Switched LAN Design

Fulvio Risso
Politecnico di Torino

Copyright notice

- This set of transparencies, hereinafter referred to as slides, is protected by copyright laws and provisions of International Treaties. The title and copyright regarding the slides (including, but not limited to, each and every image, photography, animation, video, audio, music and text) are property of the authors specified on page 1.

- The slides may be reproduced and used freely by research institutes, schools and Universities for non-profit, institutional purposes. In such cases, no authorization is requested.

- Any total or partial use or reproduction (including, but not limited to, reproduction on magnetic media, computer networks, and printed reproduction) is forbidden, unless explicitly authorized by the authors by means of written license.

- Information included in these slides is deemed as accurate at the date of publication. Such information is supplied for merely educational purposes and may not be used in designing systems, products, networks, etc. In any case, these slides are subject to changes without any previous notice. The authors do not assume any responsibility for the contents of these slides (including, but not limited to, accuracy, completeness, enforceability, updated-ness of information hereinafter provided).

- In any case, accordance with information hereinafter included must not be declared.

- In any case, this copyright notice must never be removed and must be reported even in partial uses.
Outline

- Introduction
- Design criteria
- Design of switched networks
- Debug
- Fault tolerance
Introduction

- Network is the backbone of all information system
  - If it works, nobody notices it
  - If it doesn’t, everyone complains (also the CEO)

- Please note that...
  - If something else doesn’t work properly, the problem will always be the network
  - People never blame servers, applications, ...

- Therefore...
  - Your network must be as good as possible in order to demonstrate that it’s not your fault!
Design criteria

- Focusing on L2 networks

Criteria
- Reliability
- Security
- Performance
- Additional features
- Fault tolerance
Reliability (1)

- Good cabling system is a fundamental prerequisite
  - Several faults (usually intermittent and very difficult to diagnose) may arise in case of a poor quality cabling

- Selection of network devices
  - Different families of network devices, apparently with same characteristics
  - What about redundant modules?
  - What about MTBF?
Reliability (2)

- Observance of standard specifications
  - Do not exceed the known limitations of the standards
    - Cabling
      - Particular attention is needed for fiber-optics backbones
    - Attenuation
    - Number of cascading switches
    - ...
  - Other reasons (not related to reliability)
    - Required for interoperability
    - Good for debugging (the standard says what the device should do)
Performance

- Two aspects
  - Dimensioning of network devices and link bandwidth
  - Network topology
- In both cases, an adequate traffic study is required
Performance: traffic survey

- Traffic typology
  - Client-server, peer-to-peer
  - Departmental servers, or corporate servers
    - Servers (with higher bandwidth) near users or in datacenter
    - Mostly internal to the LAN, or mostly toward the Internet
- Special events (e.g. corporate-wise conventions)
- Traffic monitoring (over different time scales) may be required
  - In case of new installations, we can try with a traffic survey of some similar companies
Performance: selection of devices / links

- Given the traffic survey, we can choose devices / links

- Selection of network devices
  - Possibility to accommodate fastest network interfaces
  - Internal switching capabilities (packet processing throughput)
    - Attention required for multicast and/or other special traffic

- Links
  - Bandwidth (perhaps the most important parameter)
  - Link type (e.g. copper, fiber, ...)
  - Other characteristics (e.g. simple fiber, armored fiber, ...)
Performance: dimensioning

- The most common approach is to over-dimension the network...
  - Inexpensive
  - Simplest to achieve
  - Simple to manage
  - No traffic engineering
  - No resource reservation

- ... and setup a continuous monitoring infrastructure in order to detect bottlenecks as soon as possible

- Often the bottleneck is the connection to the Internet, which is usually slower than the internal network
  - Cannot over-dimension the Internet connection due to cost problems
Network topology (1)

- Key decision for achieving performance, reliability, security, fault tolerance
- Unfortunately, often network topology is in some sense forced by some external constraint
  - E.g. location of the wiring cabinets
    - Interior designers seems to have more importance than network engineers
  - Network specialists must do their best anyway
Network topology (2)

- Network performance highly depends on the quality and topology of the underlying cabling system
  - Best choice: design everything at the same time
    - Wiring closets and cabinets
    - Cabling conduits
    - Link/device topology
    - Link/device dimensioning
    - Servers positioning
Logical topology (1)

Core (or backbone)

Distribution and aggregation

Access
Logical topology (2)

- Core/backbone
  - Usually between different buildings in the same campus
  - Usually concentrated in a few switches, connected to the corporate data center

- Distribution/aggregation
  - Usually within the same building (vertical wiring)

- Access
  - Usually connects hosts on the same floor (horizontal wiring)
  - User control (e.g. 802.1x, ...)
  - Reliability may not be so important

- In all cases, point-to-point links
Logical topology: backbones

- Star-based system
  - $N$ devices, $N-1$ links (with no resiliency at all)
  - Highly scalable (we can add new links from the star center or upgrade the star center in order to have more bandwidth)

- Ring
  - Very efficient in terms of resiliency
  - “Shared” bandwidth
  - $N$ devices, $N$ links (with resiliency)

- Mesh
  - Usually discouraged
  - Large number of links/devices, no clear outcome of the network in case of fault
  - Difficult to debug
Logical topology: link speed

- Important to have an adequate difference between access and distribution/core
  - Limits dropped frames in L2 network
  - QoS issues
- Usually, 100Mbps is enough
  - Most people (vendors?) prefer 1Gbps, though
Logical topology: example of a building

Floor cabinet

1Gbps / 10Gbps

Vertical wiring

Horizontal wiring

100Mbps / 1Gbps

Data center (CED)

Multilayer switch

Router

Internet
Debug

- In the common belief, the problem is always the network
  - Either in terms of reachability ("I cannot reach my web server")
    or in terms of performance ("the server is slow")
- Therefore
  - The network has to be reasonably robust
  - We must have debug facilities
    - For debugging the network (and, more important)
    - For debugging servers and clients
- Mirror (also known as "span") ports are a must
Additional features

- Power over Ethernet (PoE)
- Quality of Service (QoS)
- VLANs
- Security
  - E.g. 802.1x
Fault tolerance

- The network must be able to operate also when facing one or more failures
  - Links
  - Devices
  - Device parts
    - Interfaces
    - Power suppliers
How do we achieve fault tolerance?

- Adding redundancy on critical elements
  - Interface level
    - Parallel interfaces
    - Redundant ports
  - Device level
    - Processor
    - Power supplier
    - NICs
  - Network level
    - Additional links (i.e., alternate paths)
    - Duplicating a device (e.g., a second (backup) switch)
- Combining all of these
- Robust devices, or many devices with backup capabilities?
How much redundancy?

- Each new element has a fault probability and a cost
  - Fault probability of each element must be analyzed carefully
- Too many elements may
  - Increase fault tolerance capabilities marginally
  - Increase costs substantially

- Fault tolerance is always a compromise among
  - Real fault tolerance needs
    - How much does it cost to my organization a stop of N minutes in the network?
    - Please note that a stop of N min of the network may cause a stop of M min of some services
  - Cost
The golden rule

The fault tolerant solution must be as simple as possible and use the lowest number of redundant elements required to guarantee a “path” that is alternative to the faulty one.
Good practices: power

- Redundant devices must have an independent power supply
  - Two power units, connected to different electrical backbones
  - Two independent electrical backbones

- Systems for the management and control of the network for revealing anomalies and faults
  - Otherwise, the fault may happen and it remains unfixed because the network manager does not notice it

- Uninterruptible Power Supply systems for important devices
  - Usually 15-20 minutes with batteries
  - Then, a power generation must be activated

- Power distribution must be done with care
Good practice: cabinets

- Cabinets and data centers are often in the basement
- Check that everything is safe in case of flooding
Good practice: links

- Redundant links
  - Fiber is better, especially in backbone
    - Copper is an electric conductor
      - Lightning
      - Some electrical cable that goes in touch with networking cables
  - Armored links (if needed)
  - Fiber over long distances
    - We may have intermittent problems (link flapping)
    - A de-flapper mechanism may be extremely useful
      - Especially if RSTP is used
Good practice: devices

- Redundant devices (e.g. the star center)
- What about servers?
Good practice: redundant paths

- Link Aggregation (when possible)
- Spanning Tree
  - Network analysis of the topology in case of fault of the most critical links/devices
    - Appropriateness of the resulting topology
  - Customization of BridgeID for Root bridge and backup root bridge
Spanning Tree and Fault Reaction

- Fault Reaction in 50 seconds
  - Is this time appropriate for my network?
  - Re-convergence of other services may be higher than 50s
- In case faster reaction is needed
  - New values for timers
  - Rapid STP
- STP limits
  - Max 7 bridges (also on the topology that comes out after a fault)
  - Single spanning tree (i.e. unused resources)
    - VLANs and MST?
    - L3 routing?
Redundant backbone: example
Conclusions

- Network is the backbone of any information system
- Not easy to design a good network
- Many different aspects
  - From electrical system, to location of cabinets, to cabling, networking equipment, network topology, network protocols, air conditioning, data centers
- Perhaps the most difficult problem is to foresee all the possible faults
- Experience matters