QoS v datových sítích, IntServ a DiffServ

Tento materiál byl zpracován kompilací dvou zdrojů:

Sven Ubik: QoS a diffserv – Úvod do problematiky, Technická zpráva TEN-155 CZ číslo 6/2000
Arindam Paul: QoS in Data Network: Protocols and Standards

1. Úvod

QoS je schopnost sítě zajišťovat lepší služby vybraným přenosům prostřednictvím různých technologií (ATM, IP, směrované sítě). Jinými slovy je to vlastnost sítě, pomocí které je možné rozlišovat mezi různými třídami přenosů a chápat je diferencovaně.

QoS může být klasifikována různými metrikami, jako např.:

- Dostupnost služby – spolehlivost připojení se k síťovému zařízení.
- Zpoždění – doba potřebná pro přenos paketu mezi koncovými uzly sítě.
- Rozptyl zpoždění – změny zpoždění při přenosu posloupnosti několika paketů procházejících stejnou cestou sítě.
- Propustnost – rychlost s jakou procházejí pakety sítí.
- Míra ztrát paketů – rychlost, se kterou jsou pakety zahazovány, ztrácí se a nebo jsou znehodnoceny při průchodu sítí.
- Při přenosech se snažíme maximalizovat dostupnost a propustnost, redukovat zpoždění a eliminovat rozptyl zpoždění a míru ztrát.

2. Potřeba QoS

Potřebujeme navrhovat sítě, které zajistí:

- Přenos za pomoci více tříd služeb.
- Jsou škálovatelné – dovolují zvyšování množství síťových přenosů bez vlivu na výkonnost sítě.
- Dovoluje podporovat časově kritické aplikace.

3. QoS mezi koncovými uzly

K zajištění QoS v sítích jsou dnes realizovány tři základní modely obsluhy:

- Služby typu Best Effort (s maximálním úsilím),
- Integrované služby (Integrated services),
- Rozlišované služby (Differentiated Services).

4. Služba Best effort (s maximálním úsilím)

V tomto modelu posílají aplikace data když se jim zachce, kolik chtějí a bez vyžádání si jakéhokoliv povolení. Síťové komponenty se pokouší přenést data co nejlépe, bez omezení na zpoždění, zpoždění reakce (latency), nebo rozptyl zpoždění. Dělají to i tehdy, když nemohou data doručit, bez informování odesílatele nebo příjemce. Příkladem takovéto služby je doručování v IP sítích.

5. Integrované služby (Integrated Services - IntServ)

V případě integrovaných služeb aplikace oznámí počítačové síti své požadavky na přenos dat ve formě požadovaných QoS (Quality of services). Počítačová síť ověří zda jsou k dispozici požadované prostředky, a rozhodne, zda požadavkům vyhoví. Tato funkce je označována jako admission control (řízení přístupu). V případě, že síť nemůže požadavku vyhovět, není spojení povoleno a aplikace se může rozhodnout, zda požádá o méně náročné QoS. Pokud je požadavek přijat, musí počítačová síť informovat všechny komponenty, přes které bude probíhat přenos, aby pro dané spojení rezervovaly odpovídající objem prostředků, např. šířku pásma mezi dvěma směrovači, kapacitu fronty paketů, atd. K tomuto účelu slouží rezervační protokoly. Nejrozšířenějším rezervačním protokolem je RSVP (Resource reSerVation Protocol), který je však poměrně složitý a představuje významnou režii při řízení chodu sítě. Proto se v poslední době objevují návrhy jednodušších protokolů pro rezervaci, např. YESSIR.

Možnost specifikovat přesně QoS není u všech aplikací nutná. Řada aplikací vystačí s tím, že požadované parametry se podstatně nezhorší při změně zatížení počítačové sítě. Navíc se zvyšujícím se objemu přenášené informace je třeba snížit režii při přepínání a minimalizovat objem stavové informace ve směrovačích. Proto se v poslední době objevuje jiný způsob implementace QoS - rozlišované služby (Differentiated services – diffserv).

IntServ vychází z modelu, kdy je před přenosem zajištěna potřebná kvalita přenosového kanálu. K tomu slouží RSVP (Resource reSerVation Protocol). Pro řízení sítě se používají následující strategie:

- Udržování stavu propojení.
- Hlídání a úprava přenosu.
- Předcházení zahlcení.
- Management předcházení nebo odstranění zahlcení.
- Mechanizmus sledování výkonnosti linky.

Integrované služby rozlišují mezi následujícími kategoriemi aplikací:

- Elastické aplikace – bez požadavku na doručování. Do této kategorie zapadají aplikace nad TCP. Nejsou kladeny požadavky na omezení zpoždění nebo kapacitu spojení. Příkladem je el. pošta, http protokol, atd.
- Real Time Tolerant (RTT) aplikace – požadují omezení na maximální zpoždění v síti. Občasná ztráta paketů je přijatelná. Příkladem jsou video aplikace využívající bufferování, které před aplikací skryjí ztrátu paketů.
- Real Time Intolerant (RTI) aplikace – tato třída požaduje minimální odezvu (latency) a rozptyl zpoždění (jitter). Příkladem jsou videokonference.

K zajištění obsluhy těchto aplikací má RSVP k dispozici následující třídy služeb (Class of Service - COS):

- Garanteed Service – služba je určena pro RTI aplikace a zaručuje:
  - Šířku pásma pro přenos v rámci aplikace,
  - Deterministickou horní hranici zpoždění.

To je důležité pro interaktivní aplikace nebo aplikace v reálném čase. Aplikace mohou snížit zpoždění zvýšením požadavků na šířku pásma.

- Controlled Load Service – je určena pro RTT aplikace. Zaručuje průměrné zpoždění, ale zpoždění přenosu jednoho paketu mezi koncovými uzly není deterministické.

K zajištění IntServ se mohou použít různé protokoly. V současné době jsou rozpracovány RSVP od IETF (ReSerVation Protocol) a COPS (Common Open Policy Service), který navrhlo CISCO. RSVP slouží k přenosu rezervačních požadavků a vytváří virtuální okruhy s danými přenosovými parametry. COPS slouží k řízení možností rezervace, které mají jednotlivé směrovače. Čili říká směrovačům, které požadavky přijmout a které ne.

5.1. Zvláštní rysy RSVP

Tok dat v RSVP je sekvence zpráv, které mají tentýž zdroj, cíl a totéž QoS. V RSVP jsou zdroje rezervovány pro data v jednom směru od vysílače do přijímače. Vysílaná data jsou označována jako upstream, přijímaná jako downstream. Hostitelský systém, který chce vysílat data s požadavkem na zajištění QoS posílá speciální paket nazývaný PATH do potenciálních příjemců. Tento paket nese informace o charakteristice přenosu a vytváří tzv. stav cesty podél cesty přenosu. Zpráva PATH je přenášena z hostitelského systému do sousedního směrovače, který pošle PATH do následujících sousedních směrovačů. Cesta nebude vytvořena v případě, že se v opačném směru bude šířit zpráva PATH Error.

Poté, co přijímač obdrží zprávu PATH, vyšle v opačném směru zprávu RESV s typem požadované rezervace. Existují čtyři typy rezervací – *Distinct Reservation*, sdílená rezervace (*shared reservations*), *vildcard filter type reservation* (viz dále). V tuto chvíli mají všechna zařízení podél cesty vytvořen stav cesty a jsou si vědomy charakteristik přenosu potenciálního toku dat. RESV obsahuje aktuální QoS charakteristiky očekávané přijímačem. Různé přijímače mohou specifikovat různé požadavky na QoS pro tentýž skupinový tok dat. RESV se přenáší v opačném směru, než ze kterého přichází PATH zprávy. Tak každé zařízení podél cesty zná aktuální charakteristiky QoS pro tok dat, požadovaný přijímačem a každý usoudí samostatně který z požadavků potvrdit a který odmítnout. Pokud je požadavek odmítnut, je poslána zpráva RESV Error do přijímače, který požadavek generoval.
Poté co jsou RESV zprávy přijaty vysílačem, a nebyla přijata žádná zpráva RESV Error, pošle zdroj zprávu RESV Confirmation do všech uzlů, které to požadují. Okamžitě poté začne vysílač posílat datové zprávy. Mezilehlá síťová směrovací zařízení posílají data s využitím rezervovaných zdrojů. Na konci přenosu pošle vysílač PATH TEAR zprávu, na kterou přijímač odpoví RESV TEAR a rezervace se zruší. Zprávy typu TEAR mohou být vysílány i směrovačem z důvodu timeoutu nebo koncovým systémem při přerušení přenosu.

5.2. RSVP rezervační typy

Typy rezervací, posílaných přijímačem mohou být následující:

1. Distinct Reservation: Přijímač si přeje rezervovat část pásma pro každý vysílač. V systémech s multicast vysíláním s více vysílači je každý posílající chráněn před ostatními. Tento způsob je také nazýván Fixed Filter Style.
2. Shared Reservation: V tomto případě požaduje přijímač rezervovat část pásma pro všechny zdroje s danou skupinovou adresou.
3. Wildcard Filter Type: V tomto případě požaduje přijímač rezervaci přenosové kapacity pro všechny zdroje v multicast doručovacím stromu.
4. Shared Explicit Reservation: Totéž jako předchozí případ s tím, že přijímač určí explicitně pevný počet vysílačů.

Tunelování: V systémech, kde není podpora RSVP se vytváří oblasti s podporou RSVP a tyto oblasti se propojují pomocí tunelů. Zprávy RSVP PATH a RESV REQUEST jsou zapouzdřeny do IP paketů a předávány na další RSVP směrovač.

5.3. Řízení vstupu dat (Kontrola vstupu dat)

Policy Enforcement Point (PEP) (bod vynucení přístupu) je síťové zařízení nebo metoda na síťovém zařízení, která je schopná si vynutit omezení přístupu. Může být ve zdroji, v cíli nebo někde mezi. Local Policy Module (LPM) (lokální modul omezení) je modul schopný vynutit omezení vstupu. LPM přijímá RSVP zprávy pro svou potřebu. LPM spolupracuje s PEM (Policy Enforcement Point), který kontaktuje PDP (Policy Desicion point) (bod rozhodnutí přístupu) s požadavkem na rozhodnutí o postupu o paketu a pak posílá paket do RSVP modulu. PDP je logická entita, která interpretuje policies náležející k RSVP požadavku a formuluje rozhodnutí. To závisí na tom, kdo dostal jakou QoS, kdy, od koho kam atd. PDP vytváří rozhodnutí založené na administrativně rozhodnutých politikách, které spočívají ve vzdálených databázích jako např. adresářové služby nebo síťový souborový systém.

5.4. COPS

Standardní protokol, který používají PDP a PEP pro vzájemnou komunikaci se nazývá COPS (Common Open Policy Service). Jedná se o jednoduchý protokol typu dotaz/odpověď. COPS rozlišuje následující typy požadavků:

- Požadavky řízení vstupu (Admission Control requests) – přijme-li PEP paket, požaduje od PDP rozhodnutí o povolení vstupu.
- Požadavek přidělení zdrojů (Resource Allocation request) – PEP požaduje od PDP rozhodnutí zda, jak a kdy rezervovat lokální zdroje pro zpracovávaný požadavek.
- Požadavek předávání (Forwarding request) – PEP se ptá PDP jak modifikovat požadavek a předávat jej ostatním síťovým zařízením.

COPS vyžaduje spolehlivý přenos, proto využívá TCP. Z důvodu zajištění bezpečnosti může využívat IPSec. Protože PDP i PEP jsou stavové vzhledem k cestám nebo požadavkům rezervace, nemusí být RSVP obnovovací zprávy posílány prostřednictvím COPS. Jestliže nastane timeout pro cestu nebo pro stav rezervace, nebo se objeví RSVP zpráva TEAR, pošle PEP do PDP zprávu DELETE a stav rezervace se zruší.

6. Rozlišované služby (Differentiated services – diffserv)

Rozlišované služby se od integrovaných služeb liší zejména tím, že aplikace neoznamuje předem počítačové síti své požadavky na QoS. Použití rezervačních protokolů není nutné. Jednotlivé směrovače neudržují žádnou stavovou informaci o jednotlivých spojeních. Implementace QoS je řešena tak, že každý paket vstupující do počítačové sítě je označen značkou, která určuje třídu přenosu, poskytovanou paketu. Označování paketů probíhá pouze na vstupu do počítačové

sítě, během přenosu pouze směrovače čtou značku a podle této značky řídí způsob zpracování paketu. Počet značek je poměrně malý, maximálně desítky.

U integrovaných služeb udržuje každý směrovač stavovou informaci vztaženou ke každému spojení, u rozlišovaných služeb směrovače pouze přidělí určené prostředky každé třídě přenosu a zajišťují určitý vztah mezi třídami.

6.1. Klasifikace paketů

Rozlehlé počítačové sítě lze rozdělit na organizačně menší oblasti, které jsou řízeny lokálním administrátorem. V těchto oblastech mohou být použity různé typy směrovačů, vybavené různými protokoly pro zajištění QoS. Proto je velmi obtížné zajistit jednotné zpracování požadavků na QoS. Z hlediska rozlišovaných služeb je síť rozdělena na oblasti se samostatnou správou rozlišovaných služeb, tzv. diffserv domény. Doména obsahuje dva druhy směrovačů. Vnitřní směrovače, zajišťující spojení uvnitř diffserv domény a hranové směrovače, zajišťující značkování a odznačení paketů včetně jejich posílání vnitřním směrovačům. Hranové směrovače lze rozdělit podle funkce na ingress směrovače, zajišťující značkování paketů a egress směrovače, zajišťující jejich odznačení. Pakety vstupují do diffserv domény přes ingress směrovač, jsou přenášeny interními směrovači a vystupují přes egress směrovač. Jsou-li propojeny dvě diffserv domény, pracuje hranový směrovač současně jako egress směrovač jedné domény a ingress směrovač domény druhé. Pro pakety procházející v opačném směru plní i opačné funkce.

Klasifikace paketů probíhá v ingress směrovači. Výběr značky může být proveden na základě IP adresy odesílatele nebo adresáta, čísel portů (dle požadované služby), podle výsledků měření dynamických vlastností přicházejících dat apod. Uvnitř diffserv domény zůstává značka nezměněna, ale při přechodu do jiné domény se může změnit na jinou značku se stejným významem nebo na jinou značku s jiným významem. Pakety mohou být klasifikovány již aplikací, posílající pakety do sítě. První ingress směrovač může tuto značku pozměnit nebo zachovat.

Způsob značení paketu závisí na použité technologii nebo protokolu. Značka může být obsažena uvnitř hlavičky protokolu, pokud je na ni místo, nebo přidána vně paketu. V případě použití diffserv pro přenosy protokolem IPv4 je značka obsažena v poli TOS (type of services) v záhlaví IP. V případě použití protokolu IPv6 je značka umístěna do pole Traffic Class záhlaví. Pole, do kterého je značka uložena, se označuje DS (differentiated services) a má délku 8 bitů. Šest bitů je určeno pro vlastní značku, označovanou jako DSCP (differentiated services codepoint) a 2 zbývající bity jsou určeny pro budoucí použití.

6.2. Zpracování paketů

Zpracování paketů ve směrovačích v rámci diffserv domény lze popsat ve třech úrovních abstrakce.

Na ***nejvyšší úrovni*** je zpracování určeno službou definovanou mezi koncovými body komunikace. Služba je definována souborem parametrů popisující vazbu mezi účastníkem a sítí, tzv. dohodou na servisní úrovni (Service Level Agreement – SLA). Příkladem může být komunikační kanál, který se chová jako pronajatý okruh a poskytuje dohodnutou propustnost s nízkou latencí a malou ztrátovostí paketů. Ve skutečnosti je realizován pomocí sdílených prostředků sítě.

Na ***střední úrovni*** je zpracování paketů dáno způsobem zacházení s pakety v jednotlivých směrovačích, bez ohledu na ostatní směrovače, tzv. PHB – per-hop-behaviour. Každý směrovač zpracovává pakety nezávisle na ostatních. Stará se pouze o to, aby zpracování odpovídalo značkám paketů. Uvnitř diffserv domény nejsou udržovány žádné virtuální spoje.

Specifikace PHB je klíčovou částí diffserv. V současné době jsou standardizována dvě PHB. Expedited forwarding – EF (urychlené posílání) a assured forwarding – AF (zaručené posílání). Hodnota 0 v poli DS je využita k označení implicitního PHB, kterým je přístup best-effort (s maximálním úsilím). Specifikace PHB popisuje zpracování paketů z pohledu vnějšího pozorovatele. Určité PHB může být v rámci směrovače implementováno různým způsobem. Tato implementace je ***nejnižší úrovní*** zpracování paketů.

6.2.1. Expedited forwarding – EF (urychlené posílání)

EF PHB zajišťuje, že každý směrovač v diffserv doméně odesílá pakety zařazené do EF PHB průměrnou rychlostí alespoň rovné stanovené rychlosti. Průměrná rychlost se měří v jakémkoliv časovém intervalu delším nebo rovném době potřebné pro odesílání paketu maximální délky stanovenou rychlostí. EF PHB je vhodné pro implementaci virtuálního pronajatého okruhu.

6.2.2. Assured forwarding – AF (zajištěné posílání)

AF PHB umožňuje zařadit pakety do jedné ze čtyř tříd. Každé třídě je ve směrovačích přidělen určitý objem prostředků (velikost vyrovnávací paměti, kapacita výstupní linky). V rámci každé třídy je každému paketu přiřazena jedna ze tří priorit zahození paketu (drop precedence), ke kterému může dojít v případě zahlcení. Směrovač musí odeslat paket mající nižší hodnotu priority se stejnou nebo vyšší pravděpodobností než paket mající vyšší hodnotu priority. AF PHB se používá pro implementaci služeb, u kterých je třeba volitelná úroveň kvality přenosu.

6.3. Traffic conditioning (podmiňování přenosu)

Při provozu počítačové sítě se stává, že objem dat přicházející na určitý směrovač, směrovaných na určitý výstupní port, přesahuje okamžitou kapacitu linky výstupního portu. Pakety jsou v tomto případě ukládány do fronty ve směrovači. Proto je třeba řešit situaci, kdy dojde k přeplnění fronty (congestion management – řízení zahlcení) nebo tomu předejít (congestion prevention – předcházení zahlcení). V souladu s dohodnutým SLA (service level agreement – dohoda na servisní úrovni) je třeba upravovat objem dat, přicházejících do diffserv domény. Úprava objemu dat se obvykle provádí jen na ingress směrovači. K tomuto účelu se používají techniky policing (kontrolování, hlídání) a traffic shaping (úprava přenosu).

Značkování paketů se provádí na ingress směrovači na základě jejich klasifikace. Může být ovlivněno i měřením dynamických vlastností přicházejících dat. Zpracování paketů následující po klasifikaci se souhrnně nazývá traffic conditioning a soubor parametrů popisujících toto zpracování se nazývá Traffic Conditioning Agreement – TCA (dohoda na podmiňování přenosu).

6.3.1. Policing

Policing obvykle představuje jednoduché zahazování přicházejících paketů, které se začne provádět při splnění určité podmínky, např. vyčerpání kapacity fronty, překročení určitého objemu přicházejících dat.

6.3.2. Traffic shaping

Při použití metody traffic shaping jsou ve směrovačích pozdržovány pakety tak, že se mění okamžité množství odesílaných dat ve srovnáním s množstvím přijímaných dat. Data přicházející na směrovač ve shlucích jsou odesílána rovnoměrně rychlostí, odpovídající části kapacity výstupní linky. K tomuto vyrovnávání se používá metoda zvaná token bucket nebo leaky bucket (vědro pověření, děravé vědro). Princip metody spočívá v tom, že pověření je spojeno s právem vyslat paket nebo nějaké množství dat. Na počátku je nádoba plná. Při příchodu paketu je z nádoby odebrán určitý počet pověření a paket je zařazen do výstupní fronty nebo jinak označen. Pokud v nádobě není požadovaný počet pověření, je paket zahozen nebo označen jiným způsobem. Pověření jsou do nádoby doplňovány stálou rychlostí, pokud není nádoba plná. Metodu lze parametrizovat dvěma parametry – rychlostí doplňování, která odpovídá přidělené kapacitě výstupní linky a velikostí nádoby, která odpovídá kapacitě výstupní fronty.

6.4. Řízení front

Při správné funkci traffic conditioning na ingress směrovači by nemělo na vnitřních směrovačích docházet k vyčerpání kapacity front. Vnitřní směrovače včetně ingress směrovače implementují požadovaná PHB (per-hop-behaviour). Na ingress směrovači následuje PHB po traffic conditioning. Jednotlivá PHB jsou typicky realizována použitím více front s určitým algoritmem pro zařazování paketů do front, výběr paketů z front a jejich odesílání. Nejpoužívanější metody jsou priority queueing – PQ, class-based queueing – CBQ a weighted fair queueing – WFQ. Ve všech případech je paket vložen do jedné z front podle své značky. Metody se liší způsobem obsluhy front.

6.4.1. Priority queueing - PQ

Každá fronta má přiřazenu určitou prioritu, která je u každé fronty jiná. Fronty jsou obsluhovány podle priorit – nejprve se vyprázdní fronta s nejvyšší prioritou, atd. Nevýhodou je možnost uvíznutí dat ve frontě s nižší prioritou, kdy jsou data pozdržena natolik, že vysílací stanice je považuje za ztracená, opakuje jejich vysílání, a tím ještě zvýší zatížení sítě. Metodu lze použít pro implementaci EF PHB, kdy jedna fronta slouží EF PHB a druhá pro implicitní provoz typu best effort.

6.4.2. Class-based queueing – CBQ

U CBQ jsou jednotlivé fronty obsluhovány cyklicky, jedna po druhé. Přijde-li fronta na řadu, je z ní odeslán minimální stanovený počet slabik nebo je fronta vyprázdněna. Počet bajtů odeslaných při jedné obsluze fronty lze stanovit pro každou frontu zvlášť. CBQ může být použito pro implementaci AF PHB, kdy je třeba vytvořit čtyři fronty, pro každou třídu AF PHB jednu. Metoda musí být doplněna o mechanizmus drop precedence (priorita zahazování) jednotlivých tříd AF PHB, například metodu WRED.

6.4.3. Weighted fair queueing - WFQ)

U této metody jsou průběžně obsluhovány všechny fronty. Všechny mají stejnou prioritu. Jednotlivým frontám je přidělována část kapacity výstupní linky, odpovídající váze přiřazené ke každé frontě. Není-li kapacita přidělená určité frontě využita, může být využita pro obsluhu jiné fronty. Pro přicházející pakety je vypočten čas, kdy nejpozději má být paket odeslán. Tento výpočet vychází z počtu paketů umístěných ve frontě, do které má být zařazen a z rychlosti obsluhy fronty. Postavím se do fronty, spočtu počet lidí před sebou, zjistím kolik lidí bude obslouženo za jednotku času, a podle toho vypočtu, kdy mám přijít na řadu. Po skončení obsluhy každého paketu se určí, který z paketů čekajících ve frontách má být obsloužen nejdříve. WFQ může být spolu s WRED použito pro implementaci AF PHB.

6.4.4. Weighted-Random Early Detection (WRED) a Random Early Detection (RED)

RED a WRED jsou metody prevence zahlcení (congestion prevention). Opírají se o následující úvahu. Pokud necháme fronty zaplnit zcela pakety a potom teprve začneme pakety odhazovat, může dojít k synchronizaci zahlcení mezi více směrovači a vytváření vln, kdy dojde k zahlcení s následným snížením toku dat tak, že naopak nejsou cesty využity. Příčinou je chování protokolu TCP, který když zjistí ztrátu paketu, opakuje jeho vysílání a sníží rychlost vysílání toku dat. Tomuto jevu se snaží předejít metoda RED tím, že přesáhne-li plnění fronty určitou mez, začne směrovač odhazovat pakety náhodně vybraných TCP spojení. Pravděpodobnost odhození paketu se zvyšuje se zvyšujícím se zaplněním fronty. Tím dojde ke snížení objemu přenášených dat od některých spojení, a tím k vyrovnání toku dat. U metody WRED závisí pravděpodobnost odhození paketu také na značce přidělené při klasifikaci.