Dependability of All IP Networks:

Resiliency in Ethernet Based Transport Networks

Kari Seppänen

VTT TECHNICAL RESEARCH CENTRE OF FINLAND

Project	IPLU
Contact person at VTT	**VTT, Technical Research Centre of Finland** Kari Seppänen P.O. BOX 1000, FIN-02044 VTT, Finland Tel. +358 20 722 5610 Fax +358 20 722 7028 Email Kari.Seppanen@vtt.fi

Abstract

This report is produced as a part of IPLU (Dependability of All IP Networks) to give some basic understanding about the resiliency features and problems in using Ethernet as a carrier network technology.

Contents

Acronyms

AIS Alarm indication signal
AP Access Point
ARP Address resolution protocol
ATM Asynchronous transfer mode
BER Bit error rate
BPDU Bridge Port Data Unit
BRAS Broadband remote access server
CAM Content addressable memory
CBS Committed Burst Size
CFI Canonical format identifier
CDP Cisco Discovery Protocol
CE Customer Edge Device
CESoE Circuit Emulation Services over Ethernet
CFM Connectivity Fault Management
CIR Committed Information Rate
CLI Command line interface
CoS Class of Service
DHCP Dynamic Host Configuration Protocol
DoS Denial of Service
DTP Dynamic trunking protocol
EAPS Ethernet Automatic Protection Switching
EBS Excess Burst Size
EFM Ethernet first mile
EIR Excess Information Rate
ELMI Ethernet Local Management Interface
EoS Ethernet over SONET/SDH
EPON Ethernet passive optical network
EPSR Ethernet protection switched ring
ES Ethernet Service
ESD Ethernet Services Definitions
ESM Ethernet Services Model
ETM Ethernet Traffic Management
ETTH Ethernet-to-the-home
EVC Ethernet Virtual Circuit

GFP Generic framing protocol
GMPLS Generalized MPLS
GOE Global open Ethernet
GoS Grade of Service
GVRP Generic VLAN registeration protocol
HSRP Hot standby router protocol
HTML Hypertext markup language
IETF Internet Engineering Task Force
IFG Inter-frame gap
ILMI Integrated Local Management Interface
IP Internet Protocol
ISL Inter-switch link protocol
IVL Independent VLAN Learning
LAN Local Area Network
MAC Medium Access Control
ME Metro Ethernet
MEF Metro Ethernet Forum
MEN Metro Ethernet Network
MiM Man-in-the-middle
MPLS Multi-Protocol Label Switching
MSTP Multiple STP
MTU Media transport unit
NMS Network management system
NTLM NT LAN manager
OAM Operations, Administration and Maintenance
OAMP OAM and Provisioning
PB Provider bridge
PBB Provider backbone bridge
PE Provider Edge Device
PNNI Private Network-to-Network Interface
POP Point of presence
POS Packet over SONET/SDH
PPP Point-to-Point Protocol
PSN Packet Switched Network
PWE3 Private Wire Emulation Edge-to-Edge
QoS Quality of Service
RFI Remote fault indication
RIP Routing information protocol
RPR Resilient packet ring
RRSTP Rapid Ring STP
RSTP Rapid STP
SDH Synchronous Digital Hierarchy
SLA Service Level Agreement

SNMP Simple network management protocol
SONET Synchronous Optical Network
SP Service Provider
SSH Secure shell
SSL Secure socket layer
ST Spanning Tree
STP Spanning Tree Protocol
TC Topology change
TCN Topology change notification
TCP Transmission control protocol
TDM Time-division multiplexing
TFTP Trivial file transfer protocol
TOS Type of Service
UDP User datagram protocol
UNI User–Network Interface
VC Virtual Circuit
VID VLAN ID
VLAN Virtual LAN
VLL Virtual Leased Line
VMPS VLAN membership policy server
VoIP Voice over IP
VPLS Virtual Private LAN Service
VPN Virtual Private Network
VPWS Virtual Private Wire Service
VTP Virtual trunking protocol
WAN Wide area network
WDM Wavelenght-division multiplexing

Chapter 1

Introduction

Ethernet is considered to be one of the most prominent network technologies for the new Packet Switched Network (PSN). Its wide success in the Local Area Network (LAN) and the low cost of the equipment have inspired some Ethernet enthusiasts to claim that Ethernet would be the only viable solution. However, the current office Ethernet systems do not meet the requirements of a carrier-grade transport network and thus new, more reliable and manageable, solutions have to be developed. Ethernet is also claimed to be an easy plug-and-play technology and thus it is thought that Ethernet based carrier networks could be operated with a smaller, less skilled staff. However, it seems like all the same network management operations remain as with the other "more complex" network technologies. Moreover, it is not so clear that Ethernet will have better and easier-to-use support for network management than the other alternatives.

One very confusing thing in the discussion about "carrier-grade Ethernet" is that different vendors, operators and interest groups may have quite different meanings for that concept and sometimes it is hard to determine which flavor of "carrier-grade Ethernet" someone is talking about. Depending on the author, the "carrier-grade Ethernet" can mean anything from transporting Ethernet frames over SDH/SONET to using a bit more reliable Ethernet equipment to implement, e.g., metro networks. One quite common flavor is transporting Ethernet over MPLS based Virtual Private Network (VPN). It is clear that such arrangement is not about using Ethernet as a transport network technology but about providing Ethernet User–Network Interface (UNI) with MPLS. So it is usually very important to know what kind of "carrier-grade Ethernet" an author is referring when reading a paper or listening to a talk.

The scope of this document is limited to using Ethernet as a transport network technology. This means that different Ethernet Services like E-Link and E-LAN defined by Metro Ethernet Forum (MEF) or Ethernet over SONET/SDH (EoS) are not studied in detail. The reason for this is that resiliency in such services depends on the underlying transport or network technology (e.g., SDH or MPLS). For the Ethernet network utilizing such service, the whole service is like a single link (e.g., E-Line) or a single switch/hub (e.g., E-LAN)[1]. However, it is possible to implement, e.g., Metro Ethernet Network (MEN) as

1. It is true that there are some reliability issues arising form the interactions between Ethernet clients

defined by MEF, using Ethernet technology and such cases are relevant for this study.

The main purpose of this document is not to promote nor demote the use of the Ethernet technology in the carrier networks. The main emphasis to describe the major dependability problems existing in Ethernet today. It is widely acknowledged that Ethernet has deficiencies with respect to Operations, Administration and Maintenance (OAM), reliability, traffic management, and scalability [1]. However, I am not trying to claim that Ethernet cannot be used to build a carrier grade network infrastructure because such claim would not be justified. There are many applications where Ethernet can be a sufficient solution. Furthermore, the Ethernet technology is evolving and it is quite likely that many, if not most of the, problems will be solved. In any case, it is important to know the limitations and deficiencies of the Ethernet technology to be able to use it in the best possible way.

The structure of this document is following: the rest of this introductory chapter is devoted to define the actual role of Ethernet in the networking and to describe some possible use scenarios. In the Chapter 2 the basics of current office Ethernet technology are explained to provide necessary background to understand the short-comings of Ethernet. The next chapter (Chapter 3) describes the recent developments in the Ethernet technology for the carrier networks, while the Chapter 4 describes some of the other Ethernet related technologies. The implementation related issues are considered in the Chapter 5 and the security issues in the Chapter 6. Some thoughts about how to operate an Ethernet based carrier networks are presented in the Chapter 7 and finally the Chapter 8 concludes this document.

1.1 What is the "switched" Ethernet

Ethernet is quite often regarded as a network technology which may make sense as, e.g., a large and complex switched Ethernet installation looks very much like a network layer thing. However, if the layered network models are considered, it is clear that such conception is wrong and misleading. In the OSI reference model, Ethernet covers layers 1 (physical) and 2 (data link) and, in the TCP/IP model, the network interface layer [2]. An Ethernet network segment, no matter how complex it might be, is just a link in a network [3]. What makes things more complicated, is that Ethernet switches of today usually have at least layer 3 (network) functions (some may have even layer 7 functions). This, however, does not mean that the Ethernet technology has gained recently network layer capabilities — it simply means that such switch have a simple router integrated inside it.

The most conservative way, I think, to understand the basics of Ethernet is to consider it as a shared-media broadcast link-layer network that has some filtering capabilities. While this filtering (by MAC address or VLAN ID) may look very much like switching, Ethernet would have so many peculiarities if one tries to consider it as a switched network technology.

---

and VPN edge, especially in E-LAN service, but those issues are not considered in this document.

1.2 Use scenarios

The Ethernet technology can be utilized in many roles in carrier networks and each role has its own pros and cons. In the following, a short description of few of these roles are given with their particular concerns. The selected examples are not claimed to be the most important ones nor to cover all the relevant use cases.

One quite common application for Ethernet today is to use it to replace the ATM network connecting DSLAMs to POP/BRAS site (figure 1.1). In the future, similar cases can be considered for the fixed and mobile telephony backhaul. The common factor in this type of use scenarios is that the user does not (usually) have a peer-to-peer Ethernet layer access to the carrier's metro network infrastructure. Ethernet is used only to provide a point-to-point transport pipe between an access concentrator (e.g., DSLAM) and the central office (e.g., POP). In this scenario, the carrier does not have to consider much about protecting his/her network infrastructure from users' actions[2]. The most important factors to consider are how to provide the required reliability (e.g., 0.99995 for telephony access network) and how to operate and manage the network.

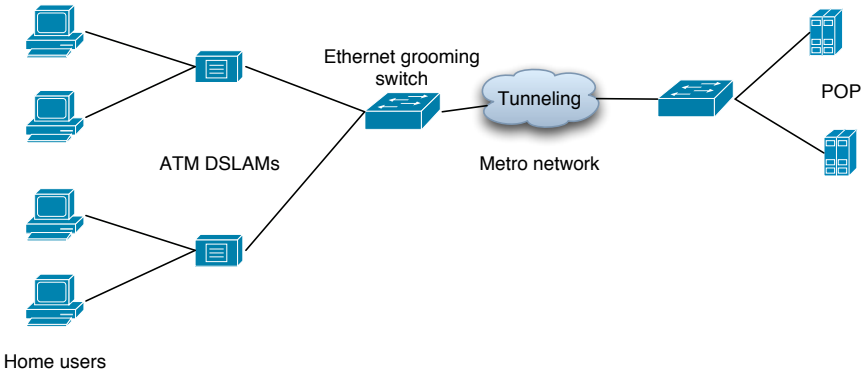

Figure 1.1: Use scenario 1: using Ethernet as an access transport solution without L2 access

Ethernet is also emerging as an access network technology, e.g., Ethernet passive optical network (EPON), Ethernet first mile (EFM) and Ethernet-to-the-home (ETTH). This kind of scenario (figure 1.2) raises more issues than the previous one: reliability and management are still issues, but now the user has a peer-to-peer Ethernet layer access to the carrier's network. In this case, it becomes very important to ensure that the user's traffic will always be isolated to protect the carrier's network. It is insufficient to rely on that all end systems are properly configured and thus unable to access carrier's equipment. Malicious attacks are only one of the threats; misconfigurations, failed or faulty software updates, etc. are also potential hazards. These problems are quite likely going be worse in the future as home theaters, household appliances and others become Internet enabled.

2. Provided that the Ethernet equipment reliably isolates user's traffic — if not, the case is completely different.

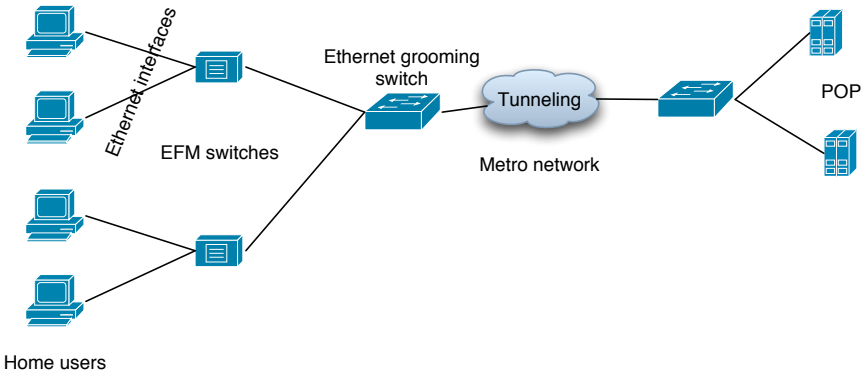

Figure 1.2: Use scenario 2: Ethernet First Mile (EFM) and similar access transport solution with L2 access

The third scenario is providing virtual private network (VPN) services for the corporate users by using Ethernet. This case has all the same issues as the previous one. However, this case is likely to add considerable amount of complexity. Each access point (AP) is likely to serve up to thousands of end systems and Gbit/s range traffic load. Furthermore, there could be a need to provide switched VPN service that is much more complex service than the point-to-point service required by the first two scenarios. Thus one of the most important issues in this scenario is likely to be the scalability of Ethernet. Moreover, corporate clients usually require some service guarantees[3] and thus the service management is also an important issue.

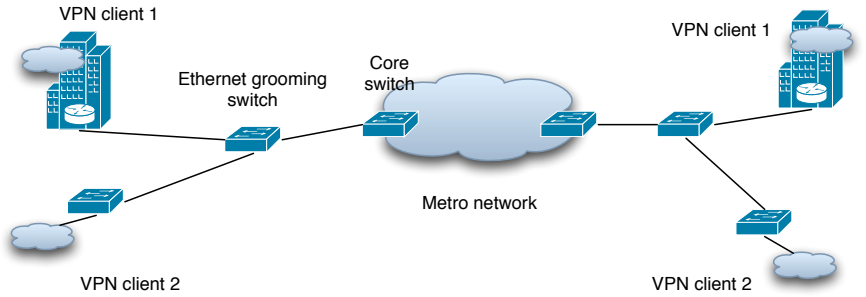

Figure 1.3: Use scenario 3: using Ethernet to provide VPN services

3. opposite to flat-rate home users

Chapter 2

Office Ethernet

To understand the primary source of the problems in applying Ethernet in the carrier networks, it is important to know the very basics of the Ethernet technology. Ethernet was designed as a shared media multiple access network for office LANs and it still carries many of the limitations of the original design. The development of the Ethernet has always been very conservative: the frame structure has remained more or less the same and the new options or functions have been strictly backward compatible. In a sense, Ethernet is still that same simple shared media network technology — the new developments have not changed this basic approach, they have merely added ways to, e.g., reduce broadcast traffic by filtering.

Accepting the shared-media-multiple-access nature of Ethernet makes it easier to understand many peculiarities and failure modes [4]. In some cases, e.g., delivering frames to unknown destination, there is no other way to operate than to fall back to shared broadcast media mode. Furthermore, certain failures may cause the same fall-back behavior[1]. It should be noted that many upper layer "helper" protocols that are used with Ethernet, e.g., Address resolution protocol (ARP) and Dynamic Host Configuration Protocol (DHCP), are based on the broadcast service. Thus such protocols usually have similar odd problems that should not exist in a true switched network.

Some of the basic conceptions in the Ethernet technology are[2, 5]

collision domain an area of a LAN where all the attached nodes can hear all the transmissions originating from the other nodes, i.e., the nodes are interconnected by a shared media like a coaxial cable or repeaters; the name originates from the fact that there can be collisions, i.e., two or more nodes can start sending at the same time

repeater a device that connects two or more physical cable segments by repeating all transmissions; the cable segments interconnected by a repeater form a collision domain

hub multi-port repeater

1. Usually this is an implementation issue: availability has been increased by scarifying performance and security.

broadcast domain an area of a LAN where all nodes can hear all broadcast transmissions originating from other nodes; collision domains are interconnected by learning bridges to form a broadcast domain

(learning) bridge a device that connects two collision domains and filters the traffic between those domains according to Medium Access Control (MAC) addresses; a bridge learns at which side of itself a specific node with a specific MAC address by listening the traffic and only the "necessary" traffic (including unknown destinations) is forwarded; prevents collisions between collision domains

switch can be considered as a multi-port bridge or a set of bridges connected by a fast internal collision domain[2]

LAN segment a domain where all nodes are interconnected at data link layer, i.e., collision domain, broadcast domain or a mix of those two

Virtual LAN (VLAN) virtual LAN segments are created by filtering traffic in switches; filtering can be based on ports or MAC addresses; filtering database define which ports (or nodes) can communicate at layer 2; only way to exchange traffic between two VLANs is to use a router to interconnect them, i.e., traffic must be delivered via network layer

(VLAN) trunking a method to transport traffic from multiple VLANs between switches; a specific Q-tag is used to label traffic from different VLANs

Spanning Tree (ST) in bridged Ethernet, meaningful learning of MAC addresses cannot be done if a bridge (or switch) hears same address from multiple collision domains; furthermore, as Ethernet frames do not have time-to-live field, broadcast and unknown destination frames would rapidly fill-up the network if there were any loops; thus loops in Ethernet LAN segment have to be avoided by creating a spanning tree and turning off all links between bridges that do not belong to that tree

Spanning Tree Protocol (STP) a protocol used by bridges to create and maintain spanning tree

forwarding base a data structure (e.g., Content addressable memory (CAM)) holding the forwarding information (MAC address → port id) learned by monitoring source addresses

Collision domain, repeater and hub can be considered to belong to the physical layer while broadcast domain, bridge and switch belong to the data link layer.

2.1 Spanning tree protocol

The problem with the bridged Ethernet networks is that a loop-free network topology is required. There are two reasons for this requirement. First of all, broadcast traffic should be sent via all (active) links except the originating link — as there is no time-to-live field in the Ethernet frame, broadcast frames would soon fill up the whole network if there were any loops. Secondly, the operation of learning bridges is based on listening

2. I think that the latter explanation makes it easier to understand, e.g., fail-open failure mode.

the traffic and collecting source MAC addresses with port information into a forwarding table. If a bridge “hears” a MAC address via more than one port, it could not learn the correct source of that address. Furthermore, in the bridged Ethernet, frames with unknown MAC address are broadcasted allover the LAN, increasing the broadcast frame problem.

It would be quite easy to solve to the problem by building only loop-free Ethernet LANs. However, it is quite common that Ethernet LANs contain redundant links, e.g., in order to increase fault tolerance (see the Figure 2.1). To over-come this problem, the Spanning Tree Protocol (STP) was developed. The purpose of STP is to create and maintain a loop-free network topology by constructing a spanning tree for a LAN.

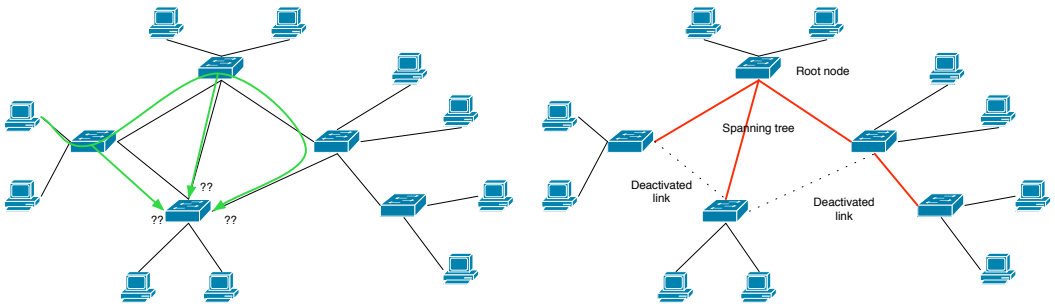

Figure 2.1: Spanning tree protocol

The operation of STP is, shortly, following [5]:

1. All links are turned off except for transmitting Bridge Port Data Unit (BPDU)
2. Each bridge/switch is assigned with a priority
3. A root bridge is selected based on priorities; after the selection only the root node generates BPDUs, the other bridges only reply to these BPDUs
4. The root bridge assigns one of its ports as root port
5. A spanning tree (ST) starting from the root port is generated by calculating least cost paths from each bridge/switch to root port
6. In all bridges/switches the ports that belong to ST are turned to forwarding mode (receives and transmits frames); all the other ports are turned to blocking mode
7. If a root failure occurs, the procedure is started form the beginning

Topology changes are managed by Topology change notification (TCN) messages that are sent to the root bridge by a bridge detecting the change. The root bridge processes TCN and sends a Topology change (TC) acBPDU to all bridges. After receiving a TC BPDU bridges flush their MAC tables after a guard time (usually 15 s) and start relearning new MAC addresses [6]. This causes usually a drop in network throughput as all frames are broadcasted along ST until new address bases are learned. Thus a single link

failure affects the whole network, that is, the very nature of Ethernet makes it hard to isolate failures [4].

2.2 Scalability, security and reliability issues in STP

There are many scalability, security and reliability problems in the standard STP. In fact, it can be said that the performance of STP is rarely acceptable in modern office LANs, no to mention other, more demanding, applications.

The most severe scalability issue of STP is the inability to utilize redundant link capacity. The links that are in blocking mode are not used to transport any traffic. Also, it is likely that the links originating from the root node will become congested as much of the traffic passes through the root node. The only performance and reliability enhancement allowed by STP is so called link bundling. In the link bundling, a set of parallel links between two switches is advertised as a single virtual link. This allows for turning all those links to forwarding mode. Furthermore, a failure in a single link of the bundle does not cause topology change.

The major security problems are related to the openness of STP interface and the usual practice to enable all protocol interfaces at all ports. In the practice, the openness means that there is no authentication nor authorization of BPDUs, that is, anyone having an access to a STP enabled interface in the network can become a bridge. This can be used for Denial of Service (DoS) attacks or for snooping traffic (more about this in the Chapter 6).

The major reliability issue is the slow reaction and convergence time of STP. As only the root bridge is able to generate BPDUs and process topology changes in normal operation mode, it takes time to detect link failures or unreachability of the root bridge. The other bridges has to rely on timers to detect a dead root bridge and to expire forwarding tables. The result of all this, means that the converge time of STP is in order of 30-60 seconds (some have reported even 3 minutes). As all the ports are turned in the listening mode during the reconfiguration, this means long service outages.

2.3 Rapid spanning tree protocol

The Rapid STP (RSTP) (originally 802.1w, now included in 802.1D) was developed to solve the slow convergence problems of the original STP. RSTP is still based on a single ST with a single root bridge, but it manages to cut down the service outage considerably.

With RSTP all bridges can generate BPDUs and process topology changes, which enables fast reaction to link failures. In practice, this means that recovery from link failures takes usually less than a second. Furthermore, bridges can detect edge ports, i.e., ports that are connected to user equipment only. The edge ports can be turned in forwarding mode immediately after a start-up or a failure. The third new feature is faster forwarding

database aging: a bridge can flush its database as soon as a failure is detected and start learning new topology. This means that bridges do not need to wait for the convergence of the LAN wide topology after a failure.

The dead root bridge detection is also faster. First of all, each bridge waits only for three hello times instead of the maximum age time to declare the root bridge dead. Furthermore, the declaring bridge can claim itself as a new root bridge without the selection process. However, the dead root bridge detection still takes few seconds (depending on the hello time setting). In large scale networks, simulation results show that dead root bridge recovery time could be up to 12–14 s [7].

In a detailed analysis [4], it was found out that the main reasons for the long RSTP convergence times are so called count-to-infinity problem and hop-by-hop port state negotiation. The former problem is actually caused by the RSTP enhancements themselves: as the bridge detecting a dead root bridge can declare itself as a new root bridge, the old reachablity information to the old root bridge may remain circulating in the network in the similar way as with, e.g., RIP, and thus persistent loops are created until path cost is incremented to infinity. In Ethernet this has far more severe results (broadcast storm), as there is no time-to-live field in the frame. The latter problem is caused by the desire to avoid loops at any cost: thus link states are changed conservatively one-by-one.

The RSTP makes the recovery times from a link failure more acceptable than the original STP. In fact, the speed of the recovery from topology change (that is link failure) is quite comparable to, e.g., PNNI. However, the root bridge is still single point of failure and the recovery time in dead root bridge cases is rather long. Moreover, RSTP does not provide anything like fast path protection nor does it solve the scaling problems related with the unused link capacity and the root bridge congestion.

2.4 Virtual LANs

VLANs (802.1Q) provide a method to create virtual LAN segments inside a physical LAN segment. It is based on assigning switch ports to different VLANs and the filtering the traffic in a way that no frames is transmitted from a VLAN to another, that is, only the end systems in the same VLAN can communicate with each other[3]. Each VLAN in one LAN segment has to be allocated with an unique VLAN ID (VID). Using VLANs does not change the interface between user equipment and Ethernet switches — in a sense, user equipment is totally unaware about the existence of VLANs.

VLANs require some extensions to the standard Ethernet to enable efficient transport of VLAN traffic between switches. The most important extension is VLAN tagging used in VLAN trunking. Adjacent Ethernet switches can detect that the link that interconnects them is a direct switch-to-switch link. After detection, such link can be declared as a trunk

3. That holds in the Ethernet layer; they can of course communicate through a network layer (L3) router.

link. The frames originating from different VLANs are then tagged with an additional 32-bit field[4], called Q-tag, containing VID. A switch, receiving Q-tagged frame from trunk link, uses VID to decide to which ports it should forward the frame.

An additional extension allows for a switch to tell the others from which VLANs they want to receive traffic. Using Generic VLAN registeration protocol (GVRP) a switch can subscribe only to those VLANs it needs. The main advantage of this scheme is that it saves trunk link capacity by pruning the distribution broadcast traffic [8].

It should be noted that there are no actual VLAN routes and that the VLAN concept should be considered just as another filtering rule-set for switches (other being MAC addresses). On the other hand, it is quite true that VLAN filtering results in some kind of implicit routes in a LAN segment in the same way as MAC learning and STP. However, broadcast and unknown destination traffic is still flooded inside a VLAN and, in fact, through the whole LAN segment if switches cannot filter traffic in trunk links.

The original VLANs have also a scalability issue: there are only 12 bits reserved for VID. As VID has to be unique within a LAN segment, only 4094[5] concurrent VLANs can be supported. This is, of course, sufficient for typical office use, but it is quite limiting if larger scale networks are considered.

2.5 Quality of service

Ethernet supports some QoS features by providing a method to mark frame priority (802.1p) using a 3-bit user priority field in the VLAN tag. There is no standard rules to process user priority field and thus its usage is usually vendor dependent [9]. It is thought that 802.1p would need IP DiffServ compatible CoS scheme with, e.g., drop precedence (DP) bit, to be really useful [10].

2.6 Management

The office Ethernet has very few Ethernet layer management related capabilities. There are some vendor specific management protocols like Cisco's Virtual trunking protocol (VTP) and Cisco Discovery Protocol (CDP) but otherwise the only management information that is available is essentially just loss-of-light or loss-of-signal indication. The common way to manage Ethernet equipment has been, and still is, to use upper-layer management systems. These systems are most often based on a specific management module inside of the equipment that provide an IP layer, e.g., SNMP, telnet and/or Web, management interface. In practice, it is quite common that there is no centralized

4. 3 bits for user priority, 1-bit Canonical format identifier (CFI), (used in transporting token ring frames), 12 bits for VID and 16 bits for original type/length field (the type/length field in the frame headers is changed to indicate VLAN tag)
5. There are two reserved VIDs

management and monitoring system: configurations are managed by, e.g., telneting to a particular device and using command line interface (CLI) to change settings, locating faults is done, e.g., by pinging the management modules.

2.7 Ethernet related protocols

There are several protocols that are used to support upper layer protocols in Ethernet networks, e.g., ARP and DHCP for TCP/IP. Many of these protocols are designed to be as simple as possible and thus they usually use Ethernet in the broadcast mode. This causes scalability and security problems in large network segments, e.g., in one study a peak rate of ARP queries in 2456 node network was 1150 queries per second [4]. This is caused by the relatively short aging timer in ARP caches. In broadcast based query services, the first reply received is usually accepted without any authentication. This makes it quite easy for an attacker to, e.g., impersonate another host by using ARPspoof.

Chapter 3

Transport network grade Ethernet

The legacy Ethernet used in intraoffice LANs does not meet the requirements that are usually imposed on carrier networks. As described in the previous chapter, there are many deficiencies in reliability, scalability and so on. There has been a lot of development work going on to overcome these problems. However, the need to preserve full backward compatibility with the office Ethernet has not made this task by any means easier. This is why, there are only more or less partial solutions and a lot of work to do.

ITU-T has many activities in defining Ethernet services for carrier networks. However, ITU-T is not defining recommendations for Ethernet transport equipment. Its approach is based on transporting Ethernet over ATM, MPLS or SDH carrier networks and thus ITU-T's activities are mostly out of the scope of this paper. Furthermore, an other group developing Ethernet services, MEF, is also working solely on service layer. This means that the IEEE SG 802.3 is the only player actually defining transport issues for carrier-grade Ethernet. However, 802.3 has limited its work in media-access control and physical-layer control leaving a management gap between services and transport[11].

3.1 Multiple spanning tree protocol

The Multiple STP (MSTP), (originally 802.1S, now included in 802.1Q) offers two enhancements to the standard STP. It allows for division of the network into regions in a way that a single link failure inside one region does not cause a need for reconfiguration of Common Spanning Tree[1] in the other regions. Furthermore, it supports a dedicated ST for each VLAN or a set of VLANs[12]. As the first option, the dedicated STs improve fault tolerance: a link failure affects only the STs using that failing link.

MSTP can be also used to implement load balancing in the network segment by planning which ST should be allocated to which VLAN. This can achieved by selecting suitable MSTP parameters[2] for each MST instance. However, the standard does not define how this should be done nor is there any automatic support for such traffic engineering. The

1. connects all switches in a MSTP Ethernet network segment
2. bridge and port priorities, port costs

same holds for the task of designing MSTP regions. Thus it seems like successful MSTP operation would benefit from a decent traffic engineering tool for optimizing the ST allocations and region planning [12, 13, 14, 15].

MSTP supports upto 64 STs in one network segment [14]. It is not so sure whether this limits affect designing optimal STs in reasonable sized network segments. However, if a complex wide area Ethernet carrier network is considered, 64 STs may imply some constraints to the traffic engineering.

MSTP has been utilized in some fault-tolerant Ethernet systems, e.g., Viking [16]. In such systems, a backup ST is allocated for each active ST and in a case of failure, traffic is switched to backup ST. It is reported that fault-recovery time in, e.g., Viking, is 400–600 ms [6]. One could envision a protection system where each protection group (one or more VLANs) are using two disjoint STs in load sharing way. In such system, the connectivity between end points should be monitored using, e.g., IP ping or Ethernet service OAM, and in a case of failure, all traffic would be transferred in the remaining active ST (such scheme has been proposed in [17]).

3.2 STP and Ethernet rings

Fast automatic protection switching based on bidirectional rings is a well known, easy to manage fault recovery mechanism that has been widely used in, e.g., metropolitan area SDH/SONET networks. It does not need any actual planning and fast automatic fail-over switching can reach recovery times in range of ten ms or so. Unfortunately, the switched Ethernet requires a loop-free network topology and thus the ring based protection mechanisms cannot be directly implemented with the standard Ethernet technology.

The only way to utilize ring structures is either to use other Ethernet like technologies and extensions or to impose some kind of MSTP based recovery scheme over rings. The former alternatives include, e.g., Resilient ring protocol and Ethernet automatic protection switching that are shortly described in the next chapter. The latter alternative requires extensive network planning to create required MSTP structures. Furthermore, in MSTP based alternative, the actual recovery operation has to be initiated by edge switches and thus 10 ms scale recovery is likely to be hard to achieve.

3.3 Carrier-grade Ethernet

The concept of “carrier-grade Ethernet” is quite often used when people talk about using Ethernet in the carrier networks. However, this concept is quite confusing as it seems to have many meanings depending on who’s talking. In some cases it means using pure Ethernet technology, e.g., 802.1ad or 802.1ah, to implement transport networks. Yet, more often “carrier-grade Ethernet” seems to mean transporting Ethernet frames over MPLS based network and, in some cases, it can mean transporting Ethernet over SDH/SONET

or GFP — or even over IP based VPNs. To add confusion, “carrier-grade Ethernet” may, in some other cases, mean just supposedly a bit better designed and built Ethernet equipment that should have “nine-fives” reliability[3].

3.4 Ethernet transport

As said in the introduction, this document is focused on using pure Ethernet in transport networks. Thus only 802.1ad and 802.1ah is considered here. The use of Ethernet as a point-to-point link-layer solution between, e.g., MPLS switches is considered in the next chapter.

3.4.1 VLAN stacking

In principle, VID could be used to separate customers in a carrier network. However, this applies only if the customers do not use VLAN tags themselves. That is, such simple scheme cannot be used to support corporate customers using their own VLAN hierarchy. The idea of VLAN stacking or Provider bridge (PB) (802.1ad a.k.a. Q-in-Q) is to add an additional VLAN tag (outer tag) between address fields and the original VLAN tag (customer tag). The outer tag is used to identify customers [10].

VLAN stacking has two serious problems that limit its scalability and reliability. First of all, the size of VID limits the number of the customers to mere 4094 in a single network segment. The other problem also limits the scalability but it may also have serious impact on the reliability. VLAN stacking exposes all the customers’ MAC addresses to carrier network. This means that each provider bridge has to learn possibly a huge number of MAC addresses [15, 8]. While the storage of a huge address data base can be technically solved, there are other issues like address aging and such. Furthermore, MAC address flooding, due to a technical error or a DoS attack, may have a serious effect on the network performance.

VLAN stacking requires also enhancements to GRVP to support millions of VLANs. These enhancements are being standardized in 802.1ak [8].

Another problem with VLAN stacking is that the outer tag is used for two functions, transport separation and customer identification [1]. This can cause some troublesome limitations to network planning.

The above-mentioned problems concern the cases where Ethernet based carrier network is used to provide Ethernet layer transport services for customers. If there is no such need, VLAN stacking can have some uses in creating hierarchy in provider’s transport network, e.g., using one tag for service ID and the other to separate different areas in the backbone layer.

3. As if that would be sufficient to guarantee carrier grade reliability of the network...

3.4.2 Provider backbone bridges

The 802.1ah Provider backbone bridge (PBB) (a.k.a. MAC-in-MAC) is meant to solve the scalability and reliability problems in 802.1ad. It provides full isolation between provider's and customers' networks. The main idea is to wrap the customer's Ethernet frame (802.1, 802.1q or 802.1ad) into another Ethernet frame. This means that the provider's backbone bridges have to manage only provider's internal MAC addresses [15]. Furthermore, Q-tags in the provider's frame can be used for service identification etc [1].

However, PBB still requires a loop free network topology making it impossible to have, e.g., load balancing inside a provider VLAN. Another problem in certain applications like VoIP, is increased framing overhead, that is up to 66 octets per frame.

3.4.3 Ethernet L2 switching

The proposed solution for load balancing is to abuse the Independent VLAN Learning (IVL) mechanism to implement L2 switching. The idea is to allocate a part of the VID address space for defining point-to-point transport paths. This is done by turning off address learning for the designated VID address range. The transport paths are then configured by defining VID/MAC pairs in each bridge in the path using a management system. The traffic forwarding is based in the fact that most of the current Ethernet switches already perform full 60-bit (VID+MAC) address lookup for incoming frames. It should be noted that VIDs do not have to be unique in IVL and thus a huge number of IVL abuse based paths can be supported in a single network segment.

The downside of this scheme is the requirement of an external management system for path management. There is no standard way to implement this but, e.g., GMPLS has been proposed [1].

3.4.4 Clock signal distribution

Ethernet as such is based on free running clocks at each source port and rapid per frame transmit clock recovery at the receiving end. Thus the Ethernet based transport networks are lacking a native clock signal distribution method. This causes troubles for the circuit emulation services over Ethernet (e.g., CESoE) that transport, e.g., E1 signals, over Ethernet. It is proposed that packet based clock synchronization methods could be used to overcome this problem, e.g., IEEE 1588 [18]. However, as far as I know, such method have not been demonstrated in a large scale network — though, it is quite likely that such system could be sufficient to some extent. In any case, the proper and reliable clock signal distribution is an issue that should be considered carefully when Ethernet based transport networks are evaluated.

It should be noted that some consider that the current adaptive clock recovery methods do not meet G.823/G.824 requirements for TDM services under many network conditions. Thus an Ethernet-PHY layer synchronization scheme is under development at ITU-T [8].

3.5 Ethernet OAM

The OAM in office Ethernet relies usually on upper layer, i.e., IP, mechanisms like SNMP, ping and traceroute. This is not sufficient for the carrier-grade transport networks for various reasons, e.g., the IP layer troubleshooting relies on the Ethernet layer, there is no per customer or per service management methods, etc. This is why Ethernet OAM mechanisms are being developed [19, 20, 21]. The Ethernet OAM concept contains three areas: service OAM, link OAM and Ethernet local management interface (ELMI). There are three different organizations developing concurrently definitions for Ethernet OAM (ITU-T, MEF, and IEEE). Fortunately, those organizations are aiming for a consistent set of recommendations and standards.

Typical office Ethernet is quite often monitored by using simple methods from a centralized management system, e.g., IP ping is used to test the reachablity of an Ethernet equipment (or actually the management modules inside it) and SNMP traps are used to track events such as “port down”. It is quite clear that such polling based system cannot have fast reactions in case of failures etc. Furthermore, locating faults can be difficult and it could be hard to detect abnormal network behavior if the SNMP traps are not properly selected.

Service OAM provides tools for managing the customer service instances like Ethernet Virtual Circuit (EVC). Its primary tasks are identifying which customers are affected by a fault, manage EVC faults and test the correct operation of an EVC. These methods are based on IEEE 802.1ag Connectivity Fault Management (CFM). CFM provides normal connectivity testing methods for Ethernet layer:

- Connectivity check messages: a type of a periodic Hello messages
- Link trace messages: traceroute for link layer
- Loopback messages
- Alarm indication signal (AIS)

These messages travel between maintenance endpoints that are located at the edge of CFM maintenance domain. Such domain could be, e.g., an access network. ITU-T is developing similar recommendations: Y.1731, G.8010 and G.8021 [8]. The main difference between 802.1ag and Y.1731 is that the former standard is limited to fault management while the latter one covers also performance monitoring [22].

Link OAM provides management tools for Ethernet links and it is being standardized in IEEE 802.3 Clause 57 standard. Link OAM contains functions: discovery, link monitoring, remote failure indication (RFI) and remote loopback. It allows for operator to monitor and manage all inter-switch links. While the link monitoring is still lacking a way to monitor exact bit error rate (BER), it provides statistics on the number of symbol coding errors that can be used to estimate BER much more accurately than using only the number of frame errors.

ELMI is a management interface for Ethernet based UNI much like ILMI in ATM. The main purpose is to provide user’s side information about the VLAN-EVC mappings, per EVC bandwidth profiles, and the status of EVC.

3.5.1 Reliability and fault modeling

Ethernet OAM copes with methods and functions for the basic OAM. It enables detection of hard faults[4] and assurance of link and path layer connectivity. However, there is another set of failures, so called soft failures[5], that are harder to detect. Actually, defining what is a soft failure and what is normal network operation, is quite hard [23]. Soft failures are usually caused by, e.g., software errors, system misconfigurations or a bad network topology. As the soft failures can degrade the network throughput considerably, it is quite necessary to have at least some basic monitoring for them.

Fault modeling relates to reliability analysis; it is quite easy to apply traditional reliability modeling for Ethernet [24]. However, such modeling can usually capture only the hard failures.

4. e.g., link or switch failure
5. e.g., broadcast storm

Chapter 4

Other Ethernets

In the previous chapter the extensions and enhancements to the standard Ethernet (802.1) were described. In the following, some other relevant Ethernet related technologies are shortly discussed about.

4.1 Resilient packet ring

Resilient packet ring (RPR) network architecture is based on two counter-rotating rings (or a set of such ring-pairs if WDM is used). It is designed to transport Ethernet frames efficiently in metro networks. However, there is no limitation to the ring size, so RPR can be used in Wide area network (WAN) too. RPR does not depend on STP and thus it allows for use of fast automatic ring protection schemes such as wrap-around. There are no dedicated protection resources, that is, both rings are used to transport traffic using shortest paths. However, if all the traffic has to be resumed after a fail-over, the traffic load has to be kept < 0.5. The RPR standard defines under 50 ms fault recovery time for 1000 km rings [7]. Similar performance levels can be achieved with vendor specific STPs, e.g., Alcatel's Rapid Ring STP (RRSTP) [8].

One big weakness in RPR is that it cannot manage faults that require traffic to be transported via another node, e.g., in a case of RPR node failure or a failure in a connection between a RPR node and an external switch[25]. Thus RPR requires other resiliency mechanisms beside itself to achieve end-to-end fault-tolerance. If fast automatic protection path switching is needed, that other resiliency method cannot be Ethernet based.

4.2 Ethernet automatic protection switching

The Ethernet Automatic Protection Switching (EAPS) was developed by Extreme Networks to enable fast fault recovery operations in ring structured Ethernet based MANs. EAPS scheme has been made public by documenting it in an informal RFC (3619) [26]

and, besides Extreme Networks, some other vendors, e.g., Allied Telesyn[1] has implemented compatible systems [27].

The main idea of EAPS is that the data frames are allowed for to be transported only into a single direction (let's call it downstream) it the ring. The RFC has different wording: it is defined that a switch has two ports, one called primary and the other secondary, and the second port is turned to blocking mode. EAPS health-check frames, generated by the master node, are allowed for traveling to the opposite direction (upstream).

If the master node detects a link fault in the ring (missing health-check frames or an explicit link-down message), it sends a control message to all nodes ordering them to flush their forwarding bases and to turn their secondary port from blocking state to forwarding state. The master node itself performs the same operations. In this way, the traffic is restored using normal bridge learning methods. The master node continues to send health-check frames to detect when the failed link is restored. When the connectivity over the ring is restored, all nodes are ordered to flush their forwarding bases and to resume normal operation mode (that is, secondary port in blocking mode).

It is possible to utilize all the link capacity in the ring by using multiple overlapping EAPS domains. In practice, this means that traffic in one set of VLANs is traveling in clockwise direction while the traffic in another set of VLANs is traveling in counter clockwise direction. Both VLAN sets have their own control traffic.

EAPS/EPSR has some short-comings that affects its usability in the carrier-grade networks. First of all, the recovery time target given in the RFC in only less than one second (it is mentioned that around 50 ms is possible to achieve in some cases). This is way behind SDH reaction times but it could be sufficient in some applications. Moreover, it seems quite likely that there will be a short broadcast storm when an EAPS ring resumes its normal operation after a link failure is restored. It is also evident that the master node is a single point of failure and there is no method to recover form master node failure.

EAPS/EPSR causes also some limitations how STP can be used in Ethernet subnets connected into EAPS ring. First of all, STP root node must be EAPS ring node and, secondly, a subnet cannot have redundant connections to the ring, that is, a subnet can be connected only into a single ring node. The first limitation is a configuration issue, but second one limits considerably the options in increasing the fault-tolerance with redundancy.

4.3 Global open Ethernet

Global Open Ethernet (GOE) is an Ethernet architecture developed mainly in Japan for the carrier-grade transport networks. The main idea seems to be in adding a GOE tag (five 32-bit fields) between address fields and customer VLAN tag [28]. In a way, GOE can be considered as an extended Q-in-Q scheme and, likewise, does not solve, e.g., the MAC learning problem. Furthermore, as IEEE is developing different scheme for transport

1. Allied Telesyn calls its EAPS implementation as Ethernet protection switched ring (EPSR).

networks (MAC-in-MAC, 802.1ah), it is quite hard to see any realistic possibilities for GOE scheme to gain any success.

4.4 Ethernet transport links

One common way to utilize the Ethernet technology is to use it as a simple point-to-point link layer protocol between IP routers or MPLS switches. While Ethernet is lacking proper link monitoring capabilities, e.g., BER estimation[2], and its multiple access properties are waste of resources in this type of applications, Ethernet ports are usually so cheap that it is quite justified to use them. However, there are some problems that should be somehow addressed if efficient transport is desired.

One problem is caused by the frame length limitation: if full length Ethernet frames are to be transported over MPLS or PPP over Ethernet, those frames have to be segmented in IP layer or rejected. The former solution causes inefficiency and the latter one transmission failures. If current Ethernet standards are strictly followed, the only solution would be adjusting MTU setting in IP end hosts[3]. This could be impractical in large networks. However, there exists some vendor specific extensions that allow for usage of longer non-standard Ethernet frames. Though, using such extensions would require uniform equipment base. To overcome these problems, IEEE is considering to add such extension to 802.1 standard.

Another problem with Ethernet transport links is lack of proper link management methods (except 10G Ethernet with WAN option that uses simplified SDH framing). Today, the only viable solution in a heterogeneous large network is to use SNMP to monitor link status and performance. In the future, it would be better to utilize Ethernet link OAM functions to manage also point-to-point links.

In some cases, the framing efficiency may become an issue, e.g, in transporting VoIP traffic over Ethernet. When mainly short data packets are transported over Ethernet, the multiple access features in the framing cause considerable overhead. Not only the 48-bit destination and source addresses are excess, but also the 96-bit Inter-frame gap (IFG) and preamble (64 bits) contribute to the overhead. This can easily lead to an situation where the throughput of the actual data is closer to 10% than 50%. There are some vendor specific extensions that can be used to, e.g., shorten IFG. Furthermore, 802.1 standard supports 16-bit addresses besides normal 64-bit ones — however, I do not know, if such option is widely supported. There is also rumors that some vendors are developing some kind of synchronous Ethernet.

In any case, it should be remembered that the efficiency of the Ethernet can cause problems at the edge of the network, e.g., $<$ 10 Mbit/s Ethernet first mile link might not be sufficient to replace E1 link to a PBX.

---

2. One recent proposal to get reliable BER figures, is to monitor 4B/5B or 8B/10B line coding IDLE symbols.
3. MTU autodiscovery process is seldom used

4.5 Ethernet as access concentrator/multiplexer

Ethernet switches are also used or will be used to replace access concentrators and multiplexers. In such applications, the role of the access switch is just labeling then upstream (user) frames with a VLAN tag and forwarding them to central office (e.g., POP/BRAS), and switching the downstream frames to the correct user port according to VID. While these operations can be taken care of by just any Ethernet switch, the security and reliability requirements have to be considered.

An Ethernet based access switch could be connected to various types of backhaul networks. In the simplest case, the connection between the access switch and the BRAS is a point-to-point connection, a e.g., dark-fiber or an Ethernet over SONET/SDH (EoS) line. Ethernet switch can also act as an access multiplexer for MPLS based VPN system. In this case, VIDs have to be mapped to MPLS labels at the MPLS edge switch. Furthermore, VLAN configurations at the access switches have to be somehow integrated into label switch path (LSP) creation and management. Depending on the MPLS VPN type, the interconnection between access switch and BRAS can be like a point-to-point link (virtual private wire service, VPWS, e.g., private wire edge-to-edge emulation, PWE3, over MPLS) or a switched service (virtual private LAN service, VPLS). In the latter case, MAC address learning must be supported at the MPLS edge devices.

In the most complex case, the Ethernet based access switches are connected in an Ethernet based backhaul network. In such case, the problems and the implementation alternatives are the same as described in the previous chapter.

The most important security requirements for an Ethernet based access switch are that a subscriber must not have any access to another subscriber's traffic nor should he/she be able to send traffic directly to another subscriber, and that the management and control interfaces are protected from unauthorized access. The first requirement can be fulfilled by allowing only upstream forwarding, that is, the traffic is forwarded only between an access port and the upstream port. Disabling all signaling protocols at access ports should take care of the second requirement.

To be sure, it would be better not to trust software based configurations to ensure security. Configuration errors and faults in software, firmware and/or hardware can easily invalidate all security measures. Thus the best solution would be to rely on equipment that is designed for this particular task, i.e., devices that, by design and construction, cannot do anything else than upstream forwarding and that cannot process any management or configuration frames arriving at an access port.

MAC flooding (see chapter 6) is one security problem that has to be somehow solved especially in VPLS or pure Ethernet based backhaul networks. If the subscribers have an Ethernet layer access interface, MAC flooding can be used for DoS attack that affect the whole backhaul network and all its users. The current solutions are usually based on configuring or pinning allowed MAC addresses to a certain port. However, it is quite unrealistic to require that operators should begin to manage subscribers' MAC addresses. It would be much better if the access switch equipment could automatically limit the num-

ber of MAC addresses to certain suitable amount (e.g., 10-20 per port) and the frequency of new MAC registrations to, e.g., one per second.

Chapter 5

Implementation issues

All types of communication equipment may have implementation failures in, e.g., software. However, in one sense, Ethernet equipment is a special case: certain failures do not cause service to stop operate — they just undo all security measures and degrade the performance. These failures are also so called soft failures that are usually quite hard to detect and pin-point. Furthermore, there are some common design principles, e.g., plug-and-play and fail-open, that cause security and reliability problems. The plug-and-play related threats are discussed about in the next chapter. In this chapter, some other issues that relate more on the design and implementation are described. These kind of problems are quite hard to find out as they are usually not documented.

It should be noted that in Ethernet, it is quite easy to snoop the make and model of the access network equipment just by looking into its MAC address. The availability of this kind of information, enables a malicious user to utilize known security holes and weaknesses. This is why it is very important to use bug-free, reliable systems and to keep their software and/or firmware up to date. It is also worthwhile to consider changing the default MAC address or using MAC rewrite if the equipment support such operation.

5.1 Software

This far it has been common that Ethernet equipment has lacked any true multitasking operating system in its control and management module. Usually SW has consisted a single monolithic binary and, in the best case, there has been some trivial cooperative multitasking. The lack of pre-emptive multitasking and memory protection between tasks has some serious security and reliability issues. This makes it quite easy to implement efficient DoS attacks as bombarding a single protocol interface will starve also other protocol instances. Furthermore, the lack of memory protection, makes it easier to utilize SW bugs by allowing for overwriting memory areas belonging to other protocol instance.

5.2 Failure modes

The common failure mode for the most of the Ethernet switches is so called “fail-open” mode. It means simply that a switch is failing back to hub mode, that is, instead of switching frames according to MAC address tables, the frames are broadcasted via every port. This failure mode defeat all the security that is thought be provided by switched Ethernet and VLANs. The main cause to use such failure mode, has been the desire to increase reliability: to have at least some network connectivity in error situations.

Unfortunately, the fail-open feature is also a perfect target for attackers hoping to gain open access for sniffing. The common attacks include overloading the switch control unit, e.g., by MAC flooding, or triggering a device specific bug. The fail-open mode is also a soft failure and thus it can difficult to detect. Therefore, at least at the access switch layer, fail-open mode should be disabled. That would, of course, create a possibility for DoS attacks but security would not be compromised and it would be easier to find attackers.

5.3 Known bugs

There have also been many known bugs in Ethernet switches and no-one knows how many bugs remain to be found or are currently known by few persons. What makes some of these bugs quite nasty, is that they can be used to gain an unlimited access to Ethernet infrastructure.

There are couple of examples about known bugs in Cisco’s equipment. In certain switches, sending a TCP packet containing string `"%%"` to Web interface at the management module will crash the whole management software. In some other systems, there were an “undocumented” TCP port and sending a linefeed character to that port will cause a hard reset to factory defaults[1] [29].

Some Ethernet switches accept Q-tagged VLAN frames from any port, i.e., the switching engine does not check if a Q-tagged frame arrived at trunk link or not. This bug is used in so called “VLAN hopping” attack, in which forged frames are inserted into a foreign VLAN via an user port. This also enables attacker to “insert” its MAC address to that foreign VLAN, and if the attacked switch relies only to MAC address lookup in forwarding, the attacker may gain partial access to that foreign VLAN.

1. ...and then you can use the default password to gain access to management module or interact with enabled-by-default STP, CDP, etc.

Chapter 6

Ethernet and security

It is quite common that the modern switched Ethernet with VLANs is considered to be reasonably secure[1]. This (mis)conception is as far from the reality as it can be — Ethernet is very insecure and the attacks against its "security features" are usually quite easy. Thus the most careful approach to the security in the Ethernet based networks is to consider such networks as broadcast networks. It is quite important to understand that many traditional network security solutions, e.g., firewalls, IPSec, are not sufficient to ensure LAN layer security [30]. However, layer 2 security is seldom considered even by security experts [31].

6.1 Threats

There are many ways to compromise availability, confidentiality and/or integrity of an Ethernet network. It is possible to use physical tampering to gain unauthorized access or to bring down the network. However, these attacks are not considered in this paper as preventing physical tampering by protecting equipment and cabling is common to all network technologies. Only remark is that, if the operator's Ethernet network support full plug-and-play autoconfiguration, gaining an unauthorized network access is easy once physical access is found. Furthermore, the broadcast nature of Ethernet makes it easy to cause damage and thus a good layer 2 attack can compromise more than just the local network [31].

6.1.1 Management system

The management interface of Ethernet devices is usually based on some IP layer interface, such as SNMP, HTML or telnet. This causes a security problem especially if users have a direct L2 access to an access network. There are ways to override the access limitations provided by VLANs and thus an attacker can connect directly the management modules in access network Ethernet equipment. This makes it impossible to protect the

1. Sometimes "the security" of switched Ethernet is used to justify plain text passwords or other weak, e.g., Microsoft NTLM, authentication methods.

management system by using firewalls. The only viable solution is to use strong IP layer authentication, authorization and accounting to control the access to management modules and to use strong encryption, e.g., SSL or SSH, to tunnel the management traffic [32].

Event reporting can be another weakness in the Ethernet based access networks. Some systems can report events to the network management system (NMS) using UDP messages. These messages can be easily spoofed, causing false alarms and possibly overloading NMS. While it is true that spoofing event messages requires detailed information about the NMS, it would be better to use other, more secure, methods for event reporting. SNMP is one alternative but no security improvements are gained unless SNMPv3 is used [32].

6.1.2 Ethernet switches

Ethernet switches are usually designed to offer an easy plug-and-play deployment by enabling, e.g., various kinds of autodiscovery methods. While these features make the life of an office Ethernet manager quite easy, they are the main source of the security threats — especially as all such features are usually enabled by the factory default settings.

Switch trunk links can be used to gain access to any VLAN in the network and thus they should be protected from any tampering. Unfortunately, the default behavior of 802.1q and Cisco's Inter-switch link protocol (ISL) is to negotiate a trunk link if the connecting device initiates a trunking protocol. By gaining access to a trunk link, the attacker can participate any VLAN in the network segment bypassing all the network layer security functions, e.g., firewalls. Furthermore, this can reverse the automatic pruning of VLANs, making any VLAN accessible in all ports of the network segment.

In an Ethernet based access or transport network, all trunk links can be designed in advance and they are usually very static. Thus all automatic trunk link negotiating procedures should be disabled and trunk links should be manually configured and managed [32].

Cisco's VTP can cause additional security problems as it has no security features[2]. By gaining access to VTP domain, an attacker can learn what VLANs are configured, become a VTP server and make VLAN changes and participate any VLAN.

Link bundling, like Etherchannel, is a security threat if an attacker can gain physical access to any link in the bundle. A link can be disconnected form the bundle and reconnected to attackers system to gain, e.g., trunk link access. Detecting such attack can be difficult as network connectivity is not lost due to automatic fault recovery offered by link bundles.

CDP is yet another threat. It broadcasts detailed device information periodically all across the network. An eavesdropper can obtain information about device name, IP addresses,

2. Well... you must know the VTP domain name to gain access. That is, however, not very hard task as VTP members broadcast that name to discover VTP-enabled devices on trunk links...

HW platform and capabilities, SW version and VTP management domain — that is, about all the information that is required to compromise the network. Thus CDP should be disabled.

STP is also an insecure protocol. There is no authentication and thus any host can generate STP BPDUs. Forged BPDU can be used for a DoS attack by causing STP recalculate ST over and over again [31]. Furthermore, an attacker can add a bridge in the network segment and get it selected as a root bridge. This can reduce the performance and the false root bridge can be used to capture and alter traffic. Unfortunately, STP is usually enabled by default and thus it should be disabled manually in all user ports[3]. Furthermore, STP topology changes should be monitored by NMS.

Some Ethernet switches can be forced to the fail-open state by flooding the MAC address table by sending large amount forged MAC addresses to it. There are some methods, like Cisco's port security and MAC hardcode, that could be used prevent MAC flooding. However, they are not realistic in large dynamic networks.

6.2 Known attacks

There are multiple known attacks, e.g., to gain trunk access or to insert a false root bridge. Most of these attacks can be utilized either for DoS attack or for enabling network layer attacks. There are also some upper layer attacks that are closely related to Ethernet, e.g., ARP spoof and DHCP spoof.

Spanning Tree Protocol (STP) attack is used to become the root bridge or other active switch in the network. It can be used also for DoS attack. Multihomed STP attack can be used to enable Man-in-the-middle (MiM) attack.

Cisco Discovery Protocol (CDP) can be used to gather information for other attacks. However, as there is no authentication, it can be used also for a DoS attack by flooding CDP tables.

Dynamic Trunking Protocol (DTP) is enabled by default in Cisco's switches. Can be used to gain trunk access.

Hot Standby Router Protocol (HSRP) can be used to become a new active router in the network. Can be used to enable MiM attack if a spoofed IP address is used in the registration.

3. There has been a bug in older Cisco devices allowing for bypassing STP-blocked ports by using 802.1x.

VLAN Trunking Protocol (VTP) can be used to delete and add VLANs from a single host by becoming a VTP server.

802.1q attack can be used to gain trunk link access. Double encapsulated frames can be used to insert traffic in other VLANs (VLAN hopping).

802.1x in some systems can be used to bypass other security measures.

Inter-Switch Link Protocol (ISL) same as DTP.

MAC flooding, a.k.a., CAM flooding, MAC overflow, or macof, can be used to force a switch in the fail-open state and thus enable other attacks.

6.3 Tools and other resources

There are some tools for Ethernet layer attacks that are freely available in the Internet. Usually these tools are made public in order to get people to realize the security risks in Ethernet and to get vendors to react. Here are couple of examples that I found by making a short trivial Google search.

6.3.1 Yersinia

Yersinia[4] (`http://www.yersinia.net/`), a framework for layer-2 attacks, is a versatile and easy-to-use tool-set that implements attacks against STP, CDP, DTP, DHCP, HSRP, 802.1q and VTP [33].

6.3.2 Dsniff

Dsniff (`http://naughty.monkey.org/~dugsong/dsniff/`) is a set of tools that can be used to capture and alter traffic in switched Ethernet network. It uses Arpspoof, Dnsspoof and MAC flooding attacks to route traffic between a victim and the default gateway to pass through attackers device (MiM attack). It should be noted that Dsniff can sniff also SSL and SSH connections by splitting the connection to two parts and presenting a false certificate or host finger print to the victim.

4. named after Yersinia pestis, the bacteria that causes plague

6.3.3 Parasite

THC-Parasite (`http://thc.segfault.net/`) allows for sniffing on switched networks by using either ARP Spoofing or MAC Flooding. THC-Parasite is intelligent and its algorithms are designed to bypass the basic switch security. THC site contains also dozens of other tools that can be used to, e.g., crack passwords once sniffing is enabled.

6.3.4 Default password lists

There are also many default password lists for various network equipment and those list are quite easy to find by using, e.g., Google. Changing default passwords should be a standard security measure but some devices have had "features" that can be used to remotely reset the device to factory defaults.

6.4 Security measures

The security problems with trunk links can be best solved by disabling all automatic trunk negotiation and management protocols. Trunk links should always be configured and managed manually or by NMS. There are some proprietary security measures like Cisco's VLAN membership policy server (VMPS) that monitors MAC addresses to control VLAN access. However, VMPS is quite useless in environments where MAC addresses tend to change often. Moreover, the VMPS database, stored at the server, is retrieved via TFTP, which is vulnerable to spoofing, alternation and denial of service.

I think that the only viable solution for secure Ethernet access interfaces, is to create a well defined Ethernet UNI that completely isolates the user from control and management planes. To be sure, such isolation should be implemented at hardware layer and there should be no configuration option, or what ever method, to disable it.

Ethernet VPN services for corporate users a bit more difficult to secure if customer VLANs are enabled. It means that the UNI has to be a trunk link which makes it hard to prevent VLAN based attacks in plain Ethernet. The only option would be disabling all dynamic trunking protocols and managing all customer VLAN configurations manually. A better option would be using Provider Bridges (Q-in-Q)[5] or Provider Backbone Bridges (MAC-in-MAC) at the network edge.

One thing that should be noted is that the source(s) of an attack can be hard to determine in an Ethernet LAN segment. The attacker(s) can use forged source MAC addresses and usually no trace back is left once a MAC address is aged out. Thus it could be impossible to determine which port and end-host an attacker has used for, e.g., Dnsspoof, Arpspoof or DoS attack.

5. I am not sure if it does provide sufficient protection, e.g., it does not directly address MAC flooding.

Chapter 7

Operating Ethernet based transport network

In the previous chapters the basics Ethernet, its new features and various problems are described. Using that information it is possible to outline some procedures that should be followed in operating Ethernet based transport networks. However, there are so many possible ways to implement such networks, using either standard or vendor specific equipment, that these guidelines have to be quite generic. Furthermore, one should have quite extensive hands-on experience on network planning and management to be able to write down detailed and well justified operations instructions. Thus the reader should consider the following as a partial check-list on avoiding the worst errors.

7.1 How to achieve complete failure

It is quite common to think that Ethernet networks are extremely simple to manage and that the Ethernet equipment is based on so well proved technology that one can select just any, preferably the cheapest, system. However, it is quite easy to identify some common errors that are very likely to cause troubles, e.g., low reliability, security risks, etc. So, here is the list how to reach for complete failure:

- Use always the cheapest equipment: a perfect way to ensure that you end up with a network full of equipment with incompatible configuration procedures to maximize the probability for human configuration errors.
  - As a bonus, incompatible STP implementations are known to produce real nice problems
- Don't purchase equipment that requires paying for 3–5 years on-site support: it ensures that you won't get replacements, spare parts nor firmware updates.
- Replace your transport network planners and operators with guys who have installed offices LANs for couple of years: as Ethernet is an plug-and-play technology, who needs overpaid staff.
- Do not bother to disable any neat autoconfiguration protocols: they are there to make your life easier and, anyway, subscribers' PCs don't understand those protocols. Right?

- There is no real need for configuration management — those devices have plug-and-play autoconfiguration.
- Anomality detection? Why? The network is operating just fine, except occasional cable cuts and dead lasers.

7.2 Default configurations

One common problem with the office Ethernet equipment is that it is usually shipped with a configuration that allows for all the possible services for all ports and disables all security features. If such equipment is used in an operator's network, it contains clear security and performance risks. Due to human errors, these risks are in their greatest if the equipment has to be manually configured. Thus it would be better to use equipment that can be initialized by down-loading a configuration file with a known safe configuration. The best choice would be using only such carrier-grade Ethernet equipment where nothing is allowed by default.

In any case, the configurations should be centrally managed and some kind of periodic rule-based sanity-checks would reveal possible configuration errors.

7.3 Role of VLANs

It is not so rare that VLANs are considered as a security feature. They are not. No-one should design and operate a Ethernet network where the security relies on VLANs. There are many ways to compromise VLAN "security" and even some common phenomena may cause complete failure, e.g., traffic overload in one switch may result in a transition to fail-open mode and thus VLAN traffic is leaked everywhere.

VLANs can be used to tag user frames, e.g, when Ethernet switches are used as access concentrators or in Q-in-Q networks. Furthermore, VLANs with MSTP can be used enable some level of traffic engineering.

The illusory security of VLANs can cause some troubles when, e.g., NMS is operating in its own VLAN and network maintenance is outsourced. In this case, it is quite possible to misunderstand the impact on security and to grant an Ethernet layer VLAN access for the external network operators. In reality, such access won't give only an access to NMS VLAN but potentially to the whole network.

7.4 Usability and scope of spanning trees

Some clear rules can be suggested for using spanning tree protocols in carrier-grade networks:

- If you can do it without STP, that would be the best solution. E.g., if an Ethernet based access network is strictly tree structured (that is, there is no redundant links that can create loops), STP can be disabled.
- Do not use standard STP, RSTP is the minimum requirement.
- Try to ensure that a high-availability, carrier-grade Ethernet switch is (always almost certainly) selected as a root bridge to reduce the possibility of dead root bridge problems.
- Use MSTP where possible to avoid fault propagation between VLANs and to allow for (some) load balancing.
- Minimize the size of (R)STP domains (or divide the network in MSTP regions) to speed up fault recovery and to limit fault propagation. E.g., if there is no redundancy in access network segments and some redundancy in metro (access-feeder) network, (R)STP can be disabled in access networks and each access network could be connected to POP/BRAS with dedicated VLAN, which allows for use of MSTP.
- Ensure that all customer ports are declared as edge ports that all BPDUs from edge ports are blocked.

One possible way to reduce the need for STP in the cases where redundant links are required, is to use fail-safe Ethernet switches and link bundles. E.g., if there is a need to have two redundant links from an access concentrator to the metro network edge switches, the normal edge switches could be replaced with fewer fail-safe switches and each concentrator could be connected to one fail-safe switch with two or more disjoint links. In that way, the redundant links can form a link-bundle and thus there is no need for STP. Furthermore, there can be automatic load-sharing between links. The downside of this idea is that the fail-safe Ethernet equipment is much more expensive that the plain vanilla switches.

It should be noted that disabling STP creates some risks. If there is a possibility to create a loop by, e.g., connecting two STP-disabled user ports together, the loop avoidance will fail and there will be a broadcast storm. One way to reduce the risk for such incidence, is to disable automatic link polarity detection in user ports. That should ensure that no-one creates a loop by mistake with a normal straight cable. Furthermore, in an access network, each port should be placed in a dedicated VLAN and default VLANs should be disabled.

7.5 Physical layer management

The current possibilities for physical layer management in Ethernet are minimal. However, it would be somehow beneficial to monitor all the available information, e.g., frame error rate, by collecting and analyzing relevant SNMP counters. In the future, many of these management deficiencies should be fixed by forthcoming OAM standards like IEEE 802.1ag and 802.3 Clause 57. Due to some limitations in IEEE standards, e.g., lack of performance monitoring, the work done in ITU-T is also relevant. At this phase, it

could be wise to design all new Ethernet installations to include support for the above-mentioned OAM functions.

Chapter 8

Conclusions

It is true that Ethernet can be used as an easy plug-and-play network technology needing very little management. However, that truth is not universal: it is applicable only to small intra-office networks. In any larger scale or in an environment with any security requirements, that does not hold. The office Ethernet technology lacks scalability, security and management features required in anything else than the most trivial carrier networks. Careless planning and commissioning of Ethernet based networks will cause nothing but problems.

The recent advantages in STP (RSTP and MSTP) as well as Ethernet OAM have increased the capabilities of Ethernet to a level where it can be used in some applications in the carrier networks, mostly in transporting data traffic, which does not need fast protection. However, providing a data-link layer peering connectivity between the user's equipment and the carrier network devices, is not any safer.

VLAN is, in many occasions, considered as technology to provide security by isolating the traffic between subscribers. However, this is based on misunderstanding the basic nature of Ethernet, that is, broadcast service. Thus using VLAN based "security" to provide Ethernet Service should be avoided. The new Provide bridge scheme (Q-in-Q), that supports stacking of VLAN tags, provides much better security. Yet, Q-in-Q has its own short-comings related to, e.g., MAC learning that allows for MAC flooding and has scalability issues. The best solutions to these problems would Provider Backbone Bridges scheme (MAC-in-MAC) that isolates the user equipment completely from the providers network.

One clear conclusion can be made when the applicability of Ethernet for carrier grade networks is considered: the common Ethernet equipment tend to promote the possibility for human errors. That is, without careful planning, configuration management and installation procedures, it is likely that some configuration options regarding, e.g, fault tolerance and security, are forgotten to be changed from default values, or are misconfigured. Furthermore, using advanced options like MSTP regions or using MSTP for traffic engineering require skilled network planning. All this is quite contrary to common wish that Ethernet would reduce the operational expenditures by allowing an easy plug-and-play operation.

Alternative ways to offer Ethernet Services were intentionally left out from this document. However, it should be noted that these alternative ways, especially Ethernet over MPLS, are, in most cases, considered to be the right way to implement “carrier-grade Ethernet” [8, 9, 10, 25, 34]. Some operators even prefer to upgrade their SDH/SONET network to NG-SDH/SONET standards to provide Ethernet services, e.g., Verizon [35].

Bibliography

[1] David Allan, Nigel Bragg, Alan McGuire, and Andy Reid. Ethernet as carrier transport infrastructure. *IEEE Communications Magazine*, February 2006.

[2] LANs and VLANs: a simplified tutorial. Avaya Labs Application Note, May 2002.

[3] Radia Perlman. Myths, missteps, and folklore in protocol design. In *The Proceedings of USENIX 2001, Annual Technical Conference*, 2001.

[4] Andy Myers, T. S. Eugene Ng, and Hui Zhang. Rethinking the service models: scaling Ethernet to a million nodes. In *The Proceedings of ACM SIGCOMM Workshop on Hot Topics in Networking*, 2004.

[5] Ethernet tutorial. Fujitsu Application Note, April 2006.

[6] Padmaraj M. V. Nair, Suku V. S. Nair, Marco F. Marchetti, Girish Chirovolu, and Maher Ali. Distributed restoration method for Metro Ethernet. In *Proceedings of the International Conference on Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies (ICNICONSMCL'06)*, 2006.

[7] Elie Sfeir, Sandrine Pasqualini, Thomas Schwabe, and Andreas Iselt. Performance evaluation of Ethernet resilience mechanisms. In *Proceedings of IEEE Workshop on High Performance Switching and Routing (HPSR 2005)*, May 2005.

[8] Lubo Tancevski, Maarten Vissers, Michael See, and Italo Busi. Evolution of Ethernet for data transport networks. *Alcatel Telecommunications Review*, (3):2–8, 2005.

[9] Jiyang Wang. Optical Ethernet: making Ethernet carrier class for professional services. *Proceedings of the IEEE*, 92(9):1452–1462, September 2004.

[10] Girish Chiruvolu, An Ge, David Elie-Dit-Cosaque, Maher Ali, and Jessy Rouyer. Issues and approaches on extending Ethernet beyond LANs. *IEEE Communications Magazine*, pages 80–86, March 2004.

[11] Junko Yoshida. Carrier-grade Ethernet faces obstacles. *CommsDesign*, June 2004.

[12] Amaro F. de Sousa. Improving load balance and resilience of Ethernet carrier networks with IEEE 802.1s Multiple Spanning Tree Protocol. In *Proceedings of the International Conference on Networking, Systems, Mobile Communications and Learning Technologies (ICNICONSMCL'06)*, 2006.

[13] M. Padmaraj, S. Nair, M. Marchetti, G. Chiruvolu, M. Ali, and A. Ge. Metro Ethernet traffic engineering based on optimal multiple spanning trees. In *Proceedings of 2nd IEEE/IFIP International Conference on Wireless and Optical Communication Networks (WOCN 2005)*, 2005.

[14] Xiaoming He, Mingying Zhu, and Quigxin Chu. Traffic engineering for Metro Ethernet based on multiple spanning trees. In *Proceedings of the International Conference on Networking, Systems, Mobile Communications and Learning Technologies (ICNICONSMCL'06)*, 2006.

[15] Xiaoming He, Mingying Zhu, and Qingxin Chu. Transporting metro Ethernet services over metropolitan area networks. In *The Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (SUTC'06)*, 2006.

[16] Srikant Sharma, Kartik Gopalan, Susanta Nanda, and Tzi cker Chiueh. Viking: A multi-spanning-tree Ethernet architecture for metropolitan area and cluster networks. In *Proceedings of the Twenty-third Annual Joint Conference of IEEE Computer and Communications Societies (INFOCOM 2004)*, volume 4, pages 2283–2294, 2004.

[17] János Farkas, Csaba Antal, Gábor Tóth, and Lars Westberg. Distributed resilient architecture for Ethernet networks. In *The Proceedings of 5th International Workshop on Design of Reliable Communication Networks (DRCN 2005)*, pages 515–522, 2005.

[18] Georg Gaderer, Roland Höller, Thilo Sauter, and Hannes Muhr. Extending IEEE 1588 to fault tolerant clock synchronization. In *Proceedings of IEEE International Workshop on Factory Communication Systems*, pages 353–357, 2004.

[19] Mike McFarland, Samer Salam, and Ripin Checker. Ethernet OAM: Key enabler for carrier class Metro Ethernet services. *IEEE Communications Magazine*, November 2005.

[20] Dirceu Cavendish. Operation, administration, and maintenance of Ethernet services in wide area networks. *IEEE Communications Magazine*, pages 72–79, March 2004.

[21] Hiroshi Ohta. Standardization status on carrier class Ethernet OAM. *IEICE Transactions on Communications*, E89-B(3):644–650, March 2006.

[22] Dinesh Mohan. NGN OAM capabilities. a slide show in ITU-T Workshop "NGN and its Transport Networks", Kobe, April 2006.

[23] Roy A. Maxion and Frank E. Feather. A case study of Ethernet anomalities in a distributed computing environment. *IEEE Transactions on Reliability*, 39(4):433–443, October 1990.

[24] Gary W. Scheer and David J. Dolezilek. Comparing the reliability of Ethetnet network topologies in substation control and monitoring networks. Schweitzer Engineering Laboratories, Inc. technical note, 2000.

[25] Ingrid Van de Voorde, Lubo Tancevski, Girish Chiruvolu, Yves T'Joens, and Jeanne De Jaegher. Carrier-grade Ethernet: extending Ethernet into next generation metro networks. *Alcatel Telecommunications Review*, (3):1–7, 2002.

[26] S. Shah and M. Yip. RFC 3619: Extreme Networks' Ethernet Automatic Protection Switching (EAPS), October 2003. informal, http://www.ietf.org/rfc/rfc3619.txt.

[27] Ethernet protection switched rings: creating the survivable Ethernet network. Allied Telesyn white paper, July 2004.

[28] Aleksandar Kolarov, Bhaskar Sengupta, and Atsushi Iwata. Design of multiple reverse spanning trees in next generation of Ethernet-VPNs. In *Proceedings of IEEE Globecom*, pages 1390–1395, 2004.

[29] Aaron D. Turner. Network insecurity with switches. a study published in the Internet, December 2000.

[30] Rinat Khoussainov and Ahmed Patel. LAN security: problems and solutions for Ethernet networks. *Computer Standards & Interfaces*, 22:191–202, 2000.

[31] Bruce Potter. Layer 2 security: in vogue. *Network Security*, pages 18–20, November 2005.

[32] Richard Wagner. Securing network infrastructure and switched networks. SANS Institute information security document, August 2001.

[33] David Barroso and Alfredo Anders. Yersinia: framework for layer 2 attacks. Black Hat Briefings, 2005.

[34] Maher Ali, Girish Chirovolu, and An Ge. Traffic engineering in Metro Ethernet. *IEEE Network*, March/April 2005.

[35] Haidar Chamas, William Bjorkman, Stephen Liu, Lily Chen, and Mohamed A. Ali. Verizon experience with NG Ethernet services: evolution to a converged layer 1, 2 network. *IEEE Optical Communications*, August 2005.