Projekt do SPS

Konfigurace a sledování provozu protokolů PIM pro směrování multicastů

Martin Škařupa, ska132
Ondřej Zábojník, zab045

Potřebná teorie pro zvládnutí projektu

Při unicastu jsou data ze zdroje přijímána pouze jedním příjemcem. V případě anycastu je příjemcem pouze jeden z vybrané skupiny. Při broadcastu přijímají data všichni účastníci a při multicastové komunikaci přijímá data pouze vybraná skupina ze všech účastníků.

Kde se vzal multicast

Dějiny implementace multicastu ve světě Internetu se datují začátkem osmdesátých let minulého století. Tehdy postgraduální student Stanfordské University Steve Deering začal pracovat na distribuovaném operačním systému Vsystem. Tento systém se původně skládal z několika počítačů propojených mezi sebou na jednom ethernetovém segmentu. Tyto stroje si posílaly zprávy pouze pomocí ethernetového multicastu. Růst projektu vyžadoval zapojení více počítačů a naštěstí (tedy teď se dívám z pohledu rozvoje multicastu) jediné volné stroje byly až v jiné budově university. Propojení s tamní sítí bylo realizované pomocí protokolu IP přes routery a tedy přímé posílání zpráv přes ethernet nebylo dále možné.

Ukázalo se tedy nutné převést komunikaci mezi počítači z ethernetových rámců do světa IP packetů. S tím se objevil problém, jak ve světě IP protokolů používat skupinové adresování, multicast. Právě vyřešením tohoto problému byl pověřen Steve Deering. Ve své práci mimo jiné navrhnul způsob, jak multicastové packety adresovat, posílat sítí, navrhl jak rozšířit link-state routovací protokol OSPF o podporu multicastingu (MOSPF), navrhl nový vektorový routovací protokol na bázi RIP (DVMRP) a navrhl základy protokolu IGMP. Své výsledky shrnul ve své doktorské práci "Multicast Routing in a Datagram Network"

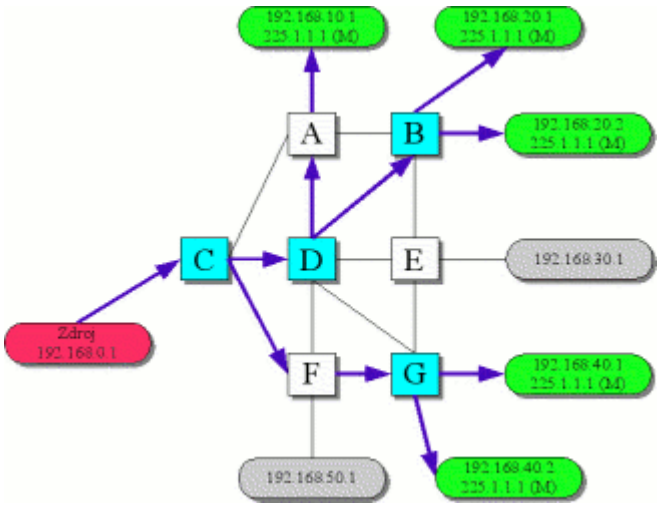
Na jeho dílo navázala řada dalších a již kolem roku 1992 vznikla první experimentální IP multicastová sít MBone. V současné době je multicast podporován především u akademických a výzkumných sítí, komerční poskytovatelé mají v této oblasti obvykle hodně co dohánět. V České republice není situace nijak specifická, o malém používání IP multicastu svedčí i fakt, ze tento druh dat si nevyměňují ani poskytovatelé v rámci propojovacího uzlu NIX.CZ.

Co je uvnitř

Na první pohled se multicastový packet v ničem neliší od běžného unicastového. Jediný rozdíl tkví v tom, že cílová IP adresa je z rozsahu adres D. Znalci si možná pamatují že dříve byly IP adresy rozdělené mezi rozsahy A, B, C, D a E. Po zavedení CIDR (classless inter-domain routing) se rozdíl mezi rozsahy A, B a C smazal, a dnes se použití adres z těchto rozsahů nijak neliší. Nicméně použití adres z D i nadále zůstala zcela odlišné. Tyto IP adresy se poznají tak, že první oktet začíná na v binárním vyjádření na 1110 a tedy adresy leží v rozsahu 224.0.0.0 - 239.255.255.255.

Ačkoliv multicastový packet je zcela stejný jako běžný unicastový, zpracování takového packetu routerem má jednu velmi výraznou odlišnost. U unicastového packetu se router pouze jednoduše rozhodne, kterým směrem ho (pokud vůbec) přepošle. V případě multicastového routování může být dalších destinací více a router musí v takovém případě takový packet zreplikovat. To přirozeně routery trochu více zatěžuje a zároveň je nutné velmi pečlivě

sledovat, zda-li se toto množení neděje nadměrně, aby protistrana nedostávala svá data zbytečně ve více kopiích a neplytvalo se tak s pásmem. Příklad průběhu packetu sítí znázorňuje následujíci obrázek.



Na tomto příkladu se routery B, C, D, G musely postarat o duplikaci packetu.

Aby síť věděla, kam má packet směřovaný na nějakou multicastovou adresu posílat, je nutné, aby se koncový stroj k používání takové adresy zaregistroval. Routery pak musí být schopny spočítat optimální cestu pro průchod a duplikaci packetů pro všechny přihlášené.

K čemu je tedy IP multicasting tedy vhodný? Především se zcela dokonale hodí pro aplikace, kdy jeden zdroj posílá velkému množství klientů stejná data, tedy hlavně jde o internetové vysílání rozhlasu či televize. Zde jsou výhody zřejmé. Vysílač posílá své vysílání pouze multicastové skupině a o duplikace se stará síť v místě, kde to je optimální. Výrazně se tím zmenšuje zatížení sítě a také zátěž vysílacího stroje. Například zmíněná burza tímto způsobem distribuuje data svým klientům. Poněkud komplikovanějším případem předchozího jsou konferenční hovory. Zapojení více uživatelů do hovoru vyžaduje v unicastovém prostředí buď N^2 vzájemných relací mezi účastníky nebo v lepším případě musí existovat centrální server, který se duplikaci dat stará. I on může být přetížen či připojen nedostatečnou kapacitou k síti. Použitím IP multicastu tento problem odpadá.

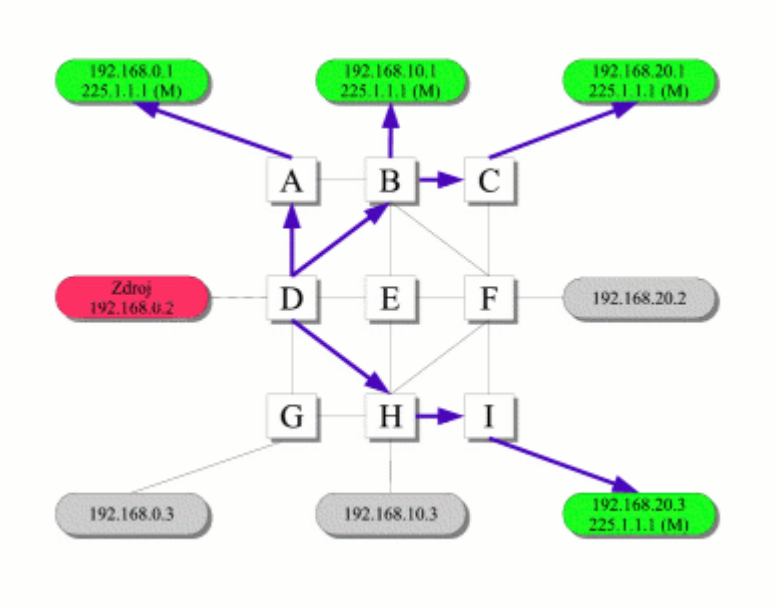
Nicméně IP multicast není žádný univerzálně použitelný prostředek. K jeho největší slabinám patří fakt, že není schopen provádět spolehlivý přenos dat, tedy není možné přes něj přenášet protokol TCP. Problém tkví právě v té spolehlivosti. TCP striktně vyžaduje opětovné posílání dat, která se ztratila po cestě. Síť by si tedy musela pamatovat packety, které již přenesla a to by neúměrně zatěžovalo routery, jiný problém by byl s přenosem potvrzení od doručení od všech příjemců. Aplikace využívající IP multicast jsou tedy převážně založené na protokolu UDP a se ztrátou informací prostě musí počítat.

Abych nemluvil jen pouze abstraktně, raději uvedu konkrétní případy aplikací, které multicast využívají. Právě díky projektu MBone dostali uživatelé pěkné dárky. Prvním je aplikace Robust Audio Tool (RAT), což je ukázkový příklad využití multicastu. Tento open source software použitelný na mnoha platformách zprostředkovává konferenční hovory. O video se pak stará příbuzná aplikace Videoconferencing Tool (VIC). Příjemnou zprávou pro internetovou většinu, jež k multicastu přístup nemá, je fakt, že obě aplikaci dokáží fungovat i v unicastovém prostředí, nicméně pochopitelně zde mají vyšší spotřebu konektivity.

Nezůstávejme ale pouze líbivých okeních aplikacích. Multicast může zjednodušit například i časovou synchronizaci. UNIXový démon NTP se je schopen synchonizovat s počítači v multicastové skupině. Správa takového systému je pak jednodušší. Například přidání dalšího časového serveru pak obnáší pouze jeho přihlášení do multicasatové skupiny. Vyhledávání okolních serverů si již pak vezme na starost síť.

Zdrojový strom

Tomuto stromu se také někdy říká strom nejkratších cest (Shortest Path Tree - SPT) a je to strom, jehož kořenem je zdroj multicastových dat, a listy tohoto stromu jsou příjemci tohoto vysílání. Příklad takového stromu vidíte na následujícím obrázku.

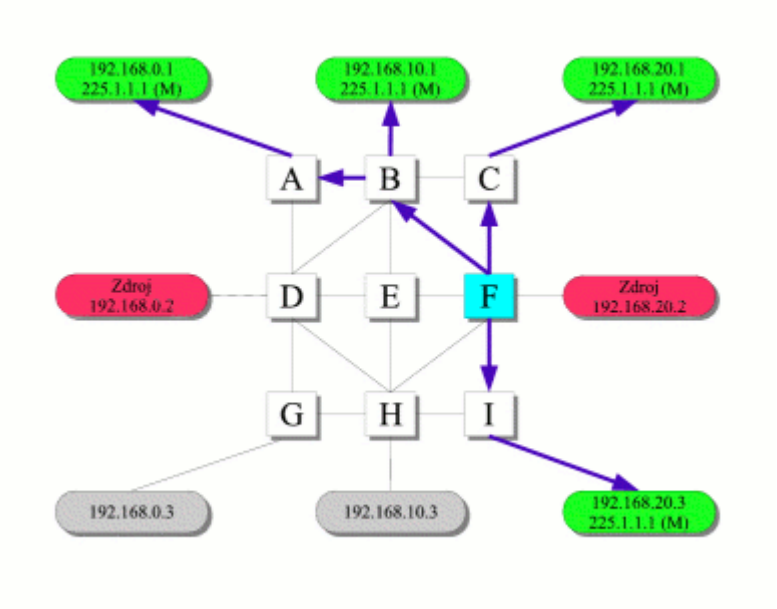


Source Tree

Pro označení stromu se obvykle používá notace (S, G), což se anglicky čte jako "S comma G", kde S je adresa zdroje a G je adresa multicastové skupiny. Náš příklad na obrázku bychom označili jako (192.168.0.2, 225.1.1.1). Je tedy zřejmé, ze pro různé zdroje posílající data do stejné multicastové skupiny existují různé zdrojové stromy. Někdy se také tento způsob práce s multicastem označuje source specific multicast (SSM).

Sdílený strom

Oproti zdrojovým stromům mají sdílené stromy kořen vždy na jednom místě, nezávisle na tom, kdo data posílá. Tomuto kořenu se obvykle říka rendesvous point (RP), a proto se také tyto stromy někdy nazývají RP stromy. Příklad takového stromu máme na následujícím obrázku.



Shared Tree

Jako označení pro tyto stromy se používá notace (*, G), anglicky "star comma G". Hvězdička označuje, že strom není závislý na zdroji multicastu. Náš příklad by se dal označit jako (*, 225.1.1.1). Jak je vidět, router F tvoří kořen stromu. Aby situace nebyla tak jednoduchá, tyto stromy jsou dvojího druhu: jednosměrné a obousměrné. Distribuce multicastu v obousměrném stromu probíhá tak, že zdroj data vysílá směrem ke kořenu a zároveň po směru stromu k listům. U jednosměrného se data pošlou unicastem ke kořenu a ten se o jejich distribuci již postará.
Tento způsob chápání multicastu má také označení ASM (any source multicast).

Nyní se vraťme k praxi a podívejme se na používané routovací protokoly. Lze je rozdělit do tří základních skupin: Link state, Dense mode a Sparse mode.

Dense mode protokoly

Protokoly této skupiny používají zdrojové stromy k doručování SSM a pracují na tzv. push principu. Tento princip předpokládá, že každý subnet má příjemce (S, G) multicastového provozu a tento provoz je tedy primárně přenášen všude. Aby se zabránilo plýtvání pásma, listy stromu, které nemají žádné příjemce pro předmětnou skupinu, pošlou směrem ke kořenu tzv. prune zprávu. Směr, ze kterého taková zpráva přišla, je pak ve stromě ořezán, až zůstanou pouze větve, které mají některé aktivní posluchače. Prune zpráva platí pouze po omezenou dobu, takže po chvíli je ji nutné opět obnovit, jinak začne nadřazený router data posílat. Naopak ve chvíli, kdy se objeví nějaký nový posluchač dané skupiny, pošle router svému nadřazenému zprávu graft a data začnou téct okamžitě.

Mezi dense mode protokoly patří například DVMRP a PIM-DM.

DVMRP (Distance Vector Multicast Routing Protocol) je jedním z nejstarších protokolů, navrhnul ho sám Steve Deering a svou koncepcí velmi připomíná unicastový RIP. Jeho přesný popis je v RFC-1075. Poměrně netypicky si sám sestavuje i unicastovou routovací tabulku a dělá to podobně jako RIP. V minulosti byl velmi využíván v síti mbone. A podobně jako RIP se v dnešní době používá už spíše z důvodů kompatibility a ve velmi malých sítích.

PIM-DM (Protocol Independent Multicast - Dense Mode) je protokol o něco mladší. Nesestavuje si sám unicastovou routovací tabulku, ale místo toho využívá tabulku od nějakého unicastového protokolu. Slůvko independent v názvu označuje, že PIM umí spolupracovat s libovolným protokolem, tedy třeba i se statickým routingem apod.

Sparse mode protokoly

Mezi sparse mode protokoly patří například PIM-SM (Protocol Independent Multicast) nebo CBT (Core Based Trees). Tyto protokoly využívají sdílené stromy pro distribuci multicastových dat a využívají tzv. pull model. Tento model předpokládá, že data se nesmějí posílat do sítě, pokud si je někdo explicitně nevyžádá. Pokud se tedy některý stroj chce připojit do multicastové skupiny, jeho příslušný router pošle zprávu join směrem ke kořenu stromu. Takto je sestavena další větev stromu a data můžou proudit. Platnost zprávy join je časově omezena, takže je nutné ji obnovovat. Pokud už v dané větvi není žádný příjemce skupiny, router pošle zprávu graft, která větev uřízne.

Důvodem pro zavedení sparse mode protokolů byla především snaha šetřit výpočetní výkon routerů. V případě, že je více přispěvovatelů do skupiny, musí dense mode protokoly počítat strom pro každý zvlášť. To u sparse mode protokolů odpadá, ale na oplátku může být jejich routing méně efektivní, a je tedy nutné velmi pečlivě volit RP.

CBT se nikdy velikého rozšíření nedočkal, naopak PIM-SM má pár implementací a je běžně v multicastových sítích používán.

Link state protokol

Jediným známým zástupcem link state protokolů je **MOSPF** (Multicast Open Shortest Path First), který je definován v RFC-1584. Tento protokol je modifikací velmi používaného unicastového protokolu OSPF. Pro distribuci multicastu také používá zdrojové stromy, nicméně nepoužívá prune/graft zprávy. Místo toho si routery vyměňují informace o rychlosti (zjednodušeně řečeno, jde o metriky, které by to měly vyjadřovat) a stavu linek vedoucích k příjemcům multicastových skupin. Každý router si tyto informace ukládá do své databáze a z ní si pak sám sestaví pro každou (S, G) skupinu strom. Při každé změně stavu nějaké linky v síti je nutné tyto stromy přepočítat. Vzhledem k tomu, že těchto skupin může být potenciálně velmi mnoho, nároky na routery mohou být v nejhorším případě značné. Snad právě z tohoto důvodu se tento protokol téměř neujal.

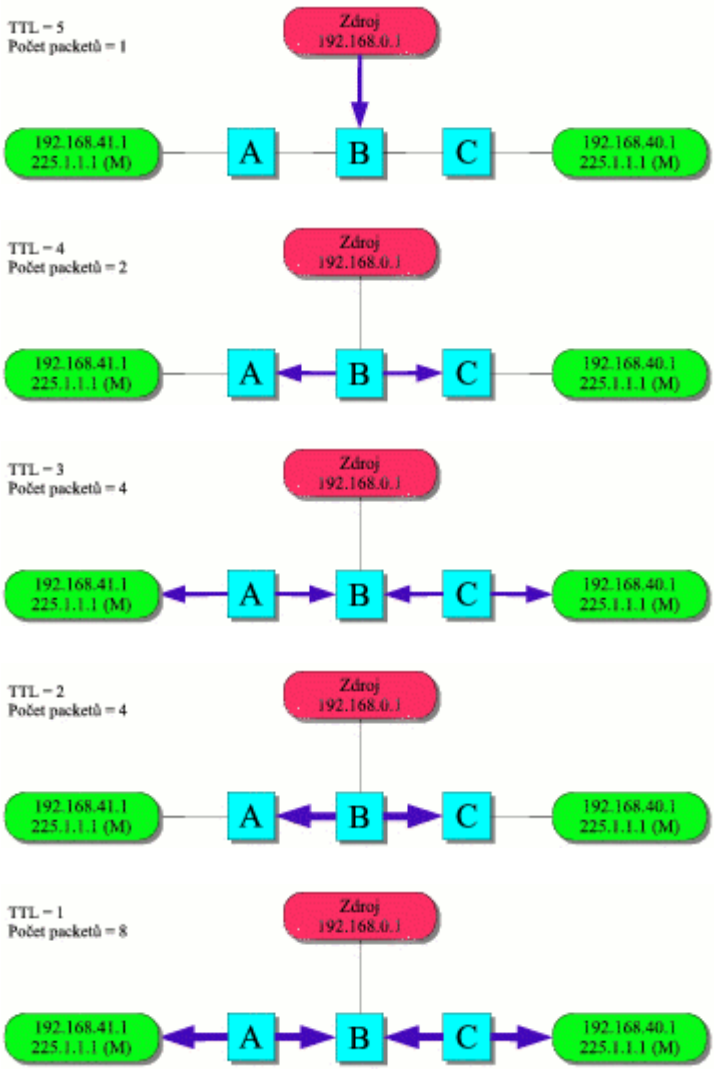
Za zmínku stojí ještě dva protokoly, a těmi jsou MSDP (Multicast Source Discovery Protocol) a BGMP (Border Gateway Multicast Protokol). Stojí trochu stranou našeho výčtu, protože se nejedná o běžné protokoly. Úkolem MSDP je umožňovat propojení více PIM-SM domén mezi sebou. Každá si přitom zachová svůj vlastní RP. Jeho definice je uvedena v RFC-3618. Naopak BGMP se více hodí pro propojování SSM domén, tedy hlavně DVMRP, MOSPF a PIM-DM. Nicméně umí i ASM, pouze vyžaduje, aby se jedna doména stala kořenem globálního stromu a ostatní s ní byly asociované. BGMP je popsán v RFC-3913.

Situace v unicastovém světě je poměrně jednoduchá. Router si pomocí dynamického routovacího protokolu nastaví routovací tabulku nebo mu ji v jednodušších připadech nastaví ručně operátor. Pokud tomuto routeru přijde packet na libovolný interface, sníží tomuto packetu nejprve parametr TTL (Time To Live) o jedna, a pokud je tento parametr stále vyšší než nula, prozkoumá routovací tabulku. Pokud v ní pro cílovou adresu packetu najde odpovídající řádku, přepošle packet na příslušný interface následujícímu routeru (next hop). Už jen fakt, že při multicastu se packety často zdvojují, dává tušit, že situace zde nebude zdaleka tak jednoduchá.

Hlavní odlišnost je v tom, že s cílovou adresou packetu router nevystačí. Jeho rozhodnutí co dál je ovlivněno ještě tím, jakou má ten packet zdrojovou adresu a z jakého interface mu přišel.

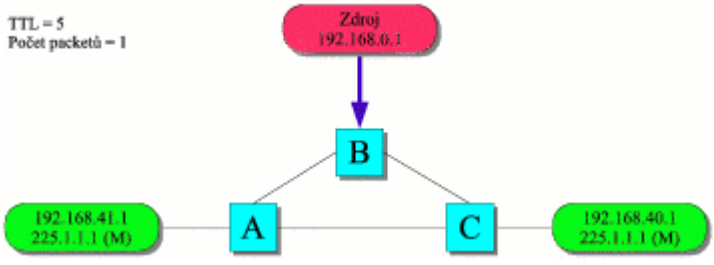
Proč je to tak komplikované? Jak již bylo zmíněno minule, při multicastovém forwardování je nutné ošetřit více věcí. Protože se packet duplikuje, je nutné pečlivě hlídat, aby těchto duplikací nebylo příliš a aby packet k příjemci došel pouze v jednom exempláři a nikoliv vícekrát. Pokud by se toto dělo, zátěž sítě by mohla být i vyšší než u obyčejného unicastu.

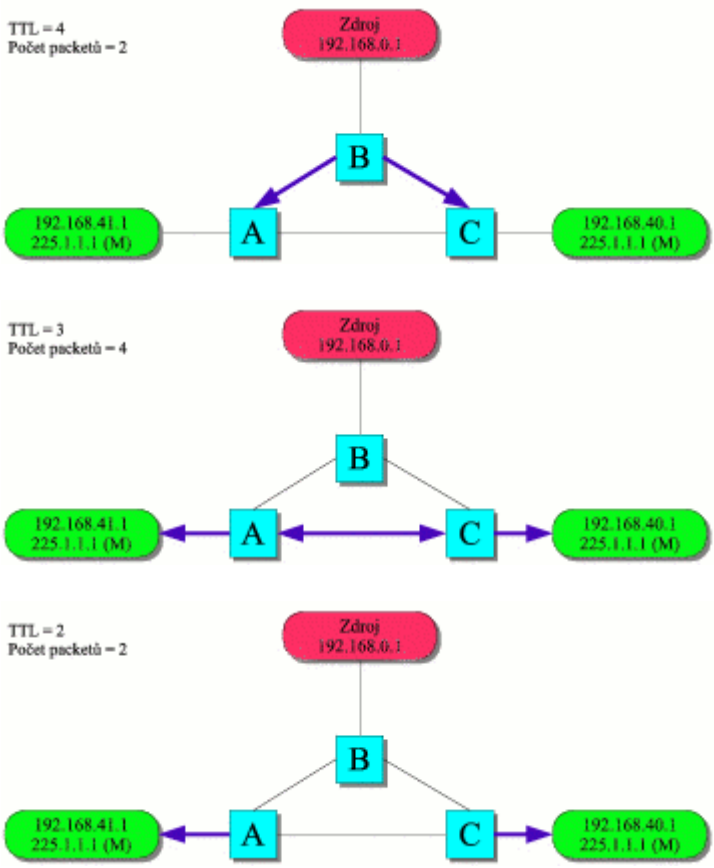
Jak by se to mohlo stát? Pokud bychom to nehlídali, tak poměrně snadno. Představme si model velmi podobný unicastovému světu, ve kterém by si každý router postavil pouze tabulku destinací, kde se nacházejí příjemci dané multicastové skupiny. Pokud by mu někdo zaslal packet určený pro takovou skupinu, prostě by jej jen rozeslal na všechny interface, kde se takoví příjemci nacházejí. Následující sekvence obrázků ilustruje, jak by taková situace mohla vypadat.



Jak vidíte, packet se neustále množí. Díky neustálému snižování parametru TTL (Time To Live) by se sice nemnožil navěky, ale rozhodně by spotřeba pásma byla několikanásobně vyšší, než je nutné. Stejně tak koncová zařízení dostanou svou zprávu několikrát, což se zbytečně zatěžuje.

Mnoho lidí jistě napadne jednoduchá úprava, že by tedy router neposílal packet zpět na interface, ze kterého ten packet obdržel. Taková situace je o něco lepší - k nadměrné duplikaci by v některých jednoduchých sítích nedocházelo, ale obecně nám to nepomůže. Tento příklad ilustruje následující sekvence.





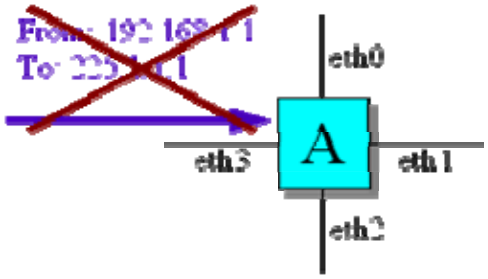
Jak je vidět, stále to ještě není ono. Sice nám v tomto primitivním konkrétním případě packety vymizely dříve, než zasáhl mocný mechanismus TTL, nicméně každý si jistě umí představit topologii, kde tomu tak nebude. Navíc i zde bylo zatížení sítě vyšší, než je nutné, a opět jsou koncová zařízení otravována duplicitními packety.

Proto byl vymyšlen způsob, kterému se říká Reverse Path Forwarding (RPF). Router si stále drží unicastovou routovací tabulku. Tuto tabulku získává buď od operátora, z klasických routovacích protokolů (nebo jejich odvozenin, jako je MBGP), případně si ji sestavuje pomocí speciálních multicastových protokolů (například DVMRP). Často může jít i o více tabulek dohromady.
Pokud pak routeru přijde multicastový packet, zjistí si jeho zdrojovou adresu a interface, ze kterého přišel.
Potom otestuje, zda-li by stejným směrem poslal i unicastový packet, jehož cílová adresa by byla rovna té zdrojové u multicastového. Pokud test neprojde, packet je zahozen, v opačném případě packet pokračuje k dalším zpracování.

Raději se na to podívejme na příkladě. Představme si router A s následující routovací tabulkou:

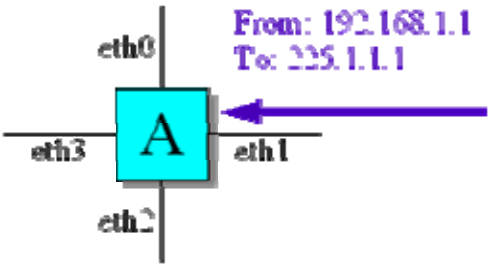
Síť	Maska	Interface
192.168.0.0	255.255.255.0	eth0
192.168.1.0	255.255.255.0	eth1
192.168.2.0	255.255.255.0	eth2
192.168.3.0	255.255.255.0	eth3

Tento obrázek ukazuje situaci, kdy RPF test (anglicky RPF check) neuspěje:



Jak je vidět, packet se zdrojovou adresou 192.168.1.1 a cílovou adresou 225.1.1.1 přijde přes interface eth3. Unicastová routovací tabulka ale říká, že packety pro 192.168.1.1 by se posílaly přes interface eth1. Packet je tedy zahozen.

Naopak tento obrázek ukazuje situaci příznivější:



Packet přijde z interface, ze kterého je očekáván, a v dalším kroku tedy bude za odměnu rozeslán dle příslušné multicastové tabulky.

Router přirozeně nedělá toto rozhodnutí s každým packetem, to by ho velmi zatěžovalo. Obvykle je takto zkontrolován pouze první packet a záznam o tomto testu je uložen do cache. Testování následujících packetů je pak již mnohem rychlejší. Tuto cache je nutné obnovovat při změně unicastové routovací tabulky.

Jak již zde bylo zmíněno, pro adresaci IP multicastových skupin se používají adresy binárně začínající na 1110, tedy decimálně vyjádřeno 224.0.0.0 - 239.255.255.255. Tento rozsah je ještě dále vnitřně členěn:

- Lokální linkové (Link-Local) adresy jsou z rozsahu 224.0.0.0 - 224.0.0.255. Packety posílané na adresy z tohoto rozsahu nesmí být forwardovány dále a tyto adresy tedy slouží pro adresování stojů v lokální síti. Obvykle mají specifický význam, ktery je zaznamenán v databázi IANA. Jako příklad uvedu jen pár z nich:
  - 224.0.0.1 - všechny multicastové stroje v dané lokální síti,
  - 224.0.0.2 - všechny multicastové routery v dané lokální síti,
  - 224.0.0.5 - všechny OSPF routery v dané lokální síti,
  - 224.0.0.6 - všechny designované OSPF routery v dané lokální síti,
  - 224.0.0.22 - všechny multicastové routery v dané lokální síti, které podporují IGMPv3 (Internet Group Management Protocol).
- Zvláštní rezervované adresy jsou z rozsahu 224.0.1.0 - 224.0.1.255. IANA je obvykle přiděluje významným síťovým službám, jako je třeba 224.0.1.1 pro NTP. Tyto adresy se již forwardují běžně.
- Adresy s administrativně omezeným rozsahem (Administratively Scoped) patří do rozsahu 239.0.0.0 - 239.255.255.255 a měly by se používat v privátních sítích. Jejich

použití je tedy obdobné jako např. u adres 10.0.0.0/8 a podobných, definovaných v RFC-1918.

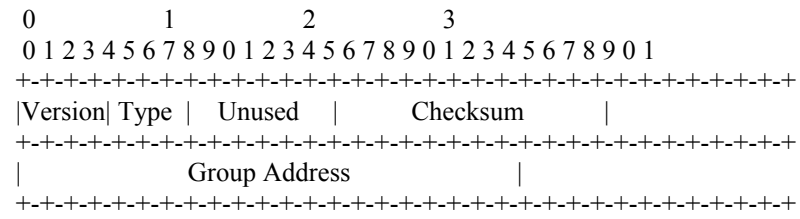
Internet Group Management Protokol

Základem pro IGMP byl Host Membership Protokol, který navrhl Steve Deering ve své doktorské práci. Verzi nula popisuje RFC-988, kdysi hojně používanou verzi 1 Deering popsal v RFC-1112, další pokračování je v RFC-2236 a zatím poslední verzi 3 popisuje v RFC-3376. Tento protokol především používají multicastoví klienti, aby signalizovali routerům své členství v multicastových skupinách. Nicméně není to jen jediné jeho využití, používá ho například i protokol DVMRP pro přenos svých zpráv a pod.

IGMPv1

Možná se zdá být trochu zbytečné popisovat protokol verze 1, který už je poměrně zastaralý, ale bohužel ještě existuje měřitelná skupina (naštěstí se již rychle zmenšující) operačních systémů, které verzi 1 používají. Je tedy nutné tuto verzi znát a rozumět jejím omezením. Ještě mnohem silněji to pochopitelně platí u verze 2.

Formát IGMP zprávy je poměrně jednoduchý. Po IP hlavičce následuje tato sekvence osmi bytů:



Význam jednotlivých políček je následující:

- Version - aktuální verze protokolu (1),
- Type - tato verze podporuje pouze dva typy zpráv:
  - Membership Query, kód 1,
  - Membership Report, kód 2,
- Unused - nepoužito, při posílání by mělo být nastaveno na 0 a při příjmu ignorováno,
- Checksum - kontrolní součet IGMP zprávy,
- Group Address - u zprávy Membership Report obsahuje toto pole multicastovou adresu, jinak by mělo být nastaveno na 0.0.0.0.

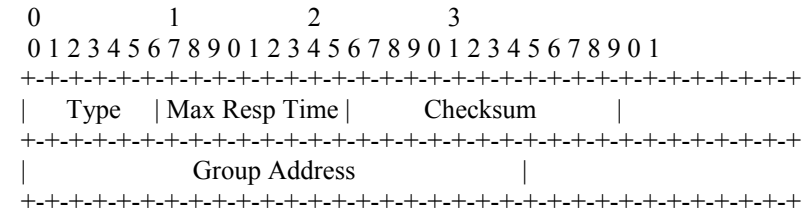
Protokol má jednoduché schéma Query-Report. Pokud se nějaký klient chce připojit k multicastové skupině a pošle zprávu typu Membership Report s vyplněným číslem skupiny na adresu té skupiny, příslušný router by ji měl zachytit a postarat se o zprostředkování. Router si průběžně kontroluje členství klientů ve skupinách a občas (jednou za 60 sekund) pošle zprávy typu Membership Query na adresu 224.0.0.1 a čeká na Reporty klientů. Protože je běžné, že klientů pro jednu skupinu je v síti více, byl navržen mechanismus, který zaručí, že pro každou skupinu dorazí pouze jediný Report. Každý z klientů, který uslyší Query, si zvolí náhodné číslo v rozsahu 0-10. Toto číslo vyjadřuje čas v sekundách, za který chce poslat Report.

Zároveň ale poslouchá, zda nějaký jeho kolega Report pro danou skupinu pošle. Pokud se tak stane, klient už nic neposílá.

V případě, že klient danou skupinu opustí, přestane jednoduše odpovídat na Query. Pokud router nedostane pro skupinu třikrát po sobě žádný Report, přestane ji posílat. Query se posílají obvykle jednou za minutu - v nejhorším případě tedy router posílá data ještě tři minuty po té, co už o ně nikdo nestojí. Tento jednoduchý mechanismus je velmi nevýhodný, pokud klient často mění členství ve skupinách. V takovém případě může síť snadno přetížit.

IGMPv2

Zprávy IGMPv2 vypadají jakoby nekompatibilně s předchozí verzí. Obsahují trochu jiná políčka:



Jejich význam je následující:

- Type - hodnoty totoho pole jsou voleny tak, aby bylo možné rozeznat zprávy IGMPv1, které na tomto místě mají políčka dvě - version a type. Protokol rozeznává následující typy zpráv:
    - Membership Query - vlastně jsou dvě, General Query a Group-specific Query, rozlišení se dělá podle obsahu políčka Group Address, číselný kód zprávy je 11, tedy zpráva vypadá podobně jako IGMPv1 Query,
    - IGMPv1 Membership Report - pro zpětnou kompatibilitu, kód zprávy 12,
    - IGMPv2 Membership Report, kód 16,
    - Leave Group, kód 17,
- Max Resp Time - používá se u Query zprávy a označuje maximální čas na zaslání reportu v desetinách sekundy. Mechanismus je stejný jako u IGMPv1,
- Checksum - kontrolní součet IGMP zprávy,
- Group Address - u zpráv Membership Report, Leave Group a Group-specific Query obsahuje toto pole multicastovou adresu, jinak by mělo být 0.0.0.0. Z toho je vidět, že zpráva General Query u IGMPv2 je až na pole Max Resp Time shodná s Membership Query u IGMPv1.

Signalizace přihlášení do skupiny je obdobná jako u IGMPv1. Novinkou je proces opuštění skupiny. Klient může poslat zprávu (a obvykle posílá) Leave Group. Router na takovou zprávu zareaguje posláním Group-specific Query do dané skupiny. Max Resp time je obvykle nastaven na 1 sekundu. Pokud je ještě někdo členem skupiny, odpoví, a router pokračuje v posílání dat.

Další novinkou je volba routeru, ktery posílá Query. Je-li na síti více routerů, je pro posílání Query vybrán ten s nejvyšším IP. Ostatní pouze poslouchají a jsou připraveni převzít jeho roli. Mechanismus pracuje tak, že pokud router uslyší Query zprávu od routeru s vyšším IP,

přestane ty své sám posílat a čeká 400 sekund, zda nepřijde další. Pokud nepřijde, vyhodnotí, že vybraný server už asi nefunguje, a začne posílat Query sám, čímž proběhnou nové volby. V IGMPv2 se Query posílá jednou za 125 sekund.

IGMPv2 je zpětně kompatibilní s IGMPv1. Pokud klienti detekují IGMPv1 router dle jeho Query, začnou posílat zprávy také v IGMPv1 a nastaví si časovač na 400 sekund. Pokud nastavený interval uplyne, začnou opět posílat IGMPv2. Naopak server je schopen obsloužit mix IGMPv1 i IGMPv2 klientů, protože umí jejich zprávy rozlišit. IGMPv1 klienti nepoznají žadný rozdíl. Pokud router detekuje IGMPv1 klienty na síti, ignoruje Leave Group zprávy.

IGMPv3

Specifikace IGMPv3 dokázala zatím ještě zhruba přehlednou situaci výrazně zkomplikovat. Problém, který se tato verze snaží řešit, je v tom, že pokud jste členem nějaké multicastové skupiny, nechcete často přijímat zprávy od všech, ale pouze od vybraných zdrojů. Proto se v IGMPv3 nepřihlašujete pouze do skupiny (*, G), ale přihlašujete se k odběru dat z konkrétního zdroje v dané skupině (S, G). Přirozeně se výrazně změnil i formát zpráv. Zprávy již nemají konstantní délku, jako tomu bylo u předchozích verzí. Přesný popis formátu by sám zabral nejméně jeden článek, takže jej vypustíme. IGMPv3 má své dvě vlastní zprávy a pro zpětnou kompatibilitu rozumí dalším třem:

- Membership Query - kód 11,
- IGMPv3 Membership Report - kód 22,
- IGMPv1 Membership Report - kód 12,
- IGMPv2 Membership Report - kód 16,
- IGMPv2 Leave Group - kód 17.

Význam zpráv je stejný jako u předchozích verzí, pouze se přidávají políčka pro členství ve skupinách. Také se trochu změnil význam políčka Max Resp Time. Nyní se jmenuje Max Resp Code a pokud je jeho hodnota nižší než 128, význam se nemění. Pokud se ale používá vyšší hodnota (první bit je 1), změní se význam na:

```
 0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+-+
|1| exp | mant  |
+-+-+-+-+-+-+-+-+
```

Max Resp Time = (mant | 0x10) << (exp + 3)

IGMPv3 Report se také neposílá do příslušné skupiny, ale na adresu 224.0.0.22.

Prvních 8 bytů Query zprávy je stejných jako u předchozích verzích, takže klienti s nižší verzí protokolu nepoznají rozdíl a odpoví zprávou Report v odpovídající verzi. Router pak zachází s každou skupinou dle verze přihlášených klientů. Pokud se tedy do stejné skupiny na stejné síti přihlásí IGMPv2 i IGMPv3 klient, router nebude provádět filtrování zdrojových adres a bude posílat všechna data skupiny. Naopak pokud klient dostane zprávu délky 8 bytů, ví, že jeho router používá IGMP nižší verze, a přizpůsobí se mu.

Multicastové ethernetové rámce se od těch běžných poznají celkem snadno. Cílová adresa takového rámce má nastaven nultý bit oktetu číslo 0 na jedničku. Tedy:

```
 oktet 0  oktet 1  oktet 2  oktet 3  oktet 4  oktet 5
+--------+--------+--------+--------+--------+--------+
|xxxxxxx1|xxxxxxxx|xxxxxxxx|xxxxxxxx|xxxxxxxx|xxxxxxxx|
+--------+--------+--------+--------+--------+--------+
 76543210 76543210 76543210 76543210 76543210 76543210
```

Speciálním případem je broadcastový packet, který má všechny bity nastavené na 1, tedy FF:FF:FF:FF:FF:FF.

Konkrétně v případě IP multicastu adresa odpovídajícího rámce začíná 25bitovým prefixem:

```
 oktet 0  oktet 1  oktet 2  oktet 3  oktet 4  oktet 5
+--------+--------+--------+--------+--------+--------+
|00000001|00000000|01011110|0xxxxxxx|xxxxxxxx|xxxxxxxx|
+--------+--------+--------+--------+--------+--------+
 76543210 76543210 76543210 76543210 76543210 76543210
```

tedy 01:00:5E a jeden bit. Pro mapování IP adresy do takovéto MAC adresy se používá zbylých 23 bitů. Jak se taková věc provede? Jak už jsme si říkali, každa multicastová IP adresa začíná bitovým prefixem 1110. Na rozlišení nám tedy zbývá 28 bitů. Mapování do MAC adresy nelze udělat beze ztráty a provede se tak, ze se prvních pět bitů z 28 uřízne a konec IP adresy se uloží do zbytku MAC adresy.

Tedy například IP 224.1.1.1 odpovídá MAC adresa 01:00:5E:01:01:01. Tato MAC adresa ovšem náleží i kupříkladu IP 224.129.1.1 nebo 225.1.1.1 a pod. Do jedné MAC adresy se tedy mapuje 2^5 = 32 IP adres.

Proč?

Je to poměrně zajímavý příběh. Na začátku devadesátých let, když Steve Deering pracoval na své práci o IP multicastingu, chtěl od IEEE delegovat 16 sousedních OUI (Organizational Unique Identifier - 24 bitový prefix z MAC adresního prostoru). Tím by získal prostor 28 bitů (každý OUI má 24 bitů a počet 16 přidá další 4 bity) a mohl by mapovat IP multicastové adresy jednoznačně. Tehdy ovšem stálo přidělení každého OUI $1000. Jeho nadřízený, Jon Postel, bohužel nebyl schopen takovou sumu vyplatit, a tedy koupil pouze jeden OUI a Deeringovi ještě přiřadil pouze jednu polovinu. Díky tomu tedy teď mapujeme více IP multicastových adres do jedné multicastové MAC adresy.

Pokud tedy nějaký počítač v síti chce poslat packet do nějaké multicastové skupiny, vypočte si příslušnou MAC adresu a v ethernetové síti takový packet pošle v odpovídajícím rámci. Příjemce zprávy už podle svého členství ve skupinách a cílové MAC může určit, bude-li rámec ignorovat, nebo jej přijme. Nicméně díky nedokonalému mapování se může stát, že rámec je předán vyšší vrstvě i v případě, kdy pro tento stroj určen není. Žádný zásadní problém kromě zbytečného vytěžování CPU to přirozeně nezpůsobuje, packet je odhalen a zahozen.

Starší a méně sofistikované ethernetové swiche obvykle rámce s IP multicastovými packety rozešlou na všechny své porty stejně jako broadcast. Tento způsob zacházení s IP multicastem je sice poměrně logický a jednoduchý, nicméně v rozsáhlejších L2 strukturách či při větších tocích může způsobovat značné problémy se zahlcováním sítě. Sofistikovanější switche tedy mívají způsob, jak IP multicast rozesílat pouze na porty, na kterých je očekáván. Způsobů, jak toho docílit, je více, nejrozšířenější z nich je IGMP snooping.

Jak už název napovídá, tato technika zkoumá obsah IGMP zpráv. Switch, který toho je schopen, tedy musí umět nejen zpracovávat L2 rámce, ale musí částečně rozumět i L3 packetům. Dále si musí udržovat ve své (asociativní) paměti tabulku IP multicastových MAC adres se seznamem portů s účastníky přihlášenými do odpovídající multicastové skupiny (a tedy vlastně odpovídajících IP multicastových skupin).

Jak to tedy pracuje?

Switch zpočátku rozesílá IP multicastový provoz pouze připojeným routerům a poslouchá veškeré IGMP zprávy. Pokud od nějakého počítače uslyší IGMP Report, přiřadí si záznam o jeho portu do paměti a začne mu posílat data pro příslušnou multicastovou skupinu. Víše bylo zmíněno, že IGMP má mechanismus, který zajišťuje potlačení zbytečných zpráv typu IGMP Report. Každá taková zpráva je za normálních okolností do sítě zaslána pouze jedním z přihlášených. Ostatní ji slyší a již dál nic neposílají. V takovém případě by ale switch nevěděl o všech přihlášených a některým by neposílal data. Z tohoto důvodu switch nepřepošle IGMP Report zpět do sítě, takže na IGMP Query donutí odpovědět všechny účastníky, protože si každý z nich myslí, že je jediným příjemcem. Pouze jeden IGMP report je poslán routerům, aby dále posílaly data pro skupinu.

Naopak, pokud se chce některý počítač odhlásit ze skupiny, pošle zprávu IGMP Leave. Switch si nemůže být jistý, zda na tom jednom portu neměl připojeno více účastníků skupiny, a tak okamžitě pošle zpět na port zprávu IGMP General Query. Pokud na portu další zájemce o příslušnou skupinu není, switch ji přestane na daný port posílat. Pokud již skupinu nepřijímá žádný klient, switch přepošle zprávu IGMP Leave i routerům.

V běžném provozu, tedy pokud zrovna nikdo neopouští skupinu či se do ní nehlásí, posílá jeden z routerů periodicky zprávy IGMP Query. Switch každou zprávu přepošle na všechny porty a odpovědi (IGMP Report) si naopak nechává pro sebe a pouze si z nich občerstvuje svou tabulku. Nakonec pošle pouze jednu zprávu na porty, na které jsou připojeny routery.

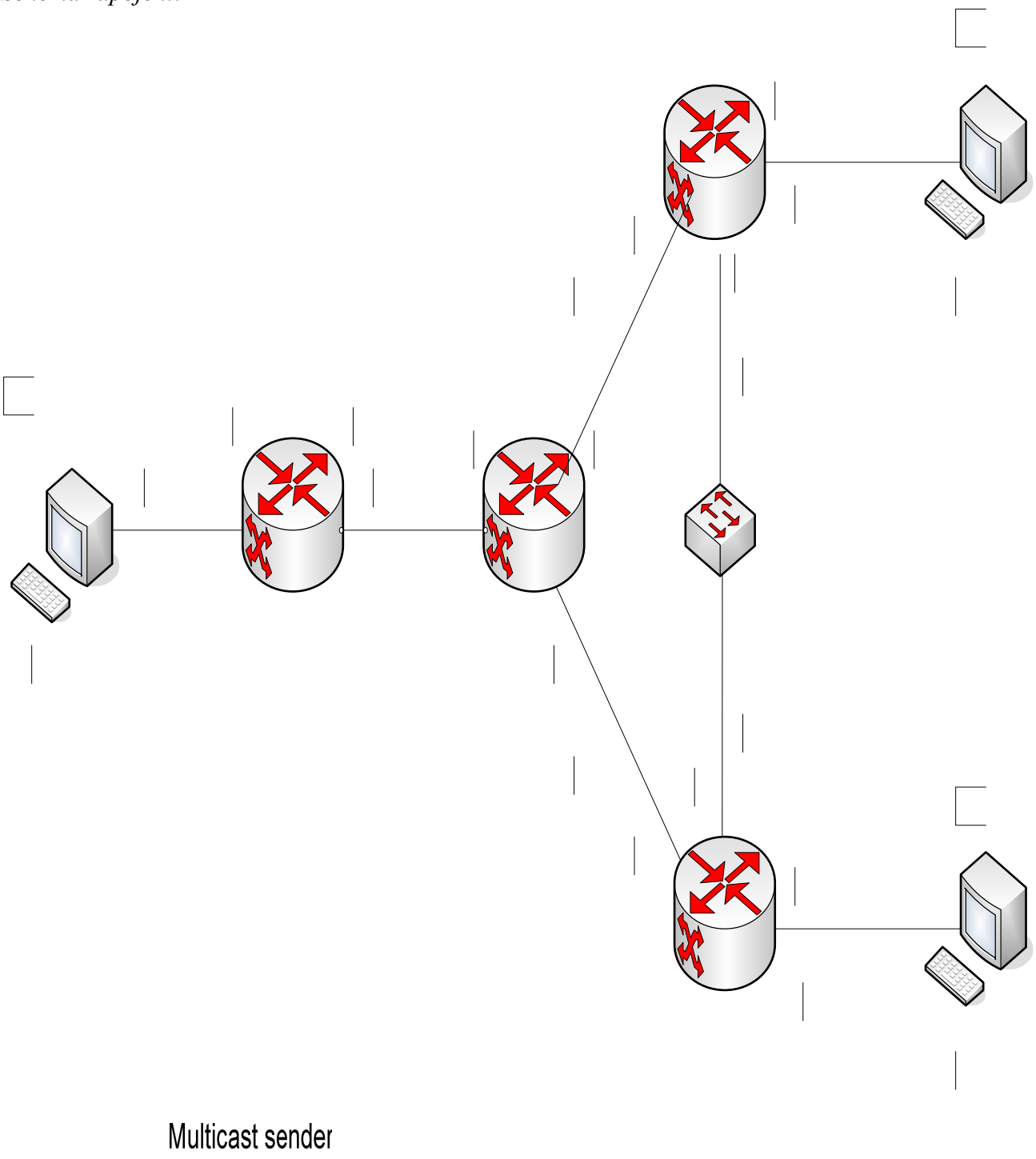
Nyní už zbývá vysvětlit poslední nejasnost. Jak switch pozná, na které porty jsou připojené routery? Pouze podle toho, že router posílá IGMP Query, to bohužel nejde, protože v lokální síti je zvolen pouze jeden router, který tyto zprávy posílá, a ostatní "mlčí". Switch si tedy musí poradit jinak. Obvykle sleduje nejen IGMP zprávy, ale poslouchá i ostatní IP packety, a pokud například uslyší packety protokolů PIM, DVMRP, OSPF, CGMP apod., pozná, že na příslušném portu je připojen router.

Jak je vidět, na druhou vrstvu nebylo příliš při návrhu IP multicastingu pamatováno a řešení v této oblasti byla až následně "uplácána“.

Schéma zapojení multicastové sítě

Cílem projektu je zístat ucelené informace o provozu muticastového protokolu PIM

Schéma zapojení:



Nastavení routerů a PC

před vlastní konfigurací se musíme přepnout do režimu *enable,* který nám umožňuje práci s rozhraními.

Nastaveni konfigurace IP adres a rozhraní na routerech
– např. na routeru RA

```
conf t                              // konfigurační režim
interface eth0                      // nastavení rozhraní eth0
ip address 10.0.4.1 255.255.255.0   // nastavení IP adresy a masky podsite
no shutdown                         //
exit                                // ukončení nastavování interface eth 0

interface eth1                      // nastavení rozhraní eth1
ip address 10.0.1.1 255.255.255.0   // nastavení IP adresy a masky podsite
no shutdown                         //
exit                                // ukončení nastavování interface eth1
exit                                // ukončení konfiguračního režimu
```

Konfigurace RIPu na routerech
– např. na routeru RA

```
conf t                  // konfigurační režim
router rip              // nastavení prtokolu RIP
network     2.0.1.0     // sit
network     2.0.4.0     // sit
                        // nastavuje se pro všechny rozhraní smerovač
```

Konfigurace PC připojeného na routery
– např. na routeru RA

```
ifconfig eth0 10.0.4.100 netmask 255.255.255.0   // nastavení rozhraní a masky
route add default gw 10.0.4.1                    // nastavení implicitní brány,
                                                 // kde budou posílány všechny
                                                 //  směřující pakety s neznámou sítí
```

Konfigurace protokolu PIM na routerech – protokol *dense mode* – např. na routeru RA

```
conf t
ip multicast-routing                // konfigurace multicatingu

interface eth0                      // nastavení rozhraní eth0
ip pim dense-mode                   // nastavení hustého módu

interface eth1                      // nastavení rozhraní eth1
ip pim dense-mode                   // nastavení hustého módu
```

Konfigurace protokolu PIM na routerech – protokol *sparse mode* – např. na routeru RA

```
conf t
ip multicast-routing                // konfigurace multicatingu

interface eth0                      // nastavení rozhraní eth0
ip pim sparse-mode                  // nastavení řidkého módu

interface eth1                      // nastavení rozhraní eth1
ip pim sparse-mode                  // nastavení řidkého módu
ip pim rp-address 10.0.1.1          // nastavení RP
```

<u>Sledování na routerech</u> – příklad sledování na routeru RA:

Příkazy pro sledování nastavení routerů v rámci multicastu.
Pro všechny routery jsou tyto příkazy stejné a vypisy u všech routerů jsou obdobné.

Zobrazení multicastové skupiny a rozhraní, na kterém se vyskytuje

```
RA#sh ip igmp groups
IGMP Connected Group Membership
Group Address    Interface        Uptime    Expires   Last Reporter
224.0.1.40       FastEthernet0/0  00:39:32  00:02:16  10.0.7.2
224.0.1.40       Serial0/2/0      00:40:33  stopped   10.0.2.1
```

Zobrazení sousedních členů

```
RA#sh ip pim neighbor
PIM Neighbor Table
Mode: B - Bidir Capable, DR - Designated Router, N - Default DR Priority,
      S - State Refresh Capable
Neighbor         Interface        Uptime/Expires      Ver   DR
Address                           Prio/Mode
10.0.2.2         Serial0/2/0      00:40:45/00:01:22   v2    1 / S
10.0.7.2         FastEthernet0/0  00:39:45/00:01:30   v2    1 / DR S
```

Zobrazení routovací tabulky

```
RA#sh ip mroute
IP Multicast Routing Table
Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,
       L - Local, P - Pruned, R - RP-bit set, F - Register flag,
       T - SPT-bit set, J - Join SPT, M - MSDP created entry,
       X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,
       U - URD, I - Received Source Specific Host Report,
       Z - Multicast Tunnel, z - MDT-data group sender,
       Y - Joined MDT-data group, y - Sending to MDT-data group
Outgoing interface flags: H - Hardware switched, A - Assert winner
 Timers: Uptime/Expires
 Interface state: Interface, Next-Hop or VCD, State/Mode
```

Zobrazení dense-mod,příchozí+odchozí rozhraní

```
(*, 224.0.1.40), 00:41:00/00:02:45, RP 0.0.0.0, flags: DCL
  Incoming interface: Null, RPF nbr 0.0.0.0
  Outgoing interface list:
    FastEthernet0/0, Forward/Dense, 00:39:59/00:00:00
    Serial0/2/0, Forward/Dense, 00:41:00/00:00:00

RA#sh ip mroute active
Active IP Multicast Sources - sending >= 4 kbps
```

Ukázka sledování pomocí Etherealu

Zde je malá ukázka výstupů z aplikace ethereal, pomocí níž sme sledovali provoz sítě.

No.	Time	Source	Destination	Protocol	Info
1	0.000000	Cisco_46:6f:c1	CDP/VTP/DTP/PAgP/U	CDP	Cisco Discovery Protocol
2	0.036957	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
3	0.862615	Cisco_46:6f:c1	Cisco_46:6f:c1	LOOP	Reply
4	1.038385	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
5	2.039827	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
6	2.863156	10.0.4.1	224.0.0.13	PIMv2	Hello
7	3.041261	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
8	4.042704	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
9	5.044134	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
10	6.045572	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
11	7.047018	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
12	8.048456	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
13	9.049895	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
14	10.051334	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
15	10.862966	Cisco_46:6f:c1	Cisco_46:6f:c1	LOOP	Reply
16	11.052791	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
17	12.054250	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
18	13.055714	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678

No.	Time	Source	Destination	Protocol	Info
77	61.124788	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
78	62.126227	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
79	62.877179	10.0.4.1	224.0.0.13	PIMv2	Hello
80	70.876966	Cisco_46:6f:c1	Cisco_46:6f:c1	LOOP	Reply
81	72.239243	10.0.4.100	234.56.78.90	IGMP	V2 Membership Report
82	72.617898	10.0.4.1	255.255.255.255	RIPv1	Response
83	72.921890	10.0.4.100	234.56.78.90	IGMP	V2 Membership Report
84	73.923303	10.0.4.100	234.56.78.90	IGMP	V2 Membership Report
85	80.877304	Cisco_46:6f:c1	Cisco_46:6f:c1	LOOP	Reply
86	82.861031	10.0.4.100	224.0.0.2	IGMP	V2 Leave Group
87	82.862120	10.0.4.1	234.56.78.90	IGMP	V2 Membership Query
88	83.878227	10.0.4.1	234.56.78.90	IGMP	V2 Membership Query
89	84.464953	10.0.4.100	234.56.78.90	UDP	Source port: 1032 Destination port: 45678
90	85.459785	10.0.4.100	234.56.78.90	UDP	Source port: 1032 Destination port: 45678

No.	Time	Source	Destination	Protocol	Info
5	18.718037	10.0.5.1	224.0.0.1	IGMP	V2 Membership Query
6	21.633734	10.0.5.1	224.0.0.13	PIMv2	Hello
7	27.065679	Cisco_0a:f9:b7	Cisco_0a:f9:b7	LOOP	Reply
8	27.650803	10.0.5.1	255.255.255.255	RIPv1	Response
9	31.505386	Cisco_0a:f9:b7	CDP/VTP/DTP/PAgP/U	CDP	Cisco Discovery Protocol
10	37.065333	Cisco_0a:f9:b7	Cisco_0a:f9:b7	LOOP	Reply
11	47.064821	Cisco_0a:f9:b7	Cisco_0a:f9:b7	LOOP	Reply
12	50.960477	10.0.5.1	224.0.0.13	PIMv2	Hello
13	55.220546	10.0.5.1	255.255.255.255	RIPv1	Response
14	57.064423	Cisco_0a:f9:b7	Cisco_0a:f9:b7	LOOP	Reply
15	67.064021	Cisco_0a:f9:b7	Cisco_0a:f9:b7	LOOP	Reply
16	70.591616	Cisco_0a:f9:b7	DEC-MOP-Remote-Con	0x6002	DEC DNA Remote Console
17	77.063562	Cisco_0a:f9:b7	Cisco_0a:f9:b7	LOOP	Reply
18	78.719470	10.0.5.1	224.0.0.1	IGMP	V2 Membership Query
19	80.575343	10.0.5.1	224.0.0.13	PIMv2	Hello
20	81.712515	10.0.5.1	255.255.255.255	RIPv1	Response
21	85.625485	10.0.5.100	234.56.78.90	IGMP	V2 Membership Report
22	85.886285	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
23	86.207942	10.0.5.100	234.56.78.90	IGMP	V2 Membership Report
24	86.887646	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
25	87.063159	Cisco_0a:f9:b7	Cisco_0a:f9:b7	LOOP	Reply
26	87.209391	10.0.5.100	234.56.78.90	IGMP	V2 Membership Report
27	87.889126	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
28	88.890609	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
29	89.892079	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
30	90.893418	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
31	91.503889	Cisco_0a:f9:b7	CDP/VTP/DTP/PAgP/U	CDP	Cisco Discovery Protocol
32	91.894885	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
33	92.896728	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678
34	93.897802	10.0.4.100	234.56.78.90	UDP	Source port: 1031 Destination port: 45678

Závěr:

Vysílač jsme umístili na PC připojené k routru RJ a na PC u routerů RA a RI jsme umístili přijímače.
Po té jsme střídavě vypínali a zapínali vysílač i oba přijímače a sledovali jejich provoz.
Jelikož PC u routeru RA a RI jsou vzhledem k topologii sítě téměř totožné uvedli jsme příklad sledování jen u RA.

Jelikož se nám nepodařilo zcela optimálně zpracovat náš projekt po praktické stránce, přikládáme popis protokolu PIM teoreticky. Tento popis se sestavuje s teorie převzaté na internetu, spolu s přepracovaným manuálem RFC 2362.
Spolu s touto teorii přikládáme i část našeho praktického řešení, konfiguraci routerů a PC, sledování pomocí debugů a ukázky výpisů v etherealu.
Cílem projektu bylo nalézt a popsat chování protokolu PIM v multicastovém režimu, ve kterém se používá. Výklad začíná celkovým zavedením do děje světa muticastů, poté se zaměříme detailněji na protokol PIM a nakonec, jak již bylo řečeno následuje RFC 2362.

Uvedená literatury(odkazy) :

http://www.lupa.cz/clanky/uvod-do-ip-multicastu/
http://www.isdn.cz/clanek.php?cid=3712

Protocol Independent Multicast (PIM)

PIM je pravděpodobně nejrozšířenějšího protokolu pro použití multicastů.
Tento protokol v sobě fakticky kombinuje protokoly dva, které sdílí jen jméno a formát hlavičky ve zprávách. Jde o **PIM Dense Mode (PIM-DM)** a **PIM Sparse Mode (PIM-SM)**. Rozděleny jsou proto, že ani jeden z protokolů nefunguje úplně dobře ve všech možných situacích. PIM-DM je doporučován k použití v rychlých LAN sítích, kde všechny, nebo skoro všechny sítě mají členy nějakých skupin, zatímco PIM-SM je určen pro použití ve WAN sítích, kde členové dané skupiny okupují pouze malou část Internetu.

PIM získal své jméno tím, že pro směrování multicastových paketů využívá unicastovou směrovací tabulku nezávisle na tom jakým protokolem je udržována. Tudíž je možno jej nazvat protokolově nezávislý.

PIM-DM

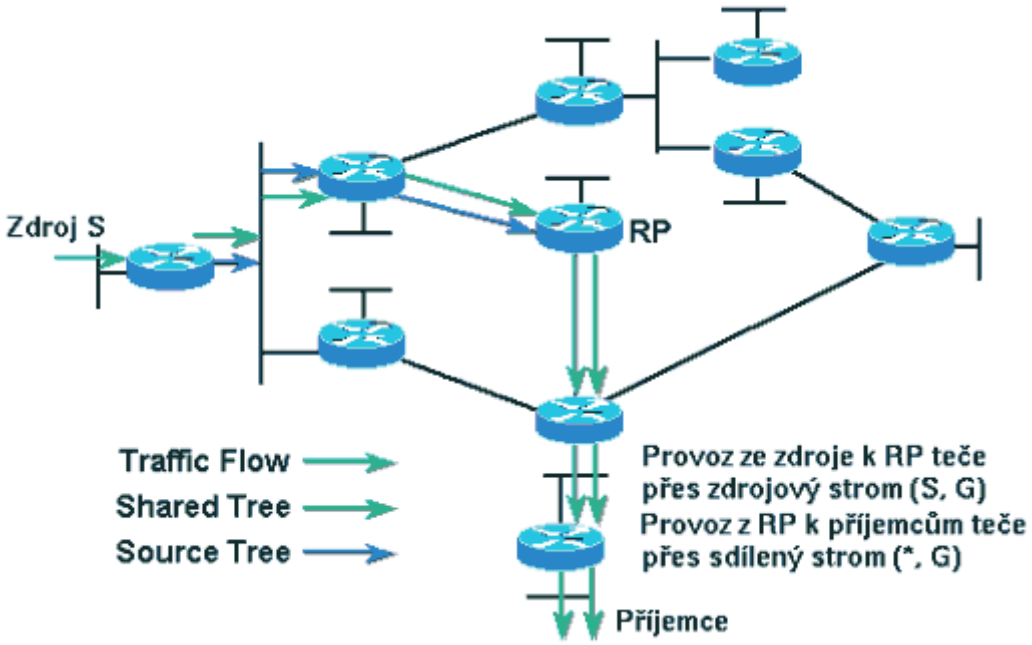
PIM-DM předpokládá sítě s nízkou odezvou, které mají nevyužité pásmo. PIM-DM je svým chováním velice podobný DVMRP, používá totiž techniku flood and prune. PIM-DM předpokládá, že normální směrovací tabulka je správná a podle ní provádí RPF testy příchozích datagramů (na rozdíl od DVMRP, který si udržuje speciální tabulku). Odřezávání nepotřebných větví je podobné jako u DVMRP. Se škálovatelností je na tom PIM-DM podobně, jako DVMRP, ačkoli navrhovaná zlepšující technika State-Refresh, která by eliminovala pravidelné flood and prune cykly, by protokolu trochu pomohla.

PIM-SM

PIM Sparse Mode je řízen poptávkou po datagramech, přitom používá jak sdílené, tak zdrojové stromy. Pro sdílený strom v tomto protokolu existuje kořen, od kterého se veškerý multicastový provoz šíří směrem k příjemcům. Tento kořen se nazývá **Randezvous Point (RP)** a může jich být v síti více (pak ovšem pro konkrétní skupinu je jeden hlavní, ostatní jsou záložní).

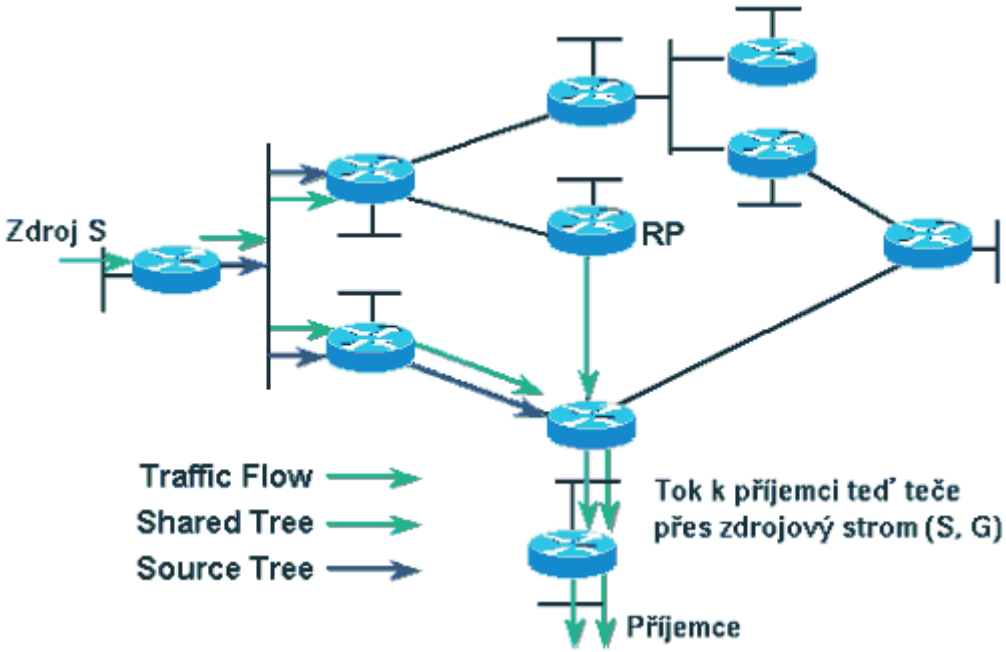
Když hostitel vstoupí do skupiny G, lokální směrovač si do tabulky poznačí záznam (*, G) a informaci šíří dále stromem nahoru směrem k RP. RP rovněž registruje tuto žádost o provoz skupiny G a tím je sdílený strom pro tuto skupinu utvořen.

Jestliže v síti začne zdroj S generovat provoz pro skupinu G, je tento provoz zachycen nejbližším směrovačem a spolu s registrační zprávou je odesílán unicastově na RP. RP se rozhodne podle toho, zda je skupina někým poslouchána, ale vždy pošle onomu směrovači zprávu, že již není třeba multicast balit do unicastových datagramů. Jestliže je ustanoven na RP sdílený strom (*, G), RP se sám přihlásí do skupiny G a požádá o vytvoření zdrojového stromu (S, G). Přijímané multicastové datagramy šíří dále dolů sdíleným stromem (*, G). Popisovanou situaci názorně vidíte na obrázku:



Přepnutí ze sdíleného stromu na zdrojový strom

Silná vlastnost PIM-SM protokolu je přepínání stromů. Pokud by doručování multicastu přes RP a sdílený strom v rámci LAN bylo neefektivní, směrovač nejblíže k příjemci se může rozhodnout, že se budou datagramy doručovat kratší cestou, než přes RP. Většinou se tak děje v závislosti na provozu. Malý provoz jde přes RP, větší přímo. Směrovač požádá o vytvoření (S, G) a současně dá RP vědět, že již nechce datagramy ze zdroje S pro skupinu G. RP to pochopí a případně sám přestane odebírat datagramy této skupiny od zdroje S. Protože nejkratší cesta platí pouze pro jeden zdroj, směrovače musí dál držet v paměti sdílený směrovací strom. Výslednou situaci v modelové síti můžete vidět na následujícím obrázku.



Nyní známe princip činnosti protokolu PIM. Příště seriál ukončíme povídáním o protokolu CBT a spolehlivém multicastu.

Detailní popis protokolu PIM-SM pomocí RFC 2362

kapitola 2
2. PIM-SM Protocol, přehled

Nejprve si povíme něco o architekturyckých komponentech potokolu PIM-SM.
Směrovač explicitně obdržuje zprávy Join/Prune pro připojení/odříznutí do multicastové větve od sousedních směrovačů kteří jsou jeho downstream členové. Roter posílá data pakety do multicastové skupiny G, na ta rozhraní, které byly přijaty Join zprávy.

A Designated Router (DR) periodicky posílá Join/Prune zprávy ke specifikovanému routeru nazývanému Randezvous Point (RP) kterí jsou aktivní členové.

Vstupem do směrovače zahrnujeme políčka jako jsou zdrojová adresa, adresa skupiny, příchozí rozhraní ze kterého byl paket přijat, seznam ochozích rozhraní, na které byl paket zaslán(časovač,paritní bit, atd.)

Vstupy jdou příchozími rozhraními směrem k RP, výstupy odcházejí směrem k sousedním downstream routrům , který poslal RP Join/Prune zprávpu. Tento stav vytváří sdílení , Rp-centered distribuční strom, který vede ke všem členům skupiny. Když se zdroj dat poprvé přihlásí ke své skupině, jeho RD pošle registrační unicastovou zprávu pro RP se zdrojovými datovými pakety zapouzdřenými uvnitř. Pokud je datová rychlost velká, Rp může poslat source-specific Join/Prune zprávu zpět ke zdroji a pakety se zdrojovými daty budou následovat výslednou odesílací stanici a cestovat nezapouzdřené k RP. Je jedno zda data přijdou zapouzdřené nebo původní, RP je na začátku rozpouzdří a pošle je dolu po distribučním stromu směrem ke členům skupiny. Pokud to přenosová rychlost zaručí, směrovač s lokálními příjemci může přijmout zdroj do nejkratší cesty distribučního stromu a může prořezat zdrojové pakety mimo centralizovaný strom.Pro pomale zdroje dat, žádny RP, ani poslední routr nepotřebuje připojit zdroj ke stromu nejkratších cest a datové pakety můžou byt dodány přes sdíleny RP-Strom.

Následující část popisuje SM operace podrobněji řídící zprávy a akce které se zpustí.

2.1. Lokální příjemce se přihlašuje do muticastové skupiny.

Pro připojení se do multicastové skupiny G, host(příjemce) conveys
přijímá informace z Internet Group Manager Protokolu(IGMP). Od tud
poznáváme(rozhodujeme) o tom, zdali je příjemce vysílač R, nebo člen skupiny G.

Když DR (např.: směrovač)obdrží upozornění na požadavek členství z IGMP pro novou skupinu G, DR vyhledá skupinový RP(formou paketu). Daný DR vytvoří pro danou multi. skupinu strom (upstream, downstream), uváděnou zde jako (*,G) zápis.Pokud není žádný další zdroj je tento packet poslán dal podle daného mult. stromu.

RP adresa je zahrnuta ve zvláštních polisměrovacích záznamech a je obsažena v periodiskych zprávách Join/Prune messages. Výstupní rozhraní jsou obsažené v IGMP členských zprávách pro nového člena. Vstupní rozhraní jsou nastavené na rozhraní užívaných pro posílaní unicastů k RP.

Když už zde nejsou žádní přímo spojení členové dané skupiny, IGMP to oznámí DR. Když DR nemá žádné lokální ani žádné downsreamy , stav (*,G) je zrušen.

2.2 Nastavení RP kořene směrovacího stromu.

Pokud je připravený (*,G) stav , pak DR vytvoří Join/Prune zprávu s RP adresou, join listem, wildcard bitem(WC-bit) a Rp-tree bitem (RPT-bit) . Wc-bit signalizuje to , že nějaký zdroj odpovídá a byl forwardován s tohoto vstupu jestliže delší dobu neodpovídál. RPT-bit signalizuje to , že tento spoj je součástí sdíleného Rp-stromu. Prune list je ponechán prázdný. Když RPT-bit je nastaven na 1 to signalizuje, že spoj je přidružený se sdíleným Rp-stromem a proto Join/Prune zpráva je propagována po Rp-stromu. Když je Wc-bit nastaven na 1 to signalizuje to, že adresa RP a downsream přijímače čekají na přijetí paketů od všech zdrojů po této sdílené cestě(stromě).

Rozhraní kterým je zpráva Join/prune přijata je přidané do seznamu odchozích rozhraní (oifs) pro (*,G). Podle tohoto vstupu každý upstream směrovač mezi přijímačem a RP pošle Join/Prune zprávu v které join list obsahuje RP.
Paked payload obsahuje Multicast-Address=G, Join=RP,Wc-bit, RPT-bit, Prune=NULL

2.3 Příjemci posílající data do skupiny

Když host začne posílat multicastové pakyty do skupiny musí jeho DR (Deasignated router) doručit každý paket k RP routeru pro distribuci po to RP-stromě.

Inicializace DR vysílačů

každý paket se zapouzdří do Register message(zprávy) a pošle se unicastem k RP dané skupiny. Rp rozpouzdří každou zprávu a rozešle pakety dounstremem ve RP-stromu.

2.4 Přepnutí ze sdíleného stromu na strom nejkratších cest

Směrovač s právě spojenými členy se nejprve připojí k Rp-stromu. Směrovač se může připojit ke zdrojům nejkratší cestou k Sp-stromu, potom přijímá pakety z tohoto zdroje přes daný Rp-strom. Doporučena taktika připojit switch k Sp-stromu potom přijímat podstatné množství datových patetů během předepsaného časového intervalu z jednotlivých zdrojů.
Pro uvědomění si této taktiky směrovče mohou sledovat pakety ze zdrojů, pro které nejsou zdroje multi. vysílačů uvedené, takové vstupy, kdy rychlost přenosu dat překoná tento nakonfigurovaný práh.

kapitola 3

3.1 Zasílají se HELLO zprávy, takže sousední směrovače poznají své sousedy.

3.1.1 Posílání HELLO zpráv

HELLO zprávy se posílají periodicky mezi PIM sousedy po každé periodě [Hello-Period]. Tato perioda je jednou za 30 sekund. Tato zpráva informuje směrovače jaké rozhraní má sousední směrovač (ALL-PIM-ROUTERS group).
Paket obsahuje prodlevu [Hello-Holdtime], po kterou si mají směrovače informaci podržet. HELLO pakety se posílají na všechny typy komunikačních spojení.

3.1.2 Přijímání HELLO zpráv

Když směrovač přijme HELLO zprávu, uloží si adresu sousedního routeru a nastaví Designated Router (DR) na to rozhraní. Designatet router se stává ten s největší adresou.

Jestliže nějaké rozhraní vede dole, routr může optimálně počkat se všemi PIM sousedy na přidružených rozhraních.

3.1.3

3.2 Zprávy Join/Prune

Zprávy Join/Prune jsou posílány pro odříznutí, nebo připojení části větve do multicastového stromu.

Jediná zprava obsahuje atributy spojení a prořezání, z nichž může byt jeden nulový. Každý zpráva obsahuje soubor zdrojových adres, sdílené nebo přímé stromy, které chce směrovač připojit nebo prořezat.

3.2.1 Posílání Join/Prune zpráv

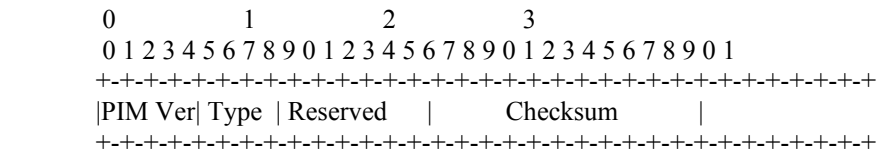
Join/Prume zprávy jsou sloučené tak , že zprava se posílá jednotlivým sousedům, N, obsahuje aktuální připojení a prořezání zdrojů, které jsou dosažitelné přes. N. Podle unicastového směrování Join/Prune zprávy jsou posílány všem směrovačům dané multi. Skupiny na směrovač na dalším skoku směrem k S nebo RP. Join/Prune zprávy jsou posílány každou Join/Prune-Periodu. V budoucnu se má představit mechnyzmy na rychlostní limit, tato kontrola by měla byt již od prvního skoku za účelem se vyhnout nadměrné režii na malých spojeních.

Periodicky se opkující posílání Join/Prune zpráv.

Router posílá periodicky Join/Prune zprávy ke každému RPF sousedovi označenému (S,G), (*,G) a (*,*,RP) vstupu.
Join/Prune zprávy jsou zasílány pouze tem RPF kteří jsou PIM sousedi.

kapitola 4
4 Formát paketu PIM

V této sekci popíšeme detaily jak vypadá formát paketu pro PIM zprávy.
Všechny PIM kontrolní zprávy mají číslo protokolu 103.
Všeobecně, PIM zprávy jsou také unicastové,, nebo multicastové hop-by-hop ke všem PIM routerovským členům 224.0.0.13



PIM Ver - PIM verze číslo je 2.

Type - Typ specifikující PIM zprávu. PIM typy jsou:

0 = Hello
1 = Register
2 = Register-Stop
3 = Join/Prune
4 = Bootstrap
5 = Assert
6 = Graft (used in PIM-DM only)
7 = Graft-Ack (used in PIM-DM only)
8 = Candidate-RP-Advertisement

Reserved – nastaven na 0. Ignored upon receipt.

Kontrolní součet je 16-bitový jedničkový doplněk z jedničkového doplňku součetu z celé PIM zprávy (vyřazující datový vstupní kanál v Registrvé zprávy) Pro počáteční kontrolní součet, kontrolní součet pole je nulový.

4.2 HELLO zpráva

Je Zasílána periodicky od routerů na všechny rozhraní.