

Mirosław Siergiejczyk
Politechnika Warszawska Wydział Transportu
Stanisław Gago
Politechnika Warszawska Wydział Transportu

ZAGADNIENIA OCENY JAKOŚCI USŁUG TELEMATYCZNYCH W TRANSPORCIE KOLEJOWYM

Streszczenie: W referacie przeanalizowano warunki, w jakich funkcjonuje system telematyki ze szczególnym uwzględnieniem licznych zagrożeń dla struktury fizycznej systemów telematyki i poprawności realizowanych przez nich usług. Zaproponowano modele zapewnienia jakości dla transmisji informacji w systemach telematyki transportu. Scharakteryzowano podstawowe mechanizmy kształtowania ruchu w celu zagwarantowanie minimalnej przepustowości aplikacjom krytycznym z punktu widzenia działalności przedsiębiorstwa. Przedstawiono metodę kształtowania ruchu. Metoda ta posłużyła do opracowania koncepcji regulowania ruchu telekomunikacyjnego, powiązanego z działalnością największego polskiego przewoźnika kolejowego – PKP CARGO S.A. Koncepcja została opracowana w taki sposób, żeby zapewnić priorytet danym pochodzącym od aplikacji krytycznych z punktu widzenia działalności tego przedsiębiorstwa.

Słowa kluczowe: telematyka transportu, usługi, jakość, modele, kształtowanie ruchu,

1. WSTĘP

Świadczenie na odpowiednim poziomie usług transportowych w zakresie przewozu towarów jest dużym problemem i wymaga styków (interfejsów) do wymiany danych (informacji) między uczestnikami łańcucha transportowego. Odnosi się to także do przedsiębiorstw kolejowych, gdzie występuje wymiana informacji pomiędzy różnymi zarządcami infrastruktury kolejowej i przedsiębiorstwami przewozowymi. Ruch prowadzony przez granice państw wymaga nie tylko wymiany danych (informacji) dotyczących przewożonych towarów, ale także danych technicznych dotyczących pociągu i wagonów. W uproszczonych przepisach, wymagana jest wymiana danych związana z prowadzeniem ruchu pomiędzy zarządcami infrastruktury, wymiana danych dotyczących przewoźnika towarów tj. organizacji transportu towarów z odpowiednim listem przewozowym i odpowiednią autoryzacją.

Wymiana danych w transporcie wykorzystywana jest w systemach zarządzania, kierowania i sterowania procesami transportowymi oraz w systemach informacyjnych. Przedsiębiorstwa kolejowe prawie od początku swojego istnienia, używały środków łączności do sprawnego prowadzenia ruchu kolejowego. Można stwierdzić, że łączność stanowiła narzędzie pracy, które w znakomity sposób usprawniało zarządzanie ruchem kolejowym. Wraz z rozwojem sieci kolejowej i postępu w technice telekomunikacyjnej narzędzie to „sięgało w coraz to głębsze warstwy procesu przewozowego” np. kierowanie czy nawet sterowanie ruchem kolejowym. Obecnie wraz z wprowadzeniem sieci pracujących z protokołami TCP/IP (ang. Transmission Control Protocol/Internet Protocol) można zaobserwować coraz to szersze wprowadzenie usług telekomunikacyjnych do procesu przewozowego i obsługi klienta (pasażera). Dla przykładu można tu wymienić bazy danych (ang. data center), rezerwowe bazy danych (ang. Recovery Data Center), wydzielone wirtualne sieci prywatne (ang. Virtual Private Network) dla poszczególnych branż kolejowych a nawet aplikacje, lokalizacje pociągów, przesyłek, zdalna rezerwacja biletów itd. Usługi te wspierają konkurencyjność przewoźników i firm kolejowych w walce o klienta. Wymienione powyżej przykładowe usługi nie mogłyby być świadczone bez odpowiedniej sieci telekomunikacyjnej (w tym przypadku bez sieci pracującej zgodnie z protokołami TCP/IP).

Istotnym zagadnieniem jest znalezienie rozwiązania transmisji danych uwzględniającego specyficzne wymagania transportu. Głównie chodzi tu o zapewnienie ciągłości świadczenia usługi (tzw. dostępność usługi), zapewnienie na wymaganym poziomie jakości świadczonych usług transmisji danych QoS (ang. Quality of Service) oraz odpowiednie bezpieczeństwo przekazywanych danych. Biorąc pod uwagę wielość aplikacji, które są i będą używane w transporcie jak i duża ilość danych używanych przez te aplikacje jedynym rozwiązaniem dla potrzeb przesyłania danych są sieci pracujące zgodnie ze stosem protokołów TCP/IP.

2. JAKOŚĆ USŁUG W SIECIOWYCH SYSTEMACH TELEMATYKI TRANSPORTU

2.1. Zależności w sieci realizującej usługi telematyczne

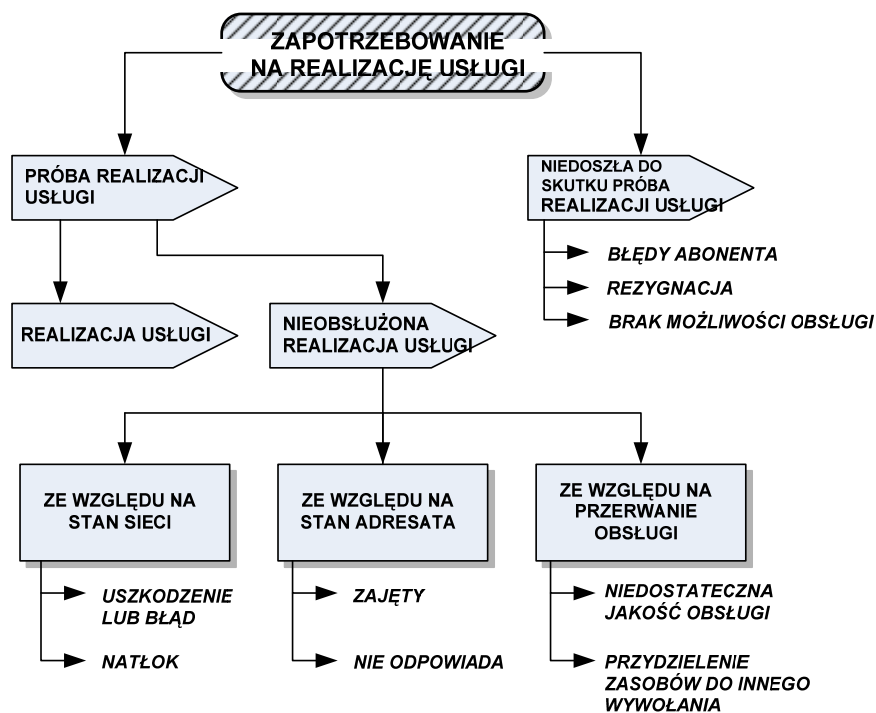
Współczesne, wysoce zdecentralizowane i o dużej skali systemy sieciowe zwiększają efektywność funkcjonowania organizacji/przedsiębiorstwa przez integrację w jednolitą całość wszystkich poziomów organizacyjnych. Jednakże, takiemu łączeniu towarzyszy wzrost poziomu ryzyka związanego z wtargnięciami do zasobów informacyjnych osób nieuprawnionych. Poziom tego typu ryzyka może zostać zmniejszony przez wdrażanie kompleksowych i wzajemnie uzupełniających się mechanizmów zapewniających optymalne wykorzystanie potencjału sieci celem uzyskania efektywnej pracy systemu.

Szybki rozwój i znaczenie usług telematycznych w transporcie wymaga określenia poziomu ich jakości QoTS (ang. Quality of Transport Telematics Service). W przypadku usług telekomunikacyjnych pojęcie jakości nie jest jednoznaczne i może być definiowane w różny sposób. Obecnie, w określeniach dotyczących jakości usług, coraz częściej

uwzględnia się ich zgodność z oczekiwaniami klienta. Zalecenie ITU-T E.800 [7] definiuje jakość usług telekomunikacyjnych jako „stopień zadowolenia użytkownika usługi wynikający z globalnego wykonania usługi”. Zatem jakość usług telekomunikacyjnych QoS jest pojęciem bardzo szerokim, odnoszącym się do: działania sieci telekomunikacyjnej, parametrów transmisji usługi w relacji od ”end – to – end”, tj. od dostawcy do odbiorcy usługi, a także pracy personelu zatrudnionego przy świadczeniu usług.

Potrzeby ze strony użytkownika sieci realizującej usługi telematyczne można przedstawić w postaci strumienia zapotrzebowań na usługi telematyczne realizowane przez określoną sieć. Ze względu na uwarunkowania wewnętrzne i zewnętrzne sieci, strumień zapotrzebowań obsłużonych jest kształtowany przez szereg zdarzeń, działań i procesów powodujących znaczną jego zmianę w stosunku do strumienia zapotrzebowań zgłoszonych. Podstawowe zależności zachodzące w sieci transmisji informacji telematycznych i ich wpływ na nie obsłużenie zgłoszonego przez użytkownika zapotrzebowania przedstawiono na rysunku 1 [14].

Rozpoczęcie realizacji usługi nie zawsze jest równoznaczne z jej zrealizowaniem. Obsługa zapotrzebowania może zostać przerwana w trakcie realizacji lub może nie być zrealizowana ze względu na stan sieci.



Rys. 1. Powiązania zachodzące w sieci realizującej usługi telematyczne [14]

2.2. Obszary oceny jakości usług telematycznych

Z punktu widzenia użytkownika systemu (sieci) najważniejsze znaczenie ma jakość usług oferowanych przez system (sieć). Jakość usług może odnosić się do różnych aspektów charakteryzujących system w oczach użytkownika jak np. psychologicznych

(zadowolenie użytkownika sieci) lub informatycznych – według technicznych norm ogólnosięwiatowych (zalecenie ITU-T E.800 Międzynarodowej Unii Telekomunikacyjnej) i narodowych (PN-93/N) [11]. Z merytorycznej analizy treści zawartych w wymienionych dokumentach normatywnych wynika uniwersalność wzajemnych powiązań stanu i własności sieci oraz jakości usług w poniżej wymienionych siedmiu obszarach.

Dostarczenie (utrzymanie) usługi – jest to właściwość określająca zdolność operatora do dostarczenia usługi oraz pomocy w jej wykorzystaniu. Miarami utrzymania usługi są następujące wielkości: średni czas dostarczenia usługi, prawdopodobieństwo błędnego rozliczenia usługi oraz prawdopodobieństwo integralności rozliczenia (zgodności z rzeczywistością).

Operatywność usługi – jest to zdolność usługi do łatwego i pomyślnego oraz zgodnego z przeznaczeniem wykorzystania jej przez użytkownika. Miarami operatywności usługi są prawdopodobieństwa: pomyłki użytkownika związanej z korzystaniem z usługi, błędnego wyboru numeru przez użytkownika, rezygnacji z próby dostępu do usługi przez użytkownika, rezygnacji z usługi w czasie jej trwania.

Zdolność usługowa – jest to możliwość zrealizowania wymaganej przez użytkownika usługi oraz jej kontynuacji bez zmniejszenia poziomu jakości w wymaganym czasie. Cecha ta zawiera cechy składowe takie jak: dostępność, ciągłość oraz integralność usługi.

Dostępność usługi – jest to zdolność do otrzymania usługi w zadanym czasie i w granicach określonej tolerancji czasowej. Miary dostępności usługi dzieli się na dwa zbiory:

- prawdopodobieństwa: uzyskania połączenia, bezbłędnego routingu, dostępności sieci, dostępu do usługi;
- średniego opóźnienia: dostępu do usługi, dotarcia do odbiorcy.

Ciągłość usługi – jest zdolnością uzyskania dostępu do usługi i jej kontynuowania w wymaganym przedziale czasowym. Ciągłość usługi zależy od poziomu jakości transmisji, niezawodności systemu, wydajności ruchowej oraz dostępności komponentów systemu. Do podstawowych miar tej zdolności zalicza się wartości następujące prawdopodobieństwa: ciągłości połączenia, ciągłości usługi, przerwania połączenia, ukończenia realizacji usługi.

- integralność usługi – jest to miara dostarczenia usługi bez nadmiernego obniżenia jej jakości. Miarami integralności usług są: czas chwilowego przerwania usługi, czas pomiędzy dwoma kolejnymi przerwaniem usługi oraz średnie czasy trwania pojedynczego przerwania oraz pomiędzy dwoma kolejnymi przerwaniem.
- bezpieczeństwo usługi – pojęcie to odnosi się do zabezpieczeń przeciwko nieautoryzowanemu dostępowi, wykorzystaniu i monitoringowi oraz celowemu wprowadzaniu zakłóceń, a także przeciwko narażeniom uwzględniającym oddziaływanie środowiska i czynnik ludzki. Zakłócenia są przyczynami przerw w realizacji usługi oraz wpływają na pozostałe wyżej wymienione właściwości usług.

Do przyczyn, które mogą wpływać na jakość usług świadczonych w sieci zalicza się:

- wystąpienie stanu niezdatności elementów wchodzących w skład sieci – wynikające z naturalnego procesu starzenia się;
- czynniki zewnętrzne – o charakterze katastroficznym, oddziałujące destrukcyjnie na elementy składowe sieci, czynniki te są zgrupowane w dwie podstawowe klasy:

- żywiołowe (naturalne), które są następstwem procesów zachodzących w przyrodzie np.: wichury, trąby powietrzne, powódzie, obsunięcia się mas ziemnych, trzęsienia ziemi, wyładowania elektrostatyczne, opady atmosferyczne itp.;
- wytworzone przez człowieka (nienaturalne), które są przejawami szczególnej działalności ludzkiej np.: zamierzone funkcjonowanie innych systemów realizowane przy braku pełnej informacji i kultury technicznej oraz celowe działania destruktora, które mogą zagrozić strukturze fizycznej i dostępności do usług sieciowych.

Wymienione powyżej obszary gwarantowanego poziomu zrealizowania usługi, implikują wydajność systemu uwzględniającą poniżej podane własności.

Wydajność ruchową systemu – jest to zdolność systemu (lub jego komponentu) do spełnienia wymagań ruchowych dotyczących jego natężenia oraz innych charakterystyk z uwzględnieniem warunków wewnętrznych systemu (np. intensywności uszkodzeń komponentów). Najczęściej stosowane miary wydajności ruchowej to: jest wartość oczekiwana czasu opóźnienia (średni czas opóźnienia) realizacji usługi, średnia liczba straconych bitów, (pakietów, ramek) przy realizowaniu zadanej funkcji systemowej.

Niezawodność systemu – jest to bezwymiarowa właściwość systemu opisująca jego gotowość do świadczenia usług oraz uwzględniająca czynniki wpływające na gotowość, takie jak: nieuszkodzalność, podatność na obsługę i zapewnienie środków obsługi.

Zdolność transmisyjną systemu – jest to możliwość systemu do odtwarzania i przesyłania sygnałów, zależna od zdolności medium transmisyjnego do przesyłania różnego rodzajów sygnałów. Miarami zdolności transmisyjnej są: bitowa stopa błędów i prawdopodobieństwo jednosekundowej bezbłędnej transmisji pomiędzy nadawcą i adresatem informacji.

W zaleceniu [7] definiuje się wydajność sieci NP (*Network Performance*), jako możliwość zapewnienia przez sieć lub jej określoną część komunikacji pomiędzy użytkownikami. W uproszczeniu stwierdza się, że zapewnienie QoS polega na dostarczeniu odbiorcom usług z określoną przez warunki umowy jakością, przy czym należy zaznaczyć, że nie chodzi tu o dostarczenie usług idealnych.

Na pozytywną ocenę jakości usługi telekomunikacyjnej widzianą przez klienta od strony technicznej sprawności działania sieci telekomunikacyjnej składają się następujące elementy:

- szybkość zestawienia połączenia;
- zestawienie połączenia z właściwym odbiorcą;
- jakość transmisyjna zestawionego połączenia (szumy, przesłuchy, tłumienie, echo, bitowa stopa błędów);
- prawidłowa taryfikacja połączenia;
- dostępność wszystkich usług sieciowych;
- możliwość szybkiego rozszerzenia listy usług.

Jeśli którykolwiek z powyższych elementów nie odpowiada warunkom umowy SLA (*Service Level Agreement*) dotyczącej poziomu jakości usług albo odpowiednim wymaganiom technicznym, użytkownik nie jest zadowolony z poziomu jakości oferowanych w sieci usług. Przyczyny pogorszenia jakości usług są bardzo różne, gdyż na połączenie realizujące określoną usługę składa się praca wielu urządzeń telekomunikacyjnych. Do częstych przyczyn wpływających na obniżenie jakości

świadczenia usługi telekomunikacyjnej można zaliczyć: wadliwe działanie zespołów funkcjonalnych centrali lub systemów teletransmisyjnych, wymiarowanie central i łączy ze zbyt dużymi stratami ruchu, pogorszenie parametrów teletransmisyjnych łączy, niewłaściwą strukturę sieci jako całości, niezadowolający stan techniczny całej sieci, nieefektywność prac utrzymaniowych, bieżące przeciążenia ruchowe sieci (obecnie w „normalnych” warunkach eksploatacji poziom takiej degradacji nie jest odczuwalny przez klienta).

3. JAKOŚĆ USŁUG TELEMATYCZNYCH W SIECIACH TCP/IP

Gwarantowana jakość usług w sieciach IP, określana także mianem QoS (ang. Quality of Service) jest jedną z najważniejszych charakterystyk wykorzystywanych dziś w Internecie. Niestety, oryginalna architektura protokołu TCP/IP nie wspiera QoS, ponieważ wymagałoby to utrzymywania we wszystkich węzłach sieci informacji dotyczących stanu każdego aktualnie obsługiwanego przepływu, wpływających na mechanizmy sygnalizacji i kolejgowania danych. Wszystkie routery, przełączniki oraz pozostałe elementy sieciowej infrastruktury znajdujące się między nadawcą i odbiorcą, musiałyby zapewnić obsługę funkcji QoS, a dodatkowo każdy z nich musiałby zarezerwować na ten cel wymagane zasoby systemowe, takie jak pasmo, moc obliczeniowa CPU (ang. Central Processing Unit) oraz pamięć operacyjna. W literaturze przedmiotu dotyczącej komercyjnego wykorzystania sieci TCP/IP można spotkać różne definicje jakości świadczonych usług (QoS), a w zasadzie problem zapewnienia QoS w sieci pakietowej sprowadza się do zapewnienia w sieci możliwości przekazu wybranych strumieni ruchu z gwarancjami dotyczącymi takich parametrów jak:

- wymagana przepływność (ang. bandwidth),
- dopuszczalne opóźnienia pakietów (ang. delay) i ich zmienność (ang. jitter),
- dopuszczalne prawdopodobieństwo strat (ang. bit error rate).

Sieć powinna zawierać zbiór technologii, które zapewnią użytkownikom przewidywalny poziom usług w tych trzech obszarach, w różnych skalach czasowych i tak:

- wymiarowanie sieci – dni,
- procedury przyjmowania nowych wywołań – minuty,
- procedury na poziomie pakietów – milisekundy.

W sieciach TCP/IP występują kilka modeli zapewnienia jakości usług QoS. Do nich można zaliczyć między innymi:

- model rezerwowania zasobów,
- model rozróżniania klas,
- model z wykorzystaniem reguł trasowania,
- model z wykorzystaniem metody kształtowania ruchu,
- model akceleracji ruchu telekomunikacyjnego.

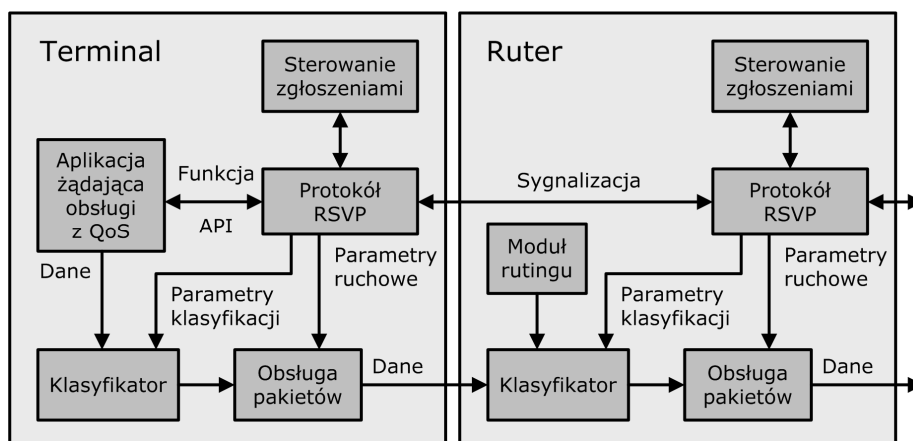
Główna różnica między tymi modelami polega na sposobie specyfikowania żądanej jakości usług tj. poprzez sygnalizację i poprzez odpowiednie oznaczanie pakietów.

3.1. Model rezerwowania zasobów

Pierwszym podejściem stowarzyszenia IETF (ang. Internet Engineering Task Force) do problemu zapewnienia usługom sieci IP określonych parametrów QoS była propozycja architektury Integrated Services (w skrócie IntServ). Stanowi ona rozszerzenie dotychczasowych mechanizmów sieci IP o możliwość realizacji usług telekomunikacyjnych wykorzystujących długookresowe strumienie pakietów z zagwarantowaniem wymaganych parametrów jakościowych. Przez strumień rozumieć należy zbiór pakietów o jednakowym źródłowym i docelowym adresie IP, identycznym źródłowym i docelowym numerze portów oraz takim samym identyfikatorze protokołu – TCP lub UDP. Głównym założeniem architektury IntServ jest zatem przekazanie informacji o wymaganiach QoS transmitowanego strumienia do węzłów sieciowych jeszcze przed rozpoczęciem transmisji. Jeżeli wymagania te nie mogą być spełnione z powodu braku wolnych zasobów, to żądanie realizacji usługi zostanie odrzucone.

Architektura IntServ składa się z trzech wzajemnie powiązanych elementów [9]:

- definicji klas usług,
- mechanizmów sterowania ruchem,
- specjalnie utworzonego na jej potrzeby protokołu sygnalizacyjnego.



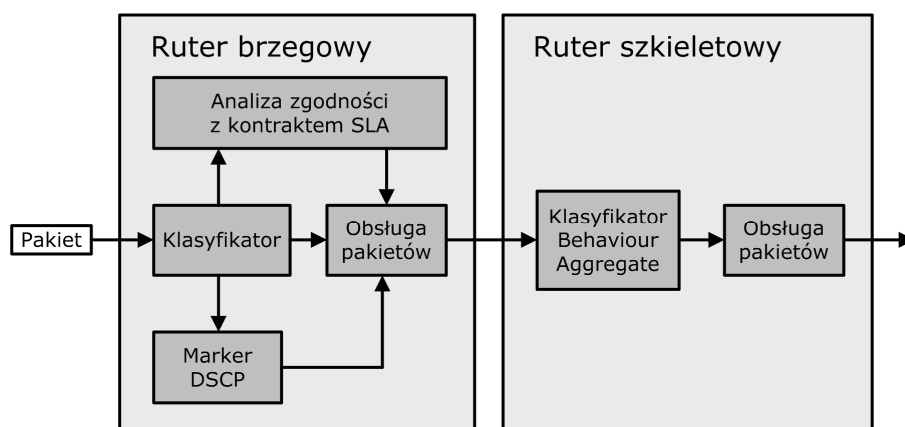
Rys. 1. Wzajemne zależności pomiędzy elementami funkcjonalnymi w architekturze IntServ [9]

Blok protokołu RSVP (ang. Resource Reservation Protocol) odpowiada za proces wymiany informacji sygnalizacyjnej wzdłuż ścieżki, po której przesyłana będzie informacja użyteczna, w celu sprawdzenia dostępności zasobów (sterowanie przyjęciem zgłoszeń) oraz ewentualnej rezerwacji zasobów. Procedura wymiany informacji sterującej ma miejsce zanim informacja użytkowa zostanie przesłana przez sieć. Należy dodać, że rezerwacja zasobów sieciowych gwarantuje żadaną przez aplikację jakość usługi tylko jeżeli generowany przez nią ruch nie przekroczy zgłoszonych wartości parametrów przepływu. Funkcja kontroli ruchu jak również ewentualnego ograniczenia jego parametrów spoczywa na węzłach sieciowych. Protokół RSVP dokonuje rezerwacji w sposób simpleksowy, a zatem utworzenie ścieżki o charakterze duplexowym wymaga zainicjowania symetrycznej rezerwacji z drugiego końca połączenia.

3.2. Model usług zróżnicowanych

Ze względu na poważne ograniczenia architektury IntServ (a zwłaszcza problemy z jej skalowalnością), stowarzyszenie IETF zaproponowało architekturę usług zróżnicowanych Differentiated Services (w skrócie DiffServ). Podstawową różnicą między tymi dwoma mechanizmami QoS jest odmienna strategia co do sposobu zapewnienia jakości transmisji. Platforma IntServ zakłada podejście sesyjne, a zatem pełną identyfikację strumienia pakietów oraz zapewnienie im właściwego poziomu QoS poprzez zarezerwowanie zasobów sieciowych (czyli pasma transmisyjnego, pamięci buforowej i mocy obliczeniowej routerów) [9].

DiffServ postuluje przeniesienie funkcjonalności na routery brzegowe sieci oraz maksymalne uproszczenie jej rdzenia, ponieważ urządzenia szkieletowe muszą obsługiwać znacznie większą liczbę strumieni danych. Zastosowano tu mechanizm polegający na tym, że routery brzegowe przydzielają pakiety do odpowiednich klas ruchu na podstawie zawartości pól ich nagłówek, natomiast węzły szkieletowe obsługują ruch z dokładnością do kilku klas reprezentujących różne poziomy obsługi. Zestaw funkcji realizowanych w węzłach architektury usług zróżnicowanych zależy od tego czy leży on na brzegu czy też w rdzeniu domeny DiffServ. Schemat działania bloków funkcjonalnych routera brzegowego i szkieletowego przedstawiony został na rysunku 2.



Rys. 2. Schemat działania elementów funkcjonalnych routera brzegowego i routera szkieletowego w architekturze DiffServ [9]

Architektura usług zróżnicowanych stanowi odpowiedź stowarzyszenia IETF na zarzuty wobec niskiej skalowalności architektury IntServ, natomiast same zasady przeciwdziałania przeciążeniom, czyli przyjmowanie zgłoszeń, szeregowanie oraz usuwanie pakietów, tak samo jak sposób routowania są w obu tych architekturach identyczne.

3.3. Wieloprotokółowa komutacja etykietowa

Dwie przedstawione architektury, a mianowicie Integrated Services oraz Differentiated Services, nie rozwiązują problemu niskiej efektywności reguł trasowania w sieciach IP. IETF w odpowiedzi na ten problem zaproponowało kolejny mechanizm sterowania

jakością usług – technikę wieloprotokołowej komutacji etykietowej MPLS (ang. MultiProtocol Label Switching). Zgodnie z zasadą działania tego protokołu każdy pakiet wchodzący w obszar domeny MPLS klasyfikowany jest w jej ruterze brzegowym LER (ang. Label Edge Router) do jednej z klas równoważności przekazywania FEC (ang. Forwarding Equivalence Class), po czym kapsułkuje się go dodatkowym nagłówkiem zawierającym etykietę z zakodowaną klasą FEC. Natomiast w ruterach szkieletowych LSR (ang. Label Switch Router) nagłówek IP nie jest już analizowany, a decyzja o wyborze następnego odcinka drogi podjęta zostaje wyłącznie na podstawie zawartości dodanej przez LER etykiety [10].

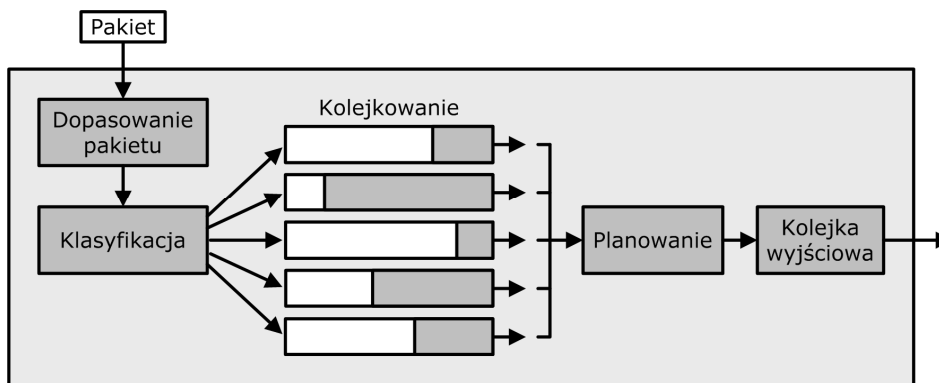
Pomimo tego, że technika MPLS ułatwia zagwarantowanie jakości obsługi poprzez kierowanie ruchu na wybrane łącza, to nie posiada procedur sterowania samymi parametrami QoS. Ponadto implementacja protokołu MPLS zazwyczaj nie działa w relacji „*end-to-end*”, co stoi w sprzeczności z podstawowym założeniem zapewnienia jakości usługom telekomunikacyjnym. Dlatego też metoda ta musi być stosowana w połączeniu z innymi mechanizmami QoS w sieciach IP, takimi jak omówione już modele IntServ oraz DiffServ.

3.4. Metoda kształtowanie ruchu

Zapotrzebowanie na przepustowość oraz parametry QoS stale rosną. Nadmierny ruch generowany przez jedne aplikacje nie może negatywnie wpływać na działanie innych, krytycznych z punktu widzenia przedsiębiorstwa, takich jak centralne bazy danych, systemy planowania zasobów przedsiębiorstwa ERP (ang. Enterprise Resource Planning), czy też systemy zarządzania relacjami z klientami CRM (ang. Customer Relationship Management).

Podstawowym mechanizmem zarządzania pasmem jest metoda kształtowania ruchu (ang. traffic shaping). Zapobiega ona powstawaniu zatorów na drodze wygładzania ruchu wprowadzanego do sieci, dzięki czemu dostawca usług ISP nie zostanie zalany pakietami i nie będzie musiał stosować twardego obcinania strumieni danych do wartości zdefiniowanej w kontrakcie SLA (ang. policing). Mechanizm ten zapewnia aplikacjom krytycznym pewną minimalną wartość dostępnego pasma, tak aby ich praca pozostawała niezakłócona.

Proces kształtowania ruchu przebiega w kilku zilustrowanych na rysunku 3 etapach, wśród których zasadniczymi elementami są klasyfikowanie oraz kolejkovanie pakietów na wyjściu z systemu, tak aby ich natężenie było równomierne bez względu na częstotliwość generowania danych w poszczególnych klasach. Ten sposób implementacji QoS realizowany jest poprzez przydzielanie pakietów do odpowiednich kategorii w oparciu o różnego typu kryteria, którymi mogą być przykładowo źródłowy i docelowy adres IP, numery portów TCP oraz UDP, wykorzystywane interfejsy, a także adresy MAC (ang. Media Access Control).



Rys. 3. Etapy procesu kształtowania ruchu

Następnie każda z klas podlega złożonemu procesowi kolejkowania, dużo bardziej rozbudowanemu niż prymitywny algorytm FIFO (ang. First In – First Out). Ostatnią fazę kształtowania ruchu stanowi planowanie porządku wybierania danych z kolejek, w taki sposób by sprawiedliwie rozdysponować poszczególnym klasom dostępną przepustowość. Etap ten przebiegać może w oparciu o kolejki klasowe oraz bezklasowe.

Kolejki bezklasowe cechuje prosta konfiguracja, co z jednej strony pociąga za sobą ich łatwą implementację, a z drugiej ograniczone możliwości priorytetowania strumieni ruchu. Do kolejek tych zaliczyć należy wspomniany algorytm FIFO oraz całą rodzinę kolejek sprawiedliwych: stochastyczne SFQ (ang. Stochastic Fairness Queuing), ważone WFQ (ang. Weighted Fair Queuing) i karuzelowe DRR (ang. Deficit Round Robin).

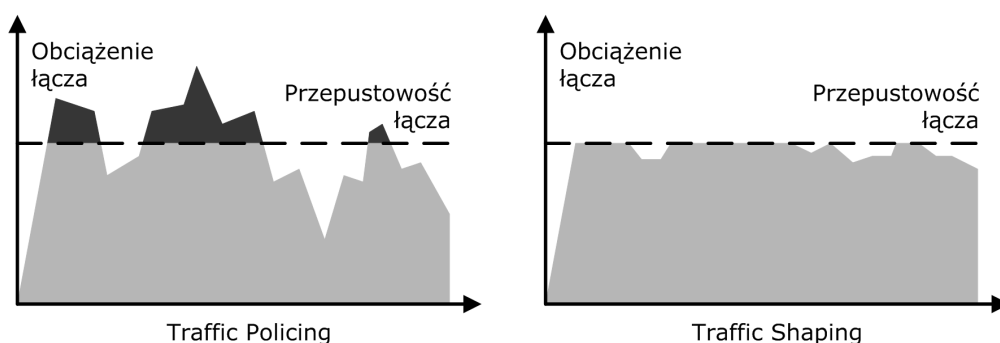
Kolejki klasowe CBQ (ang. Class Based Queuing) są dużo bardziej złożone, głównie ze względu na możliwość tworzenia hierarchii klas o różnorodnych parametrach, wśród których najważniejszym pozostaje zwykle przepustowość. Cecha ta eliminuje niedogodność niskiej konfigurowalności kolejek bezklasowych, a jednocześnie zapewnia im wysoką elastyczność. Algorytmem klasowym najczęściej wykorzystywanym w urządzeniach kształtujących ruch jest tzw. hierarchiczne wiadro z żetonami HTB (ang. Hierarchical Token Bucket) [5].

Kształtowanie ruchu umożliwia dodatkowo określanie pewnego nadmiaru danych, który zostanie przesłany bezpośrednio po rozpoczęciu transmisji (ang. burst). Metoda ta stosowana jest w przypadku krótkotrwałych transmisji o dużym natężeniu, a najlepszym przykładem obrazującym sens takiego postępowania jest ruch generowany przez protokół HTTP, gdzie żądania i odpowiedzi występują stosunkowo rzadko, lecz przenoszą dużą liczbę informacji.

Wdrożenie metody kształtowania ruchu odbywa się zwykle na styku sieci wewnętrznej przedsiębiorstwa z siecią rozległą, poprzez ulokowanie na łączu dostępowym urządzenia o nazwie „traffic shaper”. Sprzęt sieciowy obsługujący tą funkcję podzielić można na dwie grupy. Do pierwszej z nich należą przełączniki i routery wyposażone w moduły kształtowania ruchu. Druga kategoria to urządzenia dedykowane, dla których zarządzanie przepustowością jest zadaniem podstawowym. Ze względu na wysokie ceny urządzeń specjalizowanych, w niewielkich sieciach często wykorzystuje się też odpowiednio oprogramowane komputery PC.

Specyficzny przypadek kształtowania ruchu stanowi polityka jego ograniczania (ang. traffic policing), której działanie zobrażowane zostało na rysunku 4. Mechanizm ten

podobnie jak kształtowanie ruchu ogranicza przepływ danych związany z określoną aplikacją według reguł zdefiniowanych w kontrakcie SLA, jednak zamiast kolejkować pakiety przekraczające przydzieloną im przepustowość, traffic policing po prostu redukuje ten ruch (tzw. twarde odcinanie) lub zmniejsza priorytet wysyłania nadmiarowych danych. Prowadzi to do częstego retransmitowania pakietów i nieuniknionego powrotu stanu przeciążenia. Jest to zatem gorsze rozwiązanie, jednak z uwagi na swą prostotę bardzo często stosowane przez operatorów. Rezygnacja z mechanizmów kolejkowania koniecznych do kształtowania ruchu pociąga za sobą zmniejszenie czasu przesyłania danych, a także ogranicza potrzebną moc obliczeniową ruterów.



Rys. 4. Porównanie działania mechanizmów traffic policing i traffic shaping

Stosowanie mechanizmu kształtowania ruchu pozwala zredukować zatory w sieci i poprawić wydajność łącza dostępowego bez modernizacji istniejącego wyposażenia. Prawidłowe zaimplementowanie tej metody umożliwi również dopasowanie zasobów do trybu pracy aplikacji biznesowych i wymagań użytkowników. Polityka alokacji pasma (ang. bandwidth allocation) w połączeniu z kontrolą ruchu (ang. traffic control) skutecznie chronią aplikacje krytyczne oraz ograniczają wykorzystanie pasma przez te, które nie są tak istotne z punktu widzenia działalności firmy.

3.5. Akceleratory sieciowe

Akceleracja w odniesieniu do ruchu telekomunikacyjnego oznacza skrócenie czasu przesyłania danych pomiędzy dwoma węzłami sieci (np. użytkownikami, biurami lub też wyniesionymi oddziałami firmy). Realizacja tej metody zwiększenia przepustowości łącza odbywa się na drodze instalacji po obu stronach połączenia urządzeń skanujących ruch, które po wykryciu powtarzających się ciągów znaków zastępują je krótszymi, skompresowanymi sekwencjami. Efektem takiego działania jest zredukowanie objętości informacji wysyłanej do sieci, a zatem zwiększenie realnej przepustowości łącza. Do najpopularniejszych sposobów kompresji ruchu sieciowego zaliczyć należy algorytm LZS (ang. Lempel and Ziv Standard), algorytm Predictor oraz kompresję molekularną MSR (ang. Molecular Sequence Reduction) [16].

Działanie algorytmu LZS sprowadza się do wyszukiwania wewnątrz bezstanowego okna powtarzających się elementów, które zostaną następnie zamieniane na krótsze ciągi. Kompresowane są tylko te elementy, które powtarzają się wewnątrz jednego okna.

Kluczowym elementem jest zatem jego wielkość (określana programowo), ponieważ im jest ono szersze tym wyższa będzie skuteczność kompresji, ale z drugiej strony wzrośnie wykorzystywana pojemność bufora oraz czas przetwarzania danych (opóźnienie kompresji).

Wobec ograniczeń tej metody rozsądną alternatywą wydaje się być algorytm Predictor, którego działanie bazuje na przewidywaniu następnych sekwencji znaków. Wykorzystuje on w tym celu dynamiczne tablice wskaźników predykcji, którymi zastępowane są powtarzające się elementy strumienia danych. Każdy poprawnie przewidziany znak poprawia efektywność kompresji, jednak znak błędnie odgadnięty generuje dodatkowe komunikaty sygnalizacyjne, zwiększające ilość przesyłanych informacji. Kolejną wadą tej metody, z powodu której nigdy nie osiąga ona skuteczności algorytmu LZS, jest niski współczynnik kompresji dla strumieni danych, w których powtarzające się elementy są od siebie znacznie oddalone.

Cząstkowa kompresja sekwencji MSR jest najnowszym sposobem redukcji informacji, znacznie wydajniejszym od dwóch powyższych. Stanowi ona inteligentną oraz samouczącą się maszynę transkodującą, umożliwiającą bardzo szybką detekcję i kompresję powtarzających się elementów. Największe zalety tej metody to brak ograniczania odległości między pojawiającymi się symbolami, niewielkie opóźnienie oraz możliwość pracy ze strumieniami o szybkości do 45 Mb/s. Podstawą działania algorytmu MSR jest tablica cząstkowa zawierająca 256 znaków (wszystkie możliwe bajty informacji). Proces translacji między strumieniem danych a tablicą cząstkową ma naturę adaptacyjną, natomiast sam sposób działania algorytmu ulega rozszerzaniu w trakcie jego działania (tzw. samouczenie), dostosowując się do aktualnych sekwencji źródłowych.

Z porównania efektywności trzech przedstawionych tym rozdziale algorytmów jasno wynika, że kompresja molekularna stanowi najlepszy ze znanych sposobów redukcji ilości danych przesyłanych przez sieci szkieletowe. Średni współczynnik kompresji tej metody osiąga wartości rzędu 70-90%, a zatem jej poprawne skonfigurowanie może nawet dziesięciokrotnie zwiększyć realną przepływność sieci transportowej.

3.6. Gotowości transmisji informacji

Odrębnym problemem jest obliczanie gotowości przesyłanych danych (*ang. load availability*). Obliczenie tego parametru jest uzasadnione w sytuacji, gdy część danych jest przesyłana z jakiegoś terminalu różnymi połączeniami, charakteryzującymi się różnymi wartościami gotowości. Dobrym przykładem są sieci oparte na protokole IP. Wskaźnik gotowości przesyłanych danych jest średnią ważoną gotowości poszczególnych połączeń i wyraża się wzorem [1]:

$$K_{gD} = \frac{\sum_i K_{gpi} \times D_{pi}}{\sum_i D_{pi}} \quad (3.1)$$

gdzie: K_{gD} - gotowość danych przesyłanych między dwoma węzłami sieci za pomocą i połączeń,

K_{gpi} - wskaźnik gotowości i -tego połączenia,

D_{pi} - szybkość transmisji danych przez i -te połączenie,

$\sum_i D_{pi}$ - szybkość transmisji danych między węzłami (zakłada się, że połączenie jest jednokierunkowe).

Można zauważyć, że zawsze istnieje związek: $\min\{K_{gpi}\} \leq K_{gD} \leq \max\{K_{gpi}\}$.

Z punktu widzenia teorii niezawodności obliczanie takiego parametru nie jest uzasadnione. Wydaje się jednak analizując wyniki zawarte w [1], że takie podejście może mieć praktyczne zastosowanie, szczególnie przy analizowaniu usług realizowanych za pośrednictwem sieci konwergentnych wykorzystujących protokół IP (*Internetwork Protocol*).

4. KONCEPCJA WDROŻENIA TECHNOLOGII KSZTAŁTOWANIA RUCHU W SPÓŁCE PKP CARGO S.A.

W strukturze spółki wyodrębnić można Centralę z siedzibą w Warszawie oraz 23 podległe placówki: Siedziby zakładów PKP CARGO S.A. znajdują się na terenie całego kraju, co stwarza pozorną przeszkodę przy budowie sieci VPN. Okazuje się jednak, że lokalizacja tych placówek pokrywa się z rozmieszczeniem głównych węzłów sieci szkieletowej zarządzanej przez Telekomunikację Kolejową Sp. z o.o. Dlatego też w przedstawionej koncepcji wykorzystana zostanie infrastruktura tego operatora.

Rozwiązanie to jest o tyle korzystne, że uruchomiona przez Telekomunikację Kolejową w 2003 roku usługa IP VPN jest już stosowana w realizacji niektórych z wymaganych połączeń. Za wyborem tego operatora przemawia ponadto wysoki standard oferowanych przez niego kanałów cyfrowych: zgodność z zaleceniami ITU-T G.821/826, ITU-T M.2100, ITU-T G.703 oraz ITU-T G.957, wysoka niezawodność wynikająca z topologii pierścieniowej, maksymalna stopa błędów BER (ang. Bit Error Rate) poniżej 10⁻¹² i czas dostępności powyżej 99,98%.

PKP CARGO S.A. przechowuje i przetwarza olbrzymią ilość danych rozproszonych w licznych systemach wspomagających jej działalność. Aby możliwe było zarządzanie tym przedsiębiorstwem jako całością, wdrożono w nim zaawansowane narzędzia pozyskiwania informacji o zdarzeniach gospodarczych ze wszystkich swoich oddziałów. Mowa tu o aplikacjach takich jak mySAP Business Suite, czy też systemach bazujących na rozwiązaniach Oracle: handlowej obsługi przesyłek towarowych (zwanym także SWHOPT), planowania pracy pociągów towarowych (o nazwie KS-PLAN), śledzenia wagonów i przesyłek towarowych w obrębie sieci PKP (SLEDZ), a także informowania o sytuacji wagonowej w zakresie wagonów obcych na sieci PKP i wagonów za granicą (INF). Wyszczególnione systemy, zrealizowane w technologiach SAP oraz Oracle, określają zbiór aplikacji krytycznych z punktu widzenia działalności PKP CARGO S.A. - otrzymują one najwyższy priorytet w realizacji kształtowania ruchu.

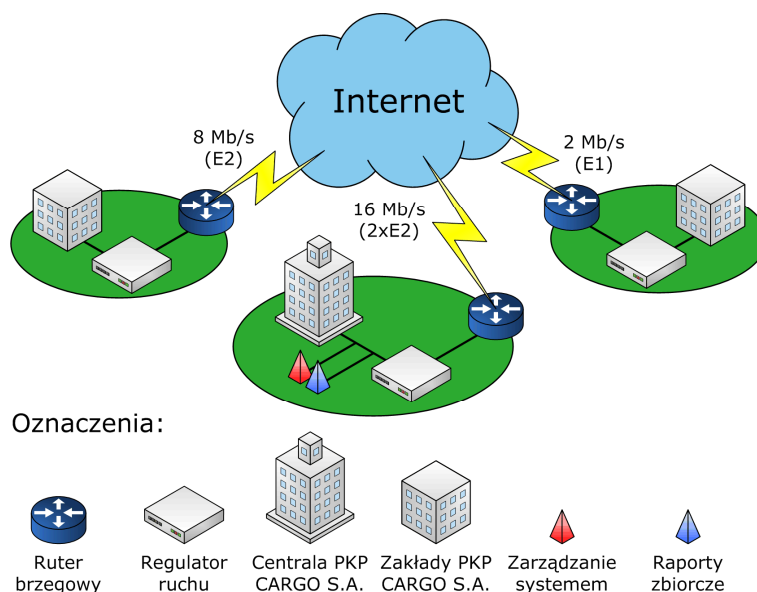
Ruch generowany przez powyższe aplikacje musi być obsługiwany w sposób natychmiastowy, aby utrzymać synchronizację baz danych. Dlatego też pakiety pochodzące od wymienionych systemów będą miały ustawiane pole TOS nagłówka pakietu IP na wartość 176 (binarnie 10110000), co zapewni im wagę ruchu krytycznego (IP Precedence równy 101), a także zminimalizuje opóźnienie przesyłania (kod typu

obsługi wynoszący 1000). Należy dodać, że omawiane przepływy cechuje tzw. wybuchowość, co oznacza, że dane przesyłane są stosunkowo rzadko, ale z dużym natężeniem. Zapewnienie określonego pasma minimalnego każdej z wyszczególnionych aplikacji jest krokiem bezzasadnym – należy raczej skupić się na zagwarantowaniu absolutnie najwyższego priorytetu obsługi.

Drugą bardzo istotną dla działalności PKP CARGO S.A. klasę ruchu stanowią przepływy skojarzone z pocztą elektroniczną i usługą VoIP. Ruch ten wynika z komunikacji pomiędzy pracownikami firmy oraz wymiany informacji z kontrahentami. Klasie tej przydzielić należy wysoki priorytet (wyższy niż domyślny ale niższy niż dla ruchu krytycznego), żeby zapewnić wymaganą sprawność operacyjną przedsiębiorstwa.

W konstruowaniu reguł kształtowania ruchu uwzględnić należy także ruch niepożądany, powiązany z aplikacjami typu P2P, a także grami komputerowymi. Wskazane jest całkowite blokowanie tych transmisji, ponieważ powodują one nagminne przeciążenia sieci. Jak pokazano na rys. 6., koncepcja przewiduje, że każda z 23 placówek posiada własną sieć LAN połączoną z siecią rozległą – Internetem. Bezpośrednio przed ruterem brzegowym każdej z tych sieci wstawiony zostanie regulator ruchu, kształtujący przepływy wchodzące i wychodzące tak, aby zapewnić niezakłóconą pracę aplikacjom krytycznym. Ponadto, w Centrali zainstalowane będą moduły zarządzania całym systemem (PolicyCenter) oraz zbierania raportów z oddalonych placówek (ReportCenter).

W przypadku realizacji sprzętowej najkorzystniejszym rozwiązaniem jest wdrożenie we wszystkich placówkach urządzeń pochodzących od jednego producenta. Niekwestionowanym liderem na rynku regulatorów ruchu jest firma Packeteer i to jej produkty powinny tu być wykorzystane. Z uwagi na odmienne zdolności przepustowe poszczególnych modeli urządzeń PacketShaper, w poszczególnych placówkach instalowane będą różne egzemplarze tej serii. Do urządzenia dokupić należy klucze programowe, pozwalające na kształtowanie ruchu o zadanej prędkości transmisji (Shaping Module) oraz klucze rozszerzające funkcje regulatora o kompresję (Compression Module) i akcelerację (Acceleration Module).



Rys. 6. Ogólna koncepcja wdrożenia technologii kształtowania ruchu w PKP CARGO S.A.

Stosowanie mechanizmu kształtowania ruchu pozwala zredukować zatępienia w sieci i poprawić wydajność łącza dostępowego bez modernizacji istniejącego wyposażenia. Prawidłowe zaimplementowanie tej metody umożliwia również dopasowanie zasobów do trybu pracy aplikacji biznesowych i wymagań użytkowników. Polityka alokacji pasma (ang. bandwidth allocation) w połączeniu z kontrolą ruchu (ang. traffic control) skutecznie chronią aplikacje krytyczne oraz ograniczają wykorzystanie pasma przez te, które nie są tak istotne z punktu widzenia działalności firmy.

5. ZAKOŃCZENIE

Warunki, w jakich funkcjonują systemy telematyki, stwarzają znaczne zagrożenie dla struktury fizycznej sieci i poprawności realizowanych przez nią zadań. Z analizy tych warunków wynika, że problem zapewnienia wymaganej jakości świadczonych usług sieciowych jest trudny i wymaga zrealizowania wielu przedsięwzięć natury organizacyjno-technicznej powodujących wydatkowanie znacznych nakładów. Aby można było problem ten rozwiązać w sposób zadowalający zarówno użytkownika jak i operatora sieci teleinformatycznej, należy go wnikliwie przebadać wykorzystując do tego celu odpowiednie modele analityczne lub/i symulacyjne uwzględniające możliwe do wystąpienia zakłócenia.

Czynniki, które zakłócają proces przekazywania informacji telematycznych, przyjęto ujmować w dwóch grupach. Do pierwszej grupy czynników zaliczamy te, które są uwarunkowane niezawodnością techniczno-eksploatacyjną elementów sieci telematycznych, uchybieniami w zakresie szeroko rozumianej kompatybilności elektromagnetycznej oraz zamierzonymi i przypadkowymi zakłóceniami. Proces funkcjonowania systemów telematyki uwzględniający tę grupę czynników można oceniać przy pomocy charakterystyk procesu ich eksploatacji. Do drugiej grupy czynników zakłócających proces przekazywania informacji zalicza się te, które wynikają z destrukcyjnego oddziaływania zewnętrznego. Czynniki te nie mają stałej powtarzalności w czasie funkcjonowania sieci teleinformatycznej i z reguły prowadzą do nieodwracalnych strat sprzętu teleinformatycznego lub długotrwałych procesów odnowy w specjalizowanych placówkach. Destrukcyjne oddziaływanie czynników zakłócających uwzględnia się przy ocenie przeżywalności sieci teleinformatycznej. Dlatego też zasadnym problemem jest – w tym przypadku – rozważanie związku z efektywnością świadczenia usług zarówno z punktu widzenia użytkownika jak i operatora sieci. Wynika z tego, że dobrym sposobem opisu zachowania się systemów telematyki transportu mogą być modele uwzględniające zachowanie się tych systemów podczas ich eksploatacji uwzględniające zarówno parametry związane z własnościami technicznym elementom sieci pozwalającymi na realizację usług (dostępność usług), jak też związane z możliwością obsługi zadań napływających do realizacji przez systemy i sieci telematyki (wydajność).

Bibliografia

1. Chołda P., Jajszczyk A.: Ocena gotowości w sieciach telekomunikacyjnych. Przegląd Telekomunikacyjny i Wiadomości Telekomunikacyjne. 2003 nr 2-3.

2. Chołda P., Jajszczyk A., K. Wajda: Koncepcja zróżnicowania usług telekomunikacyjnych na podstawie wymagań niezawodnościowych, Krajowe Sympozjum Telekomunikacji '2005, Bydgoszcz 7 – 9 września 2005.
3. Dyrektywa 2001/16/WE Parlamentu Europejskiego i Rady Europy z 19 marca 2001 r. w sprawie interoperacyjności transeuropejskiego systemu kolei konwencjonalnych - Załącznik II - aplikacje telematyczne.
4. Dyrektywa 2007/2/WE Parlamentu Europejskiego i Rady ustanawiająca infrastrukturę informacji przestrzennej we Wspólnocie Europejskiej (INSPIRE).
5. Ferrill P.: Network Traffic Shaping Tools. Processor, nr 4/2006. Wydawnictwo Sandhills, Lincoln 2006.
6. Gago S., Siergiejczyk M.: Konwergencja usług w sieciach IP dla spółek kolejowych. Materiały IV Konferencji " Telekomunikacja i Informatyka na Kolei". Polska Izba Producentów Urządzeń i Usług na Rzecz Kolei (płyta CD), Szczyrk, styczeń 2006.
7. ITU-T Recommendations E.800 . Terms and Definitions Related to Quality of Services and Network Performance Including Dependability. 1994.
8. Karczewski R.: *Integracja architektury IntServ ze środowiskiem protokołu MPLS*. V Ogólnopolska Konferencja „Internet – Wrocław 2003”, Wrocław, 4-5 grudnia 2003.
9. Papir Z.: *Ruch telekomunikacyjny i przeciążenia sieci pakietowych*. Wydawnictwa Komunikacji i Łączności, Warszawa 2001.
10. Pawełczak P., Rogowski T.: *Niedoskonałość IP szansą MPLS (cz. 1)*. NetWorld nr 11/2003. Wydawnictwo IDG, Warszawa 2003.
11. Pogrzebski, H.: Funkcje specyfikacji TAF-TSI - tryb planowania i przygotowania pociągu. TTS Technika Transportu Szynowego 2005, R. 11, nr 11, s. 66--70.
12. Polska Norma PN-93/N-50191: Słownik terminologiczny elektryki. Niezawodność; jakość usługi, Polski Komitet Normalizacji, Miar i Jakości. Warszawa 1994.
13. Siergiejczyk M., Gago S.: ADVANCES IN TRANSPORT SYSTEMS TELEMATICS. Integrated Information System in Railway Transport. Wydawnictwo Komunikacji i Łączności. Warszawa 2008. Str. 279-286.
14. Siergiejczyk M.: Efektywność eksploatacyjna systemów telematyki transportu. Prace Naukowe Politechniki Warszawskiej, seria Transport, Nr 67, Warszawa 2009.
15. Wawrzyński W., Siergiejczyk M. i inni : Sprawozdanie końcowe grant KBN 5T12C 066 25. Metody wykorzystania środków telematyki we wspomaganiu realizacji zadań transportowych. Kierownik: prof. dr hab. inż. W. Wawrzyński. Warszawa 2007.
16. Urbanek A.: Kompresja informacji w sieciach. NetWorld nr 2/2004. Wydawnictwo IDG, Warszawa 2004.
17. Zamojski W. (red.): Niezawodność i eksploatacja systemów. Wyd. Politechniki Wrocławskiej. Wrocław 1981.

ISSUES OF THE EVALUATION OF QUALITY OF TELEMATICS SERVICE IN THE RAIL TRANSPORT

Abstract: The article presents an analysis has been made of conditions in which telematics system operates, with a special consideration of numerous threats for physical structure of the telematics systems and correctness of services being provided through them. Quality indices for data transmission have been proposed (using methods and ways for quality assessment in telecommunication networks) in the transport systems telematics. Than characterize basic mechanisms of telecommunication traffic shaping for guarantee minimal capacity from the point of view of activity of enterprise critical applications has been developed. It present method of telecommunication traffic shaping. This method has been of service for elaboration of concept of adjustment of telecommunication traffic shaping, concept has been processed coherent with activity of biggest polish rail carrier.

Keywords: transport telematics, services, quality, models, traffic shaping,, exploitation process