

QoS v kostce

Přenos hlasu a živého videa vyžaduje vyšší kvalitu přenosu a vytváří tudíž potřebu Quality of Service (QoS). Bez jakýchkoli mechanismů QoS jsou pakety zpracovávány v pořadí, v jakém jsou přijímány. Když dojde k přetížení, síťová zařízení, jako jsou směrovače a prepínače, mohou zahodit pakety. Typickými místy přetížení jsou agregace, neshoda rychlostí a napojení LAN na WAN.

Bez jakýchkoli časově citlivých paketů QoS, jako je video a hlas v reálném čase, budou zahozeny se stejnou frekvencí jako data, která nejsou časově citlivá, tj. jako je e-mail a procházení webu. Zařazování paketů způsobuje zpoždění, protože nové pakety nelze přenášet, dokud nebudou zpracovány předchozí pakety.

Dva typy zpoždění jsou pevné a variabilní (jiter). Zdroji zpoždění jsou zpoždění kódu, zpoždění paketizace, zpoždění ve frontě, zpoždění serializace, zpoždění šíření a zpoždění de-jitter (vyrovnávací paměť pro uložení paketů, které se předběhly). Jitter (rozptyl zpoždění) je rozptyl zpoždění přijatých paketů.

Hlasový a obrazový provoz jsou dva hlavní důvody QoS. Hlasový provoz je plynulý a bez velkých výkyvů, ale je citlivý na poklesy a zpoždění.

Meze přípustnosti pro hlas jsou

- latence (zpoždění) $\leq 150\text{ms}$
- jitter (rozkmit zpoždění) $\leq 30\text{ ms}$
- ztráty paketů $\leq 1\%$ šířky pásma

Meze přípustnosti pro video jsou

- latence (zpoždění) $\leq 200\text{-}400\text{ ms}$
- jitter (rozkmit zpoždění) $\leq 30\text{-}50\text{ ms}$
- Loss $\leq 0,1 - 1\%$ při šířce pásma

Hlas může tolerovat jen určité množství latence (zpoždění), rozptylu zpoždění (jitter) a ztráty bez viditelných účinků. Video provoz je citlivější na výpadky než video.

Video provoz je náročnější než hlasový provoz kvůli velikosti paketů, které odesílá po síti. Video provoz je robustní, chamtivý, citlivý na poklesy a zpoždění.

Datový provoz není tak náročný jako hlasový a video provoz. Datové pakety často používají aplikace TCP, které mohou přenášet data znovu, a proto nejsou citlivé na poklesy a zpoždění.

Zásady QoS implementované správcem sítě se stanou aktivními, když na spoji dojde k přetížení. Queuing je nástroj pro správu přetížení, který dokáže před odesláním do cíle vyrovnávat, prioritizovat a v případě potřeby přeuspořádat pakety.

FIFO (First in, First out) zařazuje vyrovnávací paměti do fronty a předává pakety v pořadí podle jejich příchodu. FIFO nemá žádnou koncepci priority ani tříd provozu, a proto nerozhoduje o prioritě paketů.

WFQ (Weighted Fair Queuing) je automatizovaná metoda plánování, která poskytuje spravedlivé přidělení šířky pásma veškerému síťovému provozu. WFQ aplikuje prioritu nebo váhy na identifikovaný provoz a klasifikuje jej do konverzací nebo toků.

CBWFQ (Class Based Weighted Fair Queuing) rozšiřuje standardní funkčnost WFQ o podporu pro uživatelem definované třídy provozu. S CBWFQ definujete třídy provozu na

základě kritérií shody včetně protokolů, seznamů řízení přístupu (ACL) a vstupních rozhraní. Funkce LLQ (Low Latency Queuing) přináší do CBWFQ přísné prioritní řazení (Priority Queuing) – speciální prioritita pro hlas.

Existují tři modely pro implementaci QoS: Model nejlepšího úsilí (Best-effort), Integrované služby (IntServ) a Diferencované služby (DiffServ). U Best-effort se o nic nestaráme a věříme, že vše dobře dopadne samo.

Model architektury IntServ byl vyvinut tak, aby vyhovoval potřebám aplikací v reálném čase, jako je vzdálené video, multimediální konference, aplikace pro vizualizaci dat a virtuální realita. Tato služba je zabezpečována od konce ke konci pomocí rezervace paměti, je náročná na paměti routerů, které pro svoji potřebu blokuje, a musí se průběžně potvrzovat, že je tato služba stále zapotřebí. Pro větší sítě, jako je internet, je příliš náročná.

Model DiffServ QoS určuje jednoduchý a škálovatelný mechanismus pro klasifikaci a správu síťového provozu. Návrh DiffServ překonává omezení modelů best-effort i IntServ, lze ji použít i v internetu.

V poli DSCP (Differentiated Services Code Point) můžeme rozlišit provoz tří kategorií, a to

- Best-Effort (BE), bez priority,
- Expedited Forwarding (EF) vysoce prioritní, používaný jen pro hlas,
- Assured Forwarding (AF) s třemi úrovněmi priority a třemi úrovněmi vyhazování.

Existují tři kategorie nástrojů QoS:

- nástroje pro klasifikaci a označování (značkování),
- nástroje pro předcházení přetížení a
- nástroje pro řízení přetížení.

Klasifikace určuje třídu provozu, ke které pakety nebo rámce patří.

Metody klasifikace toků provozu na vrstvě 2 a 3 zahrnují použití rozhraní, seznamů ACL a map tříd. Provoz lze také klasifikovat na vrstvách 4 až 7 pomocí programu Network Based Application Recognition (NBAR).

Správa přetížení zahrnuje metody zařazování do fronty a plánování, kde je nadměrný provoz ukládán do vyrovnávací paměti nebo zařazen do fronty (a někdy zrušen), zatímco čeká na odeslání výstupního rozhraní.

Nástroje pro zamezení zahlcení pomáhají monitorovat zatížení síťového provozu ve snaze předvídat a vyhnout se zahlcení na běžných úzkých místech v síti a v interních sítích, než se přetížení stane problémem.

Cisco IOS QoS zahrnuje mechanismus WRED (weighted random early detection, neboli vážené náhodné včasné zjišťování) jako možné řešení zamezení přetížení (resp. zahlcení). Vyhazování pak začne od paketů s nižší prioritou anebo podle profilu DSCP.

Generic Routing Encapsulation (GRE) a Protokol IPSec Encapsulating Security Payload (ESP) přidávají další zátěž a díky ESP není vidět obsahu přenášených dat. Řešit to lze dvojitým způsobem a to

- ToS Byte Preservation (vytáhnu ven byte ToS type of service tak, že ho zkopíruji z GRE záhlaví do IP Sec záhlaví)
- QoS Pre-Classify, kdy klonuji záhlaví a klon připojím k zašifrované části.