Slovenská technická univerzita
Fakulta informatiky a informačných technológií
Ilkovičova 3, 842 16 Bratislava 4

Bezpečné Datacentrum

Vypracovali: Martin Nagy, Dávid Oros, Roman Panenka, Martin Pirháč
Študijný odbor: Počítačové systémy a siete
Predmet: Projektovanie aplikácií počítačov
Ročník: 3. – letný semester
Ak. rok: 2009/2010

Obsah

1. Úvod

1.1 Účel a rozsah dokumentu

Tento dokument je výsledkom práce na predmete Projektovanie aplikácií počítačov. Práca je vypracovaná študentmi 3. ročníka fakulty FIIT v letnom semestri. Pokrýva zadanie pre vypracovanie návrhu Bezpečného Datacentra, teda datacentra s dôrazom na bezpečnosť a vysokú dostupnosť. Bezpečnosť sa týka predovšetkým dát ale aj samotných zariadení. V našom návrhu sa zameriame predovšetkým na Open Source riešenia a taktiež aj operačný systém bude preferovaný Linux.

1.2 Redundancia a zabezpečenie

V našom projekte sme a rozhodli použiť predovšetkým redundantné technológie. Postaviť datacentrum tak, aby väčšina jeho komponentov bola redundantná. Redundantný znamená záložný alebo náhradný. Každý komponent teda bude mať svoj záložný komponent. Či už to bude server, prepínač alebo smerovač. Tak isto budú v datacentre aj záložné zdroje elektriny ak by sa primárny zdroj vypol.

Bezpečnosť však nespočíva iba v zdvojovaní komponentov, ale taktiež treba mať zabezpečené datacentrum. Miesto bude mať viacero ochranných prvkov ako bezpečnostné dvere, alarm pri vniknutí, požiarny alarm, monitorovanie priestorov spojené s pohybovými čidlami a pod.

2. Sieťová topológia a jej vplyv na dostupnosť

Jedna z prvých vecí, ktorá má dopad na dostupnosť je fyzická realizácia siete. Topológia siete má priamy dopad predovšetkým na ukazovateľ chybovosti MTBF (Mean Time Between Failures). Tento ukazovateľ hovorí o priemernom čase medzi dvoma poruchami, čo je vlastne priemerný čas, kedy sieť funguje bez poruchy. Čím je tento čas vyšší, tým je sieť menej poruchová a tým aj bezpečnejšia. Ukazovateľ MTBF vo všeobecnosti klesá s množstvom komponentov zapojených do série a stúpa s množstvom paralelne zapojených zariadení.

MTBF = celkový čas bezporuchovej prevádzky / počet porúch

2.1 Spôsoby zlyhania siete

Existujú tri typy porúch topológie, ktoré majú významný dopad na dostupnosť siete :

- Zlyhanie komponentu
- Zlyhanie systému
- Jediný bod zlyhania

2.1.1 Zlyhanie komponentu

Tento ukazovateľ hovorí o pravdepodobnosti, že daný komponent zlyhá. Chybovosť komponentu sa dá štatisticky vyjadriť ako podiel času kedy komponent fungoval a času, počas ktorého by mal fungovať.

Chybovosť = čas up / (čas up + čas down)

2.1.2 Zlyhanie systému

Zlyhaním systému sa myslia chyby zapríčinené rôznymi vonkajšími faktormi. Napríklad vytrhnutie kábla, zvýšená teplota, vlhkosť a podobne. Čím viac komponentov je v systéme, tým je väčšia pravdepodobnosť zlyhania systému

2.1.3 Jediný bod zlyhania (Single point of failure)

Jediný bod zlyhania je taký sieťový prvok, ktorého zlyhanie by spôsobilo zlyhanie celého systému. Pre dosiahnutie čo najväčšej dostupnosti siete je potrebné minimalizovať výskyt takýchto miest v sieti. Typickým prístupom pre odstránenie jediných bodov zlyhania je redundancia sieťových prvkov, ktorá umožní jednotlivé poruchy izolovať. Zlyhanie redundantného prvku potom spôsobí iba lokálnu poruchu, ktorá nemá výraznejší vplyv na správny chod systému.

2.2 Redundancia

Redundancia okrem zvýšenia dostupnosti so sebou prináša aj iné výhody. Redundantné prvky napríklad poskytujú ďalšie sieťové prostriedky, ktoré môžu byť použité na prerozdelenie záťaže a zvýšenie výkonu siete. Nasadenie redundantných komponentov do bežnej prevádzky je však na zváženie, pretože v prípade poruchy sa celá záťaž prenesie na jediné zariadenie, ktoré pravdepodobne nebude schopné zabezpečiť všetky služby v požadovanej kvalite. Okrem toho príliš veľké množstvo redundantných prvkov môže výrazne zvýšiť čas konvergencie siete v prípade poruchy.

3. Dostupnosť siete

Aby bola dosiahnutá čo najvyššia dostupnosť siete, je potrebné zamedziť výskytu porúch, správnou voľbou topológie siete. Ohľad pritom treba dbať na zabezpečenie dostupnosti sieťových prostriedkov tak na druhej ako aj na tretej vrstve sieťového modelu RM OSI a tiež na vzájomnú kompatibilitu týchto riešení. V nasledujúcich kapitolách budú podrobnejšie rozobrané spôsoby zabezpečenia dostupnosti siete na druhej a tretej vrstve.

3.1 Agregácia prenosových liniek (Trunking)

Úlohou agregácie prenosových liniek, niekedy nazývanej ‘trunking’, je zabezpečiť dostupnosť rozdelením sieťovej komunikácie medzi viaceré fyzické prenosové linky. V prípade poruchy a výpadku jednej z prenosových liniek sa komunikácia nepreruší, ale sa prenesie na ostatné linky, ktoré sú ešte aktívne.

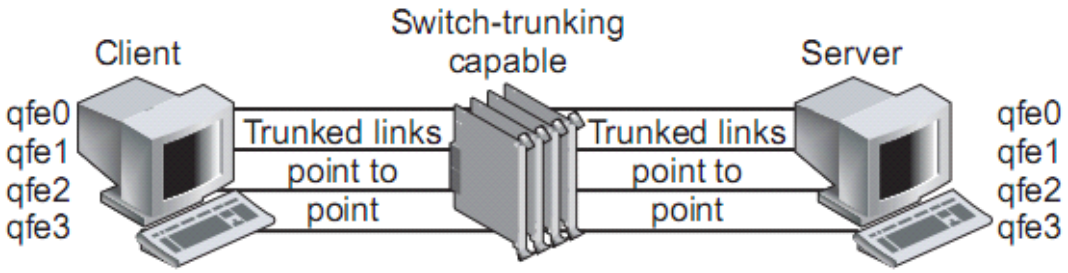
O agregácií hovorí štandard IEEE 802.3ad, ktorý zabezpečuje kompatibilitu tohto riešenia medzi rôznymi sieťovými zariadeniami rôznych výrobcov. Ako mnoho iných štandardov tak aj tu existuje viacero spôsobov k implementácií danej špecifikácie.

Pri agregácií sa namiesto fyzickej MAC adresy, ktorá je viazaná na konkrétny port sieťového zariadenia využíva logická MAC adresa, ktorá môže byť priradená viacerým fyzickým portom. Táto logická adresa bude posunutá tretej vrstve, čo znamená, že sa bude namiesto fyzickej MAC adresy vkladať do ARP odpovedí logická adresa. Všetky porty, ktoré sa zúčastnia agregácie budú teda navonok vystupovať ako jediný interface DLPI (Data Link Provider Interface). Keďže porty smerovača vystupujú navonok ako jediný port, je zaručené, že ich protokol STP (Spanning Tree Protocol) neuvedie do blokujúceho stavu v snahe zabrániť vzniku topologických slučiek.

Agregáciu prenosových liniek samozrejme musia podporovať obidve strany spojenia, ktoré budú navzájom komunikovať prostredníctvom protokolu LACP (Link Aggregation Control Protocol). Komunikácia začína výmenou správ typu ‘query’, prostredníctvom ktorých sa zúčastnené strany agregácie navzájom ohlásia a určia porty, ktoré sa majú agregácie zúčastniť. Pokiaľ majú obidve strany záujem o agregáciu, vytvorí sa trunk odoslaním správy ‘start group’, ktorá obsahuje identifikátory liniek prislúchajúcich k zúčastneným portom.

LACP protokol môže neskôr tieto porty zo skupín odstraňovať. Taká situácia môže nastať napr. pri detekcií poškodenej linky. Namiesto rozloženia záťaže medzi ostatné porty, algoritmus jednoducho celú sieťovú komunikáciu poškodenej linky prenesie na jeden z fungujúcich portov skupiny. Zberateľ (collector) poskladá komunikáciu prichádzajúcu na iných portoch a rozdeľovač (distributor) prerozdeľuje tok dát medzi porty patriace do trunkovacej alebo agregačnej skupiny.

Pre lepšie pochopenie technológie a overenie vhodnosti riešenia na zvýšenie sieťovej dostupnosti uvádzame jednoduchý príklad. V jednoduchej topológií, znázornenej na obrázku Obrázok 1 figuruje jeden prepínač (switch), jeden klientský počítač a server. Klient aj server sú k prepínaču pripojený prostredníctvom štyroch liniek. Toto nastavenie umožňuje rozdeľovanie záťaže na všetky štyri linky.



Obrázok 1 - agregácia liniek medzi klientom a serverom

Čas - 14:22:05

Name	Ipkts	Opkts	%Ipkts	%Opkts
qfe0	210	130	100	25
qfe1	0	130	0	25
qfe2	0	130	0	25
qfe3	0	130	0	25

Čas - 14:22:08

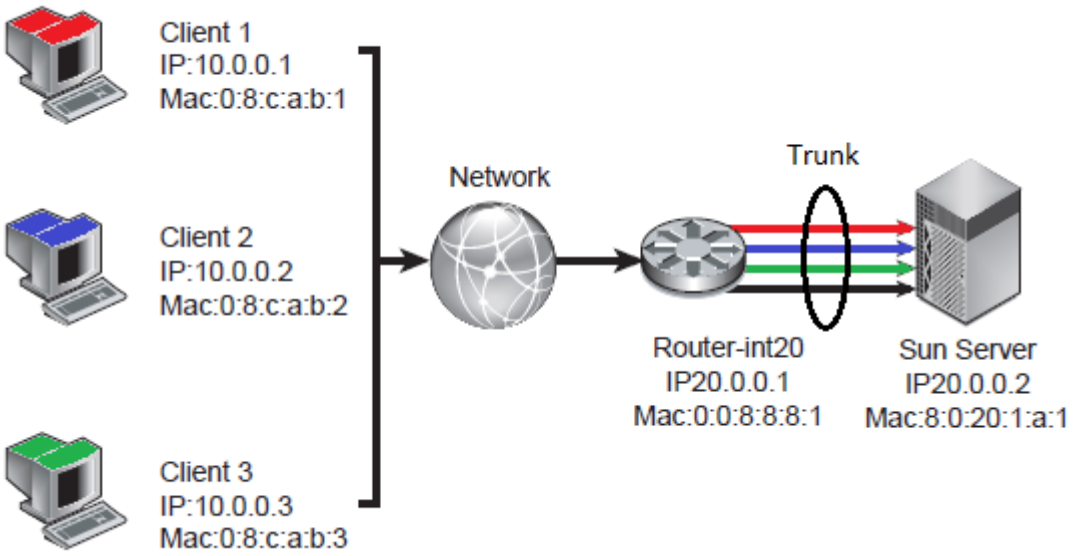
Name	Ipkts	Opkts	%Ipkts	%Opkts
qfe0	0	0	0	0
qfe1	1028	1105	100	51.52
qfe2	0	520	0	24.24
qfe3	0	520	0	24.24

Z klienta bolo vygenerovaných niekoľko testovacích TTCP (Test Transmission Control Protocol) tokov. Keď boli funkčné všetky linky, záťaž bola prerozdeľovaná rovnomerne a každý port vykazoval 25 percentnú záťaž. Po vypadnutí jednej z liniek (qfe0) sa poškodená komunikácia premiestnila na jednu z ďalších funkčných liniek (qfe1), ktorá začala vykazovať 51 percentnú záťaž. Výpadok trval 2-3 sekundy.

Keby boli všetky linky veľmi vyťažené, spomínaný algoritmus by mohol nútene zvýšiť vyťaženie jednej z fungujúcich liniek o záťaž na vypadnutej linke. Napr. ak by boli všetky linky vyťažené na 55 percent svojej kapacity a jedna z liniek by havarovala, došlo by k nasýteniu inej linky na 55 + 55 = 110 percent. Z toho dôvodu je táto technológia vhodná iba na linky spájajúce dva uzly na jednom segmente (point-to-point). Taktiež treba spomenúť nutné náklady v podobe obsadenia viacerých portov na oboch stranách komunikácie.

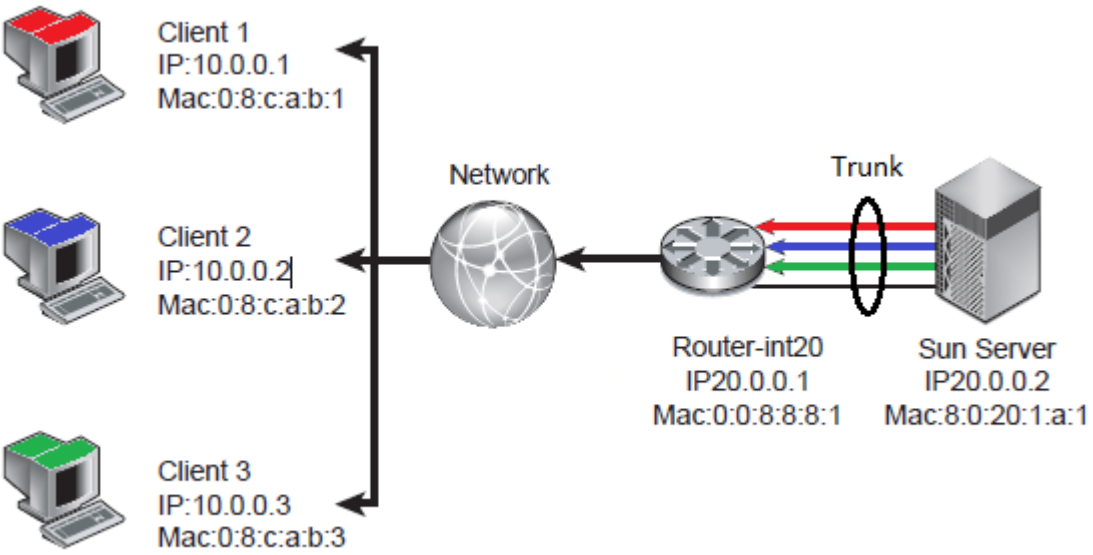
3.2 Princíp rozdelenia záťaže (Load sharing)

Pre optimálne rozdelenie dátových tokov medzi prenosové linky je potrebné zvoliť vhodný algoritmus. Môžeme napríklad využiť to, že prichádzajúce dátové toky budú s najväčšou pravdepodobnosťou pochádzať z viacerých zdrojov, preto je možné toky zaraďovať do príslušných prenosových liniek na základe ich zdrojovej IP adresy. Pre každú novú zdrojovú IP adresu sa pridelí náhodná prenosová linka algoritmom Round Robin, čím sa zabezpečí rovnomerné zaťaženie všetkých dátových liniek v trunku(Obrázok 2).



Obrázok 2

Podobný princíp rozdelenia záťaže sa použije pre dátové toky odchádzajúce zo servera s tým, že ako identifikátor sa nepoužije zdrojová, ale cieľová IP adresa. Rozdelenie záťaže medzi prenosové linky na základe cieľovej IP adresy ilustruje Obrázok 3.



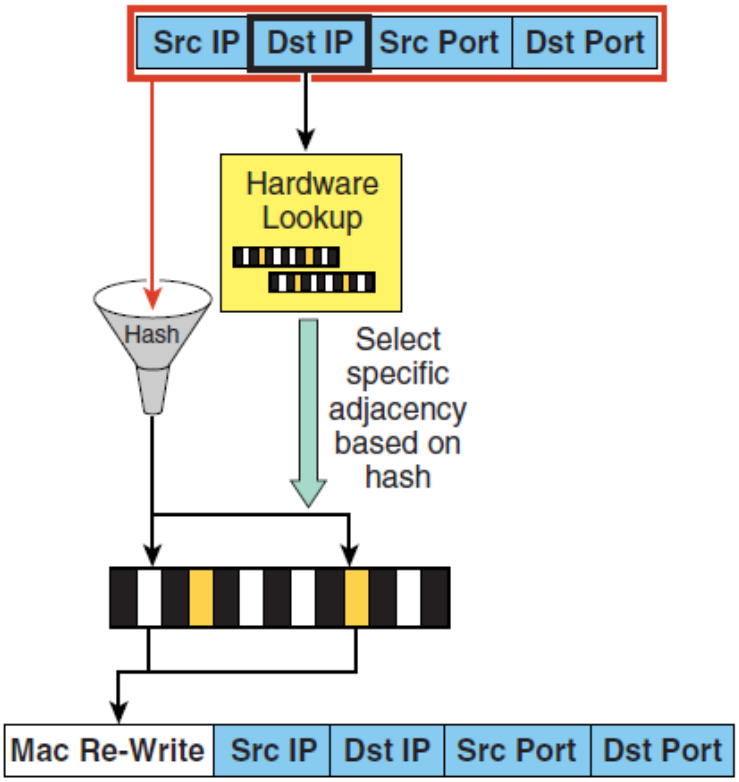
Obrázok 3

Ako vidno z vyššie uvedených obrázkov, trunk pozostáva zo štyroch prenosových liniek, pričom jeho konce sú identifikované jedinou logickou MAC adresou prislúchajúcou k DLPI. Z tohto dôvodu nie je možné použiť algoritmus pre rozdelenie záťaže, ktorý by dátové toky identifikoval na základe MAC adries.

3.3 EtherChannel

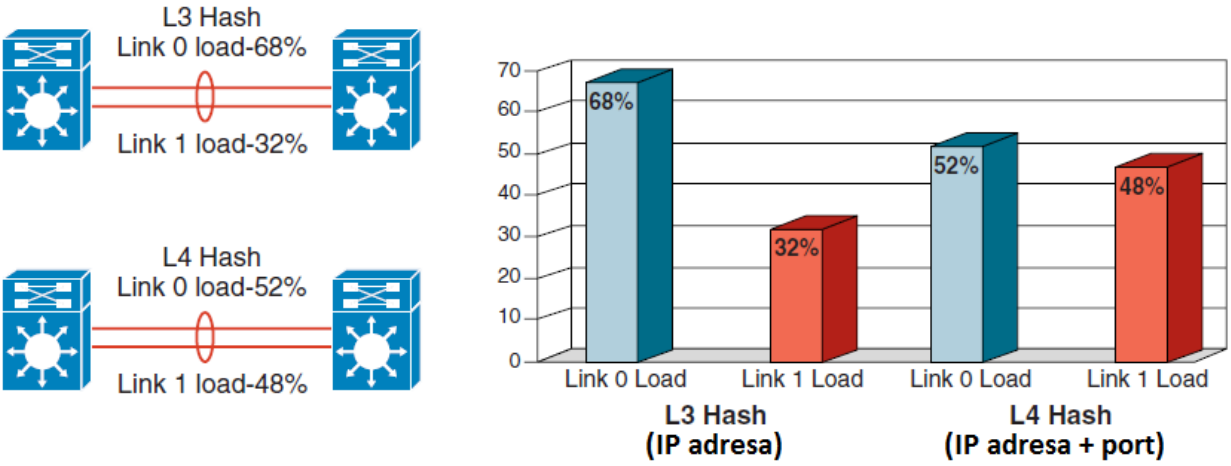
Zariadenie firmy Cisco však podporujú aj iný variant agregácie prenosových liniek nazvaný EtherChannel. EtherChannel je Cisco proprietárny štandard, ktorý využíva namiesto spomínaného protokolu LACP protokol PAgP (Port Aggregation Protocol). EtherChannel umožňuje zoskupenie až ôsmych paralelných liniek, ktoré sa navyše môžu fyzický nachádzať na viacerých moduloch prepínača. Tým sa zabezpečí zvýšená dostupnosť v prípade poruchy jedného z modulov.

Oproti IEEE 802.3ad implementácií agregácie liniek poskytuje EtherChannel výhodu aj pri rozdeľovaní záťaže medzi jednotlivé prenosové linky. Zatiaľ čo 802.3ad štandard využíva len základnú identifikáciu dátových tokov na základe ich IP adresy, EtherChannel identifikuje dátový tok aj na základe adresy štvrtej vrstvy, TCP/IP portu. Z IP adries a portov každého prichádzajúceho paketu sa vytvorí tzv. L4 Hash, na základe ktorého sa potom určí, ktorou prenosovou linkou daný paket pôjde. Každý špecifický index vypočítaný ako L4 Hash je uložený do tabuľky susedstiev, podľa ktorej sa smerujú ďalšie prichádzajúce pakety s rovnakým indexom. Proces výberu prenosovej linky na základe L4 hash-u ilustruje Obrázok 4.



Obrázok 4 - výber prenosovej linky

Na základe portu a IP adresy je možné každý dátový tok jednoznačne identifikovať, čo umožňuje rovnomernejšie rozdelenie záťaže medzi prenosové linky v trunku. Výhodu využitia čísla portov pri identifikácií dátových tokov ilustruje Obrázok 5.

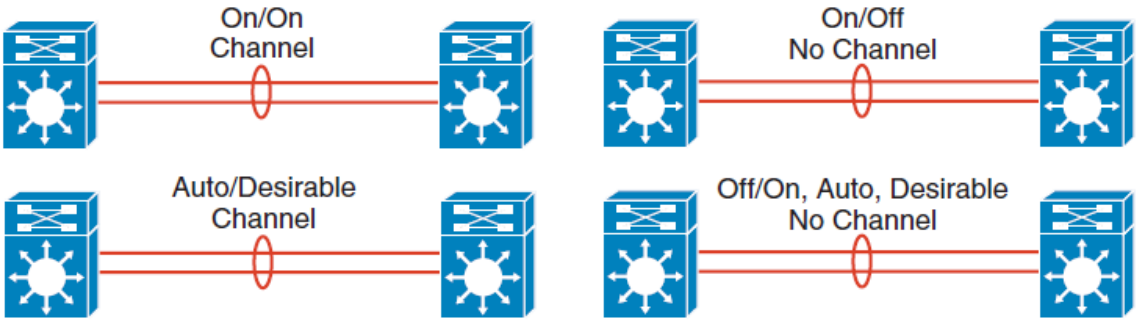


Obrázok 5 - identifikácia dátových tokov

3.3.1 Konfigurácia PAgP:

PAgP poskytuje možnosť automatickej konfigurácie EtherChannel-u medzi prepojenými prepínačmi. Porty smerovača je možné nastaviť do štyroch rôznych stavov, ktoré určujú, či a za akých podmienok sa EtherChannel vytvorí.

- On – vždy je súčasťou EtherChannel-u
- Desirable – vyzýva opačnú stranu k vytvoreniu EtherChannel-u
- Auto – stane sa súčasťou EtherChanel-u ak je k tomu vyzvaný
- Off – nikdy nie je súčasťou EtherChannel-u



Obrázok 6 - Konfigurácia etherChannelu

Typické nastavenie portov je nastavenie ‘auto‘ režimu na strane prepínača na prístupovej vrstve a ‘desirable’ na strane prepínača distribučnej vrstvy. Pri takomto nastavení je EtherChanel aktívny až po kompletnej konfigurácií a zároveň je zabezpečená konektivita pri nekompletnom nastavení kanálu. Taktiež sa tým zabráni vzniku nechcených párov a zlyhaniu protokolu STP, ku ktorému by mohlo dôjsť pri nastavení on/on. Odporúčaná konfigurácia EtherChannel-u pod Cisco IOS je uvedená nižšie.

Globálny konfiguračný režim:

```
port-channel load-balance src-dst-port
```

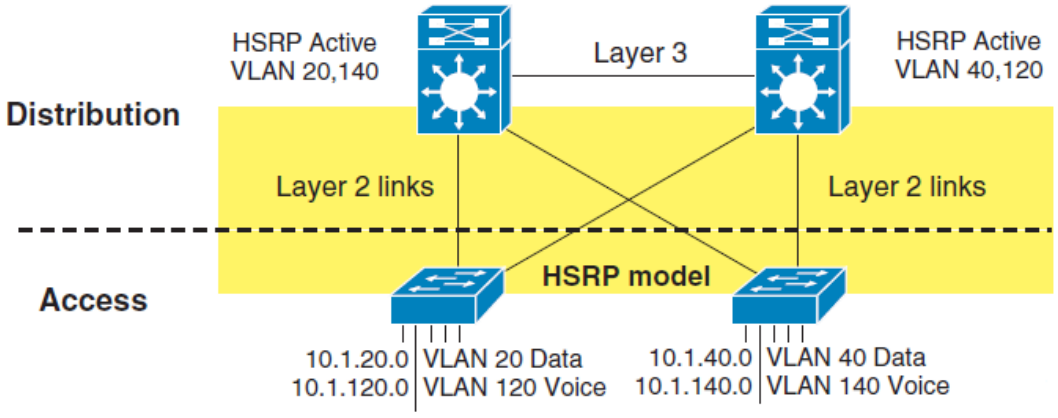
Konfigurácia interfejsov:

```
interface GigabitEthernet2/1
description to 6k-Core-left CH#1
no ip address
mls qos trust dscp
channel-group 1 mode on
!
interface GigabitEthernet2/2
description to 6k-Core-left CH#1
no ip address
mls qos trust dscp
channel-group 1 mode on
!
interface Port-channel1
description to cr2-6500-1 CHANNEL #1
ip address 10.122.0.34 255.255.255.252
mls qos trust dscp
```

3.3.2 Spanning Tree Protocol

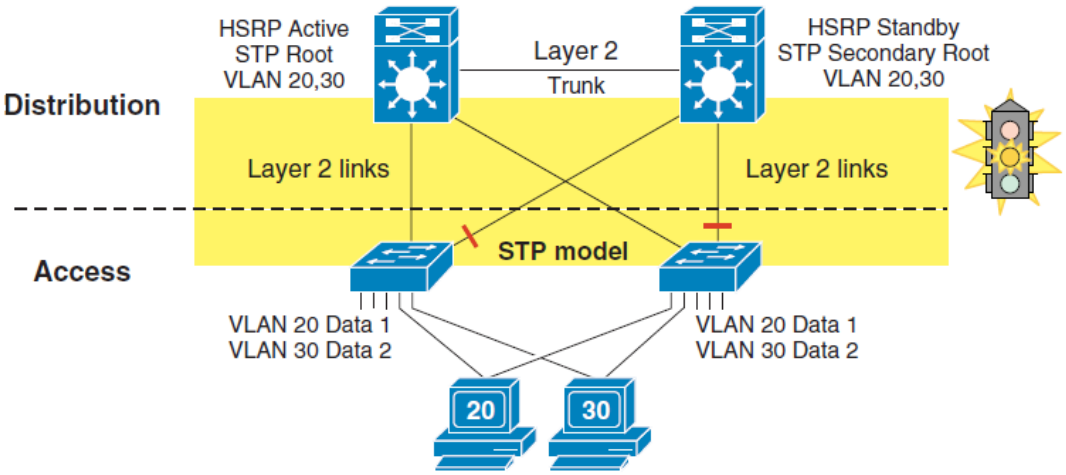
Zabezpečenie vysokej dostupnosti vyžaduje redundantné cesty, aby sa zabezpečila konektivita v prípade výpadku niektorého uzla či prenosovej linky. Úlohou STP protokolu je však blokovanie nadbytočných liniek tak, aby vznikla výsledná topológia bez slučiek. V praxi väčšina najvýkonnejších sietí, čo sa týka ich dostupnosti, spoľahlivosti a konvergencie za normálnych podmienok nepotrebuje STP protokol. Avšak STP sa stále využíva k zabráneniu vzniku neočakávaných slučiek, ktoré môže nechtiac vytvoriť používateľ na prístupovej vrstve.

V prípade že medzi L3 prepínačmi na distribučnej vrstve nie je potrebné prepojenie na druhej vrstve, topológia s redundantnými prvkami je bez slučiek. Takáto topológia je zobrazená na obrázku Obrázok 7 a zaobíde sa bez nasadenia protokolu STP.



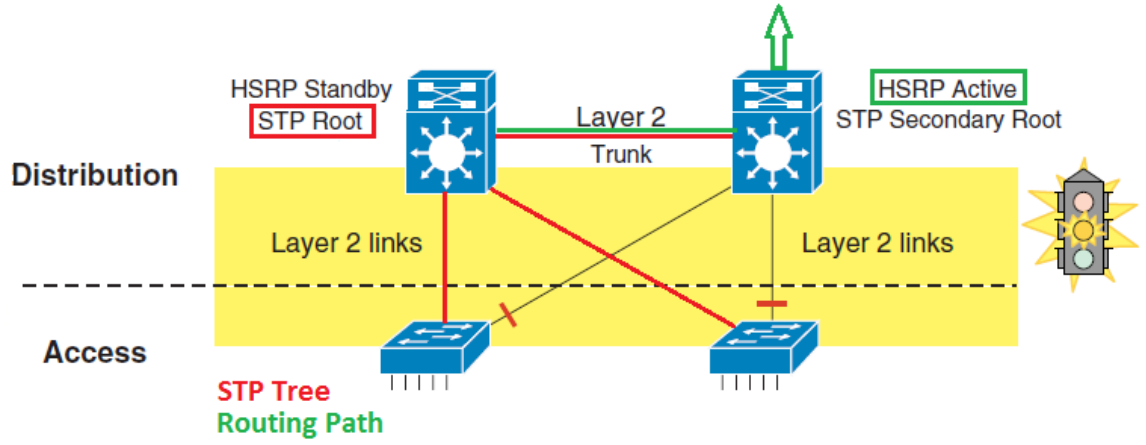
Obrázok 7 - STP prípad 1

Pre dátové centrum je však typické dvojité pripojenie serverov na prepínače prístupovej vrstvy, aby sa vylúčil ‘single point of failure’ v podobe jedinej prenosovej linky. Tiež je potrebné prepojenie na druhej vrstve medzi L3 prepínačmi na distribučnej vrstve, aby sa predišlo vzniku nechcených ciest v prípade prerušenia prepojenia so smerovačom v core resp. backbone vrstve. V takomto prípade je potrebné zabezpečiť topológiu proti uzavretým slučkám prostredníctvom STP protokolu. Zároveň je dôležité nastaviť STP tak, aby výsledná cesta korešpondovala s cestou k default gateway nakonfigurovanou v HSRP resp. VRRP.



Obrázok 8 - STP prípad 2

Pre rýchlu konvergenciu protokolu STP použijeme v našej topológií variant Rapid PVST+ (Per VLAN Spanning Tree Plus) s mechanizmom pre rýchle zotavenie redundantných liniek UplinkFast. Rapid PVST+ je nadradený ostatným variantom STP protokolu ako PVST alebo klasickému STP podľa štandardu 802.1d. Čas konvergencie tohto Cisco proprietárneho protokolu sa v sieti nášho rozsahu pohybuje okolo jednej sekundy. Ako bolo už vyššie spomenuté, pre správnu funkciu siete je nevyhnutné nastaviť Rapid PVST+ root na L3 prepínači, ktorý zároveň plní funkciu HSRP aktívneho smerovača. V opačnom prípade by sa v sieti nevytvorili optimálne cesty a prenosová linka medzi L3 prepínačmi na distribučnej vrstve by bola zbytočne zaťažená prebytočnou komunikáciou. Nesprávne nastavenie STP root prepínača a HSRP aktívneho smerovača ilustruje Obrázok 9.

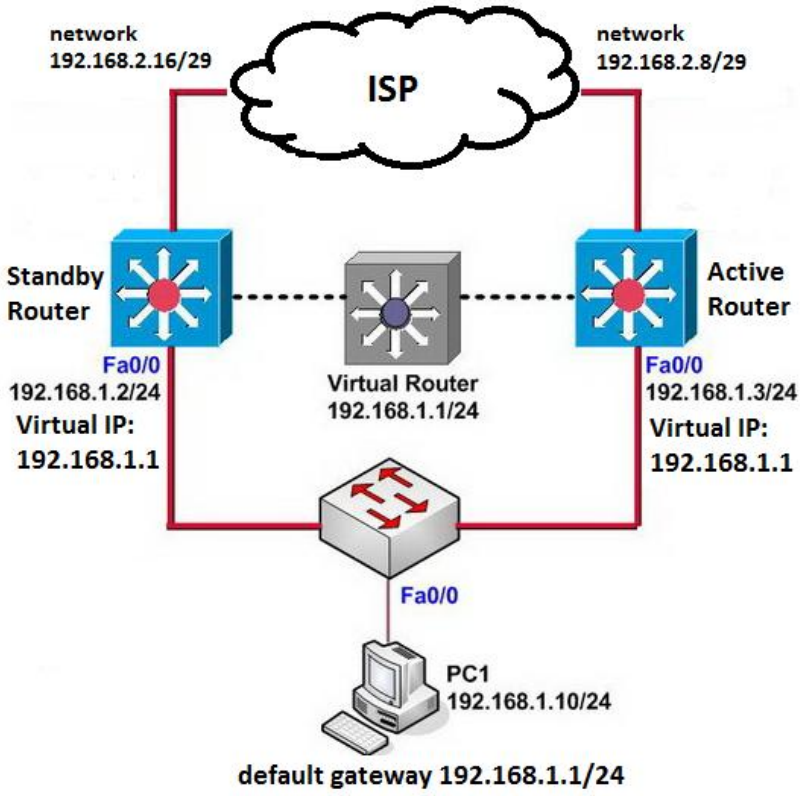


Obrázok 9 - STP prípad 3

3.4 Využitie HSRP default gateway redundancie

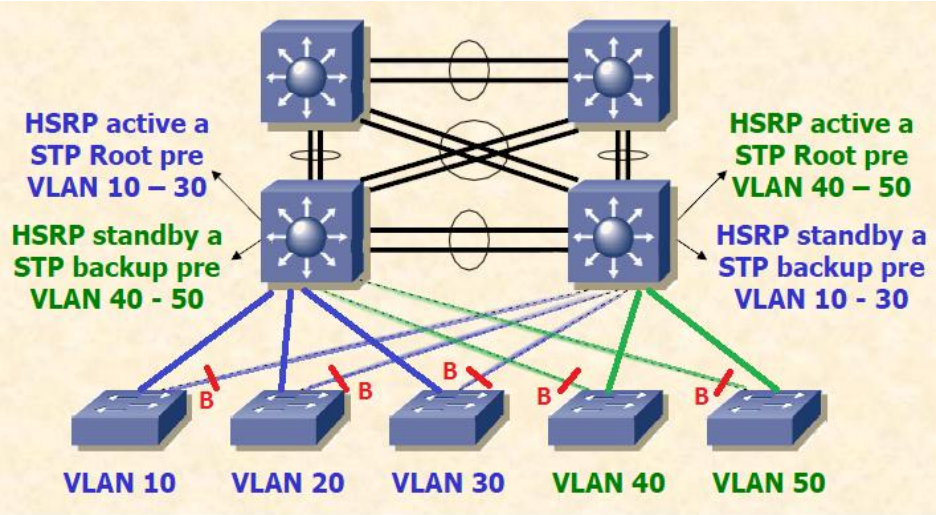
Redundancia smerovačov nastavených ako ‘default gateway’, niekedy nazývaná aj ‘first hop‘ redundancia, umožňuje dostupnosť siete v prípade zlyhania default gateway smerovača. Firma Cisco vyvinula protokol HSRP (Hot Standby Router Protocol), ktorého úlohou je zabezpečiť konektivitu všetkých staníc v prípade zlyhanie gefault gateway smerovača. Táto technika bola následne štandardizované organizáciou IETF pod názvom VRRP (Virtual Router Redundancy Protocol) ako štandardná technika poskytujúca redundanciu smerovačov.

Podstatou funkcie HSRP je že každá stanica má nastavenú virtuálnu IP adresu default gateway smerovača. V skutočnosti však máme dva smerovače s rovnakou virtuálnou IP adresou, ale o tom koncové stanice nevedia. Pri použití HSRP je jeden smerovač s najvyššou prioritou aktívny a ostatné ostanú v stave ‘standby‘, teda sú pasívne. Na ARP dotazy koncových uzlov na virtuálnu IP adresu default gateway smerovača odpovedá svojou MAC adresou vždy iba aktívny smerovač. Pasívny alebo ‘standby’ smerovač neustále monitoruje činnosť aktívneho smerovača prostredníctvom hello paketov, aby vedel detegovať jeho prípadnú poruchu a okamžite prevziať jeho funkciu. Keďže pasívny smerovač má tú istú virtuálnu IP adresu, začne v prípade poruchy okamžite odpovedať na všetky dotazy ARP smerované na default gateway svojou MAC adresou. Týmto sa zaručí, že konektivita sa obnoví okamžite po skonvergovaní protokolu STP, takže koncové stanice by poruchu siete nemali vôbec postrehnúť. Základná topológia s HSRP redundanciou je zobrazená na Obrázok 10



Obrázok 10 - HSRP redundancia

Pri voľbe aktívneho a pasívneho smerovača je potrebné aby bolo nastavenie kompatibilné s výslednou topológiou po odstránení slučiek protokolom STP. Topológia s využitím HSRP je zobrazená na obrázku Obrázok 11. Hrubo vyznačené linky predstavujú aktíve prepojenie, rozmazané linky, ktoré sú protokolom STP blokované predstavujú redundantné prepojenia, ktoré sa použijú v prípade poruchy niektorého z default gateway smerovačov.



Obrázok 11 - Topológia s HSRP

Konfigurácia dvoch smerovačov, pričom každý bude aktívny pre jednu skupinu používateľov a standby pre druhú skupinu používateľov, tak ako je to znázornené na obrázku Obrázok 11 je uvedená nižšie. Týmto sa zároveň zabezpečí aj rovnomerné rozdelenie sieťovej komunikácie medzi obidva smerovače tzv 'load balancing'.

Konfigurácia pre smerovač A:

```
hostname smerovacA
!
interface ethernet 0
ip address 10.0.0.1 255.255.255.0
standby 1 priority 110
standby 1 preempt
standby 1 ip 10.0.0.3
standby 2 preempt
standby 2 ip 10.0.0.4
```

Konfigurácia pre smerovač B:

```
hostname smerovacB
!
interface ethernet 0
ip address 10.0.0.2 255.255.255.0
standby 1 preempt
standby 1 ip 10.0.0.3
standby 2 priority 110
standby 2 preempt
standby 2 ip 10.0.0.4
```

V tomto prípade bude mať jedna skupina nastavenú default gateway IP adresu 10.0.0.3 a druhá 10.0.0.4. Keďže každý smerovač má nastavenú vyššiu prioritu pre inú skupinu, pri voľbe aktívneho smerovača sa smerovačA stane aktívnym pre skupinu 1 a smerovačB pre skupinu 2 (prednastavená priorita smerovača je 100). Príkaz '`standby preempt`', zabezpečí, že po zotavení z poruchy sa uskutoční voľba aktívneho smerovača, čo zaručí, že nastavenie sa vráti do pôvodného stavu. Napríklad ak spadne aktívny smerovačA, smerovačB sa stane aktívnym pre obidve skupiny. Aby sa po zotavení stal smerovačA opäť aktívnym pre skupinu 1 a bol dosiahnutý load balancing, je potrebné uskutočniť voľbu aktívneho smerovača. Za príkazom '`standby preempt`' sa zvyčajne uvádza ešte oneskorenie, ktoré poskytne smerovaču dostatok času aby sa po zotavení stihol naučiť potrebné smerovacie informácie a až následne sa stal aktívnym.

3.5 Zhodnotenie

V tejto kapitole boli popísané niektoré techniky riešenia redundancie ako nástroja na zvýšenie dostupnosti siete odstránením možných miest zlyhania 'single point of failure'. Na druhej vrstve bolo navrhnuté riešenie v podobe EtherChannel-u, ktoré zabezpečuje agregáciu prenosových liniek a zdieľanie ich prostriedkov (load sharing). Vzniku topologických slučiek zabráni protokol Rapid PVST+. Na tretej vrstve odporúčame využitie protokolu HSRP, ktorý rieši redundanciu smerovačov a zároveň umožňuje efektívne využitie sieťových prostriedkov vhodným prerozdelením komunikácie (load balancing). EtherChannel, STP variant Rapid PVST+, ako aj HSRP sú štandardy licencované firmou Cisco a pre ich realizáciu sú potrebné sieťové zariadenia s operačným systémom od tejto firmy.

4. Serverové riešenia

Keďže servery sú primárnym prvkom v datacentre, je potrebné zabezpečiť ich, aby pracovali čo najlepšie a ideálne bez nežiaducich výpadkov. Výpadky sa však samozrejme nedajú obmedziť na žiadne, no pokúsime sa k tomuto číslu čo najviac priblížiť.

4.1 Redundancia serverov – Clustering

Ako už bolo spomenuté, každý komponent bude mať svoj záložný, resp. istiaci komponent. Nebude to ináč ani pri serveroch. Redundancia v serverových riešeniach sa nazýva clustering.

Cluster je vlastne zhluk počítačov, najmenej však dvoch, ktoré sa navonok tvária ako jeden samostatný počítač. Tieto počítače navzájom úzko spolupracujú. Cluster môže byť založený na viacerých technológiách, podľa toho na čo je daný cluster určený.

Najbežnejší cluster v malých a stredných datacentrách je tzv. „failover cluster“. Tento typ clustera sa využíva tam, kde je potrebná vysoká dostupnosť služieb. Ďalším typom je „load balancing“ cluster, kde slovo load môžeme preložiť ako záťaž. Tento typ clusteru rozdeľuje serverovú záťaž rovnomerne na všetky počítače spojené do jednotného clusteru. Tretím typom clusteru je „compute cluster“, ktorý je využívaný predovšetkým tam, kde je potrebný vysoký výpočtový výkon. Jedným z najnovších typov clusteru je „grid computing“ cluster, ktorý je takým spojením medzi load balancing a compute clusterom.

4.1.1 Failover cluster – HA cluster

Failover cluster sa taktiež nazýva HA Cluster – High Availability Cluster, čo v preklade znamená vysoko dostupný cluster. Tento typ clusteru sa používa najmä pri serverových riešeniach, pre ktoré je dôležitý čo najmenší výpadok služieb.

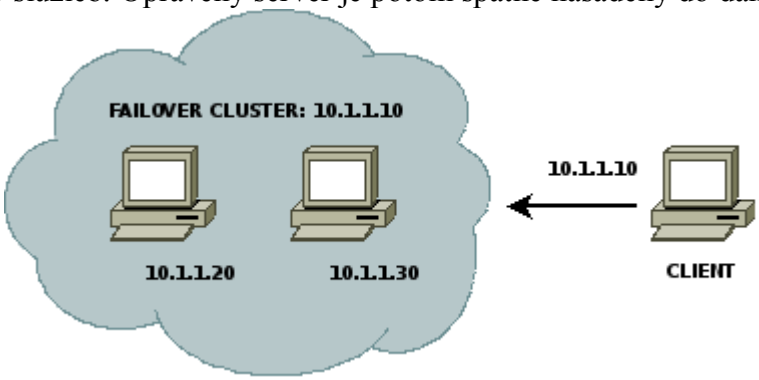
Cluster je tvorený minimálne 2 počítačmi, no horná hranica určená nie je. Pri tomto type sa však nepoužíva až toľko veľa serverov. Princíp failover clusteru spočíva v tom, že jednotlivé prvky medzi sebou komunikujú a navzájom si poskytujú informácie o ďalších. Najdôležitejšia je informácia, ktorý server je v akom stave. Server môže byť buď schopný prevádzky alebo nie, prípadne ešte môže byť schopný iba na polovicu.

Pri spustení takéhoto typu clustru sú všetky servery rovnako nakonfigurované, väčšinou sú až na malé odlišnosti identické. Servujú rovnaké služby s rovnakými dátami. Jeden z týchto serverov je určený za hlavný server a to buď nami alebo softvérom, ktorým jednotlivé servery komunikujú. Tento server sa nazýva hlavný, preto že na ňom bežia všetky služby a vybavuje všetky požiadavky. Ostatné servery označené ako záložne sú v stave pohotovosti. Znamená to, že dostavajú informácie o stave hlavného systému a taktiež odosielajú informácie o svojom stave.

Takýto cluster ako už bolo spomenuté sa navonok javí ako jeden počítač a teda má pridelenú jednu verejnú IP adresu. Avšak jednotlivé servery majú ešte aj svoje vnútorne IP adresy, práve kvôli tomu, aby boli rozlíšiteľné a mohli medzi sebou komunikovať.

Význam failover clustru spočíva v tom, že ak hlavný server sa z akéhokoľvek dôvodu stane nedostupným, jeden zo záložných serverov preberie jeho verejnú IP adresu a postupne preberie všetky požiadavky, ktoré predtým smerovali na hlavný server a obsluhuje ich ďalej. Navonok to vyzerá ako by sa nič nestalo, keďže IP adresa sa nezmenila. Môže sa stať, že výpadok bude postrehnuteľný, avšak väčšinou ide o maximálne pár sekúnd. V tejto chvíli sa hlavným serverom stal ten, ktorý prebral verejnú IP adresu a obsluhuje všetky požiadavky a služby.

Medzi časom môže byť pokazený server opravený bez toho, aby nastal rapídny výpadok niektorej zo služieb. Opravený server je potom spätne nasadený do daného clustru.



Obrázok 12 – Failover Cluster

4.1.1.1 Technológia CARP

Protokol CARP *(Common Address Redundancy Protocol)* je bezpečnou a voľne dostupnou alternatívou k protokolom VRRP *(Virtual Router Redundancy Protocol)* a HSRP *(Hot Standby Router Protocol)*. Dovoľuje viacerým systémom z tzv. redundantnej skupiny zdieľať v lokálnej sieti rovnakú IP adresu. Za normálnych okolností používa zdieľanú IP adresu iba hlavný systém a obsluhuje všetky požiadavky, ktoré sú na ňu smerované. Ak by však tento systém z ľubovoľných príčin prestal reagovať, začne zdieľanú IP adresu používať jeden zo záložných systémov v závislosti od jeho priority.
Tento protokol je jednou z mnohých technológií, ktoré dala svetu skupina vývojárov zoskupená okolo systému OpenBSD.

Jednou z hlavných výhod riešenia failover clustera s protokolom CARP je, že nevyžaduje žiadne dodatočné sieťové prepájanie členských systémov. Je tomu tak najmä vďaka skutočnosti, že všetky informácie o aktuálnom stave celého clusteru, ktoré si medzi sebou tieto systémy vymieňajú, sú po sieti prenášané šifrované.

Pri technológii CARP si môžeme zvoliť či necháme všetkým serverom rovnakú šancu stať sa hlavným serverom, to je vtedy ak sú všetky rovnako výkonne, alebo niektorý zo serverov uprednostníme zvýšením jeho priority. Zvýšenie priority niektorého zo serverov sa využíva vtedy, ak máme server, ktorý je výkonnejší oproti záložným serverom. Vtedy využívame záložne servery výhradne pri výpadku hlavného serveru.

Konfigurácia protokolu CARP je pomerne jednoduchá. Nie je založená na žiadnom konfiguračnom súbore. Parametre sa zadávajú priamo pri spúšťaní. Je samozrejme možne celé toto spúšťanie zautomatizovať skriptom. Potrebné je však vytvoriť 2 skripty, ktoré zaručia nastavenie prebratej IP adresy na dané rozhranie, s ktorým bude nový hlavný server komunikovať so svetom.

Spustenie protokolu CARP vyzerá nasledovne:

```
# ucarp --interface=eth0 --vhid=42 --pass=heslo --addr 10.1.1.10 --srcip
10.1.1.20              --upscript=/usr/local/bin/carp-up.sh             --
downscript=/usr/local/bin/carpdown.sh daemonize
```

Skript carp-up.sh obsahuje príkazy vyššie spomenuté pre nastavenie prebratej IP adresy na vopred zvolené rozhranie. Parameter "*interface*" urcuje sietové rozhranie, ktoré má program využívat na výmenu informácii o stave cluster-a. Parameter "*vhid*" a "*pass*" musia byt na všetkých členských systémoch zhodné, pretože určujú identifikátor clusteru a heslo, ktoré je použité pri šifrovaní komunikácie medzi jednotlivými členskými systémami. Parametrami "*upscript*" a "*downscript*" sú určené skripty spúšťané pri prechode členského systému na hlavný resp. záložný server. Posledný parameter "*daemonize*" zabezpečuje daemonizáciu procesu, a teda okrem iného aj uvoľnenie terminálu, na ktorom bol program spustený. V tomto prípade bude platiť šikovnejší vyhráva a stane sa hlavným serverom. Ak by sme chceli zabezpečiť, aby niektorý server bol prioritný, použijeme nasledovné prídavne parametre:

```
--advskew=0 --advbase=1 --preempt
```

Tieto parametre určujú okrem iného aj prioritu členského systému.

Vzorec na výpočet priority má tvar *(advskew/256)+advbase* a preferovaný je systém s najnižším výsledkom. Parameter "*preempt*" zabezpečuje akceptovanie takto určenej priority a striktné preferovanie členského systému s najnižším výsledkom. Všetkým ostatným systémom okrem toho, z ktorého chceme spraviť hlavný server, zmeníme hodnotu advskew na hocijakú vyššiu ako má hlavný systém a tým zabezpečíme, že boj o hlavný server vyhrá stále ten nami predurčený.

4.1.1.2 Wackamole a Spread

Wackamole je aplikačné riešenie clusterového problému prehadzovania si jednej verejnej IP adresy z jedného stroja na iný v rámci daného clusteru. Spravuje verejné IP adresy, ktoré majú byť dostupné pre svet po celý čas bez výpadku. Jednotlivé systémy clusteru sú po celý čas monitorované a ako náhle wackamole zistí, že systém je nedostupný, priradí verejnú IP adresu nasledujúcemu stroju. Táto aplikácia taktiež zabezpečuje, že v jednom čase môže byť verejná IP adresa priradená iba jednému stroju a nik iný na nej nepočúva. Pracuje so servermi v clusteri, ktoré sú navzájom priamo prepojené, teda sú na jednej lokálnej počítačovej sieti. Wackamole vytvára a spúšťa svoje inštancie na každom pripojenom systéme a vytvára tak akési združenie inštancií.

Medzi týmito inštanciami sa šíria notifikácie od nástroja, ktorá sa nazýva Spread. Spread monitoruje stav každého pripojeného serveru do clusteru a odosiela ich všetkým inštanciám Wackamolu. Takto sa udržiava informovanosť medzi jednotlivými systémami. Pri zistení, že stroj, ktorému bola priradená verejná IP adresa je nedostupný, Wackamole tento stroj nahradí ihneď iným a to tak, že pridelí danú IP adresu inému stroju v clusteri.

Spread je Open Source nástroj, ktorý poskytuje vysoko účinný systém posielania správ, ktoré dokážu pokryť aj veľké siete. Spread pracuje na princípe vlastného komunikačného kanálu pre rôzne distribuované aplikácie. Poskytuje vysoko nastaviteľný multicast tunel pre komunikáciu aplikácií. Je ním možné posielať aj skupinové správy a taktiež správy medzi vybranými jednotlivcami. Umožňuje aj posielanie správ s potvrdením o doručení.

Spread je ľahko škálovateľný z lokálnych sietí na veľké medzimestské siete. Používa distribuovaný algoritmus. Plne podporovaný systém posielania správ. Je ľahko nastaviteľný a jednoducho konfigurovateľný.

Dvojica Wackamole a Spread je veľmi obľúbená, pretože je efektívna a jednoduchá. Obe nástroje sú Open Source, takže sú finančne nenáročné.

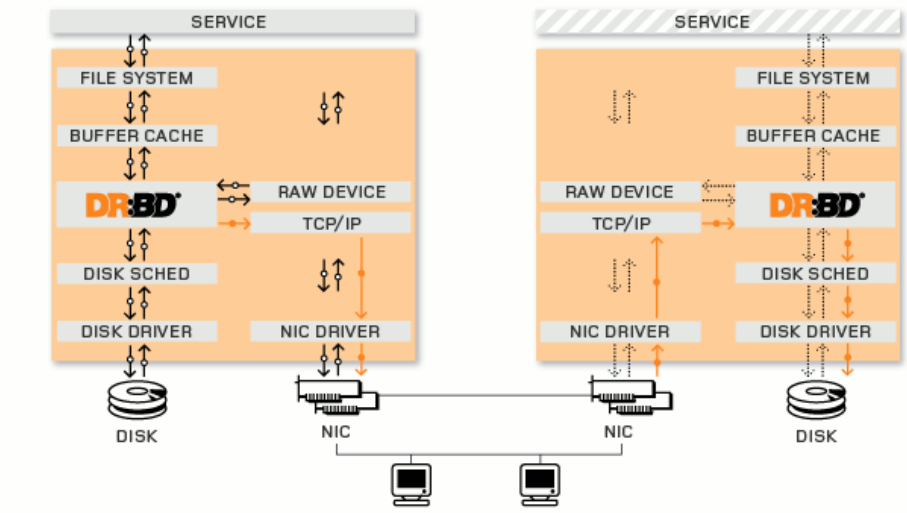
4.1.2 Dáta

Zabezpečenie redundancie služieb sme popísali vyššie, no budeme však potrebovať zabezpečiť aj dáta. Redundancia dát môže byť riešená viacerými spôsobmi. my sa však zameriame na tie najefektívnejšie v spolupráci s clusteringom serverov.

4.1.2.1 Technológia DRBD

Technológia DRBD je softvérové riešenie postavené nad fyzickými zariadeniami ako sú pevné disky a tak isto podporuje aj LVM – Logical Volume Manager. Zrkadlí bloky dát. Pomáha tak pri otázke ako uchovávať dáta pri clusterovom riešení, kde je potrebné, aby boli dáta prístupné viacerým serverom. Zapisuje však vždy iba jeden server, musíme tým pádom zabezpečiť, aby zápis zaregistrovali aj ostatné servery, ktoré sú síce neaktívne, ale ak by došlo k výpadku, musia pracovať s najnovšími dátami.

DRBD môžeme označiť ako sieťový RAID 1. Máme servery, ktorých disky sa budú zrkadliť po sieti. Takto zabezpečíme, že každý zo serverov bude mať na svojom disku najnovšie dáta.



Obrázok 13 - Ako pracuje DRBD

Na obrázku Obrázok 13 môžeme vidieť ako pracuje DRBD s dátami. Obe oranžové obrázky reprezentujú servery zapojené v HA Clustri. Hore nápisom service je reprezentovaná niektorá zo služieb, ktorá pracuje s dátami. Na obrázku sú znázornené klasické komponenty Linuxového serveru a čierne šípky reprezentujú tok dát medzi týmito komponentmi. Oranžové šípky znázorňujú tok dát ako ich zrkadlí technológia DRBD z hlavného serveru na záložný.

Možností ako DRBD zrkadlí dáta je viacero:

- real time – zrkadlenie sa deje nepretržite, zatiaľ čo aplikácia zapisuje dáta na disk
- transparentne – aplikácie nevedia o tom, že ukladajú dáta na viacero diskov, ostávajú tak v domnení, že zapisujú iba na jeden disk
- synchrónne – aplikácia, ktorá zapisuje je oboznámená, že zápis skončil, až keď sú dáta zapísané na oboch serveroch – hlavnom aj záložnom
- asynchrónne – aplikácia, ktorá zapisuje, je oboznámená, že zápis skončil hneď potom ako sú dáta lokálne zapísane na disk, ešte predtým, než sa zapíšu na záložný server

Jadro DRBD je implementované ako modul do linuxového jadra. DRBD vytvára špecifické virtuálne blokové zariadenie, čím je vlastne technológia situovaní „blízko“ vstupno-výstupného zariadenia systému. Tento princíp robí z DRBD pomerne flexibilnú a mnohostrannú technológiu, vhodnú pre zabezpečenie vysokej dosiahnuteľnosti akýchkoľvek služieb.

Nevýhodou tejto technológie je však, že ňou dokážeme pokryť iba 2 serverový cluster. Sklamaním taktiež je, že nevie odhaľovať poškodené dáta na súborových systémoch.

4.1.2.2 NAS + NFS

Ďalším možným riešením ako uchovávať dáta, je spojiť technológie NAS a NFS. Kde technológia NAS (Network Attached Storage) predstavuje úložisko pre dáta na súborovej úrovni. NFS (Network File-System) je sieťový súborový systém.

NAS je počítač pripojený do siete, ktorý slúži ako úložisko dát pre ostatné zariadenia v sieti, napríklad cluster. Väčšinou sa na takejto NAS jednotke nespúšťajú žiadne iné služby. Takýto počítač nie je stavaný, aby bol hlavným serverom pre niektorú zo služieb. Má pripojených čo najmenej zariadení a len to najnutnejšie. Počítač je spravovaný po sieti cez niektorý protokol, napríklad HTTP.

Jednotka NAS nepotrebuje operačný systém plný rôznych funkcií ako si vyžadujú servery. Používa sa tu veľmi osekaný operačný systém, napríklad FreeNAS, Open Source riešenie pre NAS jednotky, ktorý je postavený na základe operačného systému FreeBSD. Ďalšími riešeniami sú, pre úplnosť, napríklad OpenFiler a TurnKey súborový server založený na Ubuntu operačnom systéme. NAS sa skladá z väčšieho množstva pevných diskov zapojených technológiou RAID.

Najväčším záporom pri použití NAS je, že takýto systém je limitovaný svojou hardvérovou konfiguráciou. Preto sa môže stať, že pri chybe hardvéru budú všetky dáta nedostupné. Tomuto vieme predísť tak, že aj jednotlivé NAS jednotky spojíme do clusteru, ktorý sa bude medzi sebou zrkadliť, alebo môžeme využiť možnosť, že sa dáta budú zapisovať na rôzne NAS jednotky. Dosiahneme tým vlastne technológiu RAID cez sieť, čím zabezpečíme, že aj pri výpadku jednej NAS jednotky nestratíme všetky dáta. Takýto clusterovaný NAS systém využíva distribuovaný súborový systém, ktorý simultánne pracuje na viacerých počítačoch, napríklad GlusterFS.

Pri serverových riešeniach na báze Open Source využíva NAS jednotka súborový systém NFS, ktorý je najrozšírenejší sieťový súborový systém na unixových, resp. linuxových strojoch.

NFS je sieťový súborový systém navrhnutý tak, aby umožňoval pripojiť na k počítaču diskovú partíciu zo vzdialeného počítača tak, akoby bola pripojená lokálne. Dosahuje sa tým rýchle zdieľanie dát po sieti, ktoré sa navonok javí ako lokálne.

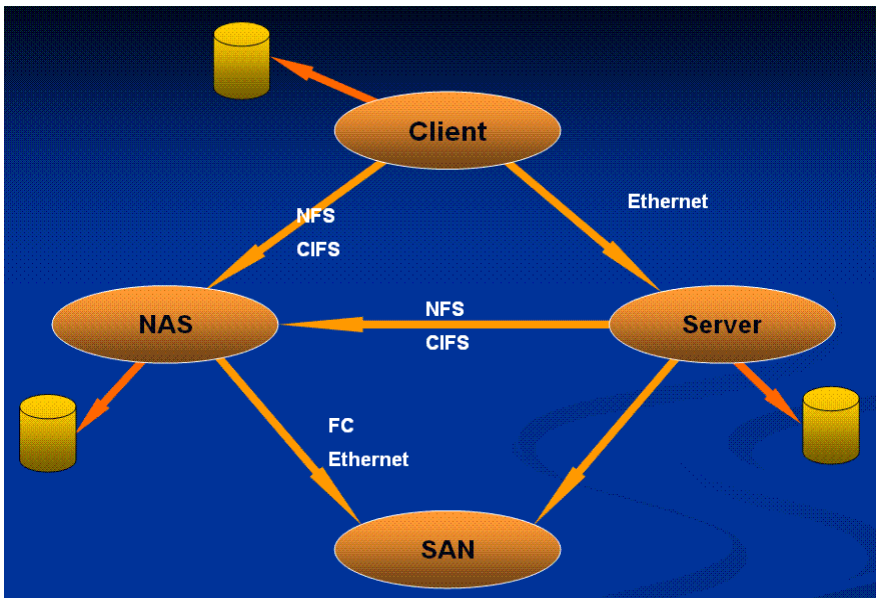
4.1.2.3 SAN

SAN (Storage Area Network) je technológia, ktorá umožňuje pripojiť externé diskové zariadenia, ako napríklad diskové polia, páskovacie jednotky a iné, k serverom tak, že sa javia ako lokálne pripojené pre operačný systém. SAN je komplexné riešenie prístupu k diskovému poľu. Je však pomerné drahé a náročne, preto ho využívajú iba veľké, väčšinou, korporátne spoločnosti.

Nad technológiou SAN je súborový systém, ktorý poskytuje súborovú abstrakciu. Takýto súborový systém sa nazýva súborový systém zdieľaného disku.

Historicky, datacentrá najprv vytvárali tzv. ostrovčeky SCSI diskov(diskových polí), ktoré boli priamo pripojené. Tieto ostrovčeky boli dedikované pre jednotlivé aplikácie a boli viditeľné ako virtuálne pevné disky (LUNy). SAN v podstate zgrupuje takéto ostrovčeky pomocou vysoko rýchlostnej siete. Aby pri zapisovaní na LUNy nedošlo ku korupciu dát zapisovaním viacerých počítačov naraz, sú potrebné SAN súborové systémy, keďže daný ostrovček sa javí pre operačný systém ako lokálne pripojený virtuálny disk.

Rozdiel medzi technológiou SAN a NAS spočíva v tom, že NAS sa pre operačný systém javí ako vzdialená jednotka, čo vlastne aj je. Potrebuje k tomu sieťový súborový systém. SAN sa však pre operačný systém javí ako lokálne pripojené úložisko dát. Napriek tomuto rozdielu môžeme tieto technológie spojiť do SAN-NAS Hybrid ako ukazuje Obrázok 14



Obrázok 14 - SAN-NAS Hybrid

4.1.2.4 SeznamFS

SeznamFS je distribuovaný súborový systém založený na FUSE, ktorý všetky svoje operácie zapisuje do binárneho logu. Tento binárny log sa potom presúva k záložnému serveru v clustri a tým je zabezpečené, že sa všetky operácie zápisu, ktoré boli vykonané na hlavnom serveri, vykonajú aj na záložnom serveri. Táto metóda sa všeobecne dá nazvať replikácia. Je možné využiť replikáciu master-master, čo znamená, že ktorýkoľvek server môže byť hlavný. To je následok toho, že každý server má vlastné serverové ID pri písaní binárneho logu.

Momentálne je dostupná verzia 0.2.x a balíčky sú zatiaľ dostupné iba pre linuxovú distribúciu Debian GNU/Linux.

SeznamFS je zložený z 2 komponentov a to zo samotného súborového systému a z administrátorskej konzoly. Celý súborový systém sa delí na vlákna nasledovne:

- worker vlákna – sú vytvorené FUSEom a obsluhujú FUSE systémové volania ako sú čítanie/zapisovanie z/do úložného miesta a taktiež do binárneho logu
- master vlákna – hlavné vlákna, ktoré počúvajú na určenej adrese a porte, obsluhujú požiadavky záložného serveru a administrátorskej konzoly
- slave vlákno – vlákno záložného serveru pripojené na hlavný server, číta z binárneho logu hlavného serveru aké operácie boli vykonané a vykoná ich lokálne, taktiež ich zapíše do lokálneho binárneho logu

Administrátorská konzola je jednoduchá základná konzola pre prácu so serverom. Obsahuje základne príkazy, ktorými dokáže zobraziť status replikácie záložného serveru, naštartovanie a zastavenie záložného serveru, prekonfigurovanie hlavného a záložného serveru a rekonfiguráciu súborového systému.

4.1.3 Databázový cluster

Dnes už takmer každá aplikácia využíva na ukladanie svojich dát databázy. Nebolo by efektívne držať na každom serveri aj databázový server. Preto sme sa rozhodli vytvoriť výkonný databázový cluster kde budú uchovávané všetky databázy. Použitá technológia bude Open Source MySQL.

4.1.3.1 MySQL replikácia

MySQL replikácia je podporovaná od verzie 3.23.15, dnes už systémy reálne pracujú s verziou MySQL 5.1. Táto replikácia prebieha asynchrónne, kde jeden server vystupuje ako hlavný a jeden alebo viac serverov ako záložný.

Princíp MySQL replikácie je podobný ako pri replikovaní dát na súborovom systéme SeznamFS. Vytvára sa teda binárny log súbor, ktorý obsahuje všetky zmeny, ktoré sa udiali na hlavnom systéme. Všetky zmeny sa udejú taktiež aj na záložných serveroch. Táto metóda je výhodná z dôvodu, že pri zmene niektorej databázy, nemusíme prekopírovať kompletne celú databázu na záložný server, ale iba sa replikujú zmeny, čím sa zvyšuje rýchlosť a znižuje záťaž jednotlivých serverov.

Výhodou replikácie je, že pri problémoch na hlavnom serveri ho môžeme kedykoľvek odstaviť a ďalej používať záložný server. Replikácia ponúka taktiež možnosť rozdeliť záťaž na jednotlivé servery. Avšak všetky operácie zápisu a teda write/update sa dejú iba na hlavnom serveri, no všetky ostatné operácie získavania informácií z databázy môžu byť vykonávané záložnými servermi, čím sa zníži záťaž hlavného serveru a zvýši sa tak odozva na požiadavky aplikácií.

Výhodou replikácie je taktiež to, že pri zálohovaní databáz môžeme využiť záložný server a môžeme nechať hlavný pracovať. Zablokujeme komunikáciu medzi hlavným a záložným serverom, vykonáme zálohovanie a komunikáciu povolíme. Servery sa navzájom zosynchronizujú a máme zálohu bez výpadku. Samotná replikácia sa za zálohovanie považovať nedá, keďže pri vykonaní deštruktívnych operácií na hlavnom servery, sa tieto operácie prejavia aj na záložných serveroch.

4.2 Zálohovanie dát

Je dôležité aj pri clusterových riešeniach zaoberať sa otázkou zálohovania. Samotné clustery, ktoré sme navrhli využívajú technológiu RAID, ktorá pri chybe niektorého z diskov neznamená stratu dát. Ďalej využívame HA clustre, kde jednotlivé servery sú takmer identické a pracujú s tými istými dátami. Pri výpadku jedného serveru teda neprídeme o žiadne dáta, maximálne tak pár sekúnd až minútu staré. Pri technológií NAS taktiež využívame clustering a teda túto NAS jednotku zdvojujeme práve kvôli tomu, aby pri výpadku NAS sme neprišli o dáta a nespôsobili tak výpadok. Pre dáta uložené v databázach využívame databázový cluster, ktorý stále obsahuje najnovšie informácie. Avšak replikácia, ktorá prebieha na databázovom clustri nie je plnohodnotné zálohovanie, prebieha zálohovanie všetkých databáz zo záložných serverov pomocou nástroja „mysqldump“, čo je integrovaný nástroj MySQL serverov. Je určený práve na zálohovanie a prenos databáz. Databázy sa pomocou tohto nástroja môžu zálohovať v ľubovoľnom intervale, kedy sa zamkne celý databázový server na záložnom serveri a vykoná sa záloha. Zamknutie je dôležité, aby sa nepoškodili dáta. Tieto zálohy sa následne po lokálnej sieti prenesú na zálohovací cluster, kde budú servery s vysokým počtom diskov určené iba pre tento účel a to uchovávať zálohy. Dôležité dáta sa taktiež môžu zálohovať obdobne ešte aj na zálohovací cluster ak je to potrebné. Tu by sme využili jednoduchý ale výkonný nástroj „rsync“, ktorý jednoducho synchronizuje dáta. Obsahuje však kontrolu synchronizovaných dát a tým zabezpečí, že synchronizuje iba súbory, ktoré boli zmenené od poslednej synchronizácie.

4.3 Zhodnotenie

V tejto kapitole sme si popísali aké technológie pre clustering sa dajú využiť v malom až stredne veľkom datacentre. Zamerali sme sa predovšetkým na riešenia Open Source, keďže celé dátové centrum sme navrhli na linuxových, prípadne unixových operačných systémoch. Riešenia popísané v tejto kapitole sú bežné používané malými a strednými IT firmami. Riešení sme uviedli viac, z dôvodu, že mnohé z nich sú kombinovateľné a tak dokážeme dosiahnuť lepšiu optimalizáciu pri návrhu clusteru pre nejakú problémovú oblasť.

5. Smerovacie protokoly

Jedným z dôležitých aspektov týkajúcich sa dostupnosti je aj schopnosť siete rýchlo sa adaptovať pri výpadku komunikačnej cesty. O rýchlu konvergenciu siete sa starajú smerovacie protokoly. V tejto kapitole porovnáme dva známe smerovacie protokoly OSPF a RIPv2.

5.1 Open shortest path first (OSPF)

OSPF vnútrodoménový stavový (interior link-state) smerovací protokol. Hlavnou ideou OSPF je schopnosť každého OSPF smerovača určiť cestu a ohodnotenie cesty ku všetkým smerovačom . Inak povedané, každý smerovač má pohľad na obraz celej topológie, čo mu umožňuje ľahko a presne nájsť najlepšiu cestu ku každému smerovaču.

Smerovače spolu komunikujú prostredníctvom LSP správ (link state packets), ktoré obsahujú tieto hlavné informácie:

- Unikátny identifikátor smerovača
- Zoznam priamo pripojených susedných smerovačov a ohodnotenie príslušných ciest
- Poradové číslo(na identifikáciu najaktuálnejších správ) a životnosť správy
- Autentifikačné informácie
- Informácie o hierarchií, prerozdeľovaní záťaže a kontrolný súčet

Po zozbieraní správ od všetkých smerovačov si vytvorí alebo aktualizuje svoj vlastný pohľad na topológiu. Najkratšie cesty určuje na základe Dijsktrovho algoritmu najkratšej cesty. Aby nedošlo k zahlteniu siete informačnými správami, smerovač nikdy neprijme LSP správy, ktoré v minulosti poslal.

Pre skontrolovanie o overenie schopnosti OSPF vysporiadať sa s výpadkami v sieti, sme nasimulovali experiment v jednoduchej sieti s pár smerovačmi a prepínačmi. Po nakonfigurovaní a odskúšaní siete sme nasimulovali výpadok jednej z liniek (vypnutím jedného sieťového rozhrania na jednom zo smerovačov). Medzi komunikujúcimi počítačmi sme spustili nekonečný ping a na smerovačoch sme pozorovali ako OSPF zareaguje. V smerovacej tabulke sa objavila fungujúca trasa za cca 2 sekundy.

OSPF je teda dobrú smerovací protokol a je vhodný pre takéto podnikové siete. Okrem mnohých dobrých vlastností, treba vyzdvihnúť schopnosť detekcie chyby v sieti a následné rýchle zotavenie siete. Treba ale myslieť na niekoľko bezpečnostných záležitostí, ktoré treba vyriešiť pred nasadením v podniku.

5.2 Routing Information Protocol (RIP)

RIP je vnútrodoménový vektorový (interior distance-vector) protokol založený na Bellman-Fordov algoritme. Hlavnou ideou RIP protokolu je schopnosť každého RIP smerovača vytvárať vlastné jednorozmerné pole, kde sú uložené počty skokov na dosiahnutie ostatných smerovačov v sieti. Pri ohodnocovaní ciest je protokol OSPF efektívnejší, keďže zohľadňuje aj rýchlosť linky.

RIP smerovače si vymieňajú informácie o ich priamo pripojených susedoch. Na základe týchto správ si budujú smerovaciu tabuľku. Ak z niektorej správy zistia kratšiu cestu k niektorému zo smerovačov, aktualizujú si smerovacie tabuľky. Keď smerovač deteguje výpadok nastaví tejto smerovacej ceste vzdialenosť na nekonečno.

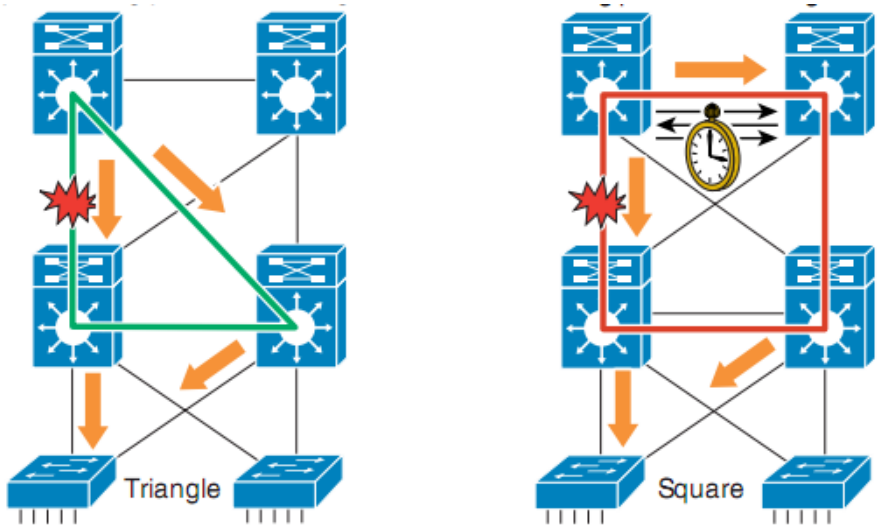
Ideálne si všetky smerovače nakoniec vymenia smerovacie aktualizácie a zostavia presnú smerovaciu tabuľku. Avšak ak je sieťová topológia navrhovaná s redundanciou, môžu vzniknúť problémy pri aktualizácií smerovacích tabuliek po vypadnutí jednej z liniek. Na predchádzanie týchto problémov existuje viacero riešení ako napr. count-to-infinity, alebo split-horizon. V klasickej verzií protokolu RIP bolo viacero nedostatkov a obmedzení. Preto bola vydaná druhá verzia protokolu RIPv2.

Podobný test na rovnakej topológií ako pri OSPF sme vykonali aj pre smerovací protokol RIP. Identifikácia chyby v sieti a následné zotavenie v sieti prevýšilo 21 sekúnd.
RIPv2 je široko dostupný smerovací protokol, ale neoslňuje svojimi schopnosťami pri problémoch v sieti ako OSPF.

5.3 Optimalizácia sieťovej topológie pre smerovacie protokoly

Smerovacie protokoly tretej vrstvy sú najčastejšie nasadzované medzi dvoma kmeňovými(core) vrstvami sietí alebo medzi kmeňovou a distribučnou vrstvou siete. Smerovanie na prístupovej (access) vrstve nebýva obvyklé. Ich hlavnou úlohou v hierarchickom návrhu siete je presmerovanie komunikácie okolo poškodenej linky alebo uzla v sieti.

Pri stavaní väčších redundantných sietí je veľkou výhodou zapájanie jednotlivých uzlov do trojuholníkovej topológie miesto štvorcovej.



Obrázok 15 - Porovnanie trojuholníkovej a štvorcovej topológie

Rozhodli sme sa teda pre trojuholníkové topológie s rovnocennými cestami do všetkých redundantných uzlov, čím sme sa vyhli nedeterministickej konvergencií. Po vypadnutí linky sa cesta zaznačí ako nepoužiteľná a komunikácie bude presmerovávaná do alternatívnej cesty, no s rovnakým ohodnotením cesty.

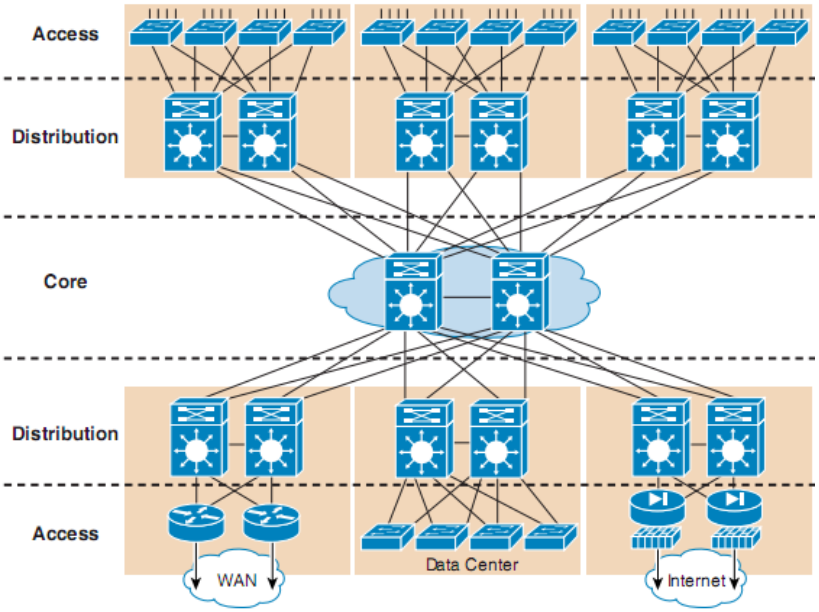
Použitie trojuholníkovej topológie je odporúčané, no často sa používa aj štvorcová topológia. Je to spôsobené buď limitovaným počtom fyzických pripojení alebo nižšou nákupnou cenou. V tomto prípade ale nie je možné získanie rovnako deterministickej konvergencie pri výpadku linky, a teda sieť nebude optimalizovaná pre vysokú dostupnosť.

6. Sieťová topológia

6.1 Hierarchický sieťový model

Hierarchický sieťový model je model vytvorený spoločnosťou Cisco v roku 1999 (pozri obrázok). Umožňuje budovať modulárne topológie použitím škálovateľných stavebných blokov, ktoré sú ľahko prispôsobiteľné potrebám rozvíjajúceho sa podniku. Modulárny dizajn robí sieť ľahko škálovatelnú, ľahko pochopiteľnú a sieťou, v ktorej je riešenie problémov jednoduchšie vďaka deterministickým vzorom sieťového toku.

Hlavné stavebné bloky sú:
- Prístupová vrstva (access)
- Distribučná vrstva (distribution)
- Chrbticová vrstva (core / backbone)



Obrázok 16 - Hierarchický sieťový model

6.1.1 Chrbticová vrstva

V typickom hierarchickom modely sú jednotlivé stavebné bloky prepojené pomocou tejto vrstvy. Tvorí hlavnú kostru celej siete a poskytuje hlavnú konektivitu pre ostatné stavebné bloky, a preto musí byť nesmierne odolná, pružná a rýchla. Súčasné hardvérovo akcelerované systémy majú potenciál doručovať komplexné služby rýchlosťou na úrovni káblov, no čo sa týka chrbticovej siete, tak by sme sa mali radiť pravidlom „Niekedy je menej viac“. Preto by mali mať zariadenia na tejto vrstve minimálnu konfiguráciu.

6.1.2 Distribučná vrstva

Táto vrstva oddeľuje prístupovú a chrbticovú sieť a poskytuje niekoľko výhod:

- agreguje uzly s prístupovej vrstvy, čím chráni chrbticovú vrstvu pred veľkou koncentráciou portov
- Logicky izoluje a ohraničuje prístupovú vrstvu pred prípadnými poruchami a výpadkami
- Poskytuje prerozdeľovanie záťaže a dvojité pripojenie na chrbticovú vrstvu, čo v prípade výpadku linky alebo uzla rezultuje v rýchlejšiu deterministickú konvergenciu
- Poskytuje prioritizáciu niektorých komunikačných tokov (Quality of service)
- Zvyčajne vystupuje ako dvojica L3 prepínačov, a teda prístupovú vrstvu môže prepínať na druhej vrstve, pričom chrbticovú aj na tretej

6.1.3 Prístupová vrstva

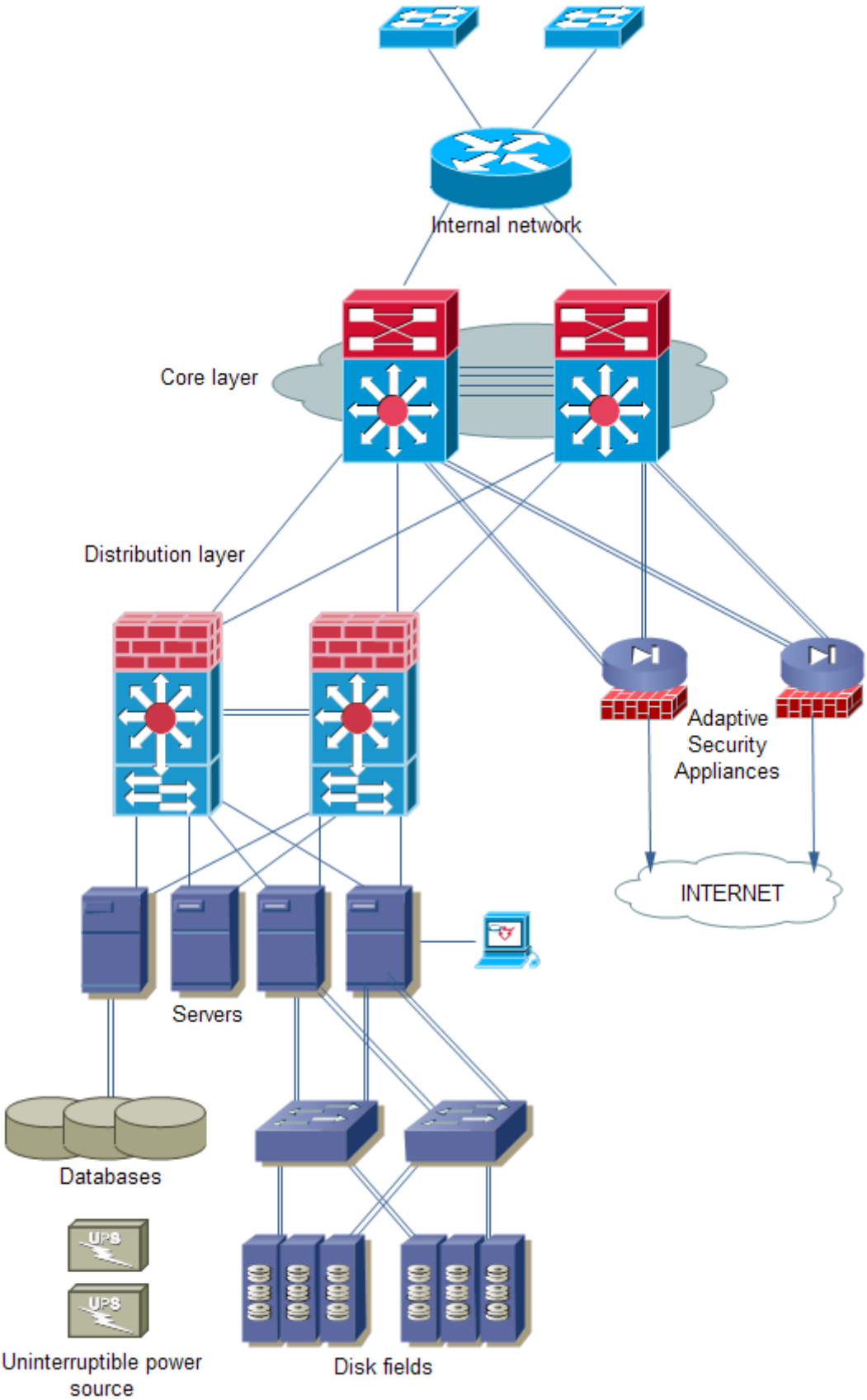
Táto vrstva je prvý vstup do siete pre hraničné zariadenia ako počítače, tlačiarne, IP telefóny. Prepínače v tejto vrstve sú kvôli redundancií pripojené na obidva distribučné prepínače. Keďže v našej práci túto vrstvu neuvažujeme a pripájame naše datacentrum priamo na distribučnú vrstvu tak ju nebudem hlbšie rozoberať.

6.2 Naša topológia

Pri návrhu sme vychádzali s hierarchického sieťového modelu, keďže v budúcnosti plánujeme ďalšie rozširovanie a zmohutňovanie nášho datacentra. To si bude vyžadovať kvalitnú dostupnú chrbticovú sieť a výkonné zariadenia na distribučnej vrstve. Pre dodržanie nepísaných štandardov vysokej dostupnosti sme siahli po viacnásobnej redundancií, ako bolo už spomenuté v niekoľkých častiach našej práce.

V topológií dominuje niekoľko redundantne zapojených prepínačov schopných prepínať na viacerých vrstvách modelu OSI a aplikačné servery s externým úložiskom dát. Pre zachovanie bezpečnosti sme do topológie zahrnuli dvojicu Cisco ASA zariadení starajúcich sa o komplexnú ochranu siete pred útokmi zvonku(zahrňujúce PIX firewall a IPS) a dvojicu softvérových firewallov na zariadeniach vo vnútri siete pred vstupom do dátového centra.

Výsledok nášho snaženia a spojením všetkých technológií spomenutých na predchádzajúcich stranách sme vytvorili topológiu nakreslenú na obrázku Obrázok 17



Obrázok 17 - Naša topológia

7. Výber hardvéru

Existuje mnoho serverových riešení, ktoré sú výkonné a ponúkajú vysokú spoľahlivosť. V našej práci sa však zameriame na tie najdostupnejšie, teda riešenia od firiem Intel a AMD. Ostatné riešenia od renomovaných firiem(IBM, HP,SUN,...) sú príliš zložité a nákladné na použitie a správu v stredne veľkom dátovom centre.

Obidve firmy ponúkajú porovnateľný výkon, vzhľadom na cenu a približne rovnaké možnosti hardvérovej akcelerácie virtualizačných technológií. Preto sú pre nás obidve platformy rovnocenné, aj keď je všeobecne známe, že AMD ponúkne za rovnakú cenu výkonnejšie riešenie, resp. viac serverov, ktorých kombinovaný výkon bude vyšší, ako výkon jedného servera od firmy Intel.

Dôležitejšie ako platforma servera je pre nás škálovateľnosť a počet podporovaných rozhraní. Jedná sa hlavne o podporu diskových polí a rôznych technológií pripojenia diskov, sieťové rozhrania a maximálna veľkosť operačnej pamäte. (1) (2)

7.1 Dátové úložiská

Prístupy k realizácii diskového poľa sa líšia hlavne fyzickým pripojením, či už k sieti alebo k samotným serverom. Existujú tri hlavné druhy realizácie diskových polí :

- DAS
- NAS
- SAN

7.1.1 DAS (Directly attached storage)

Každý server má svoje disky a na zaručenie bezpečnosti dát prebieha synchronizácia medzi servermi v clusteri v určitom intervale. Na synchronizovanie je obyčajne použité iné sieťové rozhranie, ako na komunikáciu v rámci lokálnej siete (3). Je to najmenej nákladné riešenie. Hlavné protokoly a rozhrania sú:

- ATA
- SATA
- eSATA
- SCSI
- SAS
- Fibre Channel

7.1.2 NAS (Network attached storage)

Je to vlastne server, ktorý obsahuje mnoho diskov a servery sa k nemu pristupujú cez sieť. Existujú kompletné riešenia od rôznych firiem. Prístup k dátam je na úrovni súborov, ktoré dátové úložisko ponúka serverom. Nevýhodou je, že v prípade zlyhania úložiska samotného sa všetky dáta stanú nedostupné a taktiež to, že toto úložisko má obmedzené zdroje (operačná pamäť, procesor,..) a pri veľkom počte klientov, môže dôjsť k ich vyčerpaniu. Hlavné protokoly na prístup k úložiskám typu NAS:

- CIFS
- NFS
- FTP
- SFTP
- HTTP
- UPnP

7.1.3 SAN (Storage area network)

Fyzicky môže vyzerať ako predchádzajúce riešenie, ale jeho princíp je úplne iný. Servery k diskom v dátovom úložisku pristupujú na blokovom princípe. Teda diskové pole sa tvári, akoby bolo pripojené priamo ku každému serveru. Samozrejme je tu potreba použitia špeciálneho súborového systému, aby nedochádzalo k strate integrity dát. Technológie na realizáciu tohto dátového úložiska:

- ATA over Ethernet (AoE)

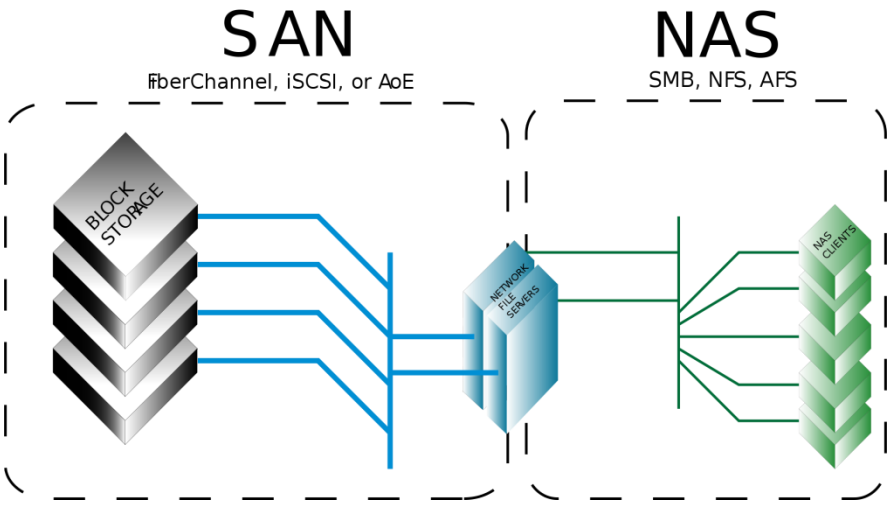
  Sieťový protokol na prístup k diskom SATA cez Ethernet. Na rozdiel od iSCSI, AoE nepoužíva vyššie sieťové vrstvy ako IP alebo TCP, čo ho robí síce nesmerovateľným, ale o veľa jednoduchším oproti iSCSI.
- Fibre Channel Protocol (4)

  Je to technológia primárne používaná na budovanie veľkých a drahých dátových centier. Môže používať metalické, ale vo väčšej miere sa používajú optické rozvody s rýchlosťou do 20 Gb/s.
- Fibre Channel over Ethernet (FCoE)

  Podobne ako AoE, tak aj FCoE nepoužíva TCP alebo IP na svoj transport. Principiálne je to vlastne iba zapuzdrenie Fibre Channel rámcov do Ethernet rámcov. Výhodou je hlavne úspora čo sa týka nákladov na kabeláž a špeciálnych optických prepínačov (5).
- ESCON
- HyperSCSI,
- iFCP

- iSCSI,

  Prenos SCSI príkazov cez Ethernet sa deje pod protokolom TCP. Znamená to, že prevádzka môže byť smerovaná cez Internet, a dovoľuje vytvárať dátové úložiská, ktoré sú od seba fyzicky vzdialené. iSCSI síce nevyžadujem žiadnu ďalšiu infraštruktúru (okrem už existujúcej sieťovej infraštruktúry), ale výkon iSCSI môže byť degradovaný, pokiaľ nie je zaručená dostatočná šírka pásma pre potreby diskových polí, poprípade oddelená podsieť.



Obrázok 18 - Rozdiel medzi zapojením SAN a NAS

7.1.4 RAID

Samozrejme v rámci diskového poľa býva realizované aj takzvané pole RAID (Redundant Array of Independent /Inexpensive Disks). Pole RAID vyžaduje minimálne 2 disky. Môže byť v rôznych konfiguráciách, pre nás však budú dôležité:

7.1.4.1 RAID 0 (Stripping)

Dáta sa rozdeľujú na dva disky, preto je toto zapojenie veľmi rýchle, ale pri strate jedného disku, sú všetky dáta nenávratne preč, keďže dostupná je iba polovica každého súboru.

7.1.4.2 RAID 1 (Mirorring)

Toto zapojenie je výhodné pre bezpečnosť dát. Disky v poli sú identické a pri strate niektorého z diskov sú dáta stále dostupné na disku, ktorý bol kópiou chybného disku.

7.1.4.3 RAID 1+0 (Mirorred strip)

Je kombináciou predchádzajúcich zapojení a vyžaduje 4 disky. Dáta sa rozdeľujú na dve časti ale každá časť sa zapíše na dva disky, teda na RAID 1. Kombinácia výkonu a bezpečnosti.

Vybrali sme kombináciu dvoch riešení aby bola zaručená rozširovateľnosť dátového centra. Každý server bude mať vlastný pevný disk, ktorý sa bude synchronizovať so sekundárnym strojom v clusteri. Okrem toho sa tu bude nachádzať diskové pole SAN, ktoré bude prepojené so všetkými servermi cez gigabitový prepínač a ďalej pripojený na iSCSI SAN úložisko .

Druhým konkurenčným riešením je iSCSI cez 10 gigabitový Ethernet, ktoré používa už stávajúcu sieťovú infraštruktúru. V našom datacentre by to znamenalo použitie optickej infraštruktúry cez 10 gigabitový Ethernet. Zariadenia pre tento typ prepojenia sú však ešte príliš drahé. Jej výhody sú však nesporné, medzi inými aj transport nad protokolom TCP, teda použitie WAN siete na prenos informácií. (6)

Fibre Channel riešenie bolo taktiež zamietnuté kvôli svojej vysokej cene, ktorá sa blíži cene 10 gigabitového Ethernet riešenia, ale nebolo by použiteľné pre LAN a WAN prevádzku ako Ethernet riešenie.

7.1.5 Hardvér pre Fibre Channel riešenie:

Fibre Channel prepínač: Brocade 300
24x 8Gb Fibre channel liniek
Cena: okolo 7 000 €

Fibre channel karty do serverov: Emulex LPE 12000
8Gb Fibre Channel
Pripojenie PCIe
Cena: 1 299 $

Fibre channel dátové úložisko: Dell/EMC CX4-120
Kapacita až 72 TB
4 porty 8Gbit alebo 4Gbit Fibre Channel a 4 porty 10Gbit iSCSI
Cena: neznáma

7.1.6 Hardvér pre 10Gbit konvergované iSCSI riešenie:

10 Gb prepínač: Dell PowerConnect 8024F
24 x 10Gb Ethernetových portov
Smerovanie na 3 vrstve v plnej rýchlosti
Webový manažment
Spaning tree protokoly
Link aggregation
Podpora IPv6 smerovania
Cena: 8 699 $

10 Gb sieťová karta: Broadcom NetXteme II 57711 10GbE
Dva konektory Konektor SPF+
Pripojenie PCIe 8x
Rýchlosť 10Gb/s
Cena: 709 $

Diskové pole : Dell EqualLogic PS6010E
Kapacita až 16 TB
Dva 10Gb SPF+ porty
Podpora RAID 5, RAID 6, RAID 10 and RAID 50
10 Gb/s pripojenie
Podpora IPv4, IPv6
Cena: neznáma

7.1.7 Použité iSCSI riešenie

1 Gb prepínač: Dell PowerConnect 2824
24x 1Gb Ethernet
Podpora STP
Podpora VLAN
Cena : 179 $

Sieťová karta: Intel® Gigabit ET NIC, Quad Port, Copper, PCIe-4
4x 1Gb Ethernet port
PCIe pripojenie
Cena: 529 $

iSCSI dátové úložisko: Dell PowerVault MD3000i
Kapacita až 30TB
2x 1Gb Ethernet port
Cena: 4 819 $ (vrátane 4x 500GB 7.2K RPM SATA)

7.2 Servery

Servery v datacentre budú zastávať tri hlavné roly (7):

- Clusterový server
  - Priorita je výkon, teda veľa RAM a silný procesor
- Databázový server
  - Veľa RAM, v ktorej budú uložené databázy
- Zálohovací server
  - Veľké diskové pole, kam sa budú zálohovať niektoré kritické súbory

Preto boli vybrané tri rôzne konfigurácie od výrobcu Dell (8):

Clusterový server: Dell PowerEdge 2970
Veľkosť 2U
2xQuad Core AMD Opteron 2374HE 2.2GHz 4x512K Cache
8GB (4x2GB), 800MHz
2x500GB 7.2K RPM, HotPlug Hard Drive
Intel Gigabit ET NIC, Quad Port
DVD-RW
Cena: 2 192 $

Databázový server: PowerEdge 2970
Veľkosť 2U
2xQuad Core AMD Opteron 2374HE 2.2GHz 4x512K Cache
16GB (4x4GB), 800MHz
2x500GB 7.2K RPM, HotPlug Hard Drive
Intel Gigabit ET NIC, Quad Port
DVD-RW
Cena: 2 395 $

Zálohovací server: Dell PowerEdge R510
Veľkosť 1U
1x Intel Xeon E5502, 1.86Ghz, 4M Cache
4GB (4x1GB), 1333MHz
4x 1TB 7.2K RPM SATA 3.5" Cabled Hard Drive
Intel Gigabit ET NIC, Quad Port
Cena: 3 409 $

7.2.1 Vybavenie siete

Distribučný prepínač Cisco Catalyst 4948
Výška 1U
L3 prepínač
Priepustnosť 96 Gbps
Až 48 10/100/1000 RJ-45 Ethernet portov
Procesor s taktom 266 MHz
Cena: 4 644 $

Prepínač v jadre siete: Cisco Catalyst 4948 10 Gigabit Ethernet
Výška 1U
L3 prepínač
Priepustnosť 136 Gbps
Až 48 10/100/1000 RJ-45 Ethernet portov
2x 10Gbit optické porty
Procesor s taktom 666 MHz
Cena: 8 499 $

Firewall: Cisco ASA -5550
Výška 1U
Priepustnosť 1.3Gbps
8x 10/100/1000 RJ-45 Ethernet portov
4x SPF porty
1x 10/100 Mbit port
4 GB RAM
Cena: 8 733$

Tienená kabeláž kategórie 5e poprípade 6

8. Zázemie datacentra

8.1.1 Elektronická požiarna signalizácia

Elektrická požiarna signalizácia (EPS) slúži na preventívnu ochranu objektov pred požiarmi tak, že opticky a akusticky signalizuje vznik a miesto požiaru. Samočinne alebo prostredníctvom ľudského činiteľa urýchľuje odovzdanie informácie o požiari osobám určeným na vykonanie protipožiarneho zásahu, prípadne uvádza do činnosti zariadenia, ktoré bránia rozšíreniu požiaru alebo priamo vykonávajú protipožiarny zásah. Zariadenie EPS má teda charakter pomocného zariadenia, ktoré je jedným z prostriedkov protipožiarneho istenia objektu.

Základná zostava EPS pozostáva z

- hlásičov požiaru,
- požiarnych slučiek,
- ústrední EPS,
- signalizačnej linky,
- doplňujúcich zariadení (signalizačné zariadenie, zariadenie diaľkového prenosu informácií, ovládacie jednotky a podobne).

Hlásiče požiaru sú prístroje, ktoré vytvárajú výstupný elektrický signál:

- samočinne, pri dosiahnutí hodnoty reakcie – samočinný (automatický) hlásič,
- vedením do činnosti osobou – tlačidlový hlásič.

Z hľadiska komunikácie s ústredňou a zapojenia do liniek možno tlačidlové a automatické bodové hlásiče rozdeliť na:

- neadresovateľné – kolektívne a
- adresovateľné – interaktívne.

Pri výbere hlásiča požiaru treba zohľadňovať:

- pravdepodobný druh požiaru,
- očakávanú veľkosť požiaru v začiatočnej fáze,
- svetlú výšku priestoru,
- podmienky prostredia a
- možnosť planých poplachov.

V závislosti od charakteru horľavých materiálov sa používajú štyri základné typy samočinných hlásičov požiaru:

- ionizačno-dymové hlásiče,
- opticko-dymové hlásiče,
- tepelné hlásiče,
- hlásiče vyžarovania plameňa.

8.1.1.1 Ionizačno-dymové hlásiče

Ionizačno-dymové hlásiče pracujú ako dymové detektory, ktoré sú citlivé na zmenu vodivosti vzduchu, vyvolanú prítomnosťou dymu. Detektorom je otvorená ionizačná komora, kde sa potrebná ionizácia dosahuje rádioaktívnym žiaričom. Hlásiče sú vhodné do prostredia, kde sa predpokladá v začiatočnom štúdiu požiaru tvorba neviditeľného i viditeľného dymu a patria medzi najúčinnejšie a najuniverzálnejšie. Nie sú vhodné do priestorov, kde vzniká dym a aerosoly pri technologickom procese. Tento typ hlásiča sa vyrába s nastaviteľným prepínačom pre tri stupne citlivosti a tiež vo vyhotovení pre priestory s nebezpečenstvom výbuchu.

8.1.1.2 Opticko-dymové hlásiče

Opticko-dymové hlásiče reagujú na nárast koncentrácie viditeľného dymu, najmä dymu svetlej farby, ktorý vzniká pri horení niektorých druhov plastov, bavlny, izolačných materiálov elektrotechnických prístrojov a pod. Používa sa na okamžitú identifikáciu tlejúcich a otvorených ohňov so vznikom dymu a je vhodný na pripojenie do kruhovej linky s ústredňami. Vyhodnocujú všetky druhy svetlých aj tmavých dymov. Odporúča sa použiť ich v kombinácii s ionizačno-dymovými hlásičmi. Ich citlivosť je pevne nastavená a nemožno ju meniť.

8.1.1.3 Tepelné hlásiče

Tepelné hlásiče reagujú samočinne na zmenu teploty okolitého prostredia. Určené sú do priestorov, kde nemožno použiť dymové hlásiče, pretože pri požiari nedôjde k vývinu dostatočného množstva dymu, resp. kde sa dym vyskytuje už z technologických dôvodov alebo kde požiar sprevádza veľký vývin tepla. Vyrábajú sa vo vyhotovení pre obyčajné prostredie a prostredie s nebezpečenstvom výbuchu. Pri voľbe hlásiča treba prihliadať na teplotu prostredia, aby nedošlo k nežiaducej funkcii hlásiča v dôsledku prekročenia maximálnej teploty pre trvalú prevádzku (60^{o}) a rýchlej zmeny teploty o viac ako + 6^{o}C za minútu. Nesmú sa umiestňovať tam, kde vy sa mohli zohrievať inými zdrojmi tepla ako je napríklad slnečné žiarenie, teplovody a podobne. Hlásič sa skladá z maximálnej časti reagujúcej na prekročenie nastavenej teploty (približne 70 ^{o}C) a diferenciálnej časti reagujúcej na rýchlosť narastania teploty okolia (približne 10 ^{o}C za l až 3 min.).

8.1.1.4 Hlásiče vyžarovania plameňa

Hlásiče vyžarovania plameňa sú vhodné do prostredia, kde sa v prípade požiaru predpokladá rýchle horenie otvoreným plameňom. Reagujú na infračervené žiarenie plameňov horiaceho dreva, plastov, alkoholu, výrobkov z ropy a podobne. Nazývajú sa taktiež hlásiče infračerveného žiarenia. Nie sú citlivé na denné a umelé osvetlenie, nereagujú na tepelné, ultrafialové, röntgenové alebo γ žiarenie. Tento typ hlásiča sa vyrába pre bežné a pre výbušné prostredie. Pre umiestnenie na potrubie pneumatickej dopravy sa používa tiež typ, ktorý reaguje na infračervené žiarenie iskier. Použitie hlásiča je obmedzené citlivosťou, vyjadrenou najväčšou dovolenou výškou umiestnenia nad stráženou plochou (≤ 25 m) a najväčšou dovolenou stráženou plochou($\leq 1000 m^{3}$).

8.1.1.5 Výber riešenia

V našom riešení sme uprednostnili adresovateľný systém elektrickej požiarnej signalizácie pred konvenčným. Tento variant má viacero výhod, napríklad presné určenie miesta požiaru, keďže ku každému hlásiču je možné priradiť názov jeho umiestnenia, či jednoduchšia a úspornejšia kabeláž.

Z hľadiska inštalácie samočinných hlásičov požiaru sme sa rozhodli pre v súčasnosti najbežnejšie riešenie v obytných budovách –optický hlásič.

Na základe týchto kritérií sme vyhľadali firmy schopné zrealizovať naše požiadavky.

Zo spoločností Securiotn Slovakia, IVTER, LITES a Jablotron sme sa rozhodli pre posledne menovanú. Keďže kvalita zariadení aj mená spoločností v danom obore sú porovnateľné, zvíťazila u nás väčšia flexibilita a možnosť zostavenia systému presne na mieru pred pevne navrhnutými realizáciami od prvých dvoch firiem a taktiež integrácia systému nepovoleného vniknutia (9).

Liters MHU 109

Maximálne 256 adresovateľných hlásičov
Určené pre stredné objekty
Cena: 32850Kč

Lites MHA 141

Tlačidlový adresovateľný hlásič
Napájaný z ústredne
Cena: 1620Kč

Lites MHG 21/161

Adresovateľný optický hlásič dymu
Reaguje na splodiny horenia na princípe detekcie rozptýleného infračerveného žiarenia
Nastaviteľná doba reakcie, citlivosť na dym
Cena: 2160Kč

Nami vybrané riešenie od firmy Jablotron

Uvažovali sme dve riešenia. Prvé decentralizované,
od firmy Jablotron.

Ústredňa JA-82 OASiS

Ponúka až 50 adries
Modulárny systém
Samotná ústredňa ponúka iba drátové vstupy
Obsahuje akumulátor
Cena: 762 €

Bezdrôtový dymový detektor JA-80S OASiS

Používa optickú komoru
Maximálna plocha pokrytia 50m^3
Výdrž približne 3 roky
Cena: 55.06 €

Hlavné mesto Slovenskej republiky Bratislava v spolupráci s Hasičským a záchranným útvarom hlavného mesta SR Bratislavy ponúka službu, ktorú plánujeme využiť v našom objekte, pult požiarnej ochrany. Jedná sa o plne automatický systém, ktorý zabezpečuje monitorovanie objektov strážených elektrickou požiarnou signalizáciou (EPS) na pulte požiarnej ochrany. Pripojenie systému EPS na pult požiarnej ochrany je riešené zariadením diaľkového prenosu (ďalej iba „ZDP“), ktoré vysiela signály zo stráženého objektu rádiom a sekundárne po telefónnej linke.

Hlavnou výhodou pripojenia na pult požiarnej ochrany je včasný zásah profesionálnej hasičskej jednotky Hasičského a záchranného útvaru hlavného mesta SR Bratislavy

8.1.2 Elektronický zabezpečovací systém

8.1.2.1 Klasické alebo bezdrôtové systémy

Klasické prvky elektrickej zabezpečovacej signalizácie (EZS) sú navzájom prepojené káblami, ktorými sa prenáša napájacie napätie a všetky informácie. Oproti tomu bezdrôtové systémy medzi sebou komunikujú rádiovo a detektory sú napájané z batérií. Spoľahlivosť a bezpečnosť oboch variantov závisí na type výrobku a nie je možné tvrdiť, že napríklad bezdrôtové systémy sú určené pre nižšie riziká. Naopak, posledné modely bezdrôtových systémov spĺňajúcich prísne európske normy pre EZS sú na takej kvalitatívnej úrovni, že za sebou nechávajú aj celú radu klasických systémov.

Aké sú hlavné prednosti klasických systémov?

Výrobky sú väčšinou lacnejšie ako bezdrôtové (pokiaľ sa nepočíta inštalačný materiál a práca) a môžu sa väčšinou kombinovať komponenty niekoľkých výrobcov v jednej inštalácii. Nie je nutné meniť batérie v detektoroch, je však potrebné vykonávať preventívne prehliadky systému.

Aké sú hlavné prednosti bezdrôtových systémov?

Samotná inštalácia je veľmi čistá (s minimom vŕtaní a sekaní) a rýchla (a teda lacná). Výsledný vzhľad interiéru potom nie je ani narušený inštalačnými lištami. Systémy sú veľmi rýchlo rozšíriteľné a dajú sa aj jednoducho odinštalovať (pokiaľ sa napr. sťahujete). Samotestujúce funkcie všetkých súčastí systému upozornia na prípadnú poruchu alebo potrebu výmeny batérií.

Mozgom každého zabezpečovacieho systému je ústredňa. Ústredňa vyhodnocuje všetky signály z detektorov a ovládacích zariadení a na základe ich analýzy a v súlade s nastavením programu rozhoduje o vyhlásení poplachu.

K 80 % vlámaní dôjde prekonaním vchodových dverí. Preto je ochrana vstupu najpodstatnejšia. Dvere by mali byť predovšetkým chránené mechanickým zámkom, aby nemohlo dôjsť k ich ľahkému otvoreniu. Samozrejme, že čím je zámok zložitejší, tým je odolnejší. O indikáciu otvorenia dverí sa postará magnetický detektor. Ten upozorní ústredňu, že došlo k otvoreniu dverí a ústredňa čaká na vypnutie systému. Pokiaľ nedôjde k vypnutiu počas nastavenej doby, dôjde k vyhláseniu poplachu narušenia objektu.

Moderné EZS sa obvykle vypínajú mocou klávesnice zadaním niekoľko-miestneho vstupného kódu alebo stlačením tlačidla diaľkového ovládača poprípade RFID kartou. Obidva systémy sú bezpečné. Možnosť zadania kódu je obvykle obmedzená na niekoľko málo pokusov než dôjde k vyhláseniu poplachu. U ovládacích kľúčeniek je zase obvykle použitý takzvaný plávajúci prenosový kód, ktorý úplne znemožňuje jeho skopírovanie.

Pre kvalitnú ochranu vnútorných priestorov pred narušiteľmi sa používajú predovšetkým infrapasívne detektory (tzv. PIR detektory). Tieto detektory sú schopné na základe analýzy teplôt v miestnosti spoľahlivo detegovať pohyb človeka v priestore. Pre rôzne aplikácie sa používajú PIR detektory s odlišnou charakteristikou, vhodné napríklad pre štandardné priestory, pre dlhé úzke chodby alebo detektory odolné voči menším živočíchom.

Nemenej dôležitú úlohu zohráva napájanie systému a jeho zálohovanie v prípade výpadku sieťového napätia. Z toho dôvodu je potrebné používať kvalitné akumulátory s vhodnou veľkosťou kapacity, adekvátnou veľkosti systému (10).

Ústredňa *Honeywell galaxy GD-264*

Vhodná pre stredne veľké inštalácie
Pamäť 1000 + 1500 udalostí
Maximálne 256 zón
Cena: neznáma

Detektor PIR *Honeywell PIR VF 5888H*

Pre vnútorné prostredie
Dosah vejár 12m, pohľad pod seba
Cena: neznáma

Prijímač bezdrôtového signálu *Honeywell G8VF*

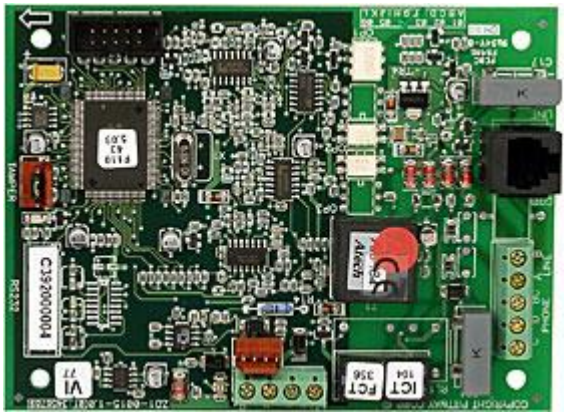
Softvérová emulácia až štyroch G8 koncerntrátorov
Podpora až 32 bezdrôtových zón
Až 30 rádiových ovládačov
Cena: neznáma

LCD klávesnica *Honeywell MK7*

Dvojriadkový displej
Slovenská lokalizácia
Cena: neznáma

Telefónny komunikátor pre PCO vo formáte ACID/SIA *Honeywell G8VF*

Telefónny komunikátor pre prenos na PCO
Pulzné/tónové vytáčanie
Cena: neznáma



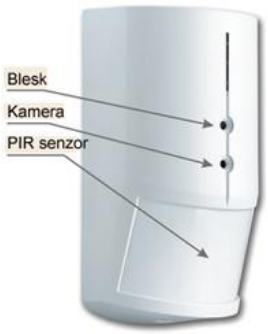
Modul pre pripojenie ústredne do TCPIP siete *Honeywell E080*

Modul pre monitoring, správu užívateľov a konfiguráciu ústrední
Pripája sa priamo na dátovú zbernicu ústredne a do siete Ethernet
Cena: neznáma

Firma Honeywell okrem už spomínaného systému ponúka aj kompletné riešenia bezpečnosti, či už požiarnej alebo zabezpečenie proti neoprávnenému vniknutiu. Sú však určené pre celé budovy a tiež informácie na stránke výrobcu sú skôr marketingové ako technické. Preto sme uprednostnili riešenie firmy Jablotron, ktoré je veľmi dobre zdokumentované.

JA-84P OASiS Bezdrôtový detektor pohybu s kamerou a bleskom

Vysoká citlivosť
Minimalizácia falošných poplachov
Kamera 160x128 bodov
Dosah blesku až 3 m
Cena: 119.65 €



Klávesnica
JA-80E OASiS

Integrovaný LCD displej
Integrovaná RFID čítačka
Cena: 77,50€

GSM komunikátor JA-80Y OASiS
Podpora hlasu a SMS
Podpora GMS/GPRS
Cena: 284.60 €

Kombinovaný komunikátor LAN + telefónna linka JA-80V OASiS
Telefonické pripojenie k centrálnemu pultu
IP pripojenie k centrálnemu pultu
Cena: 117.45 €

Modul prenosu fotiek pre ústredne JA-80Q OASiS
Prenos fotografií medzi ústredňou a detektorom s kamerou
Cena: 21.13 €

8.1.3 Rozvody elektrickej a dátovej kabeláže

Po špecifikácií komponentov, môžeme prejsť k voľbe chladenia, záložných zdrojov a generátora.

Predpokladáme začiatočný počet 50 serverov s 5 dátovými úložiskami. Tomuto faktu je potrebné aj nadimenzovať chladenie priestoru datacentra a zálohovanie.

Každý server a dátové úložisko má spotrebu zhruba 500W, samozrejme v plnom zaťažení a v špičke. Stratové teplo vyžarované odhadujeme maximálne na 250W. Dokopy je to 27 500W a 13 750W tepla. Tieto údaje ešte upravíme o ďalšie zariadenia ktoré je potrebné napájať a vyžarujú teplo(prepínače,, smerovače,..) . Teda výsledok je zhruba 31kW elektrickej energie a 16kW tepla (a to sú už tieto výsledky veľmi naddimenzované).

Datacentrum bude používať dvojitú podlahu. Jej výhody sú nielen vo využití chladenia cez tento priestor, ale aj cez vedenie káblov dátových a prúdových rozvodov touto cestou. Pri spomínanom počte zariadení, bude centrum potrebovať 5 až 7 rackových skriní.

8.1.4 Chladenie

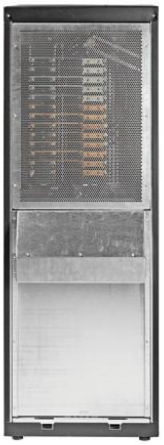
Na chladenie použijeme osvedčený systém EMERSON LIEBERT Challenger ITR, čím bude využitá dvojitá podlaha datacentra (11).

Každý rack v serverovni má nasávanie zo spodnej strany a teplý vzduch zo serverov v tej istej rade je zbieraný do chladiacej jednotky. Týchto jednotiek môže byt v rade rackov aj viac. Jednotka má veľkost klasickej rackovej skrine a je spojená s vonkajšou jednotkou umiestnenou väčšinou na streche, ktorá sa stará o výmenu tepla medzi vonkajším prostredím a chladiacim médiom.

Model Challenger ITR je schopný spracovať 23kW tepla pri prietoku vzduchu 7000m^3 /h, čo vysoko presahuje naše požiadavky, ale pre budúcu rozšíriteľnosť centra je to žiaduce. Bohužial, výrobca neudáva spotrebu systému, ale podobný systém od konkurenčného výrobcu STULZ má spotrebu rádovo 6 až 7 kW. To bude taktiež potrebné zahrnúť do požadovaného výkonu generátora.

Cenu výrobku je dostupná na požiadanie.

8.1.5 UPS



Význam batériového zdroja energie je v tom, že generátor nie je schopný nabehnúť ihneď pri prerušení dodávky energie. Preto je tu potreba systému, ktorý preklenie tento časový úsek.
My sme vybrali produkt firmy APC, a síce APC Smart-UPS VT 40kVA s výkonom 32kW/40kVA.

Má výkon 32 kW, čo však v tomto prípade postačuje, keďže klimatizácia, ani iné podporné systémy nemusia byť napájané. Tento záložný zdroj podporuje paralelné zapojenie s ďalším záložným zdrojom, čo bude pre nás výhodné pri rozširovaní centra.

Záložný zdroj sa pripája do svorkovnice medzi elektrické rozvody a napájanie jednotlivých serverov a sieťových prvkov (12).
Cena je približne **12 535.40** €

8.1.6 Generátor

Ako generátor sme vybrali produkt od renomovanej firmy Broadcrown. Pre nás je výhodný napríklad model BCJD 65-50, ktorý poskytuje 60kVA pri 220 Voltoch.

Je postavený na turbodieselovom 4 valcovom motore John Deere s vodným chladením. Spotreba motoru pri plnom výkone je 17,7 litra za hodinu. Kapacita nádrže je 155 litrov, čo znamená prevádzku centra po dobu 8 hodín, vrátane podporných systémov ako klimatizácia a zabezpečovacie systémy (13). V prípade požiadavky vyššej výdrže, je možné dokúpiť prídavné nádrže.

Cena je približne 12,198.05 $

8.1.7 Približný cenový odhad

Servery:

5x Databázový server

5 x 2395 $

5x Zálohovací server

5 x 3409 $

40x Clusterový server

40 x 2192 $

Dátové úložisko:

5x Dátové úložisko

5 x 4819 $

2x SAN prepínač

2 x 179 $

Sieťová infraštruktúra:

2x Distribučný prepínač

2 x 4 644 $

2xPrepínač v jadre v siete

2 x 8 499 $

2xFirewall

2 x 8 733$

Záložné zdroje:

1x Záložný zdroj APC

1x12 535.40 €

Záložný generátor:

1x Záložný generátor Broadcrown

1x12 198.05 $

Požiarna bezpečnosť:

1x Ústredňa JA-82 OASiS

1 x 762€

5x dymový detektor JA-80S OASiS

5 x 55.06 €

Elektronický zabezpečovací systém:

5x Snímač JA-84P
5 x 119.65 €

1x Klávesica JA-80E OASiS
1 x 77,50 €

1x GSM komunikátor JA-80Y OASiS
1 x 284.60 €

1x Kombinovaný komunikátor LAN + telefónna linka JA-80V
1 x 117.45 €

1x Modul prenosu fotiek pre ústredne JA-80Q OASiS
1 x 21.13 €

Spolu sú teda náklady na vybudovanie nášho dátového centra 160 772,6092 € s tým, že tu nie je zahrnutá cena klimatizácie, kabeláže a ani upravenie dvojitej podlahy.

8.1.8 Zhodnotenie

V tejto kapitole boli odprezentované konkrétne riešenia pre navrhované datacentrum. Vybrali sme produkty renomovaných výrobcov, čo sa podpísalo na výslednej cene. V profesionálnom nasadení sa neoplatí riskovať použitie neoverených výrobcov na úkor ceny a spoľahlivosti.

Taktiež sú niektoré systémy naddimenzované, čo sa neskôr prejaví v rozšíriteľnosti poprípade menšom zaťažení systémov.

9. Záver

V dokumente sme vypracovali návrh bezpečného datacentra s vysokou dostupnosťou a veľkým dôrazom na bezpečnosť. Odprezentovali sme niekoľko prístupov pre zvýšenie dostupnosti sietí a prišli sme k týmto záverom:

- Agregácia liniek je vhodná pre zvýšenie dostupnosti a šírky pásma iba na linkách spojujúcich dvoch hostiteľov (point-to-point)
- STP nie je vhodný kvôli pomalej detekcií porúch a následnému prispôsobeniu siete. Vhodnejšie je implementovať novšiu verziu – RapidSTP, ktorý vylepšuje obmedzenia STP
- HSRP je veľmi dobra technológia na zvýšenie dostupnosti, poskytuje rýchlu detekciu a opravu porúch v sieti, je škálovateľná a v konečnom dôsledku zvyšujú MTBF
- OSPF poskytuje rýchlejšiu detekciu výpadku linky alebo uzla než RIPv2
- OSPF ale nie je vhodná pre smerovanie na serveroch, okrem viacerých bezpečnostných rizík, môže ovplyvniť výkonnosť vypočítavaním vhodných ciest z link-state databáz

Ďalej sme si popísali aké technológie pre clustering sa dajú využiť v malom až stredne veľkom datacentre. Rozobrali sme najznámejšie riešenia ukladania dát cez sieť. Zamerali sme sa predovšetkým na riešenia Open Source, keďže celé dátové centrum sme navrhli na linuxových, prípadne unixových operačných systémoch.

Keďže sme navrhovali bezpečné datacentrum, venovali sme sa aj bezpečnosti samotných priestorov datacentra a serverovne. Preskúmali sme niekoľko zabezpečovacích systémov. Pre prípad výpadku napájania sme navrhli náhradné systémy dodávky elektrickej energie.

Na záver sme odprezentovali konkrétne riešenia pre navrhované datacentrum. Vybrali sme produkty renomovaných výrobcov a vyčíslili približný cenový odhad celého riešenia. Výsledná cena sa môže zdať ako vysoká, no v profesionálnom nasadení sa neoplatí riskovať použitie neoverených výrobcov na úkor ceny a spoľahlivosti a taktiež sme počítali s rozširovaním datacentra, čo prinesie zvýšenú záťaž.

10. Bibliografia

1. Red Hat Cluster Suite: Configuring and Managing a Cluster. [Online] [Dátum: 1. 4 2010.] http://www.redhat.com/docs/manuals/csgfs/browse/rh-cs-en/ch-hardware.html.
2. HA-Cluster with loadbalancing for Zope (and Plone). [Online] [Dátum: 1. 4 2010.] http://plone.org/documentation/kb/ha-load-balanced-cluster-for-zope-and-plone.
3. Mixing it Up with SAS, SATA. [Online] LSI Corporation. [Dátum: 1. 4 2010.] http://www.serialstoragewire.org/Articles/2007_07/developer24.html.
4. Fiber Channel. [Online] [Dátum: 1. 4 2010.] http://compnetworking.about.com/cs/fibrechannel/g/bldef_fibrechan.htm.
5. Fibre Channel Industry Association. [Online] [Dátum: 1. 4 2010.] http://www.fibrechannel.org/.
6. Failover cluster s protokolom CARP. [Online] [Dátum: 1. 4 2010.] http://www.linuxos.sk/clanok/251/index.html.
7. A High-Availability Cluster for Linux. [Online] [Dátum: 1. 4 2010.] http://www.linuxjournal.com/article/3247?page=0,1.
8. [Online] [Dátum: 1. 4 2010.] www.dell.com.
9. [Online] [Dátum: 1. 4 2010.] http://www.lites-sk.sk/000000.htm.
10. Inteligentný domový zabezpečovací a komunikačný systém. [Online] http://e-shop.jablotron.sk/index.php?zobraz=produkty&idp=253&id_kategorie=253.
11. Liebert Challenger ITR Precision Cooling System. [Online] [Dátum: 1. 4 2010.] http://www.42u.com/cooling/crac-crah/liebert-challenger-itr.htm.
12. Smart-UPS® VT. [Online] [Dátum: 1. 4 2010.] http://www.apc.com/products/category.cfm?id=13.
13. Product selector, Broadcrown. [Online] [Dátum: 1. 4 2010.] http://www.broadcrown.com/product-selector.php.
14. Deepak Kakadia, Sun Microsystems, Inc. *Enterprise Network Design Patterns: High Availability.* s.l. : Sun BluePrints™, 2003.
15. Cisco Systems, Inc. *Cisco Data Center Infrastructure 2.5* . 2010.
16. Cisco Systems, Inc. *Campus Network for High Availability Design Guide.* 2008.
17. Ing. Igor Grellneth, Phd. Prednášky k predmetu Počítačové siete II. [Online] 4. 5 2010. http://www2.fiit.stuba.sk/~grellneth/ps2-ZS2009/index.htm.
18. [Online] 6. 4 2010. www.spread.org.
19. Project SeznamFS. [Online] 6. 4 2010. http://seznamfs.sourceforge.net/documentation.html.
20. [Online] 6. 4 2010. www.drbd.org.