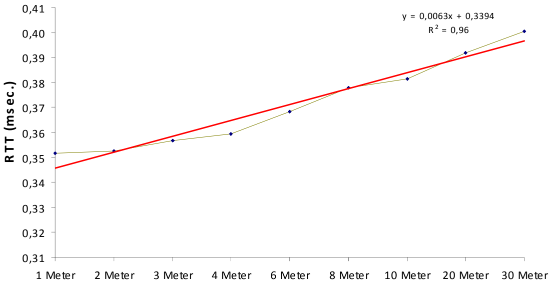


Fig. 2. RTT for several UTP category 5 pigtails.

Taking these measurements into account, we can state that the wire together with the computer processing time and the packet transmission time introduces delays of about 1 msec. every 100 meters, where about 0.34 msec. will be due to the computer processing time and the packet transmission time. We must remember that it is not allowed a wire length larger than 100 meters in fast Ethernet. For Gigabit Ethernet we have calculated the delay of the propagation time for 1 kilometre about 0.004 msec.

Once we have obtained the elapsed time due to the wire and to the computers, we can obtain the delay time introduced by the network devices. In order to do it, we have taken the average time of 100 pings (with the same size than the previous test) using two wires of 10 meters (one for each computer). Figure 3 shows the delay measurements introduced by some switches from several manufactures subtracting the RTT obtained for 20 meters in the previous test. The range of delays varied from 0.113 msec. to 1.943. The average value of all measured switches was 0.667 msec. The same methodology was used for the routers to measure their processing delay. Figure 4 shows results obtained from some routers of several manufacturers. The range varied from 0.479 msec. to 2.229 msec. The average value was 1.191. We have observed that the processing delay introduced highly depends on the switch or router manufacturer and even on its model. Switches and routers with more features introduce higher delays and, on the other hand, older switches and routers have higher delays than other devices from the same manufacturer.

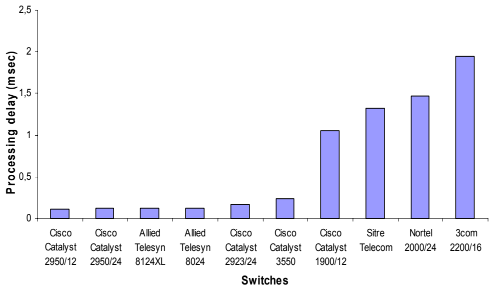


Fig. 3. RTT for several Switches.

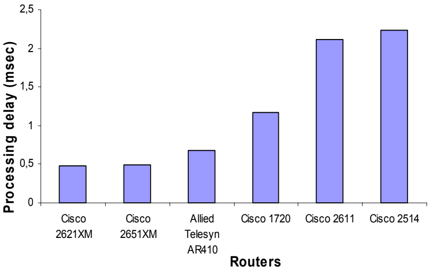


Fig. 4. RTT for several Routers.

Using the measures obtained from the real world, we can conclude that the RTT value (in msec.) in an Ethernet network follows approximately equation 2.

$$RTT = 1.1912 \cdot Z + 0.6669 \cdot Y + 0.0063 \cdot X + 0.3394 \qquad (2)$$

Where Z is the number of routers, Y is the number of switches and X is the meters of wire in the Ethernet network. Assuming that there are 2 switches at least per router in the network and supposing that there is a mean value of 30 meters of wire between any device or computer, we can calculate that there couldn't be more than 18 routers between the most remote devices in order to have a network with real time requisites.

3.2 Convergence time in Spanning Tree Protocol.

In order to check that the STP doesn't accomplish real-time responses, we have setup a testbed with three switches arranged in a triangle and test the response time when the link where the traffic goes through fails down. Figure 5 shows the network topology and the features of the computers for the testbed. In order to measure the time needed to recover a link, where the traffic is going through, we transmitted 10000 pings, each one with a difference of 200 msec. from one laptop to the other and, then, we measured how many packets were lost in the convergence time. Table 1 shows our measurements for several manufacturers and models.

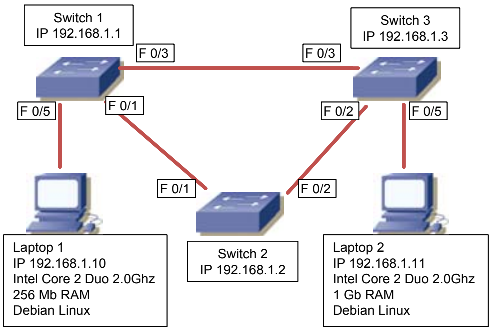


Fig. 5. Testbed for STP measurements.

Table 1. STP measurements.

	Cisco C2923XL	**Cisco C2950-12**	**HP Procurve**
Packets received	7466	7871	9943
Packet loss	25.34%	21.29%	0.57%
Time (in msec)	33954	28862	4458
Minimum value	0.156	0.136	0.107
Average value	0.356	0.305	0.295
Maximum value	1.561	4.731	1.231
Max. desviation	0.073	0.162	0.087
Convergence time (msec)	506800	425800	11400

We can now state that the convergence time in a Fast Ethernet network, when a switch fails and the path is recovered using STP, exceeds always 50 msec, so such networks don't provide real time recovery.

3.3 Round trip time in the Spanish railway network.

In order to test a MPLS network, area 3 (Madrid-Albacete-Játiva-Valencia) of ADIF corporate network was chosen. This area has 28 routers as it is shown in figure 6.

The network has two levels. The second level can be distinguished by lines in black in figure 6. Stations are connected to the rings with 2 Mb/s, to deliver service in short distances (15-60 kilometers). Above this topology we can find 1 GB fiber optic links that connect main network nodes. Lines in red form the first level, that is, nodes 0, 1, 2 and 3 (Madrid, Albacete, Valencia and Jativa). Each ring has two connected nodes to this fiber network through 100Mb/s links. Thus, when a station needs to communicate with a station of another ring or other control node, data will go through fiber network. Distances between nodes in the same ring not exceed 15 Kilometers. However, first level node links are about 250 Kilometers.

In order to know the real response time between the most remote nodes, we have gathered RTT from node 9 to node 21. Figure 7 shows the real response time obtained by 100 regular pings between those nodes. The minimum value was 23 msec., the maximum value was 98 msec. and the average time value was 24.84 msec. The 20th ping and the 94th ping peaks could be because there was sporadic excess of traffic in one of the nodes in the path, so the queue of that device needed more time to process all packets. Reference [13] shows more information about real measurements and simulations of this test bench.

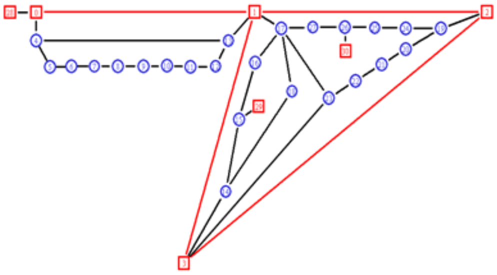


Fig. 6. ADIF corporate network - Area 3.

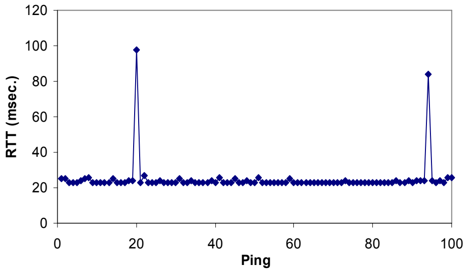


Fig. 7. Delay time from node 9 to node 21.

3.4 Convergence time when there are node failures in MPLS networks.

In order to measure the delay produced in the traffic because of a link failure, we have used the testbed shown in figure 8. All routers in the topology were Cisco 2621 XM running MPLS. All links between routers have a bandwidth of 2 Mbps except the links between the routers b and c, d and e, and f and g, which are 10 MBps. In T=0 sec. we started to send 200 UDP packets per second (one packet every 50 msec.), with a size of 64 Bytes, from the computer A to the computer B. Due to the bandwidth of the links in the network, the path followed by the packets was a-b-d-e-g-h. In t=5 sec., the link between routers e and g fails. The time needed to send the traffic through an available path is 5.4 secs. This time, the path followed by the packets is a-b-d-f-g-h. In t=15 secs. we recover the link between routers e and g and it is needed 5.0 secs to restore the old path. In both cases (the path recovery and the path restoration), the delays are quite higher than 50 msec.

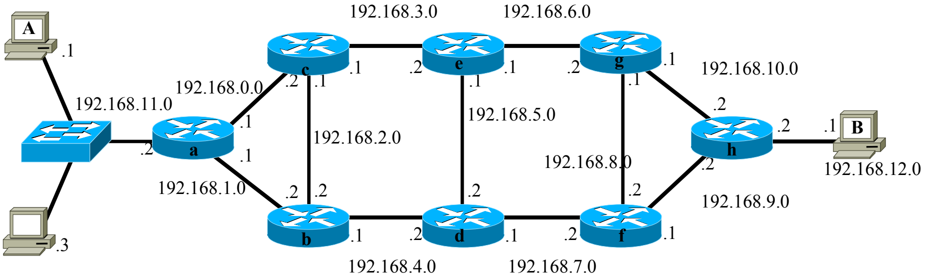
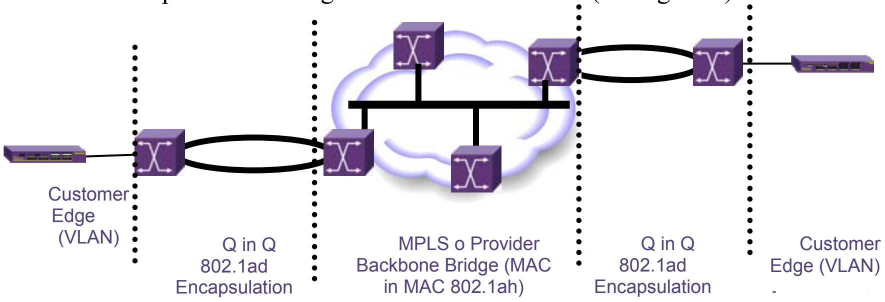


Fig. 8. Testbed used to measure the convergence time.

4 Proposal

We propose a new model for the integration of critical and non critical applications over standard Ethernet and MPLS networks. The new scenario is based on MetroEthernet corporate and integrated real-time network (see figure 9).



Customer Edge (VLAN)
Q in Q 802.1ad Encapsulation
MPLS o Provider Backbone Bridge (MAC in MAC 802.1ah)
Q in Q 802.1ad Encapsulation
Customer Edge (VLAN)

Fig. 9. Real-time corporate and integrated control network

For each application or real time traffic flow, it is assigned a separate VLAN to establish a quality of service with level 2 or 3 marking. It will be a relationship between VLANs and VPNs at the backbone. It is necessary to improve the transport network to guarantee high quality communications. The main objectives of these improvements are: guarantee quality of service to vital applications (also during fails), minimize fail occurrences and get low recovery times (lower than 50 msec., the standard on SDH systems). MetroEthernet, technology is not used to support critical applications, where it is necessary not only to adjust to QoS requirements, but high end-to-end reliability and availability. We can find many innovations about QoS aware MPLS/VPLS networks, but really there is a gap about fault-tolerance for supporting critical services. Our main goal is to provide high quality and reliability for the end-to-end path through Ethernet access networks and MPLS/VPLS core for non critical and critical real-time services.

4.1 Components of the proposed model.

The general components used in our proposed model are (i) a topology discovery protocol to build the network map, (ii) an algorithm for selecting primary and backup paths, (iii) an encapsulation method and frame format to carry data and explicit routing information, (iv) a protocol to detect and notify failures and, finally, (v) a mechanism to switch from active to backup paths.

A protocol discovery is needed to learn network connectivity. It is better to use local non bridging protocol to avoid traffic overhead. In this case, we use LLDP (Link Layer Discovery Protocol) [14], defined in IEEE 802.1AB standard. LLDP data units (LLPDUs) are sent to destination MAC address, defined like LLDP multicast address. A LLDPDU will not be forwarded by MAC bridges or switches that using IEEE 802.1D standard [15]. In order to distribute this information we have chosen SNMP, and to retrieve the initial topology, network nodes use SNMP polls to query topology MIB database and they use SNMP traps to notify topology changes. LLDP system capabilities information is used to query only neighbours which are intermediate nodes (bridge, AP, etc.) and to register end-stations (station-only).

Paths are pre-calculated with the algorithm proposed in [16]. It is a simple solution to build optimal primary and backup paths (k-paths) based on the SPF algorithm. But, in our case, we are going to include QoS and reliability considerations in a new integrated metric. All the parameters can be obtained through SNMP queries to the nodes in the path. We have not used STP algorithm because of its limitations (only one working path, no load balancing, a unique point of failure, maximum bridge diameter, etc). We have designed the metric shown in equation 3 to achieve our goals.

$$Metric = \left(k_1 \frac{delay}{bandwidth} \right) + \left(\frac{loss \cdot length}{availability} \right) \qquad \textbf{(3)}$$

The metric is formed by two parts. The first part, on the left, is related with QoS. *Bandwidth* is given by the addition of the bandwidth of the links along the path and the *delay* is the Round Trip Time required to move a packet from the source to the destination. The second part, on the right, is related with the reliability. *Availability* is the fraction of time in which nodes or links from all paths are working with full functionality, *loss* is the packet loss probability, and it is obtained from the SNMP interface counters, and *length* is the number of hops between the source and the destination. We have considered $K_1=10^3$ to get the QoS part into the desired values.

Only border network nodes build a full network map and select optimal working and backup paths, as a function of the availability and reliability based metrics. This algorithm falls under the category of proactive (pre-calculated backup paths) and global fault management (end-to-end). Really, the algorithm only have to add route descriptors (node and port identifier pairs) to a vector and have to detect if the node identifier is repeated for discarding this path. The next computation of the path can begin from the last node before the loop, and try to find another feasible segment to a destination. Our proposal allows several active paths for load balancing because the loop free topology is inherent to explicit routing. Moreover, we have not a unique point of failure; we have a faster path recovery and an increased bridge diameter.

The encapsulation method is based in QinQ or IEEE 802.1AD [17]. QinQ was originally designed to expand the number of VLANs by adding an additional tag to an 802.1Q frame. With this new tag, the number of VLANs is increased to 4096+4096. QinQ tags can indicate different information. For example, the inner tag indicates the user and the outer tag, the service. QinQ can be executed as the extension of a core MPLS VPN in the MetroEthernet to provide end-to-end VPN technology. QinQ encapsulation can be based on port or on traffic (on traffic is more flexible). QinQ encapsulation is also compatible with default-gateway standby protocols, like VRRP. Level-3 resiliency methods can be also added to our method.

Ingress nodes store in their memory explicit end-to-end primary and backup paths. If an active path fails, the ingress node decides to switch to the best alternative path. The old primary path is marked inactive, and can be deleted from their memory if a timeout is exceeded. If the recovery time of the failed node, or link, is lower than a timeout, old path is marked as an active. In this case, it is not necessary to recalculate the path. This switching mechanism is extremely simple, and there is no need for changing frame control information to notify an alternative path, because primary and backup paths are stored in ingress node's memory and intermediate nodes have not any routing information. In MSTP recovery solutions, normally end-hosts need to change the VLAN id in subsequent frames to select an alternative switching path [18].

As our proposal uses an explicit routing/switching mechanism based on the global view of the network, paths are pre-calculated from LLDP information and ingress node put full explicit path into the frame header. Since the path is predefined, the frame is routing at each node without the need for making routing decisions at every node along the path (hop-by-hop destination-based routing). Setting up paths implies that network will keep state information. It has been avoided in the past due to its potential reduce of performance, but now it can be done more efficiently because of higher-performance devices. In fact, MPLS networks use this technique to establish traffic engineering tunnels, but Ethernet networks have never use it before. Explicit routing approach is very useful for the prevention of routing loops, for traffic engineering and to adjust to QoS constraints (some constraints like the overall delay cannot be really calculated in a hop-by-hop fashion). Then, it is not necessary to work with Spanning Tree algorithm to find a unique path with no closed cycle, neither Multiple Spanning Tree routing solutions like 802.1AQ Shortest Path Bridging [19].

In order to guarantee full compatibility, we have employed the source routing transparent Ethernet operation described in IEEE Std 802.1D-2004 Annex C. However, we use QinQ tagged frame format between the ingress and egress node. In this case, when CFI bit is set (bit CFI=1), the switch performs source routing. It comprises a two-octet Route Control Field (RC) that has the following fields:

- 3-bit Routing Type (RT) fixed to 0XX. If the most significant RT bit is set to 0, the Route Descriptor (RD) field contains a specific route through the network. This is the unique combination used in our proposal, because there is no need for routing explorer frames (network map is built from LLDP information).
- 5-bit Length (LTH). Indicates the length (in octets) of the Routing Information Field (RIF). Length field values will be even values between 2 and 30 inclusive.
- 1-bit Direction Bit (D). If the D bit is 0, the frame traverses the LANs in the order in which they are specified in the RIF (RD1 to RD2 to... to RDn). Conversely, if the D bit is set to 1, the frame will traverse the LANs in the reverse order (RDn to RDn-1 to... to RD1).
- 5-bit Largest Frame (LF). For source routing frames, the LF bits are ignored and shall not be altered by the bridge. Then this field is free to use such as destination region/ring id.

- 1-bit Non Canonical Format Indicator (NCFI). If it is set, indicates that all MAC address information, that may be present in the MAC Service Data Unit (MSDU), is in Canonical format, and it is reset otherwise.

Figure 11 shows the modified E-RIF Route Control Field.

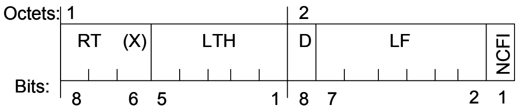


Fig. 11. E-RIF Route Control Field.

Moreover, more octets of Route Descriptors could be added (up to a maximum of 28 octets such as in IEEE Std 802.Q-2005). Each Route Descriptor has 16-bits that next node contains (7-bit code) and port destination (9-bit code) for providing routing information. This design is scalable because it has three organization levels: regions (up to 32 coded in 5-bit field), nodes (up to 128 coded in 7-bit field), ports (up to 512 coded in 9-bit field). The bridge diameter of the network is only limited by the maximum number of octets in the frame. In our implementation, end-stations transmit standard frames with no routing information included. The ingress switch is responsible to put it into the frame and the egress put off the routing information. Then, no changes are required at end-stations. Route Descriptors are concatenated and are removed from the corresponding switch. Finally, when there is no Route Descriptors in the message, the bit CFI is set to zero. Then, last switch forwards the frame to local port destination. This mechanism of putting and taking information from the frame can prevent undesirable loops in the network, because never must be two equal route descriptors in the route information fields.

In order to provide a procedure and protocol to notify link and node errors, for starting restoration actions, we can choose between two standardized methods (we can implement each one of them). Bidirectional Forwarding Detection (BFD) is an IETF draft standard that defines a method to detect errors in forwarding paths, and can be used to trigger a switch or a router to alternate interfaces or routes. It can be applied to many technologies and networks. The alternative option is to use IEEE 802.1AG standard [20]. It was designed only for Ethernet networks but it can be adapted with few modifications. Both can be used depending on the network scenario.

5 Conclusions

We have shown the need of critical and non critical real-time services integration in a unique corporate network, to optimize management and to avoid high budget. We have provided an experimental law that should be applied when it is needed a network with real-time services. Measurement results confirm that recovery times in Ethernet and MPLS networks exceed 50 msec. On the other hand, we have observed that the delay of video streams is higher when the traffic begins to use an available path after a failure than when the failed device is recovered.

We have proposed a new system to improve resiliency on critical infrastructures networks that allows path recoveries faster than 50 msec. because the proposal is based on a protection mechanism which guarantees a very small delay when a path has failed. This solution is fully compatible with Ethernet and MPLS equipments and protocols, with no need for substantial hardware or software modifications in nodes or end stations. In fact, only end switches must run this algorithm, and STP is yet not necessary to avoid loops. The key of our method is a new metric not only with QoS constraints, but also with metrics related with the reliability of nodes and links.

This work opens new research lines in MetroEthernet, Cluster and Corporate networks and it can also be a fault-tolerance and traffic engineering optimal solution.

References

1. Francisco J. Sánchez, Jaime Lloret, Juan R. Díaz, José M. Jiménez. Mecanismos de protección y recuperación en redes de tiempo real para el soporte de servicios de explotación ferroviaria. XX Simposium Nacional de la URSI. Gandia (Spain). September 2005.
2. Francisco J. Sánchez, Jaime Lloret, Juan R. Díaz, José M. Jiménez. Standardization and improvement of real time network technology on railway traffic control systems. 7th World Congress on Railway Research. Montreal. June 2006.
3. Jaime Lloret, Francisco J. Sánchez, Juan R. Díaz, José M. Jiménez. A fault-tolerant protocol for railway control systems. 2nd Conference on Next Generation Internet Design and Engineering. Valencia. (Spain) April 2006.
4. E. Harrison, A. Farrel, B. Miller. Protection and restoration in MPLS networks. Data Connection White Paper. October 2001.
5. R. Aggarwal, K. Kompella, T. Nadeau, G. Swallow. BFD For MPLS LSPs. November 2007
6. M. Amin, K. Ho and G. Pavlou. MPLS QoS-aware Traffic Engineering for Network Resilience. Proceedings of London Communications Symposium, London, UK, September 2004.
7. IEEE Std 802.17B. Resilient packet ring (RPR) access method and physical layer specifications. Pp.1-216. August 2007.
8. S. Shah and M. Yip. IETF RFC 3619. Ethernet Automatic Protection Switching. Oct. 2003.
9. G. A. Dirac. Extensions of Menger's Theorem. Journal of the London Mathematical Societ. s1-38(1):148-161. 1963.
10. A.F. Hansen, A. Kvalbein, T. Cicic, S. Gjessing, O. Lysne, Resilient Routing Layers for Recovery in Packet Networks. Proceedings of the International Conference on Dependable Systems and Networks, 2005. 28 June - 1 July 2005. Pp. 238-247.
11. D. Stamatelakis, W. D. Grover. IP layer restoration and network planning based on virtual protection cycles. IEEE Journal on selected areas in communications, Vol. 18, No. 10, Oct. 2000.
12. M. Padmaraj, S. Nair, M. Marchetti, G. Chiruvolu, M. Ali, "Distributed Fast Failure Recovery Scheme for Metro Ethernet", IEEE ICN, April 2006.
13. Hugo Coll, Jaime Lloret and Francisco Javier Sanchez, Does ns2 really simulate MPLS networks?, The 4th Int. Conf. on Autonomic and Autonomous Systems. March 2008.
14. IEEE Std 802.1AB- Draft 2 d2-2. Station and Media Access Control Connectivity Discovery. December 2007.
15. IEEE Std 802.1D. Media Access Control (MAC) Bridges. June 2004.
16. IEEE Std 802.1AD. Provider Bridge. May 2006.
17. D. Eppstein, Finding the k Shortest Paths. SIAM J. Computing 28(2):652-673, 1998.
18. S. Sharma, K. Gopalan, S. Nanda, T. Chiueh. Viking: a multi-spanning-tree Ethernet Architecture for Metropolitan Area and Cluster Networks. IEEE INFOCOM 2004.
19. IEEE 802.1AQ draft 0.3. Shortest Path Bridging. May 2006.
20. IEEE 802.1AG draft 8.1. Connectivity Fault Management. February 2007