

**Jakub Novák**

**X36MTI**

**16. listopadu 2007**

**Prezentace**

**Protokol BGP**

## BGP

Border Gateway Protocol slouží ke směrování mezi autonomními systémy (AS). Vyvinul se z protokolu EGP a v současné době dospěl k verzi 4, označované jako BGP-4 protokol.

BGP směrovače slouží jako vstupní/výstupní body do autonomních systémů. Podle toho pak jednotlivé spojení mezi směrovači dělíme na externí (jedná-li se o směrovače patřící k různým AS) a interní.

BGP směrovače si neudržují pouze adresu dalšího směrovače, přes který se lze dostat k cílovému uzlu, ale také čísla autonomních systémů, přes které cesta prochází. To umožňuje efektivnější vytváření nejlepších cest (agregování cest apod.). Další výhodou spočívá v tom, že směrovací informace nejsou periodicky vyměňovány, pouze v případě změny topologie. Pravidelně jsou vyměňovány jen KeepAlive zprávy, které jsou mnohem kratší a nezatěžují tak síť.

BGP používá TCP jako svůj transportní protokol, konkrétně pak port 179 pro navázání spojení.

Směrovače si mezi sebou posílají zprávy, které vypadají následovně:

**Maker:** slouží k autentifikaci a zajištění synchronizace. Je to nějaká hodnota, kterou přijímající strana může očekávat, tzn. spočítat podle konkrétní zvolené autentifikační metody. Zároveň slouží pro detekci ztráty synchronizace. Pokud je typ zprávy Open nebo není použita žádná autentifikace, je nastaven na hodnotu nula.

**Typ:** určuje typ zprávy, musí být jedno z následujících:

- **Open** oznamuje otevření nového spojení
- **Update** obsahuje aktualizaci směrovacích informací
- **Notification** je zaslána v případě, chyby nebo jiné mimořádné situace
- **KeepAlive** je posílána pravidelně; oznamuje, že směrovač stále pracuje

**Length:** určuje celkovou délku zprávy včetně hlavičky

[http://cs.wikipedia.org/wiki/Border\\_Gateway\\_Protocol](http://cs.wikipedia.org/wiki/Border_Gateway_Protocol)

**Border Gateway Protocol (BGP)** je dynamický směrovací [protokol](#) používaný pro směrování mezi [autonomními systémy](#) (AS). Je základem propojení [sítí](#) různých [ISP](#) v [peeringových](#) uzlech.

Směrování mezi autonomními systémy má charakteristické požadavky, které se nevyskytují v interním směrování. Směrovací tabulky obsahují stovky tisíc záznamů, nejdůležitějším kritériem nebývá vzdálenost, ale posuzují se nastavitelné parametry zohledňující například cenu a dodatečná pravidla aplikovaná v závislosti na zdroji, cíli, seznamu tranzitních autonomních systémů a dalších atributech.

Vzhledem k velkému počtu záznamu se v případě změn v topologii vyměňují pouze informace o změnách, nikoliv celé směrovací tabulky jako je tomu v případě protokolu [RIP](#).

### Autonomní systém

**Autonomní systém (AS)** je množina IP sítí a [routerů](#) pod společnou technickou správou, která reprezentuje vůči [Internetu](#) společnou routovací politiku. Uvnitř autonomního systému je pro [routování](#) použit některý z Interior Gateway Protokolů ([IGP](#)) pro routování mezi AS se používá Border Gateway Protocol ([BGP](#)).

[Routování](#) v Internetu je rozděleno do dvouúrovňové hierarchie. První úroveň tvoří routování uvnitř autonomních systémů. Pro to se používá některého z vnitřních routovacích protokolů (IGP). Typicky se jedná o protokol [OSPF](#), případně [IS-IS](#) nebo [EIGRP](#). Na základě vnitřního routovacího protokolu znají vždy všechny routery v autonomním systému nejkratší cesty do všech míst AS.

Internet by teoreticky mohl tvořit jediný autonomní systém, ale routovací tabulky by pak byly příliš velké a technicky těžko zvládnutelné. Proto je rozdělen na více autonomních systémů, mezi kterými se routuje pomocí BGP protokolu.

### Číslování autonomních systémů

Aby BGP mohl nalézt cestu do určitého AS, musí mít každý AS přiděleno jednoznačné číslo, tzv. **Autonomous System Number (ASN)**.

Číslo autonomního systému se ukládá do 16ti bitového pole, takže může nabývat 65536 hodnot. Čísla se zapisují normálně v dekadickém tvaru.

AS 0, AS 65535 a AS 23456 jsou rezervovány pro speciální účely. AS 64512 - AS 65534 jsou určeny pro privátní autonomní systémy a zbytek, tedy AS 1 až AS 64511 (bez AS 23456) je určen pro autonomní systémy na Internetu. Číslování je nestrukturované, takže čísla jsou přidělována bez nějakého dalšího systému. Správcem čísel AS je organizace [IANA](#).

Množina volných čísel pro autonomní systémy se začíná nebezpečně zmenšovat. V lednu 2006 bylo obsazených cca 40000 čísel ze 64511, přičemž dlouhodobé statistiky ukazují, že každý rok se spotřebuje cca 3500 nových čísel.

## Nový systém číslování AS

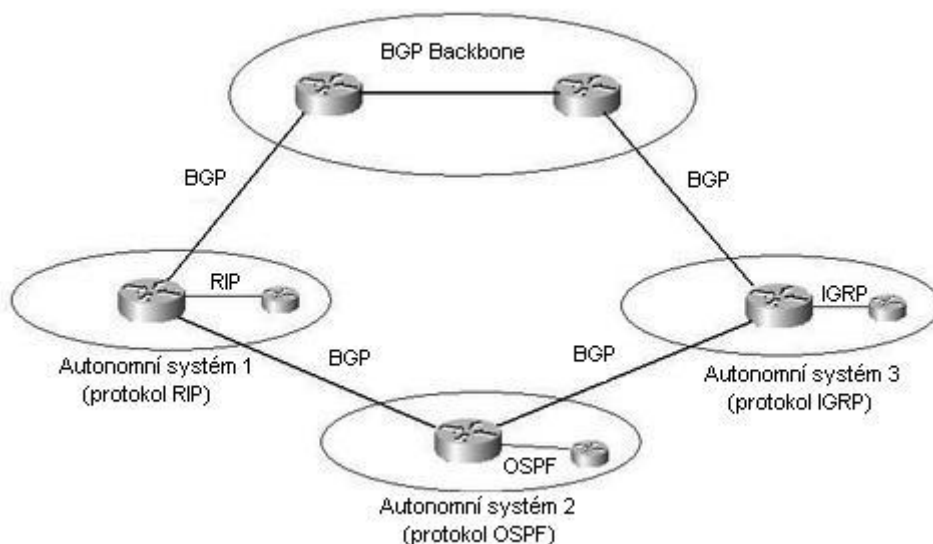
V souvislosti s nedostatkem čísel AS se připravuje přechod na 32bitové číslování AS. Nová 32bitová čísla se budou skládat z vyšší a nižší šestnáctibitové části a budou se zapisovat jako dekadická čísla odpovídající těmto částem oddělená tečkou.

Zavedení nového číslování ovšem není vůbec jednoduchá záležitost, protože je závislá na zavedení nové verze BGP protokolu, která ho bude podporovat. Vzhledem k rozsáhlosti Internetu není možné počítat s tím, že se BGP dá na celém světě vyměnit ze dne na den, a je potřeba připravit strategie, které na poměrně dlouhou dobu přechodu zajistí kompatibilitu starého a nového číslování. Základní z nich je, že stará čísla AS budou mít v novém modelu vyšší část rovnou nule, tedy například původní AS s číslem 33333 bude v novém tvaru vypadat jako 0.33333.

<http://www.owebu.cz/pc-site/vypis.php?clanek=1597>

## Protokol BGP

Následující obrázek ilustruje použití protokolu BGP pro přepojování autonomních systémů, jejichž vnitřní směrovače používají autonomně směrovací protokoly RIP, OSPF a IGRP.



Po startu si každý BGP směrovač vytvoří TCP/IP spojení se sousedním BGP směrovačem. Tím způsobem získají BGP směrovače kompletní směrovací tabulky celé sítě. Změny směrovacích tabulek nejsou vysílány periodicky, jako například u protokolu OSPF, nýbrž jsou odesílány pouze inkrementální změny v části směrovací tabulky ihned po změně stavu sítě a optimální cesty ke směrovačům. Metrika je definována jako koeficient, vyjadřující počet autonomních systémů k cílové

síti, rychlost a prioritu linky přidělenou administrátorem.  
V rámci protokolu BGP se používají následující zprávy:

**1. Open Message** – je úvodní inicializační zpráva přenášena po vytvoření TCP spojení. Musí být potvrzena zprávou „Keep Alive Message“

**2. Update Message** – zpráva pro aktualizaci směrovacích informací. Obsahuje jak informace o zrušení některých směrů a tak informace o propagaci směrů jiných

**3. Keep Alive Message** – periodicky vysílána zpráva, která kontroluje aktivitu vytvořených TCP spojení mezi BGP směrovači.

**4. Notification Message** – zpráva používána pro zrušení TCP spojení obsahující i důvod zrušení spojení. Také se používá k odesílání chybových hlášení.

### Záhlaví protokolu BGP

**M L T DATA**

16B 2B 1B n B

Význam symbolů v záhlaví protokolu BGP

| <b>Symbol</b> | <b>Význam</b>  |
|---------------|--|
| M (Marker)    | Autorizační data spojení BGP                           |
| L (Length)    | Délka zprávy BGP                                       |
| T (Type)      | Typ zprávy BGP: Open, Update, Keep Alive, Notification |
| DATA (Data)   | Datová část zpráv                                      |

Příklad zprávy „BGP Open Message“

**V AS HT BGP-id OPL OP**

1B 2B 2B 4B 1B 4B

Význam symbolů v zprávě BGP Open Message

| <b>Symbol</b>                    | <b>Význam</b>  |
|----------------------------------|--|
| V (Version)                      | Verze protokolu  |
| AS (Autonomous System)           | Identifikátor autonomního systému, který zprávu odesílá              |
| HT (Hold Time)                   | Časový interval, po jehož přesažení je trasa označena za nedostupnou |
| BGP-Id (Authentication)          | Identifikátor (IP adresa) zdrojového směrovače                       |
| OPL (Optional Parameters Length) | Délka volitelného pole parametrů                                     |
| OP (Optional Parameters)         | Volitelné parametry  |

## Technologie internetu (9.): Směrování ve velkém

ze dne: 10/01/2007 12:00:36 AM

rubrika: Internet, Technologie, Události

**Pokud jde o vnější směrování, tedy směrování mezi autonomními systémy, používá se dnes prakticky výhradně Border Gateway Protocol (BGP) verze 4. Protokol BGP umožnil přejít od centralizovaného směrování k uspořádání peer-to-peer, kdy datagramy putující mezi dvěma autonomními systémy mohou využít kteroukoli existující cestu.**

První verze BGP se oficiálně objevila v červnu 1989 (RFC 1105). Poté následovaly v poměrně rychlém sledu další verze, až se v polovině devadesátých let vývoj ustálil na verzi 4 (RFC 1771), která se s malými úpravami a řadou rozšíření používá dodnes.

### Výměna informací v BGP

Samotný protokol není, třeba ve srovnání s OSPF, příliš složitý. Výměna informací se v rámci BGP nejčastěji odehrává mezi dvojicemi hraničních směrovačů v různých AS, jež jsou spolu přímo propojeny. Relace BGP, nazývané BGP peering, jsou dlouhotrvající a k přenosu dat využívají spolehlivý transportní protokol TCP. Po navázání spojení si směrovače nejprve vymění celou směrovací tabulku, ale potom už si posílají jen informace o změnách.

Sousední BGP směrovače si, podobně jako v případě RIPv, navzájem sdělují informace o dostupnosti síťových prefixů spolu s údajem, jak je každý prefix vzdálen od původce této informace, která se odborně nazývá Network Layer Reachability Information (NLRI). Zatímco v RIPv se však jako metrika předává pouze jedno číslo, v BGP je s každým prefixem spojena celá posloupnost autonomních systémů (atribut AS\_PATH), která kompletně popisuje nejkratší cestu, jíž se lze dostat od daného směrovače k tomuto prefixu. Slovo „kompletně“ je ale třeba chápat v kontextu vyšší úrovně směrovací hierarchie – BGP se vůbec nezabývá tím, jak budou datagramy putovat uvnitř každého z autonomních systémů, které se na nejkratší cestě nacházejí.

Podívejme se na příklad, kde autonomní systém AS<sub>1</sub> obsahuje síť s prefixem 175.111.0.0/16. Oba jeho hraniční směrovače o tom informují všechny své BGP sousedy. Uvedený prefix je přitom spojen s atributem AS\_PATH, v němž je uveden jediný autonomní systém AS<sub>1</sub>. Každý ze sousedů se tím dozví, že datagram s cílovou adresou patřící do tohoto prefixu je třeba dopravit do AS<sub>1</sub>. Tato informace se ale zároveň předává dále. Například hraniční směrovač v AS<sub>3</sub> oznámí stejný prefix svému sousedovi v AS<sub>5</sub>, na začátek AS\_PATH ale přiřadí svůj vlastní autonomní systém, takže nový AS\_PATH bude v tomto případě obsahovat dva autonomní systémy – AS<sub>3</sub> a AS<sub>1</sub>. Podobný postup pokračuje dále, takže nakonec se o dostupnosti prefixu dozvědí všechny autonomní systémy v síti.

Atribut AS\_PATH zastává několik funkcí. První z nich je funkce metriky, jíž je v případě BGP vždy počet „hopů“, tedy počet autonomních systémů v AS\_PATH. V našem příkladu se tedy AS<sub>5</sub> dozví od svých sousedů o třech možných cestách k prefixu 175.111.0.0/16: první z nich odpovídá tříprvková

AS\_PATH (AS<sub>4</sub>, AS<sub>2</sub>, AS<sub>1</sub>), druhé (AS<sub>3</sub>, AS<sub>1</sub>) a třetí pak jednoprvková posloupnost (AS<sub>1</sub>). Směrovače v AS<sub>5</sub> by za normálních okolností měly zvolit cestu s nejkratší AS\_PATH, tedy přímým spojením.

Druhou funkcí AS\_PATH je ochrana proti smyčkám, které lze okamžitě rozpoznat podle toho, že se stejný autonomní systém v posloupnosti opakuje. Protokol BGP proto vyžaduje, aby směrovače ignorovaly všechny NLRI, které již v atributu AS\_PATH obsahují lokální AS daného směrovače. Znovu k našemu příkladu: AS<sub>5</sub> oznamuje po přímé lince svému sousedovi AS<sub>1</sub> dostupnost prefixu 175.111.0.0/16 s AS\_PATH = (AS<sub>5</sub>, AS<sub>3</sub>, AS<sub>1</sub>). Na oznámení však v AS<sub>1</sub> nebude brán zřetel, protože AS<sub>1</sub> se už v posloupnosti vyskytuje.

## Interní BGP

Pozorný čtenář si na obrázku jistě všiml toho, že některé autonomní systémy mají více hraničních směrovačů. Vzniká proto otázka, jak si tyto směrovače předávají informace získané z BGP, aby je mohly šířit dalším sousedům. Jednou z možností by byla jejich redistribuce do vnitřního směrovacího protokolu, například OSPF. Z několika důvodů to ale není nejlepší nápad. Předně, databáze BGP je (jak vzápětí uvidíme) obudně velická a vnitřní směrovače by se tím zbytečně zatěžovaly. Kromě toho by se redistribucí ztratily některé atributy (především AS\_PATH!) předávané jako součást NLRI, protože vnitřní směrovací protokoly je neznají. Hraniční směrovače patřící do stejného autonomního systému proto spolu komunikují také prostřednictvím BGP, přesněji jeho varianty, která se nazývá interní BGP (iBGP). Výše popsaná varianta, která slouží k výměně NLRI mezi autonomními systémy, se pak logicky nazývá externí BGP (eBGP). Interní BGP se řídí některými dodatečnými pravidly, která nejsou pro eBGP relevantní. Například kvůli ochraně proti směrovacím smyčkám uvnitř AS nesmí směrovač předávat v iBGP informace, které se dozvěděl od jiného iBGP souseda.

## Směrovací politika

Principy fungování BGP tedy není příliš těžké pochopit. Velmi obtížné – ale také zajímavé – však je, jak tyto principy využít k realizaci určité směrovací politiky vlastního autonomního systému. Toto know-how nelze jen tak někde nastudovat, správce AS se k němu musí propracovat vlastními zkušenostmi a omyly. Musíme si totiž uvědomit, že v systému BGP vystupuje velké množství relativně nezávislých účastníků, kteří si mezi sebou posílají ohromný objem informací. Podle aktuálních statistik, dostupných na serveru [www.potaroo.net](http://www.potaroo.net) pojímá globální internet v současnosti více než 21 tisíc aktivních autonomních systémů a počet prefixů vyměňovaných pomocí BGP již překročil 250 tisíc. Je jasné, že v takto složitém systému se čas od času dějí prapodivné věci, zvláště když problémy jednoho AS (ať už způsobené výpadky nebo chybnou konfigurací) mohou negativně ovlivnit celý internet.

Implicitním kritériem výběru směrovací cesty k danému síťovému prefixu, jež určuje základní způsob chování protokolu BGP, je délka cesty, tedy počet autonomních systémů v atributu AS\_PATH. Občas je ale potřeba aplikovat i jiná kritéria a směrování poupravit. Pro tento účel je k dispozici několik nástrojů, které lze v různých situacích použít.

**Základní jsou tyto dva:**

1. Filtrace některých NLRI v hraničním směrovači.
2. Umělé prodloužení atributu AS\_PATH.

První metoda upravuje směrování datagramů ven z vlastního AS: vnitřním směrovačům se některé NLRI „zatají“, čímž se pro ně dotyčná cesta stane nedostupnou a použijí jinou. Druhá metoda pak ovlivňuje opačný směr toku dat, tedy směrování z cizích AS do vlastního. AS\_PATH se uměle prodlouží tím, že se na jeho začátek uvede vlastní autonomní systém ne jednou, ale hned několikrát. Pokud bychom v našem příkladu jako správci AS<sub>1</sub> nechtěli, aby k nám AS<sub>5</sub> směřoval datagramy po přímém spoji, ale raději přes AS<sub>3</sub>, změníme AS\_PATH u prefixů předávaných autonomnímu systému AS<sub>5</sub> tak, že místo standardního jednoprvkového (AS<sub>1</sub>) pošleme (AS<sub>1</sub>, AS<sub>1</sub>, AS<sub>1</sub>), čímž se pro AS<sub>5</sub> stane posloupnost (AS<sub>3</sub>, AS<sub>1</sub>) nejkratší. Všimněme si ale, že toto naše opatření může zhatit správce AS<sub>3</sub>, pokud by autonomnímu systému AS<sub>5</sub> předával naše prefixy také s uměle prodlouženou AS\_PATH. Proto je důležité, aby správci AS mezi sebou komunikovali a své směrovací politiky koordinovali.

Další informace z [http://en.wikipedia.org/wiki/Border\\_Gateway\\_Protocol](http://en.wikipedia.org/wiki/Border_Gateway_Protocol)