Switched LAN Design

Fulvio Risso
Politecnico di Torino
Introduction

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

- 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 work properly
  - You must clever enough to have data that demonstrate that it is not your fault!
Design criteria

- Focusing on L2 networks

Criteria

- (A) Reliability
- (B) Security
- (C) Performance
- (D) Modularity and extensibility
- (E) Debugging
- (F) Additional features
- (G) Fault tolerance
(A) 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
    - E.g. Are you sure that your cables will follow the shortest path when connecting point A to point B?

- 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
    - ...

(B) Security

- Network isolation
  - VLANs
  - Access Control Links (at various level)
- 802.1x
(C) Performance

- Two aspects
  - Dimensioning of network devices and link bandwidth
  - Network topology
- In both cases, an accurate 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 (frames per second)
    - Attention required for multicast and/or other special traffic
    - Necessity of QoS capabilities (e.g. hw queues on interfaces)

- Links
  - Bandwidth
  - Link aggregation
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 fault tolerance 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 and Spanning Tree

- Have you considered that the actual topology depends on the configuration decided by the STP?
  - Customize Bridge ID for better Root bridge selection and Designated Port Selection
  - Do not forget to design the network in order to perform well also in case of the most critical failures (e.g., root bridge)

- PVST may be another option
Logical topology and 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
- Vertical wiring
- Horizontal wiring
- 1Gbps / 10Gbps
- 100Mbps / 1Gbps
- Data center (CED)
- Multilayer switch
- Router
- Internet

1Gbps / 10Gbps

Internet
(D) Modularity and Extensibility

- Links
  - Link type (e.g. copper, fiber, ...)
    - GBIC, for a better choice of the proper physical technology
  - Other characteristics (e.g. simple fiber, armored fiber, ...)

- Devices
  - Fixed format
  - Chassis-based
  - Hardware-based vs software based
    - Impacts performance as well

- Configurability
  - Fixed features, no field-upgradable operating system
  - Field-upgradable operating system (with new features)
(E) Debugging (1)

First, adopt standard technologies
- The standard specifies how the device should operate
- Required for interoperability as well

Second, be prepared in case users complain about the network
- 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
Debugging (2)

- Remember that, 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")
- Be prepared to demonstrate that it’s not your fault
  - Network traffic live recording
    - In case regulations allow you to do so
  - Monitoring tools
Debugging (3)

- Prevention is better than cure
  - 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
    - E.g., automatic network reconvergence (STP)
(F) Additional features

- Power over Ethernet (PoE)
- Quality of Service (QoS)
- VLANs
(G) 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
- So, better to duplicate only the weakest elements
- 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 can replace 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
- 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
  - Different distribution lines for network and other users (e.g. lights)
    - Are you sure that you have no stoves connected to your distribution line?
  - Multiple lines for network devices for redundancy
    - What about everything under the same differential switch, which may go off?
Good practice: cabinets

- Cabinets and data centers are often in the basement
- Check that everything is safe in case of flooding
  - Do you have water pumps in order to keep your datacenter safe?
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

BP: 24576
BP: 28672

Internet
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