Professional paper / Stručni rad
Manuscript received: 2015-09-21
Revised: 2016-05-20
Accepted: 2016-05-31
Pages: 13 - 23

Achieving Optimal Redundancy in a Small Business Network

Silvio Papić
University College Algebra, Zagreb, Croatia
silvio.papic@racunarstvo.hr

Abstract: Redundancy is a highly desirable element of the network, but sometimes it is not so easy to find the optimal level of redundancy that will ensure satisfactory performance of the entire network and at the same time being affordable and easy to implement. Redundancy in the network can be performed in several ways, which can be compatible and co-exist in the network, but can also be mutually exclusive. Although at first it seems that the redundancy depends only on the size of the network, this is not entirely true. One of the main factors to take into account is the complexity of the network and the importance of services and applications that serve the company's business and its customers. Network redundancy can be achieved in various ways. STP for example is the most basic mechanism but has several major drawbacks like underutilization of some links. Also link aggregation mechanisms could be used, which usually means investing in optical transcievers which implies a certain cost, and if one is using switch stacking it will further increase costs. For gateway redundancy one of FHRP protocols can be used which is certainly desirable. This paper will describe the redundancy in Small and Medium networks with few hundred clients and one way how to ensure redundancy by using various mechanisms such as the STP protocol, link aggregation, implementation of FHRP for gateway redundancy and using stackable switches with short analysis of recovery times for these mechanisms.

Keywords: STP, link aggregation, stacking, FHRP

INTRODUCTION

Generally speaking redundancy in the network is the ability to retain all network functionality despite the failure of specific, but not critical number of individual elements of the network. The problem that many small businesses face is how to design a network to be redundant while being within the available budget. Small and Medium Enterprises that have a large number of end users, but because of the nature of their business do not have a large data center or complex technologies, nor do they have IT staff that can do complex tasks related to the design, planning, maintenance, and troubleshooting the network. Therefore they hire external companies to do this kind of work for them or initially invest a lot of money in a solution that is too complex and that may never be fully utilized. It is interesting that this situation is not uncommon in practice. A great number of small users do not have the process of planning, monitoring and maintenance of network infrastructure, rather it is mostly ad-hoc approach which causes various problems in the network and in the company's business. Because of the ad-hoc network infrastructure such companies constantly lag behind the needs of the business and what their network or IT system as a whole should support. The importance of choosing the design initially and appropriate technologies, as well as people with the necessary skills is extremely important. When the network infrastructure is planned and structured many problems can be avoided in the beginning, the occurrence of performance problems can be minimized, and their resolution can be significantly accelerated.

Redundancy in a network is one of the basic elements that guarantee uninterrupted operation of the network, but how to know how much redundancy is enough, and when it is too much? Some say it's never enough redundancy and it would be best to do all possible systems redundant. This may at first seem to make sense, but hidden costs should be taken in consideration. The obvious thing is that the initial financial costs using this approach are very large and probably not justified with regard to the company's business objectives. However, the financial costs are not the only aspect that should be considered. The big challenge in designing the optimal redundant network is to ask the right questions with regard to the requirements of the company's business in order to avoid wasting of limited resources.

One approach is very simple, is to use a generic solution from a network equipment vendor, which uses best practice to suit a particular type of company. The problem with this approach is that it is a generic solution that does not take into account all the specifics of the IT system of specific companies. Solutions that suit most have never been optimized for specific environments that vary from company to company. It would be better to use individual approach that takes into account all aspects of the company and based on this information design a custom solution. While this will be a custom solution tailored to a particular user it will still basically be prepared according to the best practice template so that it will retain all the benefits of a generic solution. The difference will be only in fine tuning of generic model to needs of a particular company so that it could make the best of its network with regard to available resources. In this context, the quote below will be current for a long time.

»In the Internet era, reliability is becoming something you have to build, not something you buy. That is hard work, and it requires intelligence, skills and budget. Reliability is not part of the basic package.»[2]

Joel Snyder – Network World 1/10/2000
»Reliability: Something you build, not buy«

In addition to costs, which are in most cases the first factor in the decision, there is a whole series of questions that companies should ask when designing redundant networks, and those questions vary from company to company. No matter what kind of network is concerned, the approach is basically the same and one should use the following steps [6] when analyzing the situation and asking the key questions.

Risk Identification: This is the critical first step. Its objective is the identification of risks, including those within and external to the network or IT system

Risk Impact Assessment: In this step, an assessment is made of the impact each risk event could have on the network. Typically, this includes how the event could impact costs, performance and other relevant factors of the business. Assessment should be made based on the probability each risk event will occur. Sometimes best source for information is the experience of the IT staff.

Risk Prioritization: In this step one should combine the first two steps in order to derive a most critical to least critical rank-order of identified risks for the purpose of allocating critical resources.

Risk Mitigation Planning: This step involves the development of mitigation plans designed to manage, eliminate or reduce risks to an acceptable level.

For example, if it is concluded that the critical element of the network is L2 connection between customers two locations and it is important to provide an alternative link to connect the two locations of the company, perhaps the solution is to change the current ISP to one that has fiber infrastructure that goes around the city.

After careful reflection, defining risk and answering questions that arise from this process, as well as using the four steps that are mentioned above only then can one start the realization of the solutions.

Below are listed some of the questions that one should ask when choosing network design, of course, these questions will vary from case to case, but in general most of them should be asked:

- What are the critical elements of the network?
- What is the cost of downtime if one does not implement redundancy in the network?
- How acceptable is a partial breakdown of the network?
- What is the cost relative to the benefits to the business?
- Will the additional redundancy introduce too much complexity in the system and increase the duration of downtime?
- Are there external influences on business such as failure or maintenance of the ISP network?

- What is the biggest problem, from experience: Link failures, failure of equipment or the human factor?
- Is it important to ensure scalability of the infrastructure?
- Is it critical that employees have uninterrupted access to the network?
- What is the amount of data passing through the network and what kind of traffic occurs in the network?
- Does one need a network that supports virtual desktop infrastructure?
- What applications are used in the company, and to what extent these applications are vulnerable to failures of links or devices?
- How to organize the process of monitoring and documenting the network?
- Is there a redundant power supply for critical elements?
- Where the network equipment will be located, how many rooms, and who can access the equipment?
- What processes should be documented to facilitate network management?
- Does the company's business depends on the particular type of hardware?
- In the case of critical devices if there is readily available replacement equipment?
- What are the requirements for power supply and cooling equipment?

WHAT SHOULD BE CONSIDERED

As a basis for functional and purposeful network, design is of paramount importance and directly contributes to the successful operation or failure of the network.

»No amount of magic knobs will save a sloppily designed network«[1]

Paul Ferguson—Consulting Engineer,
Cisco Systems

For a well-designed network following criteria must be met:

- Designing the network with resiliency in mind
- Using technology to identify and eliminate single points of failure
- Having processes in place to reduce the risk of human error

In addition to these three basic criteria that must be considered when building an efficient network and which must complement each other, one needs to take into account elements that will ensure the successful implementation of the design.

- Physical infrastructure
- Topological/protocol hierarchy
- Scaling and Redundancy
- Addressing aggregation (IGP and BGP)
- Policy implementation (core/edge)
- Management/maintenance/operations
- Cost

There are different ways in which network can be designed. The network can be designed on the basis of three-layer hierarchical model or collapsed core model or any

other model, which is now considered best practice. However it should be remembered that no matter which model is used, one needs to take into account the specificities of the organization for which the network is designed for.

DESIGN

In the case of Small and Medium Enterprise networks one can choose a collapsed core design, slightly altered using stackable switches instead of two separate core switches. In this case, stackable switches can provide maximum redundancy in the network. In the case of smaller networks generally it is not justified to use the full three-layer hierarchical model with the access, distribution and core layers. The reasons for this are the additional costs for the distribution layer equipment that generally is not required, additional network complexity, higher maintenance costs and so on. Due to the fact that there is no distribution layers there is fewer points of potential failures in the network. Also, if the funds intended for the distribution layer are re-directed to other equipment, organisation can get better and more reliable equipment which in turn contributes to the resilience of the network.

In order to make core of the network redundant and fast, perhaps the best solution is the use of stackable switches which has the benefit of eliminating STP protocol. If stackable switches are used as L3 devices FHRP can also be eliminated with maintaining adequate level of reliability. Although it is possible and even simpler and cheaper to use L3 stackable switches as routers this is not a good solution. The reason is that besides routing the traffic one need to have various security elements implemented in the network, such as filtering traffic based on UDP / TCP ports or applications, anti-virus protection and so on. For the purposes of security one should certainly use Firewalls, and today more and more so-called Next-Generation Firewalls. In this case it would be better to use two firewalls and configure them to work together using one of the FHRP protocol or if the budget permits configure them to work in HA (High Availability) mode, and thus ensure minimal downtime in the event of failure of one of the firewalls or links to the Internet.

Ultimately the topology for this network will resemble a star topology, but without the single points of failure in the core network. Also to keep the benefit of collapsed core design every VLAN should be restricted to one access layer switch stack. For VLAN communication all the VLANs can be terminated at the core stack but this can be complex because routing has to be used for communication with the rest of the network and the Internet. Another more simple solution is to terminate all the VLANs at the firewall and have all the benefits of traffic filtering to have more secure network. If firewalls are used in HA mode than it is even simpler and more reliable solution, but as it is mentioned before it is more expensive because this kind of feature is generally licensed or comes with higher models of firewalls.

TECHNOLOGIES USED FOR REDUNDANCY

Technologies that can be used to provide redundancy in the network are different and depending on the combination of these technologies different convergence times after a failure of a link or device will be achieved. When choosing the technology one needs to be careful, because after implementation, it can be very difficult and dangerous for the company's business to make changes in the production network. Redundancy in the network can be implemented using a combination of the four technologies that will be briefly described below in this document. These are STP (Spanning Tree Protocol), Link Aggregation, FHRP (First Hop Redundancy Protocols) and switch stacking.

STP

Today it is more or less pointless to talk about the original IEEE 802.1D STP version or CST (Common Spanning Tree) when talking about network redundancy. The reason is very slow convergence after failure of a link that takes 30-50 seconds, depending on the size of network and it can even cause problems in the convergence. In the case of smaller networks CST would also do its part, but it would be quite inefficient. For larger networks with more complex patterns of network traffic and applications sensitive to disruption of communication CST should be avoided. In addition to the slow convergence problem is also the fact that the CST does not recognize VLANs, in fact CST treat the entire network as one VLAN and thus a large number of links in the network is blocked which is certainly not something one would want in a network.

The problem of slow convergence can be solved by implementing the IEEE's RSTP (Rapid STP), but less than optimal forwarding of traffic still remains as RSTP still builds one STP tree for the entire network. A better version of STP is Cisco PVST + (Per VLAN Spanning Tree Plus), which has the same slow convergence time as STP, but enables better use of links in the network. This is because of the ability to share traffic according to VLANs by making a link that is blocked for one VLAN, to be in forwarding state for another VLAN. This is also useful in the event of link failure, because in this case negative impact on the network is much smaller than in the case when STP is used.

Since the convergence is per VLAN link failure affects only VLANs whose forwarding link has failed. In addition one has the ability to configure protection against network attacks using STP BPDU Filter, Root Guard, BPDU guard, loop guard functionality as well as *portfast* option allowing us to have stable STP topology, which is essential for a functional network. These mechanisms are beyond the scope of this paper, but should be considered when implementing PVST+. The problem that can occur is when there is a large number of VLANs configured on a switch. In this situation switch CPU could be overloaded which can cause other problems in the network even inability to converge.

If the goal is to have short convergence times, and at the same time have more STP instances Cisco's PVRST + (Per VLAN Rapid Spanning Tree Plus) can be used which is sort of a combination of IEEEs RSTP and Cisco's PVST + protocol. PVRST + allows us to have a short convergence and the possibility of an independent convergence of the network

for each VLAN, which allows us more flexibility. With all the additional functionality that are listed under PVST + this would be a good choice for a stable STP topology. The problem that can occur is also linked to the number of instances of STP. Because PVRST + supports a separate instance of STP for each VLAN, and also has shorter convergence time, the CPU is significantly burdened and switch can be easily led to overloading.
Recommendations for implementation of PVRST + is never to configure more VLANs than required in a network, otherwise it could lead to 100% CPU load and ultimately bring down the network. However, if one want redundancy based on STP and one have lot of VLANs (hundreds), the solution is the use of IEEEs MST protocol (Multiple Spanning Tree) that allows single instance of spanning tree protocol for a group of VLANs which ultimately results in fewer instances of spanning tree protocol in a network and still have the ability to use all links. Though it may seem that STP is something that should be used in a network, sometimes it is not the case. STP should not be excluded completely, but redundancy in a network should not be based on STP.
Depending on the complexity of the network STP can be a valid solution, but for example, in data center networks STP should not be used at all. Some of the reasons are inefficient use of links and slow convergence for such environments with very complex topologies that are difficult to troubleshoot. STP could be used as failback mechanism in the event of a loop, but everything should be done in order to prevent the occurrence of loops. Also one should implement redundancy in the network using other mechanisms like switch stacking and link aggregation.

LINK AGGREGATION

Large amounts of traffic that is circulating through the network is not uncommon even in smaller networks, and one way to increase throughput of the network is to use link aggregation. Although it seems that the main benefit of link aggregation is higher bandwidth, the primary thing that is accomplished is in fact link redundancy between switches in a network. Link aggregation can also be used for computers or servers (NIC teaming), which certainly adds to the total resilience of the network.
There are different ways of performing link aggregation, some of which are standard, such as LACP (Link Aggregation Control Protocol), and some are owned by a variety of manufacturers such as Cisco, Juniper, Avaya, Huawei etc. No matter which protocol is used, one needs to know the capabilities of the solution and whether aggregated links will be able to take the burden of network traffic in the event of failure of critical elements such as links or devices. If aggregated link has insufficient bandwidth with respect to the traffic it does not fulfill the purpose for which it was designed, because in case of failure of the main link the impact will be similar to a situation where there is no link redundancy. This effect is amplified if there is no QoS (Quality of Service) mechanisms in the network which means no prioritization of traffic, which is especially problematic for real-time applications and essential communications.

GATEWAY REDUNDANCY

The gateway is one of the most important elements of any network and therefore it is necessary to ensure maximum availability of such devices. A way in which availability of the Gateway is ensured is to make it redundant, which means that there are at least two devices which simultaneously perform the function of the gateway in a network. Protocols that can be used for this purpose are HSRP (Host Standby Router Protocol) from Cisco and VRRP (Virtual Router redundancy protocol) which is standard and is supported by other manufacturers of networking equipment.

With these protocols one can configure at least two devices to work as a single gateway, without the need to change anything on the user computers. The idea is that all computers, according to their VLAN, are configured with the same IP address for gateway. This IP address is configured on two or more routers at the edge of the network and these routers are responsible for forwarding traffic coming to that IP address. At any time, one of the routers is primary gateway, responsible for forwarding traffic, and the other serves as a backup. If a router that is the primary gateway fails, the role of the primary gateway is assumed by another router that is configured to be his backup.

These protocols do not protect only against failures of the entire device, but they can be configured to track certain links or availability of services so that in case of their unavailability backup device can take the role of the primary gateway and communication can continue. Potential problem with these two protocols is that basically only one device is active at a given time, and the other serves as a backup. Of course a manual traffic load balance can be configured so that one router is the primary gateway for one part of VLANs, and the other router is the primary gateway for another part of VLANs, which is similar to manual load balancing in case of STP protocol. But if the goal is to have real loadbalancing there is another protocol called GLBP (Gateway Loadbalancing Protocol) which can solve this problem. GLBP is cisco propriatery protocol which allows us to simultaneously use all the routers that act as gateway so that the traffic between them is loadbalanced. This is especially useful solution in large networks that have multiple exits to the Internet. Otherwise, in most cases it is sufficient to use HSRP or VRRP.

STACKABLE AND MODULAR DEVICES

Using stackable switches is a very interesting and viable option in achieving redundant network, even though equipment that supports this type of features can be significantly more expensive than conventional switches. Stacking is the idea that two or more devices can be configured to operate as a single device by connecting backplanes of these devices using special stack cables or in some cases, using fiber optic connections. The advantage of using fiber optic connection is that stacking can be achieved between devices over distances of many kilometers, for example fiber ring around the city which is composed of stackable devices all working as one. In addition to stacking devices for the purpose of network redundancy it is possible to stack their power supply, which further increases network resilience.

In networks that are not large and can implement collapsed core network design it can be a very profitable long-term investment which has the advantage of eliminating STP which is relatively slow and can cause problems for network convergence after a failure. Networks using stacking technology combined with link aggregation can achieve more stable and predictable failover in case of failure of any device or any link. Moreover these networks are easy to expand and maintain and offer greater performance. Although in certain specific situations stack could ultimately cause some negative effects, mainly it is very flexible, resilient, easy to deploy and scalable solution, especially for a relatively small networks.

For stacking devices, technologies that are used are mainly owned by individual vendors, such as StackWise and VSS (Virtual Switching System) from Cisco or IRF (Intelligent Resilient Framework) from HP, which makes sense because these technologies are optimized for specific operating system and hardware which ensures maximum stability and reliability as well as performance. Ultimately when selecting stackable solution, devices price will be the key factor because requirements for performance in smaller networks are generally met using equipment of any manufacturer.

An alternative to using stackable devices is the use of modular devices. Although such a solution is also flexible and scalable, it can be more expensive, since the cost of buying chassis alone is significant. Also in the event that network is not expanding as planned organisation can get into situation where it has the equipment that is paid for, and it is not actually used.

PRACTICAL APPLICATION

So far it is shown what are the important elements when deciding to implement redundancy in a network, as well as what technologies are available that will enable us to accomplish this task. In the next section it will be presented how should optimally redundant network for the proposed environment look like and what are the benefits of this approach. Topology is shown on the picture bellow.

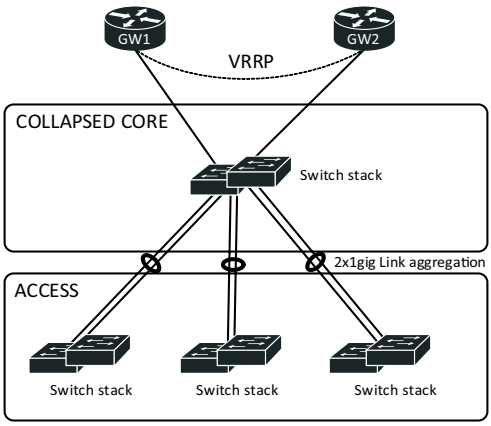

Figure 1: *Design proposal.*

Proposed network is based on collapsed core design with stackable switches at the core and at the access layer of the network. Stackable core is only layer two, and all the VLANs are terminated at the firewalls so that one can make most out this devices and their technologies. Redundancy between firewalls for every VLAN is achieved using VRRP with interface tracking options and timers configured to values less than one second.

As it is said earlier it is possible to use L3 stackable switches as gateways, but it is not the best solution. It would be much better to use dedicated devices for routing like routers or even better firewalls. Both of these firewalls need to be connected to different switch in the core stack and over that link firewalls can be configured to use VRRP for gateway redundancy or even some custom high availability technology. Access and core switches are connected using two aggregated fiber links. The number of link that is aggregated depends on the amount of traffic that is generated by computers or servers that are connected to a specific switch. On the access layer common ratio for downlink vs uplink is 20:1 [4], but this is something that will depend on the type of the environment. For the vast majority of cases that I had the opportunity to witness in practice it is possible to go with 40:1 or more.

Most of the Small and Medium Enterprises in Croatia are still not using virtualization and shared storage systems so most of the traffic is still only between the inside network and the Internet or other locations, and bandwidth for Internet link is in almost all cases lower that 100Mbps so there is really no need for ratio of 20:1. Proposed solution uses 48 port Ethernet POE (Power over Ethernet) stackable switches which can scale as needed. POE feature is important for devices such as IP phones or cameras. Redundancy at the end device level is practically unjustified and very expensive to achieve except for important devices such as servers where NIC teaming can be used in order to maximize server availability. In the event of access switch failure one option is to replace the failed switch.

Another temporary solution is to reconnect end devices to one of the working switches manually. Because of this, and because of the scalability it would be good to have some spare ports on every switch. Every switch in the access layer stack is connected to a different switch in the core stack so that if any of the core switches fail communication with the rest of the network and the Internet is maintained. Connecting a network like this will eliminate STP entirely which is good because now loop free topology with very short failover time can be achieved. Initially costs of building this kind of solution are somewhat higher than using nonstackable switches without link aggregation, but very quickly it will show the benefits of stability, scalability and resilience to failures with much less administrative work.

CONVERGENCE TIMES

For this kind of deployment convergence times are very short, but they will vary depending on the number of devices, device model and manufacturer, configuration, load of the switches etc. Because stacking uses a mechanism that keeps both devices in sync

with each other in case of a failure of one device another device will take over seamlessly (as low as 50 milliseconds) [5] with no disruption to attached hosts. If VRRP for gateway redundancy is being used timers can also be very short so that failover can be very fast, even good enough for voice traffic. If one of the uplinks on an access switch fail failover time is 250 milliseconds or less [3] which is good enough for most of the traffic. And because this is a failure of only one of the links at least half of the traffic on the affected switch and the rest of the network is undisrupted. About the same failover times could be achieved if only one core switch is used, but then there would be no redundancy. If two core switches are being used but not in a stack configuration then the complexity of the design is increased with VRRP on the switches or even STP which would result in underutilized equipment and generally slower failover time.

CONCLUSION

Everything that is mentioned above is applicable in various small, mainly client networks without complex environments like data centers. This approach will enable stable, easy to manage, resilient and scalable network. There are many different ways in which redundancy can be achieved and there is even more factors that affects the decision process, but it is very important to correctly define risks involved and to ask the right questions which will enable more accurate insight in the critical elements that need attention. In addition if organization invests in quality equipment than it can be confident that the investment will enable best support to business which, at the end of the day, is the only thing that matters.

REFERENCES

[3] Berkowitz , H. (2002). *Building Service Provider Networks*. New York: Wiley.

[2] Snyder, J. (2000). Reliability: Something you build, not buy. *Network World*, no. 9, p. 39

[6] (2015-09-12) www.cisco.com/c/en/us/td/docs/ios/cether/configuration/guide/ce_lnkbndl.html

[4] (2015-09-11) www.cisco.com/c/en/us/td/docs/solutions/Enterprise/WAN_and_MAN/QoS_SRND/QoS-SRND-Book/QoSDesign.html#wp998242

[5] (2015-09-09) h17007.www1.hp.com/docs/reports/irf.pdf

[1] (2015-09-11) www.mitre.org/publications/systems-engineering-guide/acquisition-systems-engineering/risk-management