Budování sítě v datových centrech

Ing. Pavel Danihelka
pavel.danihelka@firma.seznam.cz
Network administrator

Obsah

- **Úvod**
- Hardware
- Škálovatelnost a propustnost
- Zajištění vysoké dostupnosti
- Bezpečnost
- Load balancing
- Návrh architektury sítě
- Závěr

Představení služeb Seznamu

- návštěvnost Seznamu přes 2M uživatelů denně
- více než polovina služeb v top 10 českého internetu
- široké portfolio služeb
  seznam.cz, vyhledávání, firmy.cz, mapy.cz, obrazky.cz, email.cz, novinky.cz, sport.cz, super.cz, lide.cz

Požadavky na síť

- vysoká dostupnost
  zhruba 99.999% tj. výpadek max 30s za měsíc
- redundance síťových prvků a datových cest
- rychlá konvergence sítě
  řádově sekundy
- škálovatelnost
  neustálý růst, kolem 1000 serverů
- velká síťová propustnost
  odchozí provoz přes 2Gb/s

Obsah

- Úvod
- **Hardware**
- Škálovatelnost a propustnost
- Zajištění vysoké dostupnosti
- Bezpečnost
- Load balancing
- Návrh architektury sítě
- Závěr

Hardware

- přepínače (switch)
- konvertory
- kabeláž
- směrovače (router)
- ukázka jednoduchého zapojení

Přepínače (switch)

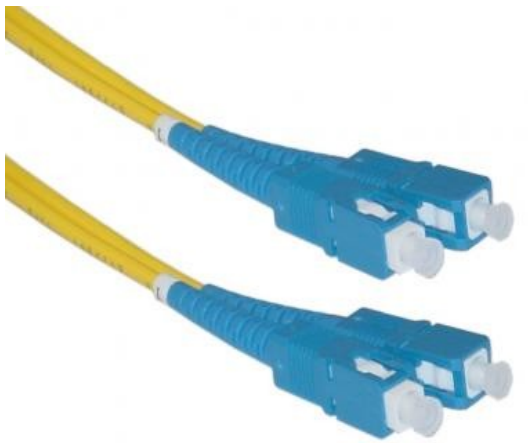
- L2 prvek, rack mount
- zapojení koncových serverů
- možnost stohování switchů
- management, SNMP, VLAN

Konvertory

- GBIC, SFP
  - 1 GigE, metalické/optické
- X2, XENPAK
  - 10 GigE, optické

Kabeláž

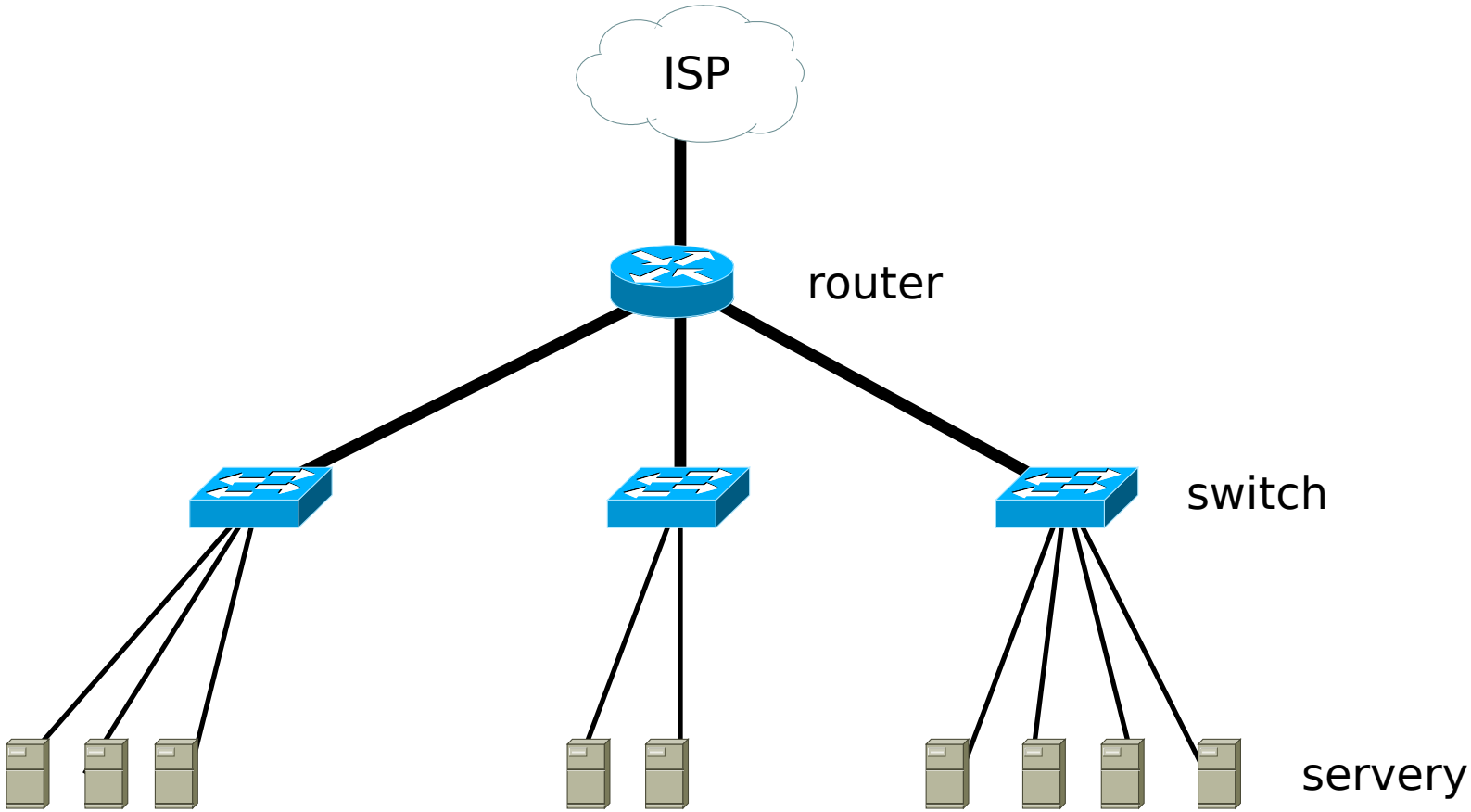
- metalická kabeláž
  kategorie 6 - 1000BASE-T
- optická kabeláž
  LC/SC konektory
  singlemode/multimode



Směrovače (router)

- L3 prvek, redundantní zdroje, modulární design
- centrální bod sítě
- propoje ke switchům a k providerovi
- směrování paketů

Ukázka jednoduchého zapojení



Obsah

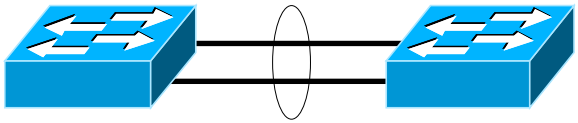
- Úvod
- Hardware
- **Škálovatelnost a propustnost**
- Zajištění vysoké dostupnosti
- Bezpečnost
- Load balancing
- Návrh architektury sítě
- Závěr

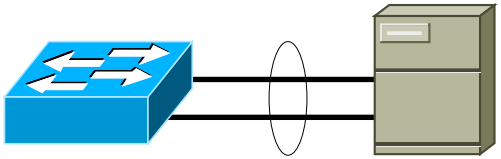
Škálovatelnost a propustnost

- 10 GigE po optice (IEEE 802.3ae)
- 1 GigE po optice (IEEE 802.3z) a metalice (IEEE 802.3ab)
- link aggregation (IEEE 802.3ad)
- dynamické směrování
  OSPF, IS-IS, EIGRP
- propojování sítí - BGP

Link aggregation

- IEEE 802.3ad, Etherchannel (Cisco), Trunking (Sun), Bonding (Linux)
- L2 protokol, agregace až 8 linek
- rozvažování provozu na základě hashe
  dst-ip, dst-mac, dst-port, src-dst-ip, src-dst-mac





OSPF (Open shortest path first)

- RFC 2740
- dynamický směrovací protokol
- používá Dijkstrův algoritmus k výpočtu nejkratší cesty
- používá se pro směrování ve vnitřní síti
- konvergence kolem 1s

OSPF - pokračování

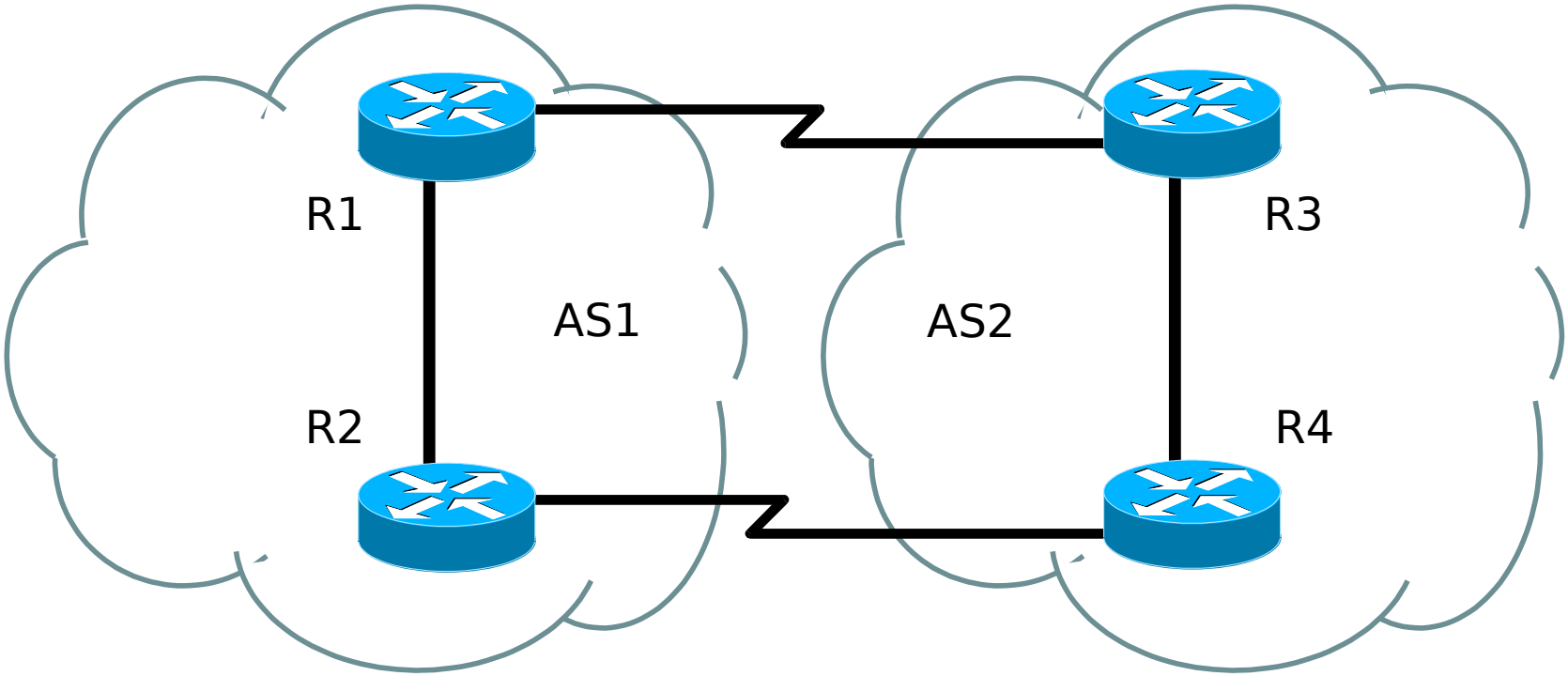
- dobrá škálovatelnost
  rozdělení sítě na oblasti (area 0)

BGP (Border gateway protocol)

- RFC 4271
- směrovací protokol
- používá se pro směrování mezi AS (autonomní systém)
- peerování, posílání směrovacích informací
  přes 200k route v internetu

BGP - pokračování

- směrování provozu
  lokální preference, MED



Obsah

- Úvod
- Hardware
- Škálovatelnost a propustnost
- **Zajištění vysoké dostupnosti**
- Bezpečnost
- Load balancing
- Návrh architektury sítě
- Závěr

Zajištění vysoké dostupnosti (High availability)

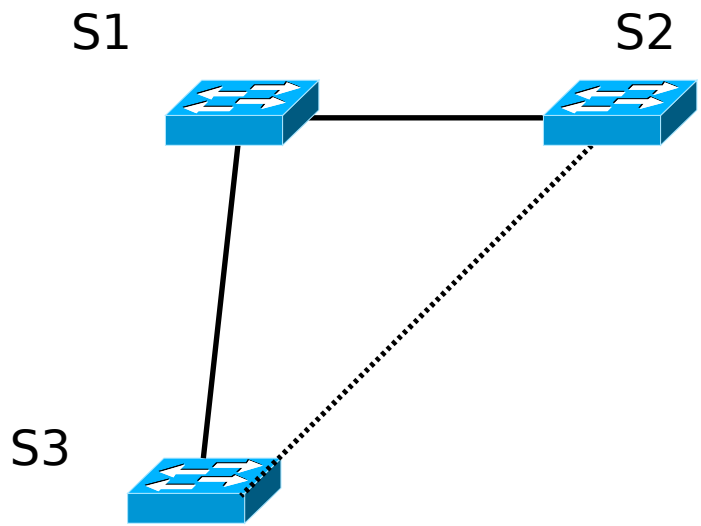
- VRRP
- STP (spanning tree protocol)

VRRP

- VRRP (Virtual router redundancy protocol) RFC 3768
- IP adresa “sdílená” mezi více boxy
- komunikace přes multicast, priority
- konvergence cca 3s

STP (Spanning tree protocol)

- L2 protokol, BPDU rámce každé 2s
- Rapid-PVST+ (802.1w)
- stromová topologie, redundance datových cest
- root bridge, path cost
- strom nejkratších cest
- konvergence kolem 1s



Obsah

- Úvod
- Hardware
- Škálovatelnost a propustnost
- Zajištění vysoké dostupnosti
- **Bezpečnost**
- Load balancing
- Návrh architektury sítě
- Závěr

Bezpečnost

- paketový filtr – ACL
- stavový firewall, IDS
- autentizace přes Tacacs+
- L2 mechanismy ochrany sítě

Obsah

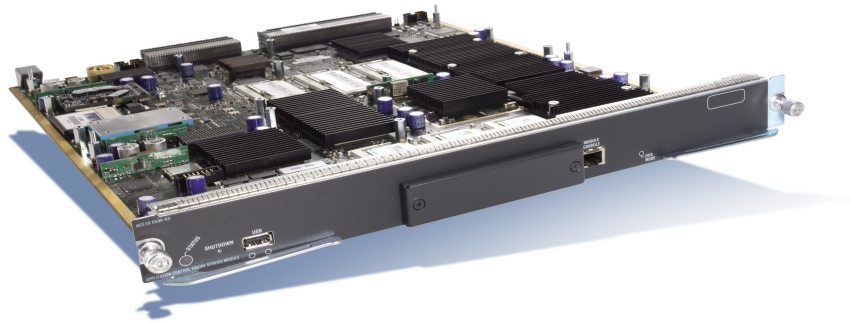
- Úvod
- Hardware
- Škálovatelnost a propustnost
- Zajištění vysoké dostupnosti
- Bezpečnost
- **Load balancing**
- Návrh architektury sítě
- Závěr

Load balancing

- motivace – rozkládání provozu na více fyzických serverů
- DNS round robin
  - jednoduché řešení
  - spotřeba IP adres, řešení výpadků strojů
- samostatné HW řešení
  Barracuda Networks, Cisco, Juniper, Nortel

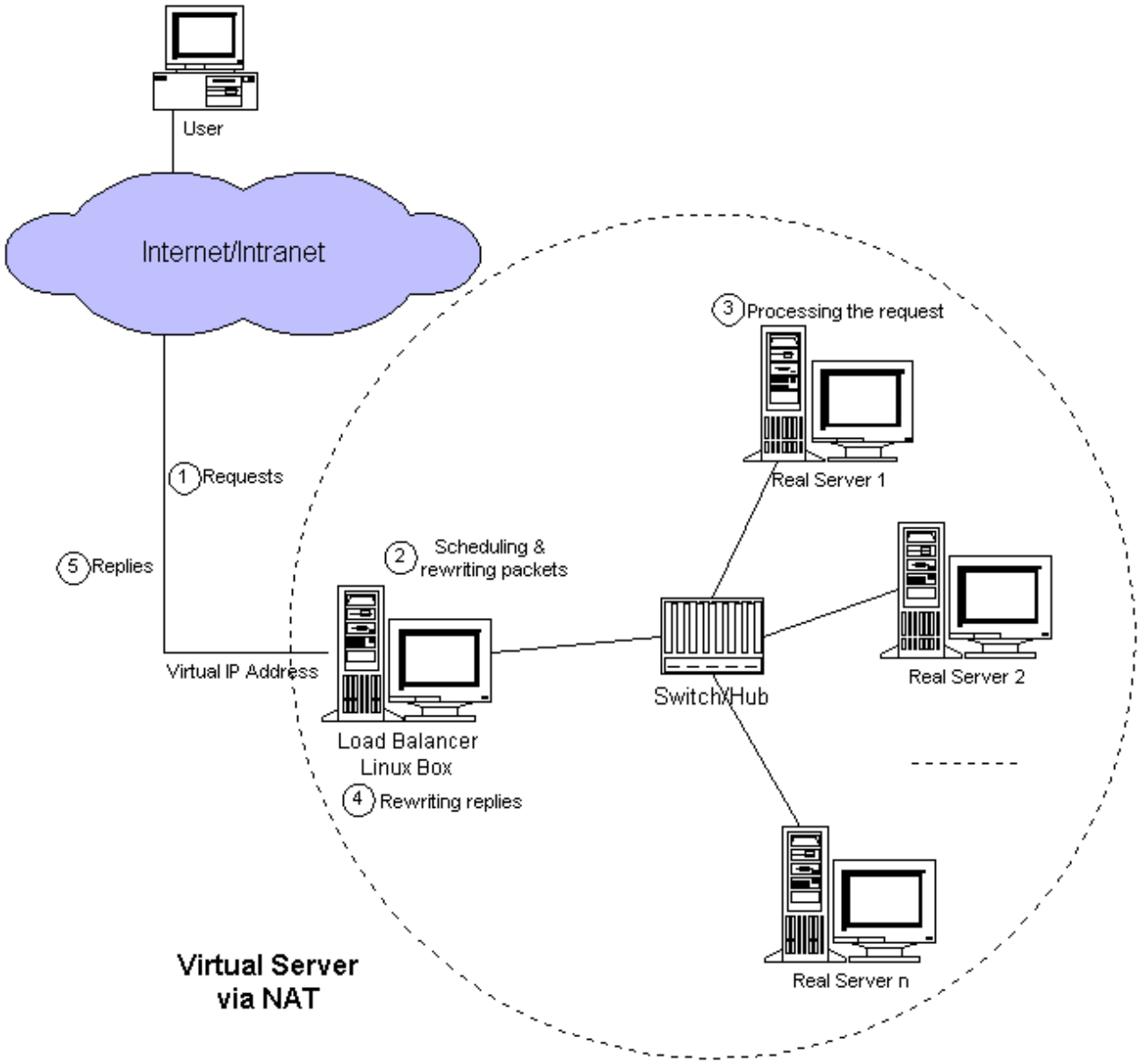
Load balancing (pokračování)

- integrované HW řešení
  Cisco ACE modul, CSM modul
- softwarové řešení
  LVS - http://www.linuxvirtualserver.org



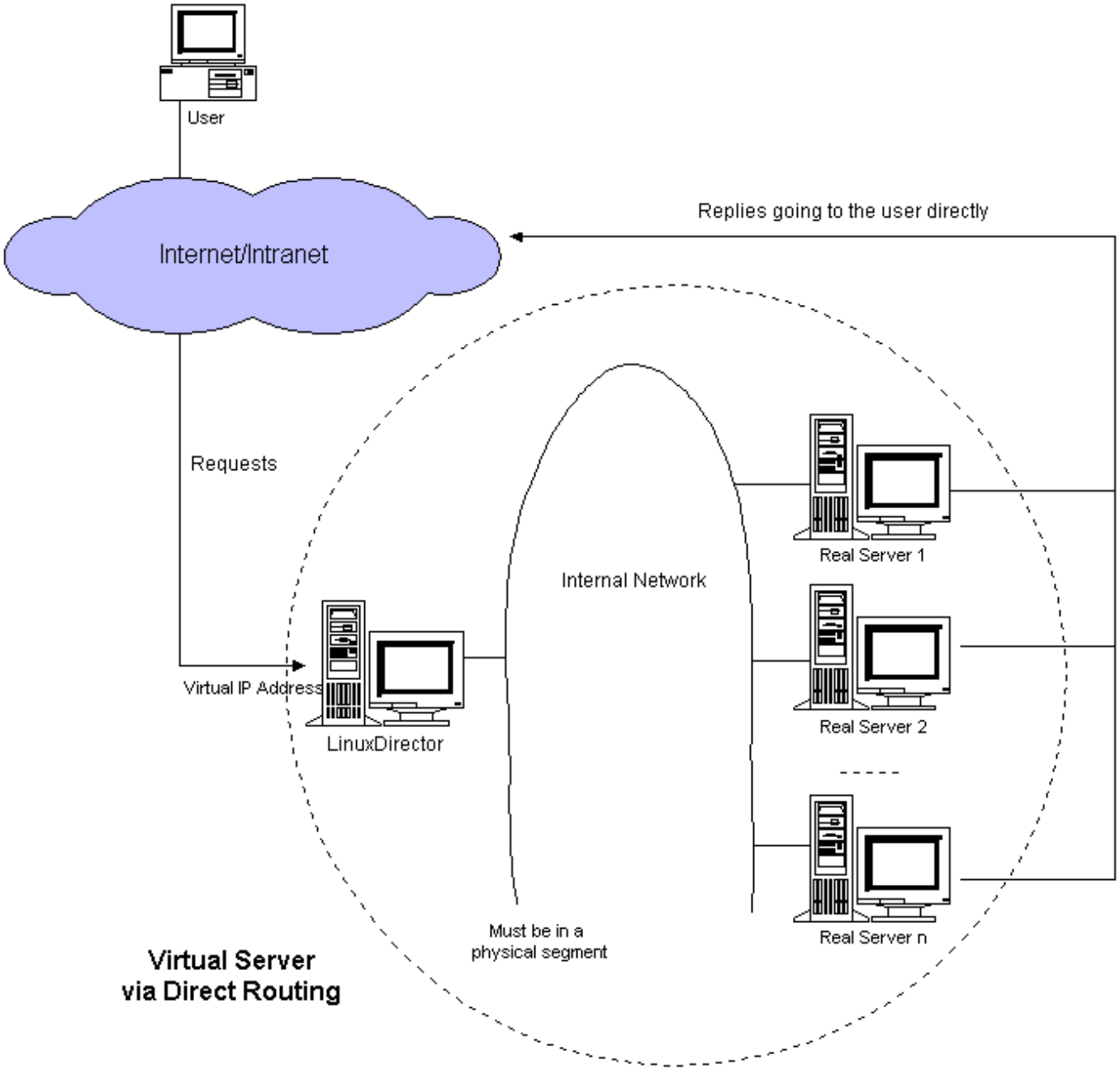
Load balancing - popis

- L4-L7 switching
- kontrola živých služeb
- režim NAT
  - použití pro veřejné služby


User
Internet/Intranet
1 Requests
5 Replies
Virtual IP Address
2 Scheduling & rewriting packets
Load Balancer Linux Box
4 Rewriting replies
3 Processing the request
Real Server 1
Switch/Hub
Real Server 2
Real Server n
Virtual Server via NAT

Load balancing – popis (pokračování)

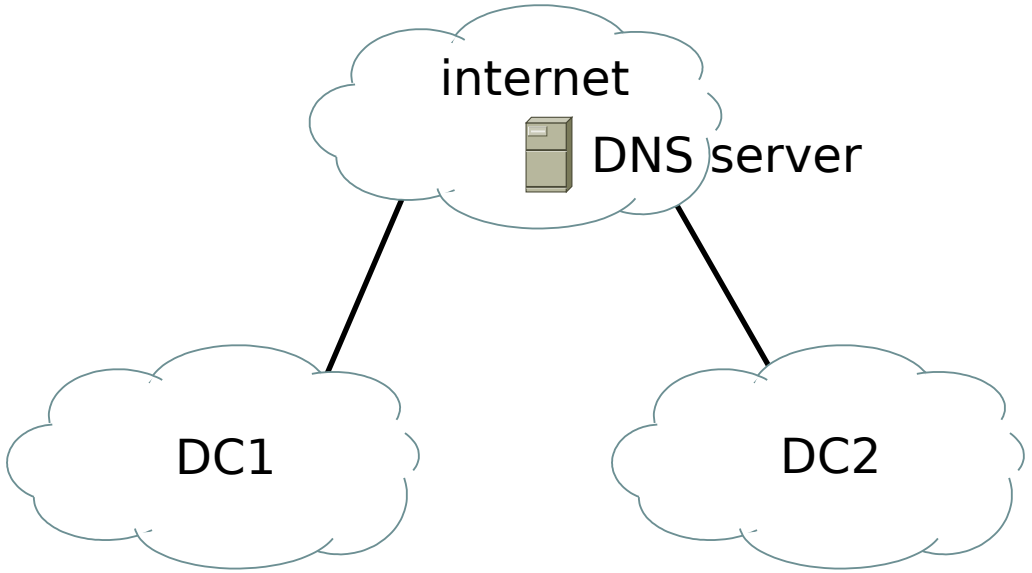
- režim Direct routing
  použití pro interní služby
  nižší HW nároky na load balancer



Virtual Server via Direct Routing

Load balancing – více datacenter

- v každém DC jiné IP rozsahy
- funguje podobně jako DNS round robin
- kontrola živých služeb na DNS serveru



Obsah

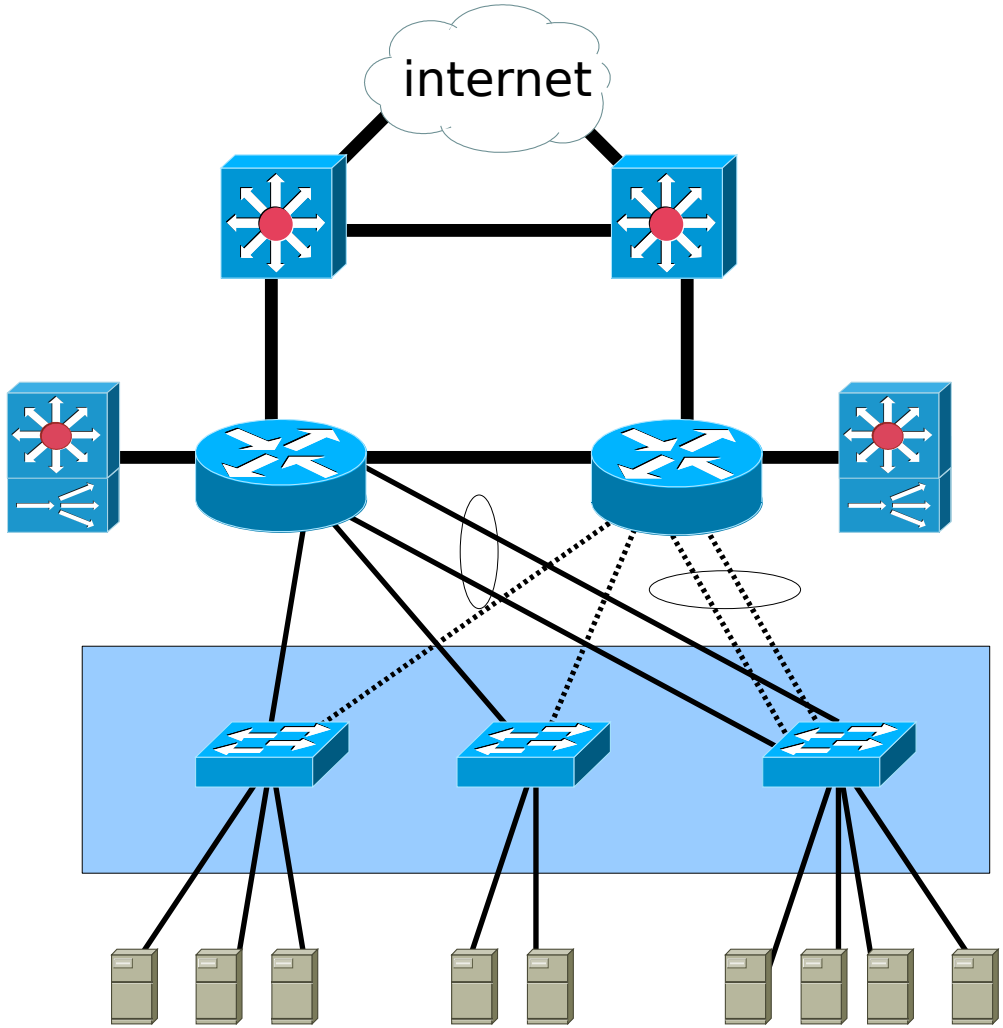
- Úvod
- Hardware
- Škálovatelnost a propustnost
- Zajištění vysoké dostupnosti
- Bezpečnost
- Load balancing
- **Návrh architektury sítě**
- Závěr

Obecný návrh architektury sítě

- přístupová vrstva
  konektivita pro servery, STP
- agregační vrstva
  řesení přístupu, HA, load balancing, VRRP
- páteřní vrstva
  škálovatelnost, HA, rychlá konvergence, propustnost

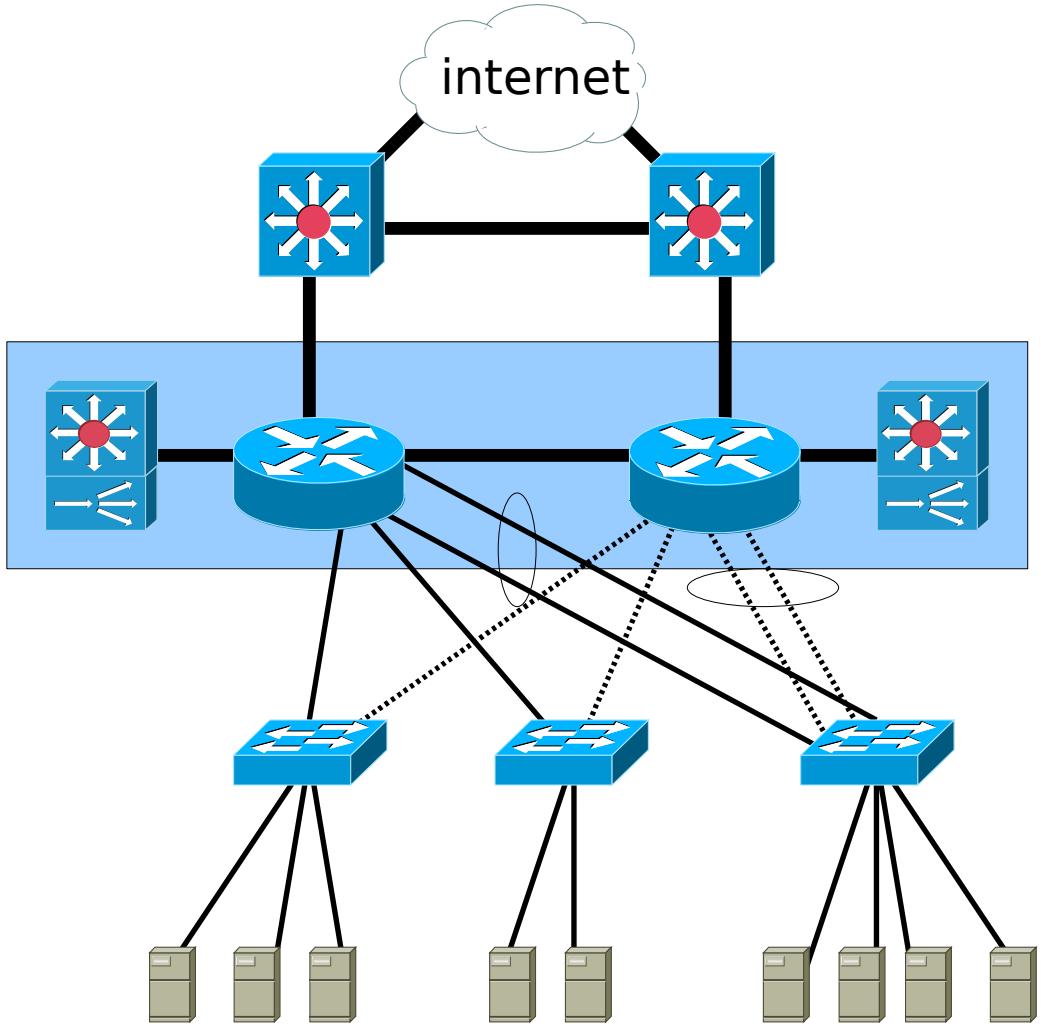
Návrh přístupové vrstvy

- rozdělení sítě do VLAN
- link aggregation propoje
- redundantní cesty



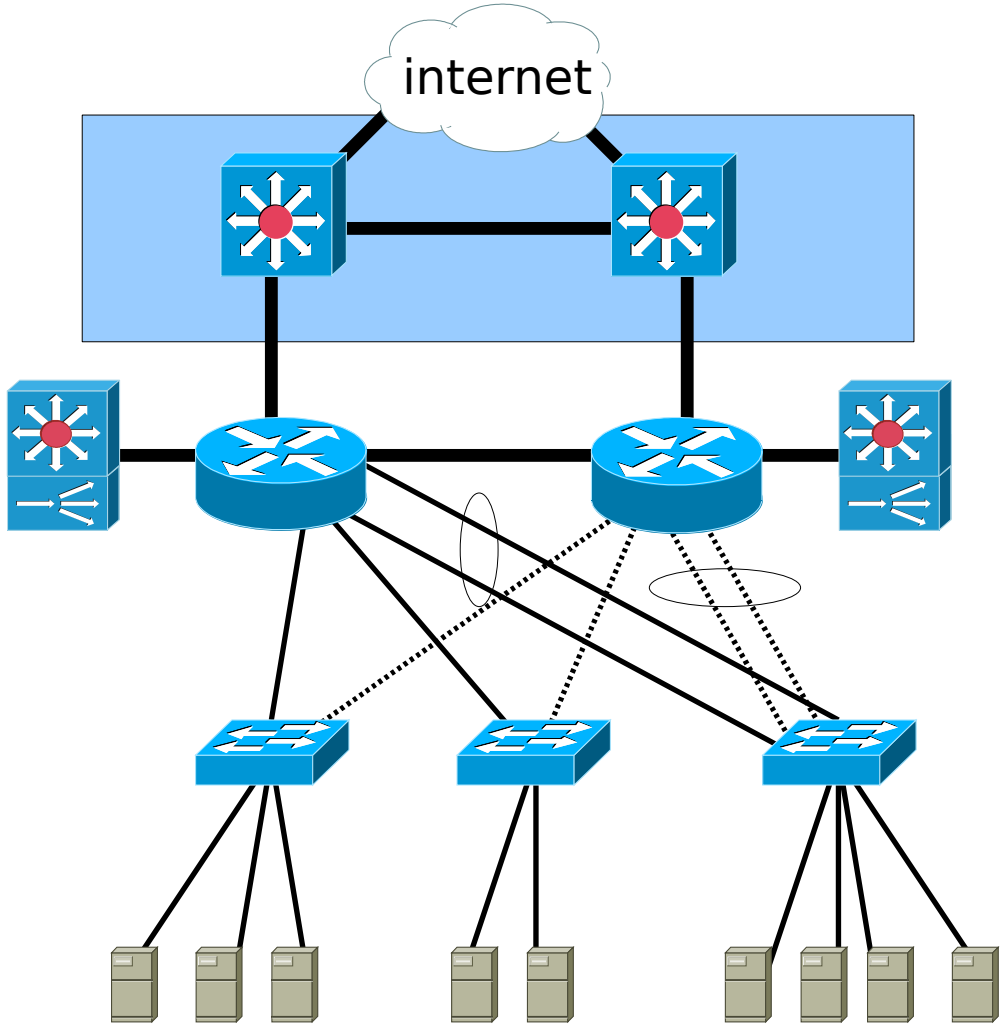
Návrh agregační vrstvy

- VRRP
- L2, Rapid-PVST+
- load balancing, firewall, SSL



Návrh páteřní vrstvy

- ne vždy je potřeba
- 10 GigE
- vše na L3
- směrování přes OSPF a BGP



Obsah

- Úvod
- Hardware
- Škálovatelnost a propustnost
- Zajištění vysoké dostupnosti
- Bezpečnost
- Load balancing
- Návrh architektury sítě
- **Závěr**

Trendy

- širší použití 10 GigE
- zvyšování hustoty serverů – blade servery, virtualizace
- vyšší integrace služeb do síťových prvků
  IDS, SSL, load balancing

Dotazy?

pavel.danihelka@firma.seznam.cz

Odkazy

- Seznam blog
  seznam.sblog.cz
- Vývojáři
  vyvojari.seznam.cz