

PA159

Počítačové sítě a jejich aplikace I

Luděk Matyska

FI MU, podzim 2007

Organizace předmětu

- Přednášky nejsou povinné
- Předpokládají se vstupní znalosti z předmětu PB156
- Loňská verze přednášek na webu
- Zkouška pouze písemná
 - plánovány cca 3 řádné termíny a 2 opravné
 - prototypové písemky budou k dispozici
- Studijní literatura
 - RFC a další, bude uvedeno u příslušných pasáží
 - RFCs: <http://www.zvon.org>, <ftp://ftp.fi.muni.cz/pub/rfc>

Cíle předmětu

- Poskytnout pokročilý pohled na oblast počítačových sítí a jejich aplikací
- Probírané oblasti
 - Architektura sítí - rekapitulace
 - využití formalismů pro specifikaci protokolů
 - Správa sítí
 - Bezpečnost
 - Kvalita služeb
 - Sítě a multimédia
 - Virtuální spolupráce s využitím sítí

Rekapitulace předpokládaných znalostí

Model ideální sítě

- Transparentní pro uživatele/aplikace
 - pouze end-to-end vlastnosti
- Neomezená propustnost
- Nulové výpadky
- Nulový rozptyl zpoždění (jitter)
- Dodržení pořadí paketů
- Nemůže dojít k poškození dat

- Slouží pro klasifikaci vlastností skutečných sítí

Reálné sítě

- Mají vnitřní strukturu, která ovlivňuje přenos dat
- Omezená kapacita (šířka pásma)
- Některé sítě
 - Variabilní zpoždění
 - Výpadky, poškození dat
 - Přeuspořádání paketů
 - ...

Požadované vlastnosti

- Účinnost
 - maximální využití kapacity
- Férovost (fairness)
 - stejný přístup ke všem tokům dat (v rámci dané třídy kvality služby)
- Decentralizovaná správa
- Rychlá konvergence při reakci na změnu stavu
- Multiplexing/demultiplexing
- Spolehlivost – alespoň obvykle

Požadované vlastnosti (2)

- Řízení toku dat – alespoň obvykle
 - ochrana proti zahlcení přijímající stanice
 - ochrana proti zahlcení sítě (ochrana síťových prvků před vyčerpáním kapacity a bufferů)
 - *de facto* globální optimalizační problém
 - odpovídá systému se zpožděnou zpětnou vazbou

Zábrana nebo řízení zahlcení

- Zábrana zahlcení
 - nikdy nedojde k zahlcení
- Řízení zahlcení
 - včas detekováno
 - rychlá reakce vedoucí k jeho odstranění

Implementace funkcionality

- End-to-End
 - požadovanou funkcionalitu je možné zajistit pouze se znalostí a prostřednictvím koncové aplikace
 - např. bezpečnost
- Hop-by-Hop
 - opakováním určité funkcionality uvnitř sítě je možné dosáhnout výrazného zvýšení výkonu na úrovni každého dvoubodového přenosu (např. kontrolní součty)
- E2E lze aplikovat rekurzivně
 - jak pro uzly sítě tak i pro jednotlivé vrstvy
- E2E vs. HbH - záleží na úhlu pohledu (HbH v jednom pohledu je E2E v druhém)

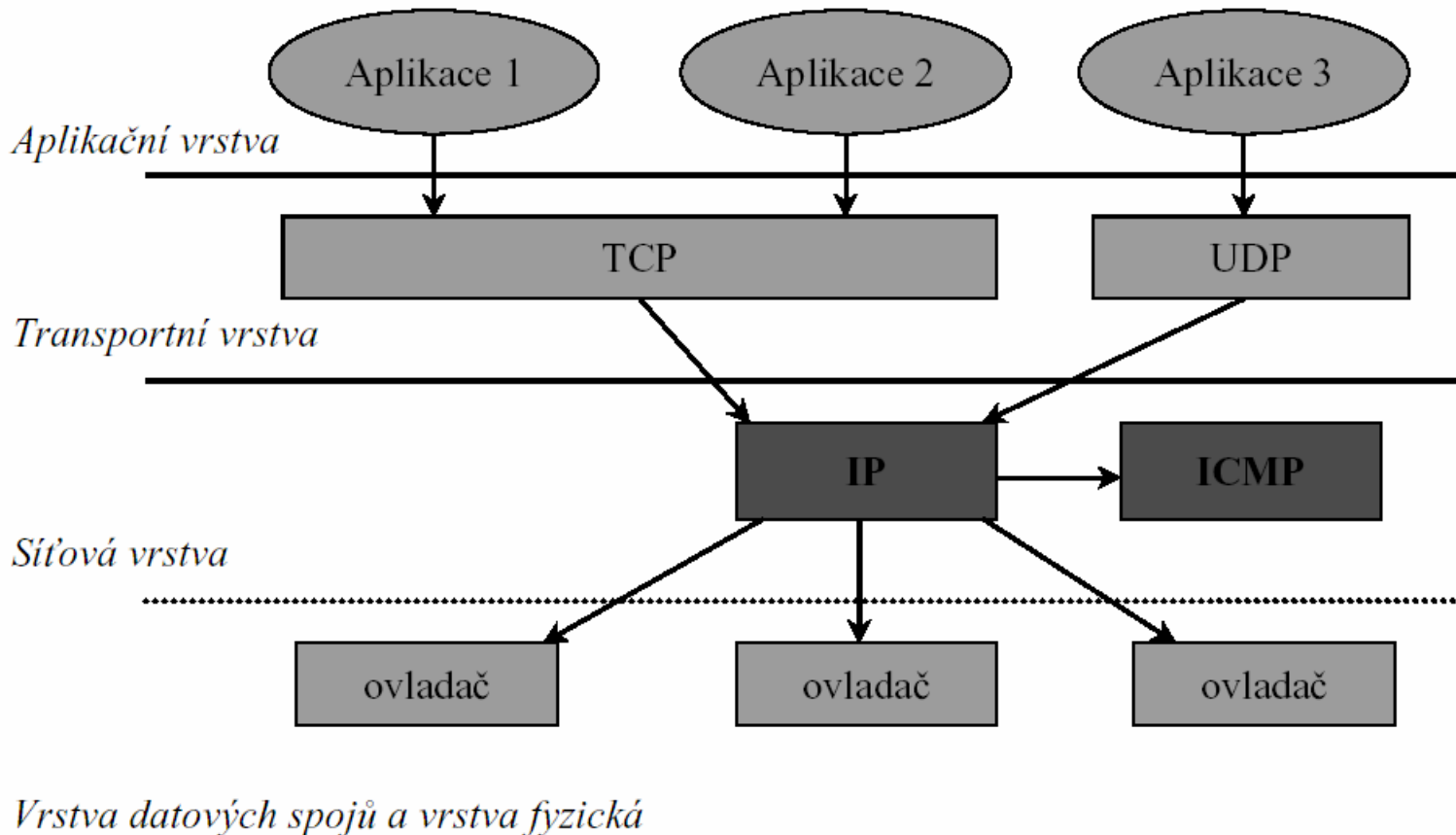
Přístupy k budování sítí

- Telefonní
 - přenos signálu o přesně definovaných parametrech (z pohledu přenosu dat)
 - connection-oriented, stavová (=>okruhy)
 - automatická kvalita služby (QoS)
- Počítačové
 - variabilní data
 - connectionless (bezstavové)
 - dělení dat na segmenty => pakety, buňky...
 - problém s realizací kvality služby (QoS)
 - není definováno spojení, každý paket může putovat jinou cestou, ...
 - best-effort

Kvalita přenosových služeb

- Platí pro spojované služby
- Dohodne se při navázání spojení
- Parametry
 - zpoždění při navazování a rušení spojení
 - pravděpodobnost neúspěchu požadavku na spojení
 - chybovost přenosu dat
 - propustnost (minimální, průměrná, špičková)
 - zpoždění (minimální, průměrné, špičkové)
 - rozptyl zpoždění (jitter)

Model přesýpacích hodin



Fyzická vrstva a vrstva datových spojů

- Typy transportních médií
 - optické, metalické, bezdrátové
- Přístup k přenosovému médiu
 - Ethernet, Token-Ring
- Kódování dat
 - limituje využití šířky pásma, může zvyšovat robustnost
- Multiplexování
 - Wave Division Multiplexing (WDM, optické spoje)
 - Time Division Multiplexing (TDM, SDH/SONET)

Pakety

- Dělení dat kvůli multiplexování
- Délka paketů ovlivňuje celkové chování sítě
 - dlouhé pakety => vyšší výkon jednotlivých spojení
 - krátké pakety => lepší férovost

 - pakety s variabilní délkou => lepší přizpůsobení se charakteru přenášených dat
 - pakety s fixní délkou => zajištění kvality služby

Pakety (2)

- Rozklad příliš dlouhých zpráv na menší části = fragmentace
 - číslování fragmentů kvůli opětovnému skládání
 - možno provádět rekurzivně
- Typické velikosti paketů/buněk
 - ATM: 53B (z toho 5B rezie!)
 - Ethernet: 1500B
 - Gigabit Ethernet: až 9kB resp. 16kB

ATM

- „Cesta do pekla je dlážděna dobrými úmysly“
- Zaměřeno na QoS
- Okruhy + krátké buňky = velká režie
 - udržování stavové informace
 - páteří směrovače běžně zajišťují >10.000 spojení
 - fragmentace dat
 - cca 10% režie (5B/53B)
- Hranice běžné aplikovatelnosti: 622 Mbps – 1.2 Gbps
- Dnes už neperspektivní
- MPLS trochu navazuje

Přepínání

- Přepínání okruhů (spojované sítě)
 - vytvoření okruhu
 - přepínání dle vytvořených cest
 - např. telefonní sítě
- Přepínání paketů (nespojované sítě)
 - store and forward
 - cut through

Přepínání u nespojovaných sítích

- Backward learning algorithm
 - přepínač/můstek se učí nasloucháním na sdíleném médiu a sleduje zdrojové adresy
 - zasílá podle cílové adresy
 - informace stárne (zapomenutí v řádu desítek sekund)
- V případě vytvoření cyklů => hledání (nejmenší) kostry ((minimum) spanning tree)
 - distribuovaný algoritmus
 - robustnost proti výpadkům
 - nevyžaduje centrální organizaci

Spanning Tree

- Cílem je některé porty můstků nepoužívat
- Volba kořenového vrcholu stromu (nejnižší adresa)
- Postupný růst stromu – nejkratší vzdálenost od kořene (preferenci mají uzly s nižší adresou, pokud více možností)
- Nalezené „nejlepší“ cesty (definují aktivní porty můstku)
- Vypnout všechny ostatní porty

Spanning Tree (2)

- Každý můstek posílá periodické zprávy
 - vlastní adresa, adresa kořenového můstku, vzdálenost od kořene
- Když dostane zprávu od souseda, upraví definici „nejlepší“ cesty
 - preferuje kořen s menší adresou
 - preferuje menší vzdálenosti
 - při stejných vzdálenostech preferuje nižší adresu

Spanning Tree (3)

- Inicializace
 - Na začátku si každý můstek myslí, že je kořenem
 - pošle konfigurační informaci na všechny porty
 - Následně můstky posílají jen „nejlepší“ konfigurace (cesty)
 - přičti 1 k vzdálenosti, pošli na porty kde je stále „nejlepší“ cesta
 - vypni zasílání dat na všech ostatních portech

Spanning Tree (4)

- Formát zprávy:
 - <kořen, vzdálenost ke kořeni, můstek>
- Informace „stárne“
 - selhání kořene vede eventuálně k volbě nového kořene
- Rekonfigurace neprobíhá okamžitě (je tlumena)
 - prevence vzniku dočasných cyklů

L2 prvky

- Můstky (Bridges)
 - přemostění sdíleného média (oddělení provozu dvou segmentů sítě)
- Přepínače (Switches)
 - víceportové můstky
 - používají L2 adresování => L2 směrování
 - v současnosti nejpoužívanější forma propojení lokálních sítí
 - možnost kaskádování
- Z pohledu vyšších vrstev (např. IP) vytvářejí uniformní transparentní prostředí

L2 prvky (2)

- Problematika budování rozsáhlejších sítí na L2
 - přepínací tabulky rostou s počtem uzlů sítě (tedy počtem stanic, nikoli jen přepínačů/můstků)
 - neumožňuje logické oddělení lokálních sítí
 - netvoří hranice pro broadcastový provoz
 - pomalá konvergence - limitovaná hledáním (minimální) kostry

Síťová vrstva (L3)

- Doprava dat mezi uzly (obecně umístěných v různých sítích)
- Connectionless vs. connection-oriented
- Internetové protokoly pro síťovou vrstvu
 - IP - IPv4 a IPv6
 - ICMP (pro IPv4 i IPv6)
- Směrování (routing)
 - přehazování paketů pro dosažení cíle
 - získávání informace o topologii sítě

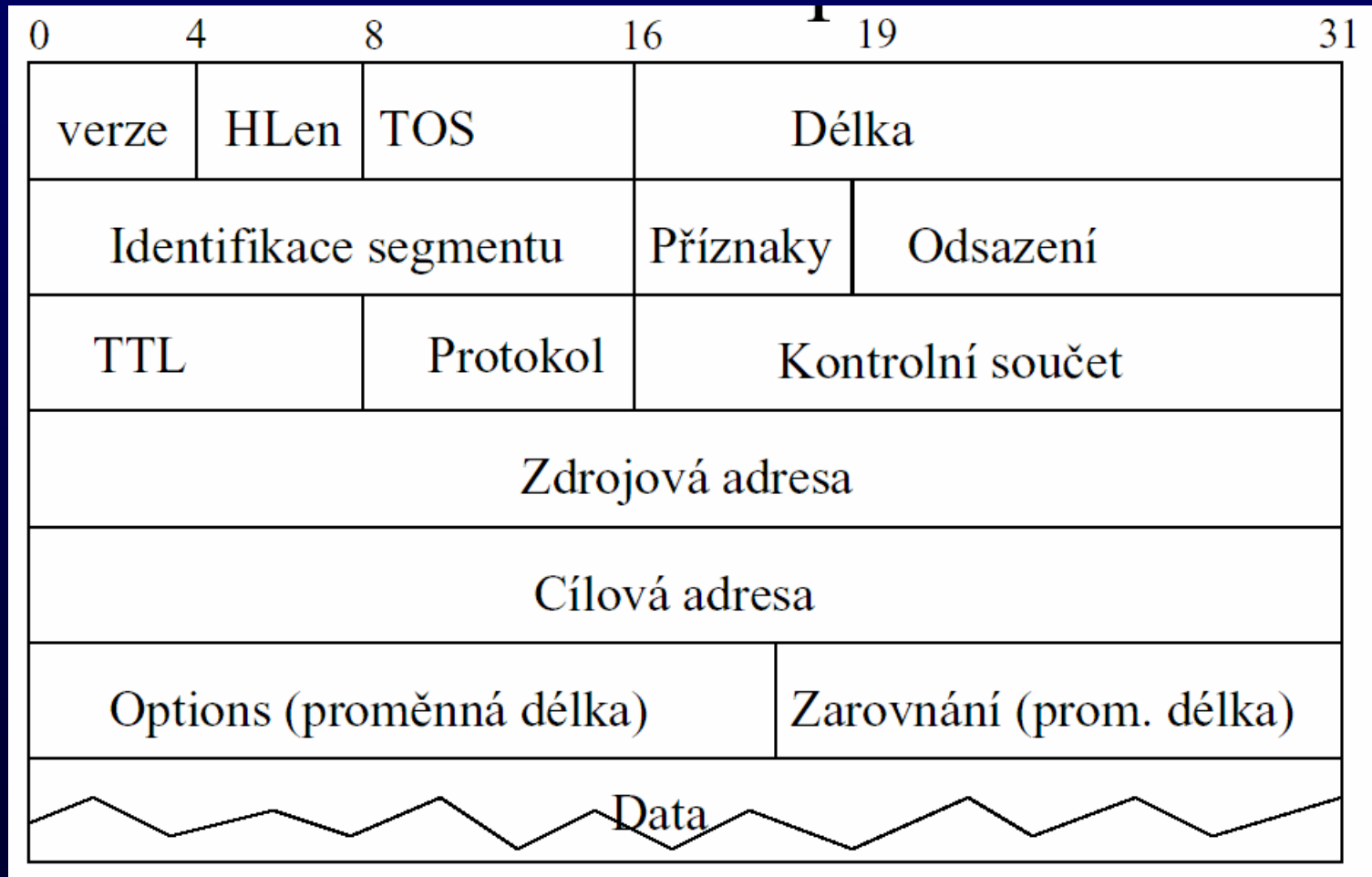
Modely síťové vrstvy

- Doručení datagramů
 - connectionless přístup, best-effort, nespolehlivé
 - není zaručeno doručení paketu
 - pakety se pohybují sítí nezávisle
 - může se využívat více cest kvůli vyvažování zátěže => problém s přeuspořádáním paketů
- Virtuální kanály
 - connection-oriented přístup, dohodnutá kvalita služby
 - signalizace pro vytváření a řízení kanálů
 - pakety jedné relace prochází jednou cestou

IP

- Internet Protocol
- Definován ve dvou verzích:
 - IPv4 (RFC 791)
 - IPv6 (RFC 2460)
- Globální hierarchické adresní schéma
 - 32 bit pro IPv4, 128 bit pro IPv6
 - mapování na adresy nižší vrstvy (ARP, RARP)
- Přenos sítí – směrování

Formát IPv4 paketu



Formát IPv4 paketu (2)

- Legenda
 - HLen = počet 32-bitových slov
 - TOS = Type of Service (nepoužíván)
 - Délka = délka paketu ve slabikách
 - Protocol = protokol vyšší vrstvy
 - CRC = detekce poškození
 - TTL = Time to Live
- Práce s TTL
 - každý směrovač sníží hodnotu o 1
 - TTL=0 => paket zahozen
 - ochrana před nekonečnými cykly a zbloudilými pakety

IPv6

- 128-bitové adresy
 - 340282366920938463463374607431768211456
($3 \cdot 10^{38}$) jedinečných adres
- jednodušší a flexibilnější hlavička se 64-bitovým zarovnáním
- podpora real-time provozu (značkování toků (*flow label*))
- směrovače nesmí fragmentovat
- podpora autokonfigurace (vylepšená obdoba DHCP, přímo součást protokolu)
- rozšíření hlaviček
 - bezpečnost (AH + ESP, povinná součást implementace ;o)
 - mobilita

Fragmentace na IP vrstvě

- IPv4
 - umožňuje (i rekurzivně) fragmentovat
- IPv6
 - vrátí chybu, v níž oznamuje max. přípustnou délku
 - fragmentace se musí odehrávat pouze na vysílači
- Problémy s fragmentací:
 - ztráta fragmentu = ztráta paketu
 - zpomaluje přenos
- Detekce nejmenšího maximálního fragmentu
 - algoritmus Path MTU Discovery
 - problém dynamických cest

ICMP

- Internet Control Message Protocol
 - RFC 792
 - řídicí informace pro IP protokol, který doprovází
- Použití
 - odhalení chyb při přenosu paketů
 - obsahují část IP paketu, který chybu způsobil
 - zjišťování stavu sítě
 - většinou se zpracovává jinak, než IP
 - např. omezení rychlosti posílání ICMP paketů
- Ochrana proti rekurzivnímu generování
 - na ztrátu ICMP paketu se nereaguje

ICMP zprávy

- *Destination unreachable*
 - destinace může být: podsít', uzel, protokol, port
- *Redirect*
 - oznámení o kratší cestě
- *TTL expired*
 - odpovídající IP paket dosáhl stavu TTL=0
- *Echo request/reply*
 - ping
 - traceroute (vlastně ping s proměnným TTL)

Směrování (routing)

- Problém:
Nalezení cesty mezi dvěma uzly, která splňuje zadané omezující podmínky.
- Ovlivňující faktory
 - statické: topologie
 - dynamické: zátěž a chování sítě v daném okamžiku
- Pakety prochází řadou směrovačů mezi vysílačem a přijímačem

Směrování (2)

- Statické směrování
 - předem nadefinované, vhodné pro statickou topologii
 - jednodušší, méně flexibilní
- Dynamické cesty
 - složité (distribuované) algoritmy
 - nutná aktualizace směrovacích tabulek
 - protokol pro aktualizaci tabulek
 - adaptabilní na výpadky a další dynamické změny prostředí
 - nezaručuje pořadí doručení

Směrovací schémata

- *Distribuované* nebo *centralizované*
- „*Krok za krokem*“ nebo *zdrojové* (source-based)
- *Jedno* nebo *vícecestné*
- *Deterministické* nebo *stochastické*
- *Dynamický* nebo *statický výběr cest*

- Metody použité v Internetu jsou psány kurzívou

Požadované vlastnosti směrovacího algoritmu

- Správnost
- Jednoduchost
- Efektivita a škálovatelnost
 - minimalizace množství řídicích informací (~5% provozu!)
 - minimalizace velikosti směrovacích tabulek
- Robustnost a stabilita
 - nezbytný je distribuovaný algoritmus
- Spravedlivost
- Optimálnost
 - „Co je to nejlepší cesta?“

Problém globálního pohledu

- Globální znalost je problematická
 - je složité ji získat
 - když už se to povede, tak není aktuální
 - musí být lokálně relevantní
- Rozpor mezi lokální a globální znalostí může způsobit
 - cykly (černé díry)
 - oscilace (adaptace na zátěž)

Reprezentace sítě pro směrování

- Sít' reprezentována jako (typicky orientovaný) graf
- Uzly
 - adresy
- Hrany
 - ohodnocení = cena komunikace
 - cena = délka fronty
 - cena = $1/\text{přenosová_kapacita}$
- Výkonnostní charakteristiky směrovacích algoritmů
 - minimalizace počtu skoků (jednotná cena)
 - minimalizace ceny

Rozhodování směrovacích algoritmů

- Okamžik
 - při uzavírání spojení (= vytváření okruhu)
 - spojované služby, virtuální kanály (ATM ev. i MPLS)
 - při příchodu dat (paketu)
 - nespojované služby, datagramy
- Místo
 - jediný uzel => centralizované algoritmy
 - každý uzel => distribuované algoritmy

Dynamické směrovací algoritmy

- Centralizované
 - stav se posílá do centra
 - centrum posílá tabulky uzlům
- Izolované
 - každý uzel sám za sebe (bez informací od okolních uzlů)
- Distribuované
 - kooperace uzlů

Izolované směrování

- Náhodná procházka
 - vysoká robustnost
- Extrakce informací z procházejících paketů
- Záplava (broadcast)
 - kopie všem kromě zdroje
 - mimořádně robustní
 - optimální
 - enormní zátěž sítě
- Vyžaduje zpětnou vazbu

Dynamické směrování

- Periodická výměna směrovacích informací
- Dynamická výměna tabulek
 - možnost dočasné nekonzistence
- Hierarchie směrování
 - sítě sítí
 - implicitní (default) cesta k neznámým cílům
 - hierarchie směrovačů
 - směrování k sítím (inter-AS)
 - směrování uvnitř sítí (intra-AS)

Metriky pro dynamické směrování

- Definice optimality
- Výměna informace o vzdálenosti
- Hledání minimální kostry
- Distance Vector (DV)
 - počet přechodů do cíle
 - možnost zavádění politik (inter-AS - např. BGP)
- Link State (LS)
 - dostupnost sousedů

Metriky pro dynamické směrování (2)

- Způsob výběru ceny
 - propustnost, zpoždění, ztráty
- Statické metriky
 - hop-count
 - optimalizace ruční editací cen
- Dynamické
 - závisí na zátěži, zabraňují přetížení
 - mohou oscilovat (nezbytné tlumení!)
- Např. ARPANET
 - původně počet paketů ve frontě, nověji průměrné zpoždění

Směrování Distance Vector

- Předpoklad
 - každý směrovač zná pouze cestu a cenu k sousedům
- Cíl
 - směrovací tabulka pro každý cíl v každém směrovači
- Idea
 - řekni sousedům svoji představu tabulky
- Inicializace
 - Distance Vector = <Cíl; Cena>
 - sousédé: známá cena
 - ostatní: nekonečno (resp. hodnota definovaná jako nekonečno, pro RIP např. 16)

Aktualizace Distance Vector

- Periodicky posílá svou tabulku sousedům
- Pokud je cesta v získaném DV zvětšená o cenu cesty k danému sousedovi lepší než stávající uložená, tak se nahradí ve vlastní tabulce
- Konverguje pro statickou topologii
- Nebezpečí zacyklení
 - A (<Internet;2;B>) — B (<Internet;1>) — Internet
 - vypadne spoj: B — Internet
 - B získá tabulku (<Internet;3;A>)
 - => dělení horizontu – směrovač nesdílí cestu zpět uzlu, od něhož se o ní dozvěděl (problém zůstává pro složitější topologie)

RIP

- DV vektor používá jako hodnotu vzdálenosti počet uzlů
- Nekonečno = 16
 - nelze použít pro sítě s minimálním počtem hopů mezi libovolnými dvěma uzly >16
- Aktualizace každých 30s
 - trigger – změna stavu hrany
 - timeout – 180s
- RIPv1 definován v RFC 1058
- RIPv2 (přidána autentizace) v RFC 1388

Směrování Link State

- Předpoklad
 - každý směrovač zná pouze cestu a cenu k sousedům
- Cíl
 - směrovací tabulka pro každý cíl v každém směrovači
- Idea
 - šíří se topologie, cesty si počítají směrovače samy
 - fáze 1:
šíření topologie (broadcast)
 - fáze 2:
výpočet nekratší cesty (Dijkstra)
- Rychlejší konvergence a složitější algoritmus než DV

Algoritmus Link State

- Směrovač udržuje databázi LS a periodicky posílá LS pakety (LSP) sousedům
 - Obsah LSP:
 - Identifikátor uzlu
 - Cena spojů k sousedům
 - Pořadové číslo (SEQNO)
 - Doba platnosti (TTL)
- Každý směrovač posílá pakety dále kromě toho, od něž informaci dostal
- Spolehlivost zajištěna potvrzením

Zvláštní stavy Link State

- Výpadek spoje nebo směrovače => odstranění starých údajů
 - SEQNO definuje nová data
 - Nové LSP s cenou rovnou nekonečnu
- Restart směrovače
 - SEQNO=0 a nové LSP s TTL=0
- Obnova spojení po rozpadu sítě
 - Synchronizace LS databází

OSPF

- Open Shortest Path First
- Nejpoužívanější LS protokol současnosti
- Rozšíření
 - autentizace zpráv
 - směrovací oblasti: další úroveň hierarchie
 - load balancing: více cest se stejnou cenou

Hierarchie směrování

- Subnetting
 - místo síť:uzel je síť:podsíť:uzel
 - řeší problém nedostatku adres
- Masky sítí/podsítí
 - $S := \text{source}; D := \text{destination}; M := \text{netmask};$
if $(D \& M) = (S \& M)$ then
 local_net;
else
 next_hop_or_default_gw;
end if;

Hierarchie routování (2)

- Hierarchické členění Internetu => autonomní systémy (AS)
 - odpovídají administrativním doménám
 - 16-bitový identifikátor
 - škálovatelnost routování
 - Inter-AS (EGP, BGP-4)
 - Intra-AS (RIP, OSPF)
- Typy AS:
 - koncové (stub) AS
 - multihomed AS
 - transit AS

Inter-AS směrování

- hraniční směrovače
 - sumarizují a zveřejňují vnitřní adresy
 - aplikují politiky směrování
- interní směrovače mohou využít implicitní (default) cesty
- jádro sítě nepoužívá implicitní cesty
 - směrovač musí znát cestu ke všem AS (dnes cca 150 MB tabulek)
- EGP – DV přístup
- BGP – Path Vector + CIDR

CIDR

- Classless Inter-Domain Routing
- Agregace souvislých bloků s délkou 2^n
 - např. 191.11.16.0 – 191.11.31.255 = 191.11.16/20
 - redukce velikosti směrovacích tabulek
- Směrovací tabulky obsahují cesty k prefixům
 - v případě vícenásobné shody se použije nejdelší prefix

BGP

- Základní vlastnosti
 - Path Vector směrování
 - možnost definic politik směrování
 - šíření informací pracuje nad TCP (ostatní protokoly UDP!)
 - používá CIDR
- Path Vectors - podobné DV
 - posílá celé cesty (nejen koncové uzly)
 - lepší detekce cyklů
 - implicitně preferovány kratší cesty (mohou zasáhnout politiky)
 - pouze dostupnost, žádná další metrika

BGP (2)

- Politiky
 - kombinace nejlepších lokálních pravidel nemusí představovat globální optimum
 - asymetrie cest
 - obchodní rozhodnutí
 - lokální rozhodnutí definují
 - výběr cesty
 - zveřejnění interních podsítí
- „Kdo má prioritu: řídicí nebo uživatelská data?“

TCP

- Transmission Control Protocol
 - teď jen stručně, podrobněji (přes)příště
- Služba s definovanými parametry
 - pokud je spojení, tak žádné ztráty, přeuspořádání či poškození dat
- Dělení na segmenty (MSS = Maximum Segment Size)
- Potvrzování
 - kumulativní příp. selektivní (SACKs)
 - piggybacking
 - duplikované potvrzení

TCP (2)

- Jedno potvrzení na 2 segmenty
 - pokud je přenos dost rychlý
 - timeout 500 ms
- Řízení toku pomocí okna
 - řídí pouze odesílající, informace jsou také od přijímajícího (rwnd)

TCP (3)

- 4 algoritmy:
 - Pomalý start (Slow Start)
 - exponenciální nárůst
 - Zábřana zablčení (Congestion Avoidance)
 - lineární nárůst – exponenciální pokles
 - Rychlá retransmise (Fast Retransmit)
 - detekce výpadku pomocí duplikátních potvrzení
 - Rychlé obnovení (Fast Recovery)
 - přeskočení SS po FR

UDP

- Nespojovaná služba
- Prostý přenos paketů s kontrolním součtem
- Nezajištěná
 - odpovědnost na aplikaci
- Hlavička
 - porty odesílatele a příjemce
 - délka
 - kontrolní součet

Výhled na (přes)příště

- Úvod do abstraktní protokolové (AP) notace
 - M. Gouda: Elements of Network Protocol Designs, 1998
- IP a směrování
 - AP notace
 - praktická implementace
- TCP a UDP
 - AP notace
 - praktická implementace