

Prof. nzw. dr hab. inż. Mirosław Siergiejczyk,
Dr inż. Stanisław Gago
Politechnika Warszawska

JAKOŚĆ USŁUG TELEINFORMATYCZNYCH W TRANSPORCIE KOLEJOWYM

SPIS TREŚCI

1. Wstęp
2. Charakterystyka usług teleinformatycznych kolei
3. Jakość usług w sieciowych kolejowych sieciach teleinformatycznych
4. Jakość usług w sieciach TCP/IP
5. Zakończenie

STRESZCZENIE

W artykule przeanalizowano warunki, w jakich funkcjonują systemy teleinformatyki ze szczególnym uwzględnieniem licznych zagrożeń dla struktury fizycznej systemów teleinformatyki i poprawności realizowanych usług. Zaproponowano modele zapewnienia jakości dla transmisji informacji w systemach teleinformatyki transportu. Scharakteryzowano podstawowe mechanizmy kształtowania ruchu w celu zagwarantowania minimalnej przepustowości aplikacjom krytycznym z punktu widzenia działalności przedsiębiorstwa. Przedstawiono metodę kształtowania ruchu.

1. WSTĘP

Struktura oraz sposób funkcjonowania przedsiębiorstwa kolejowego powodują występowanie specyficznych potrzeb łącznościowych. Potrzeby te są zaspokajane przez systemy łączności telefonicznej, instalowane w sposób dostosowany do lokalizacji różnych obiektów kolejowych, tworzące tak zwane sieci wydzielone. Występowanie sieci wydzielonych w infrastrukturze łączności przedsiębiorstwa PKP wynika głównie ze stałego dążenia do zwiększania stopnia bezpieczeństwa ruchu pociągów oraz dbałości o operatywność zarządzania przedsiębiorstwem.

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 jest wymagana 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 jest wykorzystywana 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, zdalną rezerwację biletów itd. Usługi te wspierają konkurencyjność przewoźników i firm kolejowych w walce o klienta. Wymienione 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 dużą ilość danych używanych przez te aplikacje, jedynym rozwiązaniem na potrzeby przesyłania danych są sieci pracujące zgodnie ze stosem protokołów TCP/IP.

2. CHARAKTERYSTYKA USŁUG TELEINFORMATYCZNYCH KOLEI

Dzięki rozległej infrastrukturze i stosowaniu nowoczesnych technologii, sieć teleinformatyczna kolejowa oferuje szeroki wachlarz usług, które można podzielić na następujące grupy:

- transmisja danych – od sieci lokalnych LAN do usług IP opartych na VPN/MPLS (*Virtual Private Network / Multiprotocol Label Switching*),
- dzierżawa łączy – analogowych, kanałów cyfrowych i optycznych,
- usługi szerokopasmowego dostępu do Internetu,
- telefonia stacjonarna – usługi telekomunikacyjne obejmujące wykorzystanie zarówno technologii tradycyjnych (telefonia analogowa i cyfrowa – ISDN), jak i przesyłania głosu przez sieci internetowe – VoIP (*Vois over IP*).

Wymienione grupy usług są oferowane pełnemu spektrum odbiorców – poczynając od operatorów telekomunikacyjnych, przez odbiorców biznesowych, na klientach indywidualnych kończąc.

Dla klientów klasy operatorskiej kolej oferuje usługę hurtowego zakańczania ruchu w sieci oraz tranzytu ruchu głosowego do sieci innych operatorów. Umożliwia to posiadanie wielu punktów styku z innymi operatorami (stacjonarnymi oraz komórkowymi). Dla klientów-operatorów realizuje również szerokopasmowy dostęp do Internetu. Jest on dostarczany na zasadzie symetrycznej „rury” bitowej o zakontraktowanej przepływności. Kwestie związane z zarządzaniem ruchem znajdują się w gestii klienta. Jest to rozwiązanie wykorzystywane głównie przez ISP (*Internet Service Provider*) oraz operatorów telewizji kablowych CATV (*Cable TV*).

Ze względu na rozległą, ogólnopolską sieć transmisyjną, kolej często jest dostawcą łączy. Jest to usługa wykorzystywana zarówno przez innych operatorów, jak i klientów biznesowych. Dzierżawa łączy jest realizowana w dwóch wersjach:

- dzierżawa kanałów cyfrowych – udostępnienie przezroczystego bitowo łącza cyfrowego i zapewnienie na nim transmisji o stałej szybkości w zakresie od E1 (2 Mbit/s), przez E3 (34 Mbit/s), STM-1 (155 Mbit/s), STM-4 (622 Mbit/s) do STM-16 (2,5 Gbit/s),
- dzierżawa kanałów optycznych – udostępnienie pojedynczego kanału optycznego (tzw. Lambda) w systemie DWDM.

Oprócz terminacji ruchu dla operatorów, kolej świadczy usługi głosowe dla klientów korporacyjnych oraz indywidualnych. Klienci biznesowi mogą zostać podłączeni do sieci głosowej za pomocą łączy cyfrowych w technologii ISDN 30B+D. Umożliwia to prowadzenie 30 niezależnych rozmów telefonicznych jednocześnie. Z wykorzystaniem centrali PABX oraz wspomnianej usługi jest możliwe realizowanie bezpośredniej i darmowej komunikacji wewnątrz firmy. Możliwy jest również dostęp cyfrowy ISDN 2B+D, dzięki czemu klient może prowadzić dwie niezależne rozmowy. W rejonach, gdzie realizacja dostępu cyfrowego do usług głosowych jest niemożliwa, TKTelekom oferuje również klasyczny dostęp analogowy. Ze względu na rozbudowaną infrastrukturę teleinformatyczną, część usług głosowych jest realizowana przez kolej z wykorzystaniem technologii VoIP. Pozawala ona na prowadzenie rozmów telefonicznych z wykorzystaniem łączy internetowych. Usługa ta nazywa się ToIP (*Telephone over Internet Protocol*).

Oprócz usług głosowych, kolej jest w stanie zapewnić klientom biznesowym również dostęp do Internetu. Jeden ze sposobów realizacji, oferowany w szerokim zakresie przepustowości, umożliwia dostęp do krajowych i zagranicznych zasobów sieci. W usłudze

też szybkość transmisji danych może stanowić dowolną wartość od 1 Mb/s. Prędkość w obydwu kierunkach jest symetryczna, a dostarczane łącze dostępowe jest zakończone stykiem Ethernet. Kolej umożliwia również dostęp do Internetu z wykorzystaniem technologii ADSL. Usługa jest realizowana z wykorzystaniem kabli miedzianych, służących do realizacji tradycyjnej usługi głosowej. Dostarczane jest asymetryczne łącze, którego przepływność może być w zakresie od 128 kbit/s do 6 Mbit/s. Dla abonentów korzystających z telefonii stacjonarnej, jest również możliwy dostęp *dial-up* do sieci Internet.

Najbardziej rozwiniętą ofertę dla biznesu kolej oferuje w zakresie transmisji danych. Duża część usług z tej grupy jest oparta na technologii MPLS (*Multi Protocol Label Switching*). W tej technologii *routing* pakietów został zastąpiony przez tzw. przełączanie etykiet. Uogólnieniem techniki przełączania etykiet na inne technologie zwielokrotnienia jest GMPLS (*Generalized MPLS*). Zastosowanie MPLS umożliwia klientom biznesowym powiązanie lokalnych sieci LAN (*Local Area Network*) w pełni zarządzaną sieć WAN (*Wide Area Network*).

Sieć MPLS daje możliwość stworzenia teleinformatycznej sieci VPN (*Virtual Private Network*). VPN jest wirtualną siecią, przez którą płynie ruch w ramach sieci prywatnej pomiędzy klientami końcowymi za pośrednictwem publicznej sieci, takiej jak Internet. Określenie „wirtualna” oznacza, że sieć istnieje tylko jako struktura logiczna, działająca w sieci publicznej. Sieć VPN wiąże centralę firmy ze wszystkimi oddziałami i pracownikami, zarówno w kraju, jak i za granicą. Sieć MPLS umożliwi dodatkowo realizowanie usług opartych na takich mechanizmach, jak:

- TE (*Traffic Engineering Tunnels*),
- AToM (*Any Transport over MPLS*),
- VPLS (*Virtual Private LAN Services*).

Pierwszy mechanizm umożliwia zestawianie ścieżek MPLS o zadanych parametrach jakościowych. Daje to możliwość zarządzania poziomem QoS. AToM pozwala na przesyłanie siecią MPLS technologii innych niż tradycyjne IP. Dzięki VPLS możliwe jest emulowanie sieci LAN dla rozległych organizacji. Dodatkowo sieć MPLS zapewnia mechanizmy bezpieczeństwa – tunele IP Sec oraz zabezpieczenia danych z zastosowaniem szyfrowania wewnątrz sieci. Kolej realizuje również transmisję danych z wykorzystaniem technologii Ethernet, ATM (*Asynchronous Transfer Mode*) oraz FR (*Frame Relay*).

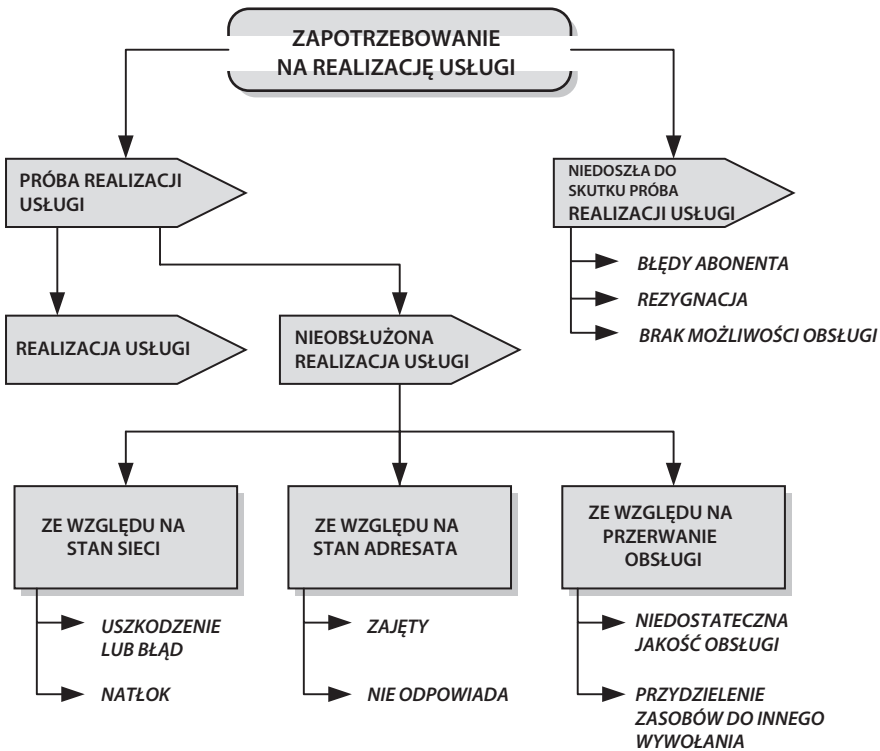
W sieci kolejowej jest realizowana również szeroka gama usług dodanych. Są one świadczone z wykorzystaniem sieci IN (*Intelligent Network*) firmy Alcatel. Zarówno dla klientów indywidualnych, jak i korporacyjnych, jest możliwe między innymi tworzenie wydzielonych grup abonentów, przenoszenie wywołań, transferowanie połączeń, wykonywanie połączeń konferencyjnych oraz prezentacja numeru wywołującego. Na platformie inteligentnej są również realizowane usługi 0800 oraz 0801, czyli połączenia bezpłatne oraz połączenia o obniżonej płatności. Są to rozwiązania wykorzystywane przez klientów biznesowych, którzy mają własne infolinie.

3. JAKOŚĆ USŁUG W SIECIOWYCH KOLEJOWYCH SIECIACH TELEINFORMATYCZNYCH

3.1. Zależności w sieci realizującej usługi teleinformatyczne

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 teleinformatycznych w transporcie wymaga określenia poziomu ich jakości QoTS (ang. *Quality of Transport Telecommunications 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]



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

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 *end-to-end*, tj. od dostawcy do odbiorcy usługi, a także pracy personelu zatrudnionego przy świadczeniu usług.

Potrzeby ze strony użytkownika sieci realizującej usługi teleinformatyczne można przedstawić w postaci strumienia zapotrzebowań na usługi teleinformatyczne realizowane przez określoną sieć. Ze względu na uwarunkowania wewnętrzne i zewnętrzne sieci, strumień zapotrzebowań obsługiwanych 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 teleinformatycznych i ich wpływ na nieobsłużenie zgłoszonego przez użytkownika zapotrzebowania, przedstawiono na rysunku 1 [14].

Rozpoczęcie wykonywania 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.

3.2. Obszary oceny jakości usług teleinformatycznych

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 ogólnościowych (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 takie cechy składowe, 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 następujące wartości 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, wcześniej 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, powodzie, 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 destruktorów, które mogą zagrozić strukturze fizycznej i dostępności do usług sieciowych.

Wymienione obszary gwarantowanego poziomu zrealizowania usługi, implikują wydajność systemu uwzględniającą następujące 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: 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 adresem 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 wymienionych 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).

4. JAKOŚĆ USŁUG W SIECIACH TCP/IP

4.1. Pojęcie jakości usług w sieciach pakietowych

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 kolejowania 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, jak np.:

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

W sieciach TCP/IP występuje kilka modeli zapewnienia jakości usług QoS. Można do nich 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. przez sygnalizację i przez odpowiednie oznaczanie pakietów.

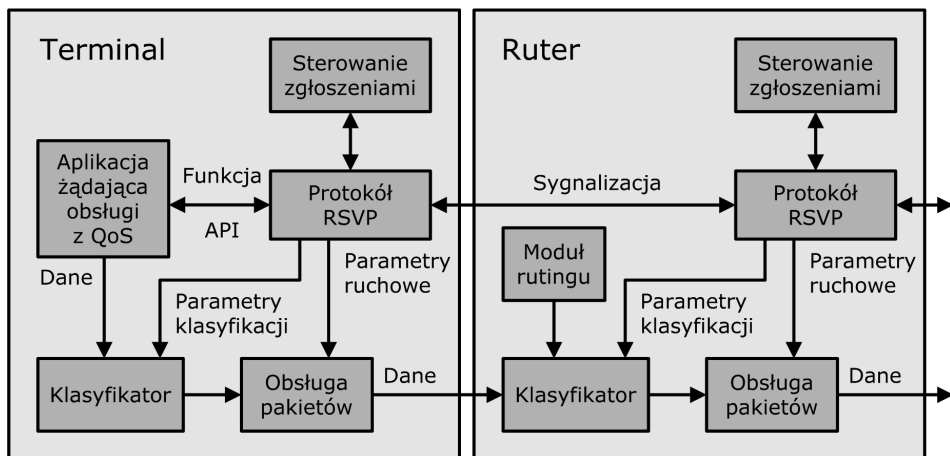
4.2. Model rezerwowania zasobów

Pierwszym podejściem stowarzyszenia IETF (ang. *Internet Engineering Task Force*) do problemu zapewnienia usługom sieci IP określonych parametrów QoS, była propo-

zycja 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 należy rozumieć 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 te wymagania nie mogą być spełnione z powodu braku wolnych zasobów, to żądanie realizacji usługi zostanie odrzucone.

Architektura *IntServ* składa się z trzech wzajemnie powiązanych elementów (rys. 2) [9]:

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



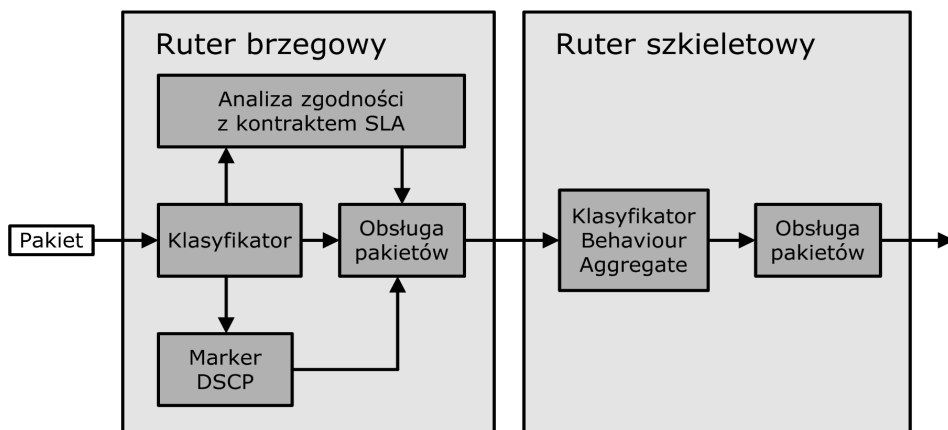
Rys. 2. Wzajemne zależności pomiędzy elementami funkcjonalnymi w architekturze *IntServ* [9]

Blok protokołu RSVP (ang. *Resource Reservation Protocol*) odpowiada za proces wymiany informacji sygnalizacyjnej wzdłuż ścieżki, po której będzie przesyłana 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 żądaną przez aplikację jakość usługi tylko wtedy, 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 dupleksowym wymaga zainicjowania symetrycznej rezerwacji z drugiego końca połączenia.

4.3. 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 sposobu zapewnienia jakości transmisji. Platforma *IntServ* zakłada podejście sesyjne, a zatem pełną identyfikację strumienia pakietów oraz zapewnienie im właściwego poziomu QoS przez zarezerwowanie zasobów sieciowych (czyli pasma transmisyjnego, pamięci buforowej i mocy obliczeniowej ruterów) [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 nagłówków, 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 przedstawiono na rysunku 3.



Rys. 3. 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, a także sposób rutowania są w obu tych architekturach identyczne.

4.4. Wieloprotokołowa komutacja etykietowa

Dwie przedstawione architektury, a mianowicie *Integrated Services* oraz *Differentiated Services*, nie rozwiązują problemu niskiej efektywności reguł trasowania w sieciach IP. W odpowiedzi na ten problem IETF zaproponowało kolejny mechanizm sterowania jakością usług – technikę wieloprotokołowej komutacji etykietowej MPLS (ang. *Multi-Protocol Label Switching*). Zgodnie z zasadą działania tego protokołu, każdy pakiet wchodzący w obszar domeny MPLS jest klasyfikowany 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 przez kierowanie ruchu na wybrane łącza, to jednak nie ma 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 ta metoda musi być stosowana w połączeniu z innymi mechanizmami QoS w sieciach IP, takimi jak omówione już modele *IntServ* oraz *DiffServ*.

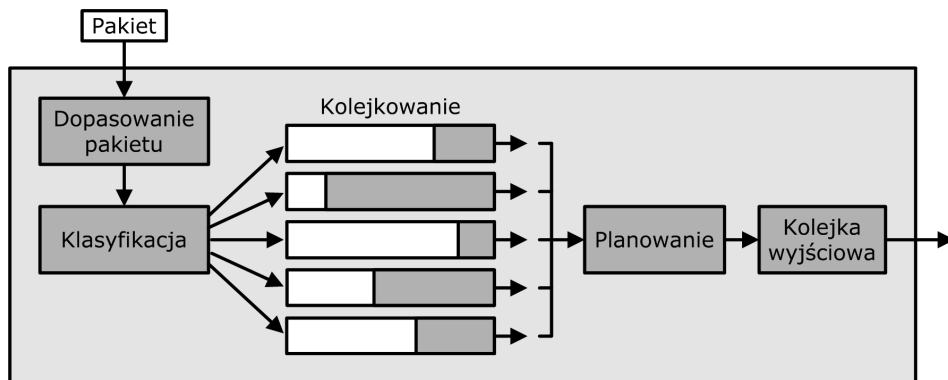
4.5. Metoda kształtowanie ruchu

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 pozostała niezakłócona.

Proces kształtowania ruchu przebiega w kilku zilustrowanych na rysunku 4 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 jest realizowany przez przydzielanie pakietów do odpowiednich kategorii według różnego typu kryteriów, 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. 4. 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 fazą kształtowania ruchu jest planowanie porządku wybierania danych z kolejek w taki sposób, by sprawiedliwie rozdysponować poszczególnym klasom dostępną przepustowość. Etap ten może przebiegać z wykorzystaniem kolejek klasowych oraz bezklasowych.

Kolejki bezklasowe cechuje prosta konfiguracja, co z jednej strony pociąga za sobą ich łatwą implementację, a z drugiej ograniczone możliwości priorytetowania strumieni ruchu. Do tych kolejek należy zaliczyć 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*).

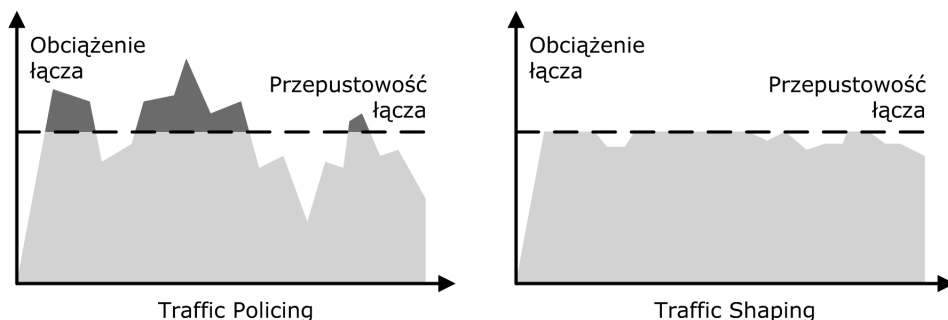
Kolejki klasowe CBQ (ang. *Class Based Queuing*) są dużo bardziej złożone, głównie ze względu na możliwość tworzenia hierarchii klas o różnorodnych parametrach, wśród których najważniejszym pozostaje zwykle przepustowość. Cecha ta eliminuje niedogodność niskiej konfigurowalności kolejek bezklasowych, 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 jest stosowana 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, przez ułożenie na łączu dostępowym urządzenia o nazwie „*traffic shaper*”. Sprzęt sieciowy obsługujący tę funkcję można podzielić 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 zostało zobrazowane na rysunku 5. 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. 5. 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.

4.6. 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 polega na przewidywaniu następnych sekwencji znaków. Wykorzystuje on w tym celu dynamiczne tablice wskaźników predykcji, którymi są zastępowane 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 wymienionych. 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 w tym rozdziale algorytmów 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.

4.7. Gotowości transmisji informacji

Odrębnym problemem jest obliczanie gotowości przenoszonych 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 przenoszonych danych jest średnią ważoną gotowości poszczególnych połączeń i wyraża się wzorem [1]:

$$K_{gD} = \frac{\sum_i K_{gpi} \times D_{pi}}{\sum_i D_{pi}}$$

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,

$\sum D_{pi}$ – 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. Analizując wyniki zawarte w [1] wydaje się jednak, że takie podejście może mieć praktyczne zastosowanie, szczególnie przy analizowaniu usług realizowanych za pośrednictwem sieci konwergencych wykorzystujących protokół IP (*Internetwork Protocol*).

5. ZAKOŃCZENIE

Warunki, w jakich funkcjonują systemy teleinformatyki, 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 ten problem rozwiązać w sposób zadawalający 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, przyjęto ujmować w dwóch grupach. Do pierwszej grupy czynników zaliczamy te, które są uwarunkowane niezawodnością techniczno-eksploatacyjną elementów sieci teleinformatycznych, uchybieniami w zakresie szeroko rozumianej kompatybilności elektromagnetycznej oraz zamierzonymi i przypadkowymi zakłóceniami. Proces funkcjonowania systemów teleinformatyki, uwzględniający tę grupę czynników, można oceniać za pomocą 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 teleinformatyki transportu mogą być modele symulujące zachowanie się tych systemów podczas ich eksploatacji, uwzględniające zarówno parametry związane z własnościami technicznymi elementów sieci pozwalającymi na realizację usług (dostępność usług), jak też związane z możliwością obsługi zadań napływających do realizacji przez systemy i sieci teleinformatyki (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., Wajda K.: *Konceptcja zróżnicowania usług telekomunikacyjnych na podstawie wymagań niezawodnościowych*. Krajowe Sympozjum Telekomunikacji, 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”, Wydawnictwo Sandhills, Lincoln 2006, nr 4.
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”, 2003, nr 11. Warszawa, Wydawnictwo IDG, 2003.
11. Pogrzebski H.: *Funkcje specyfikacji TAF-TSI – tryb planowania i przygotowania pociągu* „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*. Warszawa, Wydawnictwo Komunikacji i Łączności, 2008, s. 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 in: *Metody wykorzystania środków telematyki we wspomaganiu realizacji zadań transportowych*. Sprawozdanie końcowe grant KBN 5T12C 066 25. Kierownik: prof. dr hab. inż. W. Wawrzyński. Warszawa 2007.
16. Urbanek A.: *Kompresja informacji w sieciach*. „NetWorld”, Warszawa, Wydawnictwo IDG, 2004.
17. Zamojski W. (red.): *Niezawodność i eksploatacja systemów*. Wrocław, Wydawnictwo Politechniki Wrocławskiej, 1981.