Recommendations for a redundant campus network

Best Practice Document

Produced by UNINETT led working group
on campus networking
(UFS114)

Authors: Gunnar Bøe, Vidar Faltinsen,
Einar Lillebrygfjeld
December 2011

Document No: GN3-NA3-T4-UFS114
Version / date: December 2011
Original language: Norwegian
Original title: “Anbefalt feiltolerent campusnett”
Original version / date: 2011-12-16
Contact: campus@uninett.no

UNINETT bears responsibility for the content of this document. The work has been carried out by an UNINETT led working group on campus networking as part of a joint-venture project within the HE sector in Norway.

The research leading to the results of this report has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 238875, relating to the project 'Multi-Gigabit European Research and Education Network and Associated Services (GN3)'.

Table of Contents

Executive Summary

This document presents recommendations for the configuration of a redundant campus network. A campus network may be divided up into three parts: the core network, the distribution network and the access network. Recommendations are made for each of these parts.

In the case of a core network, we recommend a structure consisting of two equipment rooms with completely separate electrical and cooling systems. We also recommend that the network originating at these equipment rooms be a redundant fibre-optic structure. The core network should consist of at least two core switches, configured with redundant BGP connection to the NREN. There should also be redundant connections to distribution and/or edge switches. A traditional configuration may be chosen, based on IS-IS/OSPF, VRRP/HSRP and Rapid Spanning Tree (RSTP), or one could invest in a virtual, proprietary core using link aggregation (IEEE 802.3ad) directed to distribution and edge switches. Each of these configurations has advantages and disadvantages.

In the distribution network we recommend the use of Rapid Spanning Tree (RSTP, IEEE 802.1w). New, improved protocols such as TRILL and IEEE 802.aq have now been standardised, though equipment support is at present not satisfactory. If one has a large number of VLANs, MSTP (IEEE 802.1s) should be considered, as it provides good opportunities for load-sharing of traffic in groups of VLANs (with each group running RSTP). Alternatively, one may use proprietary systems with spanning trees per VLAN.

The distribution network may be designed in various ways. This recommendation focuses on four different scenarios, explaining the advantages and disadvantages in each case.

In the access network it is important to configure edge switch ports as “edge ports”. This ensures that switching on and off terminal equipment does not trigger new spanning tree computations.

Servers should be connected redundantly to the network using two different switches (typically so-called “top-of-rack switches”). A commonly used and recommended configuration is to use Ethernet Bonding on Linux servers and Ethernet Teaming on Windows servers. These should be configured in active/active mode, so that outgoing traffic from the servers uses both connections, providing assurance that the fault tolerance functions. An alternative to bonding or teaming is to use multichassis link aggregation, but this will require proprietary systems on the switches. For virtual servers connected to switches in a blade system one must arrange redundant connections to the blade system.

1 Introduction

This document presents a recommendation for the configuration of a redundant campus network. In general, campus networks may be divided into three layers:

- Core network
- Distribution network
- Access network

The different layers will be dealt with the in the following chapters.

This document does not take into account the capacity of the various links in a campus network, which is independent of design and may be upgraded gradually as required. In general we can say that new implementations should currently support 10 Gbps Ethernet in the core network and 1 Gbps out towards the periphery of the network.

We have attempted to keep the recommendation general and independent of suppliers, although a number of specific references are made to certain equipment types. It should be mentioned that the supplier references are not exhaustive. Since equipment supplied by Cisco and HP dominates in the HE sector in Norway, these are the most commonly described.

2 The core network

In the core of larger campus networks, routing is achieved by means of a dynamic routing protocol such as IS-IS or OSPF. Redundant routing between the campus network and the outside world is by means of BGP. Smaller campus networks without a redundant core network use static routing only.

2.1 A single equipment room

This scenario is typical of many colleges in the Norwegian HE sector (the larger universities and some colleges have a more redundant design). Figure 1 shows a typical network configuration in such a case.

There is only **one** equipment room on the campus:

- This is the point of connection for the NREN. In many cases the connection is redundant: in other words there are two separate routes leaving the campus. It is important that the NREN is redundant, but this is not the main focus of this recommendation.
- The institution's core switch is located in this equipment room. This is an L3 switch which provides local network routing and packet filtering in addition to L2 switching. The routing configuration is done simply with static routing.
- In many cases, the central servers are connected directly to this switch. Further, the switch is the root of an optical fibre-based tree structure with connections to telecommuications rooms in which distribution switches and/or edge switches are located.

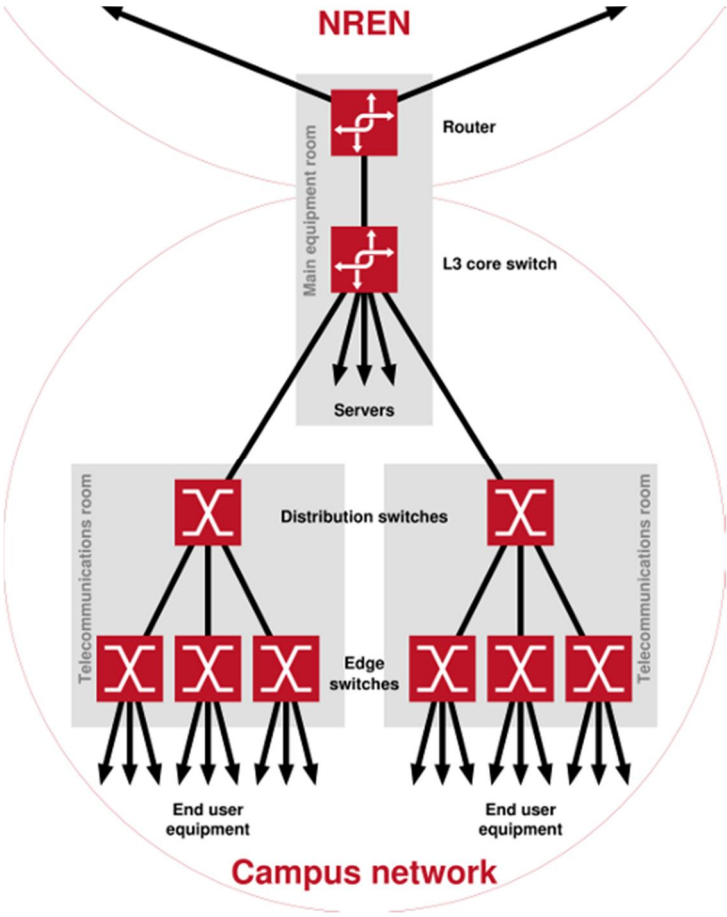


Figure 1: Non-redundant campus network

One of the strengths of such a configuration is its **low complexity**. Operating such a system does not call for specialised knowledge of dynamic routing protocols and so on.

2.1.1 Weaknesses

From a fault-tolerance perspective, this configuration has many weaknesses. Ideally, there should be no single points of failure, but the following are present here:

- Only one core switch. If this should fail, most of the campus network will be put out of action.
- Only one outgoing route from a given telecommunications room. If this connection should fail, the environment associated with that room will be without network connectivity.

2.1.2 Proposed improvement measures

If a non-redundant structure is to be retained, a number of measures may be taken to mitigate the weaknesses:

Focus particularly on electrical supply and cooling

Experience shows that it is defects in the electrical supply or cooling systems which cause the majority of the operational problems. The management modules of core routers and switches are in themselves reliable and rarely fail (though this can happen). Hence one should prioritise the following:

1. Provide a redundant power supply to core switches and potentially also to other switches.
2. Provide a redundant electrical supply to equipment rooms, using UPS and diesel-powered generators if necessary. UFS 107 [1] makes recommendations in this respect.
3. Install a redundant cooling system, cf. UFS 108 [2].
4. Provide a well-functioning monitoring system in which any failure of central components results in an immediate alarm via SMS. See UFS 128 [3] for monitoring requirements. In addition to general monitoring of network equipment, the following should also be included:
    - Monitoring of the electrical supply and cooling situation. In the case of redundant power supply, if one power fails, an alarm must be provided.
    - The UPS units must support SNMP, so that an immediate alarm is provided in the event of failure of the public supply grid, enabling the rapid implementation of extraordinary measures.
    - In the same way, an alarm must be sent in the event of a rapid temperature rise in the equipment room. The use of WeatherGoose [4] or a similar system is recommended.

In addition, though this is fortunately rarely necessary:

- Provide a means of early fire detection; cf. UFS 104 [5].

Keep a supply of spare parts

- Keep a dedicated supply of spare distribution and edge switches.
- Consider also keeping spares for fibre-optic and TP modules for core switches. At the very least, have smaller standby switches available which can be phased in if a core switch module should fail.
- If necessary, consider a special agreement with your supplier to ensure rapid delivery of spare parts (subject to an assessment of the costs involved).

Redundant management module

Even with the above-mentioned measures in place, the system is still vulnerable, especially in the event of a failure of the management module in the core switch. In reality the entire campus network will then be out of action and delivery of a replacement may take some time. The following additional measures may mitigate the situation:

- Acquire a redundant management module which is put into operation as a hot standby.
- Keep an extra fan module of the correct type in storage. If a fan module fails, the entire switch will fail.

If replication of state between the routing processes on the two cards is supported, this provides an additional advantage in that software upgrades can be carried out seamlessly, without downtime. One management module can be in operation while the other is being upgraded, and vice versa.

Redundant core switch in the same room

Additional redundancy can be obtained by installing an additional core switch in the same room. Although this does not provide all the advantages of two separate equipment rooms, it definitely helps. Redundant core networks are dealt with in detail in the next section.

2.2 Two equipment rooms

Upgrading to two equipment rooms provides substantially improved fault tolerance. This allows for:

- Two separate NREN connection points with two separate NREN routers.
- Two separate core switches with possibilities for redundant structure for distribution switches and edge switches.

Important points with regard to separate equipment rooms:

- Low probability of simultaneous electrical power failure in both rooms. NB: Ideally, each room should be supplied from two separate transformers.
- Low probability of fibre-optic failure putting the network out of action. Redundant fibre-optic structure means that at least two separate failures must occur for the system to fail.
- Very low probability of simultaneous cooling problems occurring in both rooms.
- Very low probability of simultaneous fire outbreaks or other catastrophes in both rooms.

2.2.1 Fully redundant model

Figure 2 shows an example of a fully redundant model with two separate equipment rooms, a dedicated server room and redundant fibre paths to all telecommunications rooms. While this is the ideal configuration from a fault tolerance perspective, it is very expensive. Based on an assessment of cost-effectiveness, we do not recommend such a configuration in the HE sector, but this model is included to illustrate how the full potential could be achieved.

The following elements are included:

- Two separate equipment rooms.
- Two separate NREN routers and two separate core switches. Redundant routing using BGP is recommended. For details see UFS 132 [6].
- Redundant L3 design on campus using IS-IS or OSFP.
- Redundant default gateways for all subnets using VRRP, HSRP or similar.
- Redundant L2 structure with low convergence time (RSTP).
- Each distribution and edge switch has connections to two separate switches, which entails comprehensive use of optical fibre.
- Servers are placed in a dedicated room. There is redundant access from each server to two separate server switches. Each server switch has two separate connections to each core switch.

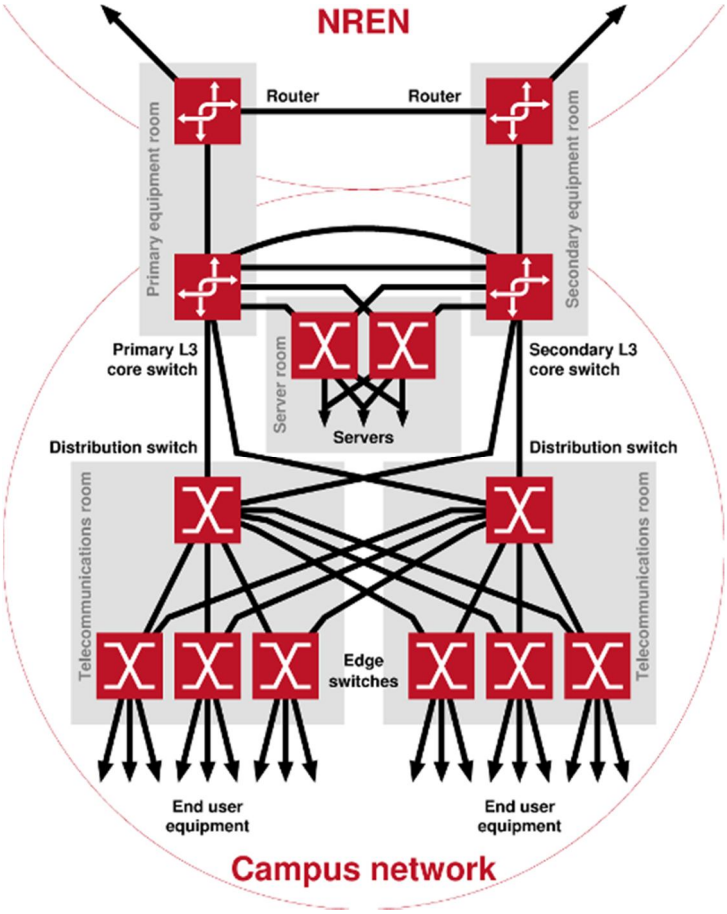


Figure 2: Fully redundant core network

2.2.2 Recommended model – traditional design

In the configuration we recommend, the degree of redundancy is somewhat moderated, cf. Figure 3. It includes the following elements:

- Two separate equipment rooms.

- Two separate NREN routers and two separate L3 core switches. Two redundant BGP sessions are set up, one over the primary connection and the other over the secondary connection. The setup can be either active/passive or load shared across the two connections. The latter will implicitly ensure that both connections work properly at all times, which is good. It will on the other hand trigger more L2 traffic between the two core switches.

- In addition an internal BGP peering between the two L3 core switches must be set up. It is critical that this BGP session always is up; if not traffic forwarding in and out of campus will malfunction. To ensure a robust internal BGP peering, two redundant L3 paths between the primary and secondary L3 switch should be established. Ideally these paths should follow separate physical guideways.

- Further a dynamic internal routing protocol is needed. Recommended candidates are either IS-IS or OSFP. OSPF has best equipment support, whereas IS-IS can manage both IPv4 and IPv6 in the same routing process, which is an advantage.

- Redundant default gateways for all subnets using VRRP, HSRP or similar protocols. The primary default gateway should be the same as the primary BGP router (in cases where BGP is set up active/passive).

- Layer 2 rings should be constructed between distribution switches and core switches for the sake of redundancy. This requires spanning tree (RSTP) in order to break loops. The root of the spanning tree should be set at the primary core switch. Spanning tree blocks during normal operation will then be as shown in Figure 3. More information about the distribution network can be found in Chapter 3.

- This configuration calls for two fibre-optic connections from each telecommunications room:

    - An ideal solution involves separate, redundant fibre-optic cables from each telecommunications room to the two respective equipment rooms.

    - A more modest solution is to lay fibre-optic cable to only one of the equipment rooms and instead provide plenty of fibre connections between the two equipment rooms. Both the principal and the redundant path are then patched to the associated equipment room, while the redundant path is patched onward to the other equipment room. In this sort of configuration the network will be able to survive a power cut or cooling problems in one of the equipment rooms, and probably also a minor fire outbreak. However, the configuration is vulnerable to breaks in fibre-optic cables.

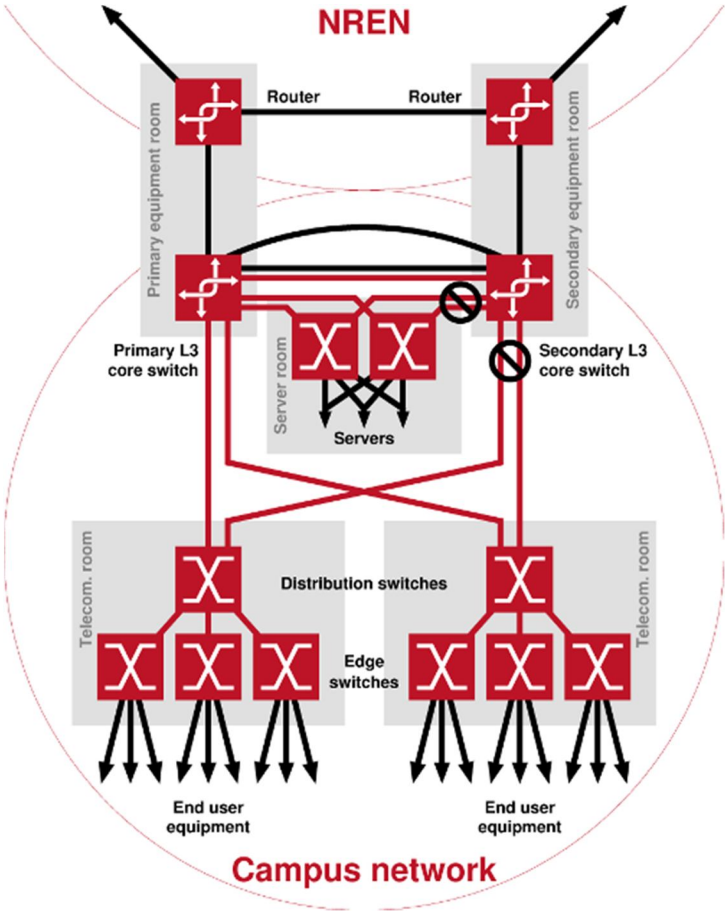


Figure 3: Recommended core network

- Servers placed in a dedicated room. Redundant access from each server to two separate server switches. Each server switch has two separate connections to each core switch. If a blade system is used, the same applies. Server setup is handled in more detail in Chapter 4.

When migrating from a design with one core switch (cf. Section 2.1) to a design with two core switches, it is important to separate all servers to dedicated switches. Otherwise true redundancy is not achieved. The same applies to end users. End user connections should not be made directly to a core switch. Core switches should only have fibre-optic connections to other switches, as well as to the NREN routers.

Neither do we recommend the use of service modules, such as wireless controllers, in the core switches, since this locks the design more firmly to a particular platform in the future. In general it is not wise to combine too many functions in the same device.

2.2.3 Recommended model – virtual design

If the campus network consists of two (and no more than two) core switches as outlined in Section 2.2.2, the same physical topology can be achieved in a different logical manner by employing virtualisation in the core. The principle is that two physical chassis are connected together using a set of dedicated fibre-optic connections which form a proprietary backplane between them, the two units behaving logically as a single one. This has a number of advantages, since it simplifies the design while retaining the desired degree of redundancy.

Figure 4 illustrates a design using virtual core switches.

All suppliers (Cisco, HP, Juniper) which are party to the Norwegian HE sector's joint agreement for network electronics can supply such a configuration. Examples of implementations are Cisco VSS (Catalyst 6500 platform), Cisco vPC (Nexus platform), Juniper virtual chassis (EX4200, EX4500, EX8200), HP IRF (A-Series, formerly H3C). To date the HE sector in Norway has had little experience of such configurations apart from some implementations of Cisco VSS (Virtual Switching System).

A challenge presented by such virtualisation is that it will be proprietary to each supplier. This restricts the type of hardware which can be selected, including the type of management module and line card. For example, Cisco VSS calls for the newest Sup720 management module and line cards in the 67xx series. The virtualisation can also lead to hidden complexity and hence a more complicated environment for fault-finding. The virtualisation technologies are a relatively recent development and are not as well tested as the IS-IS/OSPF, VRRP and spanning tree protocols.

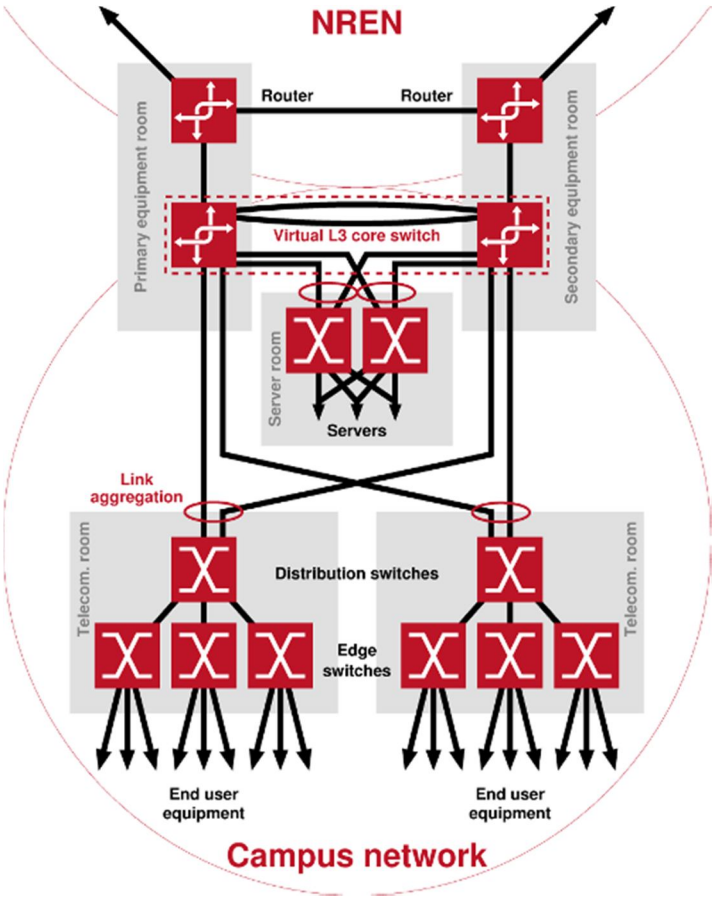


Figure 4: Setup with virtual core switches

A virtual core configuration has, however, many interesting advantages:

- There is no longer any need for dynamic interior routing protocols like IS-IS or OSPF, since there is now only one logical core switch. BGP is needed as before to connect with the NREN.
- There is no longer any need to configure a redundant default gateway with VRRP or a similar protocol. This redundancy exists implicitly in this design.
- Redundant connection to distribution/edge switches is achieved simply by using link aggregation (IEEE 802.3ad). Distribution/edge switches will be physically wired to both core switches, though they logically "believe" this is the same switch, and one can implement link aggregation for load sharing and redundancy without any problems. Hence one is not dependent on spanning tree to achieve redundancy in the distribution network. A spanning tree should still be in operation, but only to detect unintentional loops in the topology and block them if necessary. More information about spanning tree protocol can be found in Chapter 3.

3 The distribution network

The distribution network connects the access network with the core network, in other words the connection from the access switches to the backbone network. No routing takes place in the distribution network. IEEE 802.1q is used in the trunks of the distribution network to transport several VLANs through the same connection.

3.1 Spanning tree

The spanning tree protocol is used in the distribution network to detect any loops and then break them. The original spanning tree standard from 1990, IEEE 802.1D (STP), is conservative and has very slow convergence. Cisco introduced early proprietary mechanisms to improve this (port fast, uplink fast, backbone fast). In 2001 the Rapid Spanning Tree Protocol (RSTP), IEEE 802.1w was introduced, and radically improved the convergence time in connection with topology changes in the distribution network (typically from 30-50 seconds down to a few seconds).

Work is in progress to create standards which depart completely from the spanning tree algorithm and instead adopt a link state algorithm based on IS-IS. IEEE 802.1aq is one such protocol, while another is IETF's RFC 5556, TRILL (Transparent Interconnection of Lots of Links). It is too early to recommend these solutions, as the support provided by the suppliers is not yet adequate.

At present we recommend the use of IEEE 802.1w, Rapid Spanning Tree (RSTP).

However, the situation is somewhat more complex than this. Should one run a joint spanning tree agent for all VLANs, one instance per VLAN, or a hybrid configuration? Alternatively, can one manage without spanning tree in some cases?

To consider this last possibility first: If the distribution network is a straightforward tree structure which does not contain any loops, a spanning tree is in principle superfluous. We say "in principle" because loops may arise because of unintentional interconnections, and in such cases one needs a mechanism for detecting and breaking loops. Nevertheless, implementations exist which solve this using edge switches (e.g. HP's loop protect mechanism) which are less CPU-intensive than spanning tree. In general it will be an advantage if all access ports can be signed out of the spanning tree and loops can be detected in some other way.

CPU processing can actually be a problem when using spanning tree if one has a very large number of VLANs and each VLAN has its own spanning tree instance whose topology must be computed. Note that no standard exists for per-VLAN spanning tree, only proprietary implementations. Cisco's solution is called PVST+[1], while Juniper's is called VSTP. The HP A-Series also uses PVST+. The counterpath is Common Spanning Tree (CST), in which there is a shared spanning tree instance for all VLANs. The original IEEE 802.1q standard was based on the use of CST, which has clear weaknesses from a load sharing perspective.

Let us consider a realistic example in which an edge/distribution switch has two uplinks to the respective core switch as shown in Figure 5, with 1000 VLANs traversing the trunks: using CST, all the VLANs will use the same trunk, while the other will be entirely unused.

[1] The predecessor of PVST+ is PVST. PVST is only supported over ISL trunks, while PVST+ is supported on 802.1q trunks.

If, on the other hand, we configure one spanning tree instance per VLAN, we can ourselves balance the transmission of VLANs in each trunk. The disadvantage is as previously mentioned CPU processing. Looked at rationally, it is undesirable to carry out 1000 topology computations for two different scenarios. This is precisely the background for MSTP, Multiple Spanning Tree (IEEE 802.1s)[2]. MSTP allows us to assemble our VLANs into groups of spanning tree instances, with rapid spanning tree running in each group. The example in Figure 6 shows a configuration with two groups.

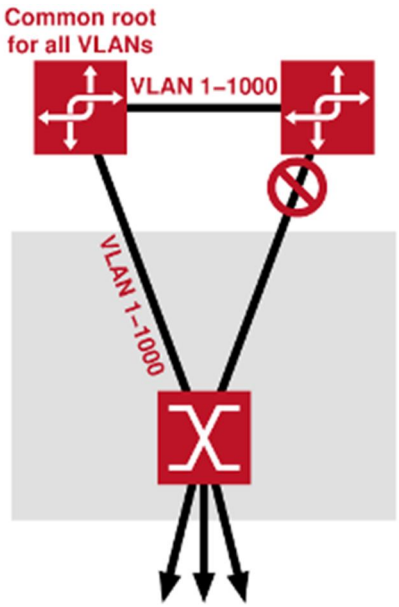


Figure 5: CST provides no load sharing

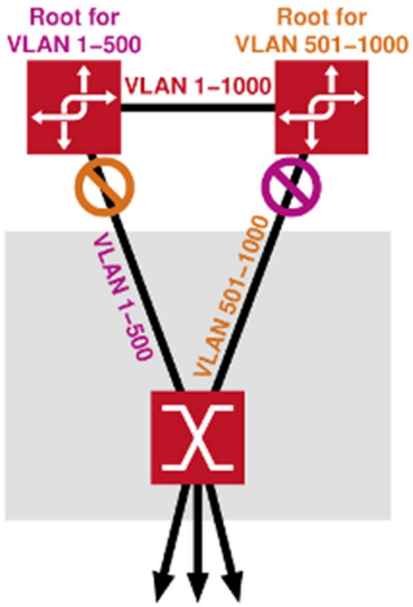


Figure 6: MSTP collects VLANs into groups

The weakness of MSTP is that it requires some extra planning and configuration, and when introducing new VLANs it will involve more work. There are currently no major MSTP implementations in the HE sector in Norway even though MSTP is very suitable for hybrid environments.

If one has an implementation consisting of core and distribution switches from Cisco or Juniper with HP at the edge, one can run PVST+ or VSTP in the core and CST at the edge.

3.1.1 Link Fault Management / Unidirectional Link Detection

IEEE 802.3ah defines Link Fault Management (LFM) as a protocol for detecting links which are not bidirectional. Juniper has implemented LFM, while Cisco and HP have an equivalent proprietary implementation, Unidirectional Link Detection (UDLD). It is recommended to configure LFM or UDLD on the links in the distribution network for the purpose of shutting down ports with unidirectional links. The occurrence of unidirectional links can have unfortunate consequences, since the various switches will get different pictures of the topology and this can lead to unintentional loops and an undesirable traffic pattern.

3.2 Link aggregation

Link aggregation is defined in IEEE standard 802.3ad[3] and is a method used to bundle several links between two switches (or between a switch and a server) to create a logical connection, thereby increasing the capacity

[2] MSTP has also been incorporated in IEEE 802.1q since 2005.

[3] IEEE 802.3ad was introduced in 2000. The IEEE 802.1 workgroup has subsequently taken over the job of link aggregation and in 2008, 802.1ax was introduced, which defines medium-independent link aggregation (i.e. no longer restricted to Ethernet connections). Note also that the predecessor to IEEE 802.3ad was Cisco's proprietary EtherChannel.

of the connection, cf. Figure 7. One may, for example, bundle four 1-Gbps connections to achieve a capacity of 4 Gbps. Load sharing is done in both directions and the algorithm for traffic distribution is based on the source and destination IP addresses (and possibly TCP/UDP ports). The Link Aggregation Control Protocol (LACP) is used for dynamic negotiation of the size of the bundle. Among other things, LACP makes sure broken links are automatically removed from the bundle in fault situations, which is important.

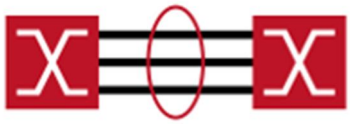


Figure 7: Link aggregation

Link aggregation has no effect on topology. It is merely a bundling of several links to create a larger connection between two switches. However, one should be aware that the STP port cost may need to be adjusted manually to achieve correct balancing (Cisco and Juniper do this automatically, but HP does not).

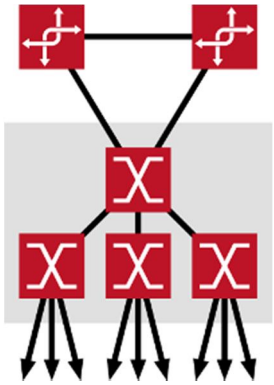
More information about proprietary multichassis expansions to link aggregation can be found in Section 4.2.2.

3.3 Redundant distribution

In this section we will discuss four different ways to implement a redundant distribution network. Each design has its advantages and disadvantages.

3.3.1 Redundant distribution switch in each room

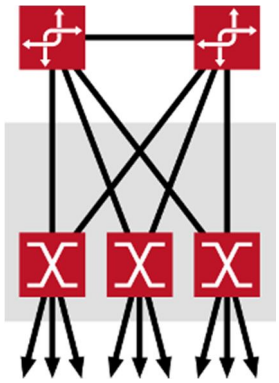
Here there is a central distribution switch in each telecommunications room which has a redundant fibre-optic connection to two different core switches. The edge switches have only one uplink, which is via a twisted pair cable (1 or 10 Gbps) located in the room.



- Advantages:
  - Edge switches only need TP ports, allowing the use of inexpensive switches.
  - The implementation requires only two fibre-optic cables leaving the room.
- Disadvantages:
  - There is no redundancy at the edge switch level.
  - This may often lead to many superfluous ports in the distribution switches.

3.3.2 Redundant edge switches connected directly to the core

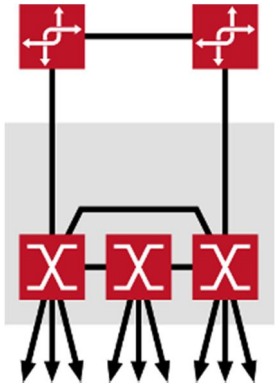
Here there is no distribution switch, but each individual edge switch is connected redundantly directly to the core switches.



- Advantages:
  - Does not need a dedicated distribution switch.
  - Provides redundancy at edge switch level.
- Disadvantages:
  - Requires a lot of costly fibre-optic cable from each telecommunication room.
  - Requires two fibre-optic ports in each edge switch and makes a greater demand on fibre-optic ports in the core switches.

3.3.3 Redundant edge switch stack

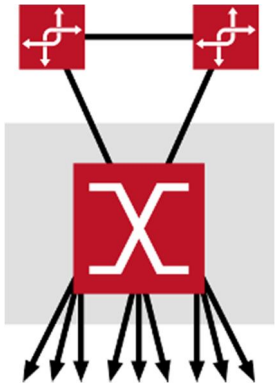
In this case, the edge switches are stacked and the stack is redundantly connected to the core network.



- Advantages:
  - Does not require a dedicated distribution switch.
  - Requires only two fibre-optic cables leaving the room.
- Disadvantages:
  - Results in sub-optimal traffic pattern from edge switch to edge switch, i.e. traffic goes through a chain of several switches.

3.3.4 Redundant, modular edge switch

Here a larger, chassis-based edge switch is used which serves the entire telecommunications room.



- Advantages:
  - Tidy implementation and simplified administration (a single switch for the entire room).
  - Does not need a dedicated distribution switch.
  - Requires only two fibre-optic cables leaving the room.
- Disadvantages:
  - Failure of the entire switch (i.e. the management module) leads to widespread network unavailability. One should definitely have spare parts available, combined with effective monitoring.

4 The access network

The access network consists of the connections between terminal equipment and edge switches. This includes server and client connections. We provide the broadest description for server connections.

Redundancy in relation to services can be achieved without making every individual server redundant. This solution may be good enough. Nevertheless one should seriously consider a server network setup which provides a redundant network connection to each server.

4.1 Edge ports

It is not desirable that frequent link transitions from end-user machines trigger new computations of the RSTP topology. The way to solve this is by configuring all such ports as edge ports[4]. An edge port will transition directly to the forwarding state after link and will not wait for a spanning tree computation, which in itself is a major advantage (one obtains immediate access to the network). As additional protection against unintentional loops created by terminal equipment, an edge port will either be disabled or lose its edge port status and become part of the spanning tree topology if BPDU packets are detected behind the port. This may happen if a switch is connected or an unintentional loop is created.

HP has an additional function called “loop protect”, which is recommended. This will be helpful in situations where unintentional loops are created and *no* BPDU packets are detected. This may occur, for example, in cases where cheap switches which erroneously reject BPDU packets are connected to the network. HP’s “loop protect” sends out certain other control packets which will pass through such equipment and the switch can then detect whether the same packets enter by another port.

4.2 Redundant top-of-the-rack switches

An effective architecture is to set up two independent top-of-the-rack switches located in the server room racks. Each server is always wired to two such switches. The same applies to blade systems: two uplinks are connected to two different switches. The capacity may be 1 or 10 Gbps, as required.

This is the physical principle: it remains to be determined whether load sharing and/or redundancy shall be implemented. There are several ways of achieving this goal. A fundamental challenge is that the link aggregation standard (IEEE 802.3ad, cf. Section 3.2) for bundling several Ethernet links requires the same device (switch or server) at each end of the bundle, and this is *not* what we want. Hence, at present one is forced to use proprietary solutions, and most suppliers offer possible configurations. We can divide these into three main categories: bonding/teaming, multichassis link aggregation and stacking.

[4] Cisco designates this “port fast”.

4.2.1 Bonding/Teaming

The simplest configuration from a network perspective is to use “Ethernet Bonding” on Linux servers or “Ethernet Teaming” on Windows servers. In both cases, the two network cards are connected to switch ports on two separate switches (in the same VLAN, of course). No special configuration of the server switch ports is necessary (with one exception, as mentioned below). Traffic *leaving* the servers can be run with load sharing through the two network cards (active/active). One may also run in so-called “active/passive” mode, but this is not recommended because in that case it is not possible to verify the redundancy in normal operation.

Each network card has its own unique MAC address, but communicates using the same sender IP address. Traffic to the server will in fact *not* be load-shared, but depending on what is stored in the ARP cache of the router (or other senders on the subnet) it will be forwarded to this MAC address.

A problem which may arise with this configuration is that one of the top-of-the-rack switches may lose its uplink(s), i.e. its contact with the outside world. The connected servers will not understand this and will continue to send load-shared traffic to this “black hole”. To avoid such a situation, Cisco and Juniper have a feature called “uplink tracking”. “Uplink tracking” will disable the link to the servers if the uplinks fail. It is expected that other suppliers will be able to provide this function in the near future.

4.2.2 Multichassis link aggregation

Several suppliers offer a proprietary expansion of link aggregation (IEEE 802.3ad, LACP) to enable distributed connection between a server and the network. A workaround will be to use a virtual core configuration as described in Section 2.2.3, but most often one does not have this type of equipment in the server room.

Other implementations which do not involve constructing a virtual chassis but keep the two switches as separate administrative units are:

- Cisco’s multichassis LACP (mLACP). This is unfortunately only supported on the Catalyst 6500 platform.
- Cisco also offers an implementation for the Nexus platform which uses a virtual port channel (vPC).
- Juniper’s MC-LAG, which is supported on the EX platform.
- HP offers for its E-Series a solution called “distributed trunking” (dt-lacp). Distributed trunking has the advantage that it will run on many hardware platforms, including the smaller ones[5].
- Other suppliers may have equivalent solutions.

The main principle is the same in all cases: a dedicated proprietary connection is set up between the two switches which are to form the multichassis link aggregation configuration. In the case of Cisco’s 6500, this cross-connection is called an interchassis communication channel (ICC), while Juniper calls it Internet Chassis Control (ICCP). HP calls the connection InterSwitch-Connect (ISC).

Note that Cisco’s and Juniper’s configurations provide an active-standby setup in communication with the server; in other words only one of the server connections is used at any time. HP’s distributed trunking uses load sharing and spanning tree. In many situations, incoming and outgoing traffic in the servers will in fact be via the ISC backbone.

4.2.3 Stacking

It is also possible to achieve a redundant top-of-the-rack configuration in a server room by using switch stacking. Most suppliers offer a possibility of stacking several individual switches. Usually these can then be administrated as a single unit, which in itself can be an advantage. The traffic pattern may be a disadvantage. A large, chassis-based switch will normally have a backplane with higher capacity than that provided by a

[5] dt-lapc is supported on the 3500, 5400, 6200 and 8200 platforms in HP’s E-Series (formerly Procurve).

stacking setup. However, from a fault-tolerance perspective, the various switches in the switch stack are completely separate and hence provide a greater degree of redundancy with regard to the server.

4.3 Virtual switches

When virtualisation is used, the virtual servers are connected to a virtual switch. This simplifies the configuration from the server. If one also ensures that the virtual switch has redundant uplinks, one will maintain a high level of fault tolerance. If the virtual server environment is in a blade system, one should ensure that there are two redundant uplinks leaving the blade system.

4.4 Redundant services

In addition to server redundancy, all critical services should themselves be redundant. We will not deal with this in detail, but will mention two services closely related to the basic network operation that are of fundamental importance:

- DNS: Ensure that you have at least two redundant resolvers. If the computers in the network lose access to DNS, the network will in effect become unusable.
- DHCP: Naturally, DHCP service is recommended for all client networks. It is both rational and practical to centralise the DHCP service, but it should also be implemented redundantly.

4.5 Client connection

Clients should be connected to the network using edge switches. Naturally, there is no redundancy in this connection, as it cannot be justified. To reduce the amount of wiring one may choose to configure one twisted pair outlet per office and use a local office switch. The disadvantage of this is that one has an additional, typically cheap, component in the network which possibly cannot be monitored either. It is recommended only to use switches in telecommunications rooms which have richer functionality, are more reliable and can be monitored effectively. Reference is made to UFS 105 [5], which deals with the recommended configuration of switches in campus networks. Functionality that should be considered implemented with regard to end-users is, among other things:

- IEEE 802.1X, which requires login to get network access, cf. UFS 133 [8] for details.
- DHCP snooping, which prevents false DHCP servers from corrupting the enviroment of the local network.

See also the treatment of edge ports in Section 4.1.

References

[1] UFS 107: Power Supply Requirements for ICT Rooms
http://www.terena.org/activities/campus-bp/pdf/gn3-na3-t4-ufs107.pdf

[2] UFS 108: Ventilation and Cooling Requirements for ICT Rooms
http://www.terena.org/activities/campus-bp/pdf/gn3-na3-t4-ufs108.pdf

[3] UFS 128: Framework Conditions and Requirements for Network Monitoring in Campus Networks
http://www.terena.org/activities/campus-bp/pdf/gn3-na3-t4-ufs128.pdf

[4] WeatherGoose
http://www.itwatchdogs.com/product-detail-weathergoose_ii-1.html

[5] UFS 104: Fire Prevention Requirements for ICT Rooms
http://www.terena.org/activities/campus-bp/pdf/gn3-na3-t4-ufs104.pdf

[6] UFS 132: Fault-tolerant Internet connectivity with BGP
http://www.terena.org/activities/campus-bp/pdf/gn3-na3-t4-ufs132.pdf

[7] UFS 105: Recommended Configuration of Switches in Campus Networks
http://www.terena.org/activities/campus-bp/pdf/gn3-na3-t4-ufs105.pdf

[8] UFS133: Recommended Configuration of 802.1X in Fixed Networks
http://www.terena.org/activities/campus-bp/pdf/gn3-na3-t4-ufs133.pdf

Definitions

Client port: A port on a switch which is connected to client machines in the network. This also includes servers, printers and other terminal equipment. Such ports have a number of properties which differ from those of network ports, in other words ports which are connected to other network components (routers, switches or base stations).

Core switch: A switch which is located in the core of the network and to which users are generally not directly connected; primarily a high-capacity connection to other switches and servers.

Distribution switch: A switch which handles aggregated traffic from a number of edge switches and connects it to core switches.

Edge switch: A switch located in the periphery of the network, closest to the users.

L2: Layer 2 of the OSI stack. Switches on Layer 2 cannot interpret IP addresses, but operate with MAC addresses.

L2+: Some switches have the ability to interpret the various characteristics of IP headers and higher layers. DHCP snooping is an example of such a function. This functionality is designated L2+.

L3: Layer 3 is the network layer, which recognises IP addresses at which the routers operate. Some switches can also perform routing. These are referred to as L3 switches.

NREN: National Research and Education Network