

Laboratorium nr. 8

Symulacja i badanie algorytmów QoS

1. Quality of Service

Gwarantowana jakość usług (Quality of Service) jest to zestaw technologii zapewniających odpowiedni (przewidywalny) poziom usług w kontekście przepustowości (ang. bandwidth), opóźnień (ang. delay) i zmienności opóźnień (ang. jitter).

Warunki panujące w sieci np. współczynnik utraty pakietów oraz łączne opóźnienie wpływają na poziom usług realizowanych w sieci. Bez względu na to, czy przesyłane są pliki z serwerów ftp, przeglądane strony WWW, odbierana poczta, czy przeprowadzana jest wideokonferencja, każdy transmitowany pakiet traktowany jest w ten sam sposób. Routery będące między nadawcą i odbiorcą traktują wszystkie pakiety identycznie, więc w przypadku znacznego obciążenia łączy może dojść do znacznego pogorszenia jakości transmisji strumieniowej wywołanej wzrostem utraty pakietów oraz opóźnień.

Tabela 1 prezentuje wpływ ograniczeń transmisji na różne rodzaje usług internetowych. Ruch generowany przez różne usługi jest wrażliwy w mniejszym lub w większym stopniu na występowanie określonego ograniczenia.

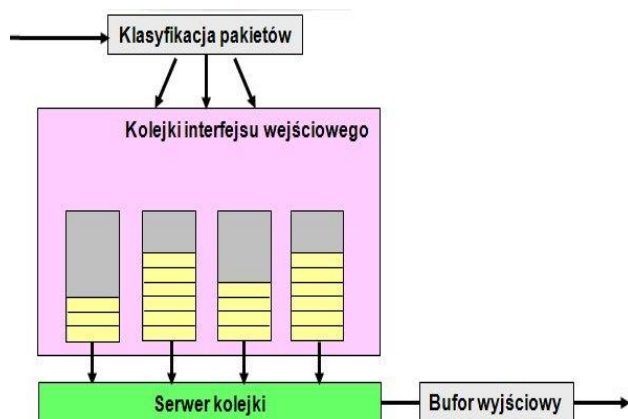
Rodzaj ruchu	Ograniczenia			
	Szerokość pasma	Utrata pakietów	Opóźnienie	Fluktuacja pakietów
Głos	bardzo mały	średni	duży	duży
Email	mały	duży	mały	mały
Transfer plików	duży	duży	mały	mały
Wideo konferencja	duży	średni	duży	duży

2. Kolejowanie pakietów

W celu zapewnienia odpowiedniego poziomu świadczonych usług stosuje się m.in. różne mechanizmy szeregowania pakietów, które zapobiegają powstawaniu skutków chwilowych

przeciążeń w sieci. Wadą mechanizmów kolejkowania jest to, że w ogóle nie sprawdzają się na łączach stale maksymalnie obciążonych, ponieważ dochodzi wtedy do ciągłego przeciążania buforów.

Na rysunku 1 przedstawiony został ogólny schemat działania algorytmów kolejkowania. Pakiety klasyfikowane są na podstawie określonych kryteriów np. typ usługi, pochodzenie, port, protokół i trafiają do specjalnych kolejek. Stamtąd kierowane są w określonej kolejności do bufora wyjściowego.



Rys. 1. Ogólny schemat mechanizmu kolejkowania [1]

Algorytm kolejkowania FIFO jest jednym z najczęściej stosowanych we współczesnych sieciach. FIFO – First In, First Out, oznacza pierwszy na wejściu, pierwszy na wyjściu. W przypadku przeciążenia łącza wszystkie nowoprzybyłe pakiety umieszczane są w pojedynczym buforze, gdzie czekają na transmisję. Gdy tylko zmniejszy się obciążenie łącza bufor opróżniany jest w kolejności przybycia pakietów. Jeśli przez dłuższy okres łącze jest obciążone, w krótkim czasie może dojść do przepełnienia bufora. Polityka odrzucania pakietów określa, które pakiety mają zostać odrzucone, czy nowoprzybyłe, czy usunięte z kolejki inne pakiety.

Kolejkowanie FIFO nie sprawdza się w obsłudze aplikacji krytycznych oraz wrażliwych na opóźnienia, ponieważ niektóre źródła mogą doprowadzić do zajęcia całego pasma, a w konsekwencji do jego znaczącego ograniczenia dla innych.

W kolejkowaniu priorytetowym tworzone są cztery kolejki o różnych priorytetach: wysokim, średnim, normalnym i niskim. Do każdego przybywającego pakietu przypisuje się odpowiednią klasę priorytetu. Klasa priorytetu może być zależna od szeregu czynników np. źródłowego adresu IP, wartości bitów Type of Service, numeru portu, interfejsu wejściowego, itp. Możliwe jest również uwzględnianie rodzaju protokołu i nadawanie wyższych priorytetów protokołom czasu rzeczywistego (RTP i RTCP). Priorytety ustalane są na podstawie odpowiednio zdefiniowanych reguł przez administratora sieci. Kolejki opróżniane są w kolejności od tej o najwyższym priorytecie, do tej o najniższym. Kolejka o niższym priorytecie obsługiwana jest dopiero, gdy kolejka o wyższym priorytecie będzie całkowicie opróżniona. Oznacza to, że jeśli jest dużo pakietów o wysokim priorytecie może dojść do dużych opóźnień w kolejkach o najniższych priorytetach. Pakiety należące do tej samej klasy priorytetu obsługiwane są metodą FIFO.

Kolejka o najwyższym priorytecie zapewnia bardzo efektywną obsługę, jednak przy pewnym nakładzie czasu w związku z klasyfikacją pakietów. Kolejowanie priorytetowe bardzo dobrze sprawdza się do obsługi aplikacji wymagających absolutnego pierwszeństwa, nie zapewnia jednak możliwości dopasowania się do warunków panujących w sieci oraz zastosowania protokołu rezerwacji zasobów RSVP.

Kolejkowanie Custom Queuing umożliwia sterowanie pasmem na interfejsach routera. Gwarantuje, że żaden typ ruchu nie przekroczy skonfigurowanego procenta przydzielonej mu przepustowości, nawet w przypadku bardzo dużego obciążenia. CQ opiera się na siedemnastu kolejkach, z czego wykorzystywane jest szesnaście.

Kolejka zerowa jest kolejką systemową, posiada najwyższy priorytet i wykorzystywana jest przez pakiety o znaczeniu specjalnym. Pozostałe pakiety trafiają do kolejek od 1 do 16. Klasyfikacja przyporządkowania pakietów do kolejek odbywa się na takiej samej zasadzie jak w Priority Queuing. Każda kolejka ma z góry określoną wielkość w bajtach. Po wytransmitowaniu wszystkich danych z kolejki router przechodzi do następnej, odbywa się to zgodnie z algorytmem round-robin. Udział w paśmie może być rozstrzygnięty na podstawie liczby bajtów pobranych w danej chwili z kolejki lub na podstawie liczby pakietów. Wszystkie kolejki mogą mieć niezależnie od siebie ustawioną wartość limitu bajtów możliwych do jednorazowego wysłania. Rozwiązanie takie pozwala na sterowanie wielkością pasma dla każdego rodzaju pakietów.

Kolejkowanie CQ zapewnia przy małym ruchu dynamiczną alokację zasobów oraz pozwala na dowolną rezerwację przepływności dla ruchu priorytetowego bez zagrożenia zagłodzenia innego rodzaju ruchu. Podobnie jak w przypadku PQ, także i to kolejkowanie nie pozwala na zastosowanie protokołu RSVP.

Algorytm WFQ jest jednym z najbardziej złożonych metod kolejkowania, zapewniający sprawiedliwy przydział łącza wszystkim rodzajom ruchu. Stosowany jest przede wszystkim w przypadkach, gdy w sieci znajdują się źródła o dużym zróżnicowaniu natężeń. W sposób dynamiczny rozróżnia pakiety danych (połączenia strumieniowe, ftp, VoIP itp.) i zarządza odrębnymi kolejkami dla każdego rodzaju. Do klasyfikacji ruchu oraz przydzielania łącza używane są wagi (priorytety) ustawiane na podstawie pola IP Precedence.

Mechanizm WFQ analizuje nagłówki IP i odczytuje wartość priorytetu z przedziału 0 – 7. Następnie przydziela proporcjonalnie więcej pasma transmisjom z wyższym priorytetem. Powoduje to, że w pierwszej kolejności obsługuje ruch o najwyższym priorytecie, przykładowo związany z aplikacjami strumieniowymi, a dopiero w dalszej pozostałe pasmo jest sprawiedliwie dzielone między innymi aplikacjami. Dane o mniej krytycznych parametrach, przekazywane są później niż te bardziej wrażliwe na ograniczenia w transmisji.

Kolejkowanie WFQ zapewnia bardzo dobrą adaptację do warunków panujących w danej chwili w sieci, a jeśli w systemie działa protokół RSVP, to potrafi on wykorzystać funkcje WFQ do ustalania kolejności pakietów oraz gwarantowania pasma.

CAR (Committed Access Rate) jest mechanizmem kolejkowania umożliwiającym jawne określanie limitu przepustowości dla ruchu określonego typu. W przypadku CAR kryterium klasyfikacji pakietów może być również numer interfejsu wejściowego, wartość pola IP Precedence pakietu, czy zapis w

liście kontroli dostępu, precyzujący, o jaki strumień danych chodzi. Zadaniem administratora jest określenie działania podejmowanego w momencie spełnienia przez pakiet danego kryterium klasyfikacji i przekroczenia dopuszczalnej dla danego strumienia szybkości transmisji.

3. Przeciążenia w sieciach

Przeciążenia występujące na łączach mogą doprowadzić do sytuacji w których strumieniowe transmisje multimedialne przestają być użyteczna dla osoby oglądających lub prowadzących wideokonferencje. Dzieje się tak, ponieważ pierwotna koncepcja przesyłania danych w sieci Internet nie brała pod uwagę transmisji w czasie rzeczywistym. W celu rozwiązania tego problemu stworzone zostały specjalne metody służące do zapobiegania przeciążeniom.

Jednym z najczęściej stosowanych mechanizmów przeciwdziałania zatorom w sieci jest RED (Random Early Detection). Wykorzystywany jest w sytuacjach ciągłego, długotrwałego przeciążenia węzłów w sieci. Mechanizm RED, tuż przed wystąpieniem zatoru, rozpoczyna losowe odrzucanie pakietów w celu zmniejszenia ruchu generowanego przez niektóre źródła. Rozwiązanie takie pozwala na lepsze wykorzystanie pasma oraz zapobiega wystąpieniu globalnej synchronizacji.

Działanie mechanizmu RED opiera się na wyznaczeniu średniej długości kolejki i przyrównywaniu jej zajętości do dolnego i górnego progu. W sytuacji, gdy średni rozmiar kolejki znajduje się poniżej minimalnego progu, to żadne pakiety nie są odrzucane. Wypełnienie kolejki na poziomie między minimalnym i maksymalnym progiem powoduje losowe odrzucanie pakietów z odpowiednim prawdopodobieństwem. Czym bardziej kolejka jest zajęta, tym więcej odrzucanych jest pakietów. Przekroczenie górnej granicy skutkuje odrzucaniem wszystkich pakietów. Konsekwencją takiego działania jest większe prawdopodobieństwo odrzucania pakietów ze źródeł generujących większy ruch.

Modyfikacją RED jest mechanizm WRED (Weighted Random Early Discarding). W przeciwieństwie do zwykłego reagowania na przeciążenia, WRED potrafi rozpoznawać klasy ruchu zapisane w IP Precedence. Pozwala to na oddzielne definiowanie prawdopodobieństwa odrzucania ruchu w zależności od jego istotności. Przykładowo w mniejszym stopniu odrzucane będą pakiety generowane przez aplikacje czasu rzeczywistego, a w większym pozostałe.

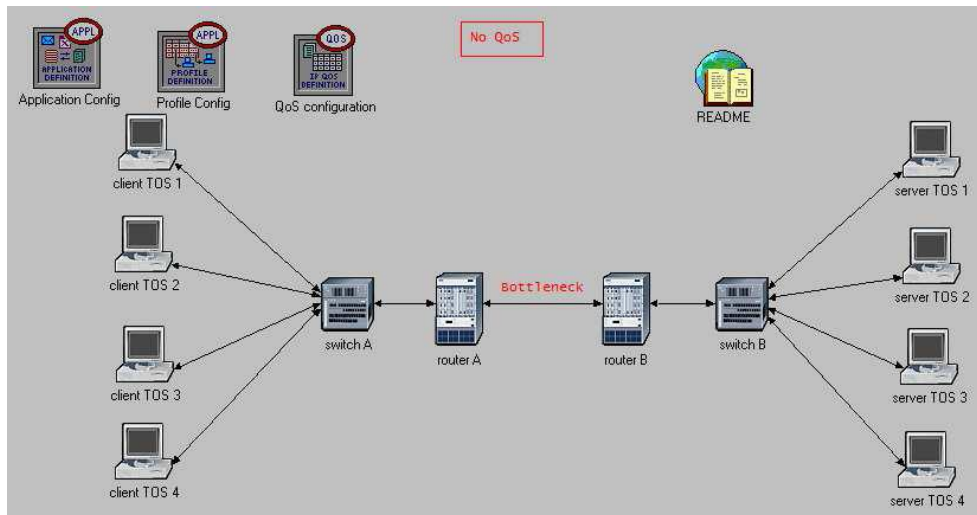
Obecnie obserwuje się coraz większą integrację usług wykorzystującą sieci IP. Oprócz przesyłu danych, których stałość strumienia nie jest krytyczny pojawiło się wiele usług w których czas błędzenia pakietu stał się bardzo ważnym parametrem. Sytuacja taka ma miejsce w przypadku wykorzystania sieci IP do prowadzenia rozmów telefonicznych lub wideokonferencji. Należy zatem zadbać, aby pakiety danych wymagające terminowego dostarczenia do odbiorcy posiadały wyższy priorytet. W tym celu stworzony system standardów stosowanych w telekomunikacji zwanych QoS (Quality of Service). Oprogramowanie OPNET IT Guru umożliwia również symulację sieci z zaimplementowanymi algorytmami QoS. Celem ćwiczenia jest zbadanie sieci wspierającej QoS przy różnych metodach segregacji i priorytetyzacji pakietów.

4. Wykonanie ćwiczenia

Podczas realizacji ćwiczenia należy przeprowadzić szereg symulacji w oparciu o gotowy projekt o

nazwie IP_QoS. W poszczególnych scenariuszach należy wykonać następujące symulacje i analizy:

1. Otworzyć scenariusz: Scenarios / Switch To Scenario / Baseline. Wykonać symulację i otworzyć wyniki analiz. Zwrócić szczególną uwagę na czas dotarcia pakietu End-To-End Delay w usłudze wideokonferencji oraz poszczególnych typów serwisów (TOS1 – TOS4), wykorzystanie łącza Router A – Router B,



2. Otworzyć scenariusz FIFO (First In First Out) i wykonać symulację. Skomentować wyniki.
3. Otworzyć scenariusz Priority Queuing (PQ) i wykonać symulację. Skomentować wyniki.
4. Otworzyć scenariusz Custom Queuing (CQ) oraz Custom Queuing with Low Latency Queue (CQLLQ) i wykonać symulację. Skomentować wyniki. Jak wpływa kolejekowanie użytkownika na jakość innych połączeń ?
5. Otworzyć scenariusz Weighted Fair Queuing (WHQ) oraz Weighted Fair Queuing with Low Latency Queue (WHQ-LLQ) i wykonać symulację. Skomentować wyniki.
6. Za pomocą dwóch ostatnich scenariuszy porównać dwa rodzaje sieci podstawową (Network without CAR) oraz sieć wyposażoną w CAR (Committed Access Rate). Skomentować różnice w założeniach dotyczących sieci oraz rezultatach symulacji.

[1]. CISCO Technical Update - Mechanizmy QoS w sieciach IP