STUDIA INFORMATICA 2001
Volume 22 Number 1 (43)

Ryszard WINIARCZYK
Instytut Informatyki Teoretycznej i Stosowanej PAN

MODELE I ARCHITEKTURY QoS W SYSTEMACH MULTIMEDIALNYCH

Streszczenie. W artykule przedstawione zostały podstawowe pojęcia dotyczące jakości usług systemów multimedialnych. Omówiono mechanizmy umożliwiające specyfikowanie oraz realizację zadanego poziomu usług. Przedstawiono przykłady architektur multimedialnych środowisk operacyjnych.

QoS MODELS AND ARCHITECRURES OF MULTIMEDIA SYSTEMS

Summary. The article presents QoS terminology in the context of multimedia systems. Mechanisms for specification and maintenance of QoS is described. Same examples of multimedia middlewares that have emerged in the literature is presented.

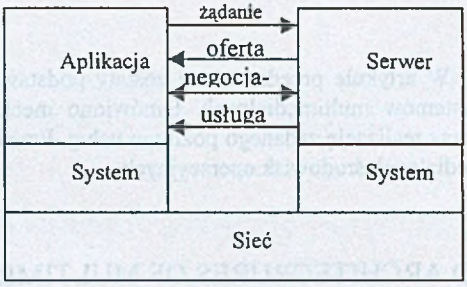
1. Wprowadzenie

Pojawienie się nowych technologii sieciowych, a zwłaszcza sieci ATM, umożliwiło transmisję informacji multimedialnej gwarantując jednocześnie wymagany poziom jakości transmisji (np. przepustowość, opóźnienie, stopa błędów) [1]. Sieci tego typu wyposażone są w odpowiednie mechanizmy zarządzania przeznaczone do przydzielania na bieżąco zasobów sieci tak, by zapewnić ustalony poziom jakości transmisji przez cały okres jej trwania. Sieć jako usługodawca analizuje „zamówienie" klienta i w przypadku pozytywnej oceny zawiera „kontrakt" na zrealizowanie zamówionej usługi. W przeciwnym przypadku usługa może być zrealizowana za zgodą klienta tylko w zakresie możliwości sieci.

Opracowane mechanizmy zarządzania przepływem informacji z gwarancją jakości na poziomie sieci [9] nie zapewniają bezpośrednio wymaganej jakości usług na poziomie aplikacji multimedialnych wykonywanych w stacjach końcowych wyposażonych w klasyczne modele

systemów operacyjnych ogólnego przeznaczenia czy nawet w systemy czasu rzeczywistego [1,6,8]. Brak odpowiednich mechanizmów zarządzania pośredniczących pomiędzy aplikacją a środowiskiem zarządzającym siecią może powodować rozbieżność pomiędzy wymaganym poziomem usług aplikacji a jakością transmisji gwarantowaną przez sieć.

Z tego względu na początku, a zwłaszcza w połowie obecnej dekady w kilku ośrodkach uniwersyteckich podjęto badania nad opracowaniem takich mechanizmów, które na podstawie specyfikacji jakości usług aplikacji dokonywałyby rezerwacji zasobów stacji końcowych i sieci oraz nadzorowały ich wykorzystanie przez cały czas pracy aplikacji. Elementy (mechanizmy) pośredniczące w nadzorowaniu jakości usług pomiędzy współpracującymi aplikacjami tworzą tzw. architektury QoS.



Rys. 1. Model klient - serwer w architekturze QoS
Fig. 1. Client- server models in QoS architecture

2. Informacja multimedialna

W systemach komputerowych akwizycja i udostępnianie informacji realizowane są poprzez różne nośniki (media), jak: tekst, grafika, dźwięk, itp. Struktury danych tworzą odpowiednio pojedyncze lub grupowe obiekty tekstowe, graficzne czy dźwiękowe. Pomiędzy obiektami może występować relacja czasowa porządkująca ich akwizycję lub prezentację. Grupa takich obiektów tworzy strumień. Zależne od czasu strumienie mogą być statyczne lub dynamiczne. Jeżeli czasy ekspozycji obiektów są równe, to obiekty te tworzą strumienie ciągłe, jak obrazy telewizyjne w standardzie PAL czy NTSC. Obiektami statycznymi są natomiast tekst i grafika.

Grupa zintegrowanych strumieni, w której występuje co najmniej jeden strumień dynamiczny, tworzy strumień multimedialny.

Szczególną cechą mediów ciągłych w odniesieniu do statycznych jest to, że obok dużych wymagań na częstotliwość akwizycji i odtwarzania informacji dopuszcza się możliwość utraty części informacji.

Główne wymagania dotyczące strumieni multimedialnych można opisać poprzez następujące parametry: przepustowość, opóźnienie transmisji wahania opóźnienia (jitter), niezawodność transmisji (dopuszczalna stopa błędów), zachowanie relacji pomiędzy częstotliwością akwizycji i częstotliwością odtwarzania informacji.

Z uwagi na integrację strumieni różnych typów listę parametrów charakteryzujących strumienie multimedialne należy poszerzyć o warunki synchronizacji takie, jak:

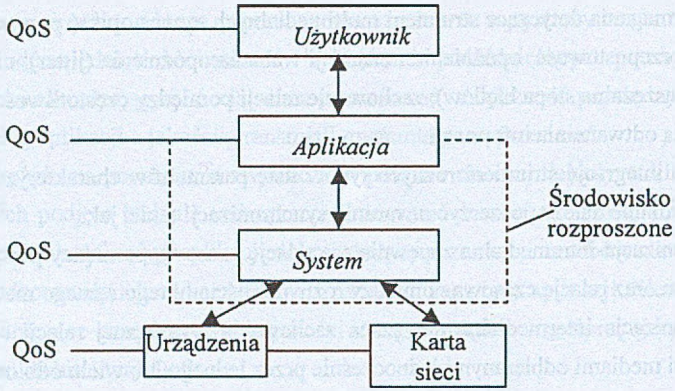
- synchronizacja intramedialna zapewniająca relację czasową pomiędzy porcjami danych jednego medium oraz relację czasową pomiędzy różnymi ujściami tego samego medium;
- synchronizacja intermedialna dotycząca zachowania wymaganej relacji czasowej pomiędzy różnymi mediami odbieranymi jednocześnie przez jednego lub wielu odbiorców;
- synchronizacja zdarzeń zapewniająca odpowiednie współwystępowanie w czasie zdarzeń generowanych przez system z informacją multimedialną..

Systemy informatyczne zdolne do przetwarzania strumieni multimedialnych nazywamy systemami multimedialnymi. Ze względu na subiektywność odbioru przez użytkownika a także z uwagi na techniczne aspekty przetwarzania informacja multimedialna może być prezentowana na różnych poziomach jakości.

Aplikacje bądź użytkownik żądając pewnego poziomu jakości usług systemu multimedialnego muszą określić swoje wymagania w formie listy parametrów stanowiących o jakości usługi. Spełnienie zadanego poziomu wymagań (wartości parametrów) jest możliwe poprzez rezerwację i udostępnienie odpowiednich zasobów przez środowisko operacyjne. Środowiska operacyjne wyposażone w mechanizmy specyfikowania jakości usług, rezerwacji i udostępniania zasobów nazywane są architekturami QoS.

2.1. Specyfikacja QoS

Problem specyfikacji jakości usług jest zagadnieniem kluczowym w rozwoju multimedialnych środowisk operacyjnych. Dotychczas nie opracowano jednolitego sposobu definiowania jakości usług aplikacji multimedialnych, tak jak to ma miejsce w przypadku usług sieci ATM. Jedna z ogólnych definicji podaje [3], że „QoS reprezentuje zbiór ilościowych i jakościowych charakterystyk multimedialnego systemu rozproszonego niezbędnych do zapewnienia funkcjonalności aplikacji". Nietrudno zauważyć, że z tej definicji wynika potrzeba specyfikacji jakości usług QoS na wszystkich poziomach systemu multimedialnego, a zwłaszcza na poziomie użytkownika, aplikacji, systemu operacyjnego i sieci.



Rys. 2. Poziomy specyfikacji QoS
Fig. 2. QoS specification levels

Na poziomie użytkownika jakość usług oceniana jest najczęściej subiektywnie, a uzyskiwana poprzez zadawanie parametrów aplikacji takich, jak: rozmiar, kontrast, jasność czy częstotliwość odtwarzania obrazów. Zapewnienie odpowiednich wartości tych parametrów wymaga od systemu operacyjnego udostępnienia procesora oraz pamięci, a od sieci zapewnienia przepustowości transmisji tak, by aplikacja mogła spełniać oczekiwania użytkownika. Specyfikacja jakości usług dotyczy zarówno strumieni multimedialnych, jak i komunikacji pomiędzy mechanizmami sterującymi przepływem strumieni.

Ze względu na różnorodność zastosowań multimedialnych, pomimo wielu podejmowanych prób [13,14], nie opracowano jeszcze ogólnego języka specyfikacji jakości usług.

Dotyczy to zarówno specyfikacji na poziomie użytkownika, jak i kolejnych poziomów występujących w środowiskach multimedialnych. Wymienione wyżej architektury QoS posiadają odmienne specyfikacje kategorii i charakterystyk usług.

Można jednak określić ogólne cechy umożliwiające klasyfikację aplikacji multimedialnych. W przypadku mediów ciągłych są to [1]:

- izochroniczność: obejmuje aplikacje, które generują lub odbierają dane ze stałą częstotliwością. Strumień informacji multimedialnej można scharakteryzować poprzez częstotliwość ramek obrazów, maksymalny rozmiar ramek, średni rozmiar ramki oraz dodatkowo maksymalne opóźnienie pomiędzy czasem wysłania ramki przez nadawcę a wyświetleniem jej przez odbiorcę;
- sporadyczna intensywność (bursty): dotyczy tych aplikacji, w których występuje potrzeba transmisji dużej serii danych. W przeciwieństwie do aplikacji izochronicznych w tym przypadku nie występuje okresowość transmisji, a dane są wysyłane i odbierane w blokach o zmiennej długości. Specyfikacja QoS zawiera maksymalny rozmiar przesyłanych bloków oraz minimalną przepustowość lub maksymalne opóźnienie dla każdego bloku.

Dodatkowo można wyróżnić klasę aplikacji wymagających małych opóźnień, a w szczególności np. krótkich czasów odpowiedzi RPC lub informacji sterującej ACK. Podobnie jak w przypadku aplikacji typu „bursty" jakość usług może być określona przez maksymalną długość przesyłanych komunikatów i akceptowalne opóźnienie. Takie wymagania stawiane są między innymi przez obliczenia symulacyjne w systemach rozproszonych.

Specyfikacja QoS definiuje ilościowe i jakościowe kryteria oceny i realizowanych usług. Kryteria ilościowe obejmują parametry mierzalne, jak: długość pakietu, prędkość transmisji, czas przetwarzania zadania, itp. Jakościowe kryteria wyrażają oczekiwania odnośnie do realizowanych usług lub występowania pewnych ich cech, jak uporządkowanie transmitowanych danych, mechanizm szeregowania zadań czy synchronizacja intermedialna [8].

2.2. Mapowanie QoS

Mechanizmy mapowania (odwzorowania) QoS realizują funkcje automatycznej translacji specyfikacjami QoS pomiędzy różnymi poziomami systemu multimedialnego (np. system operacyjny, warstwa transportu, sieci). Odwzorowania parametrów realizowane są według różnych schematów: jeden w wiele, wiele w jeden oraz wiele w wiele. Niezależnie jednak od złożoności odwzorowania proces translacji powinien być określony prostymi operacjami arytmetycznymi, co zapewnia efektywność mechanizmów mapowania.

Ponadto istotną cechą mechanizmów translacji jest dwukierunkowość mapowania parametrów pomiędzy warstwami systemu. Spełnienie tego warunku wymaga jednak dodatkowej informacji ze względu na możliwość pojawienia się niejednoznaczności odwzorowania. Na przykład, gdy na podstawie specyfikacji aplikacji obejmującej częstotliwość i wymiar obrazów zostaje określona wymagana przepustowość sieci, to zachowanie jakości usługi przy zmianie przepustowości można uzyskać zmieniając częstotliwość obrazów lub zmieniając rozmiar obrazów. W tego typu przypadkach specyfikacja zawiera, oprócz wartości parametrów, reguły postępowania w sytuacjach niejednoznacznej translacji. Mapowanie parametrów może być ustalane statycznie na podstawie stałych (średnich wartości, "najgorszego przypadku") parametrów na poszczególnych warstwach systemu realizującego usługę opierając się na analizie opóźnień w całym torze komunikacji pomiędzy aplikacjami [1] lub dynamicznie poprzez tzw. sieci aktywne [9]. Mechanizmy mapowania zapewniają struktury danych oraz zarządzanie tymi danymi dla potrzeb procesów ustalających „zobowiązania" poszczególnych warstw systemu do realizacji jakości usług na poziomie aplikacji lub użytkownika systemu.

2.3. Zarządzanie QoS

Problemy i mechanizmy zarządzania jakością usług, a dokładniej zasobami w celu zapewnienia odpowiedniego poziomu jakości, można podzielić na trzy podstawowe grupy: poziom użytkownika-aplikacji, poziom systemu operacyjnego-platformy systemowej, poziom sieci.

Na poziomie użytkownika-aplikacji zarządzanie realizowane jest poprzez definiowanie scenariuszy aplikacji multimedialnej. Scenariusz zawiera wymagania odnośnie do zagospodarowania przestrzeni konsoli multimedialnej oraz relacji czasowych pomiędzy mediami[10]. Osprzęt multimedialny umożliwia strojenie odbioru, a funkcje monitorowania wyznaczają na bieżąco wymagania jakościowe dla niższych warstw systemu.

Na poziomie platformy systemowej mechanizmy zarządzania obejmują przede wszystkim szeregowanie procesów (wątków), zarządzanie pamięcią dla potrzeb buforowania danych wysyłanych i otrzymywanych. Ponieważ nie opracowano dedykowanych platform operacyjnych do zarządzania procesami i pamięcią wykorzystuje się mechanizmy zarządzania systemów czasu rzeczywistego-zwłaszcza do odniesienia aplikacji izochronicznych.

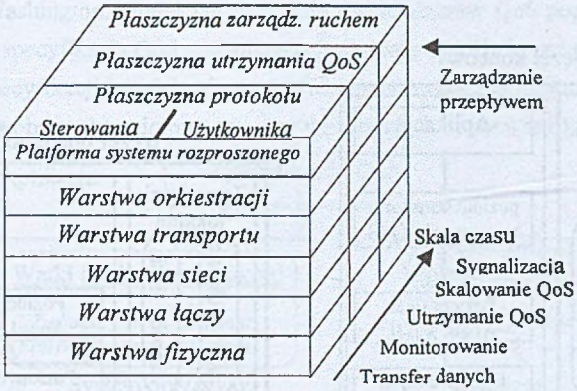
O ile tego typu mechanizmy umożliwiają sterowanie opóźnieniami pomiędzy kolejnymi porcjami danych w strumieniach, to utrzymanie jakości opóźnień na poziomie użytkowników aplikacji końcowych wymaga bilansowania opóźnień w całym torze przepływu danych [1].

Podstawowym problemem w zapewnieniu ustalonego poziomu jakości usług jest niejednorodność wykonywanych współbieżnie aplikacji. Z punktu widzenia zarządzania przepływami izochronicznymi występują sporadyczne żądania zasobów, zakłócające rytmikę obsługi strumieni ciągłych.

Zarządzanie jakością usług na poziomie sieci w przypadku sieci ATM jest określone poprzez dobrze określone klasy jakości usług, jak CBR czy VBR-rt. Podstawowym celem multimedialnych środowisk operacyjnych, pośredniczących pomiędzy aplikacją i warstwami sieci, jest dostosowanie wymagań do wybranej klasy usługi sieciowej.

3. Architektury QoS

Opracowanie modelu informatycznego środowiska umożliwiającego realizację usług o ustalonej jakości usług na poziomie użytkownika jest od kilku lat przedmiotem badań w wielu ośrodkach uniwersyteckich [4]. Jedną ze znaczących propozycji jest model QoS-A [2] przedstawiony na rysunku 3.



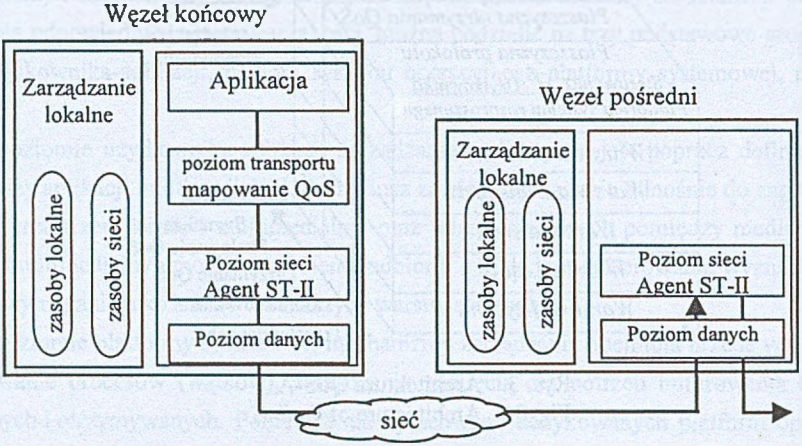
Rys. 3. Architektura QoS-A
Fig. 3. Architecture of QoS-A

Model QoS-A jest wielowarstwową strukturą mechanizmów sterowania i zarządzania zasobami zapewniającymi jakość usług. Kluczowymi pojęciami modelu są: przepływy (flows), reprezentujące generowanie, transmisję oraz odbiór pojedynczego strumienia danych; kontrakt, pełniący rolę umowy pomiędzy użytkownikiem i dostawcą, określającej poziom jakości usług; zarządzanie przepływem, realizujące funkcje monitorowania i utrzymania jakości na zadanym poziomie.

Najwyższa warstwa modelu reprezentuje platformę realizacji aplikacji wyposażoną w usługi komunikacji i specyfikacji QoS. Poniżej poziomu platformy umieszczona została warstwa orkiestracji, zawierająca mechanizmy synchronizacji przepływów strumieni. Kolejne niższe warstwy są zgodne z modelem architektur sieciowych ISO/OSI. Zarządzanie realizowane jest w trzech płaszczyznach pionowych. Płaszczyznę protokołu podzielono ponadto na dwie podpłaszczyzny zaznaczając tym odmienną charakterystykę przepływu danych i sterowania. Płaszczyzna utrzymania QoS zawiera w kolejnych warstwach funkcje zarządzania zasobami, natomiast płaszczyzna zarządzania przepływem reprezentuje mechanizmy kwalifikacji i rezerwacji zasobów oraz monitorowania poziomu jakości realizowanych usług. W warstwie orkiestracji zadaniem utrzymania QoS jest zapewnienie synchronizacji pomiędzy przepływami, podczas gdy w warstwie transportowej zadanie to polega na zapewnieniu parametrów danego przepływu takich, jak: przepustowość, stopień strat czy opóźnienie.

Ostatnia płaszczyzna architektury QoS-A, zarządzanie przepływem, odpowiada za zestawienie przepływu oraz rezerwację i sterowanie zasobami.

Innym przykładem jest środowisko HeiRAT opracowane przez ENC IBM, przedstawione na rysunku 4.



Rys. 4. Architektura HeiRAT
Fig. 4. Architecture of HeiRAT

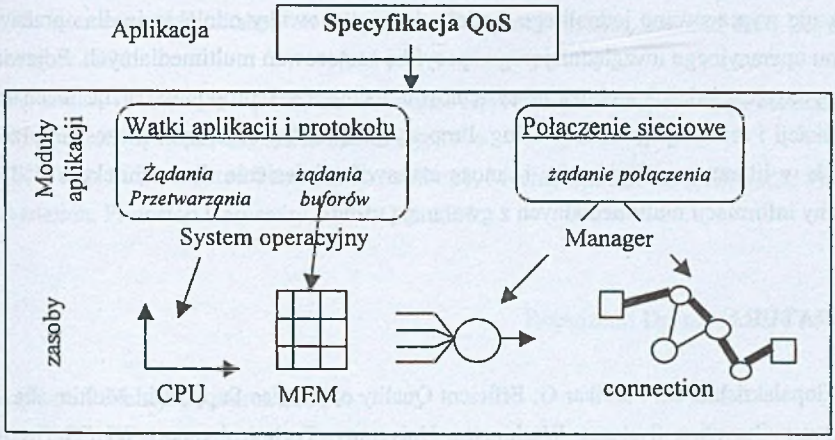
Architektura HeiRAT składa się z dwóch podstawowych konfiguracji: węzłów końcowych (źródłowego i docelowego) oraz węzła pośredniego. Aplikacje źródłowa i docelowa, agenci realizujący protokół rezerwacji zasobów oraz moduły lokalnego zarządzania zasobami uczestniczą w procesie negocjacji poziomu usług QoS. Wymagania aplikacji źródłowej odwzorowywane są przez warstwę transportową na parametry warstwy sieciowej, a następnie generowane są komunikaty w celu zestawienia połączenia. Moduł lokalnego zarządzania zasobami, w każdym węźle ścieżki, po otrzymaniu komunikatu oblicza zapotrzebowanie QoS i rezerwuje odpowiednią ilość zasobów, po czym za pośrednictwem agenta informuje węzeł inicjujący o gotowości transmisji ustalonego strumienia. W przypadku braku możliwości spełnienia wymagań wysyłany jest do stacji inicjującej komunikat zwrotny i zwalniane są wszystkie zasoby.

Dla potrzeb określenia jakości przesyłanych strumieni zdefiniowano trzy podstawowe parametry: przepustowość, opóźnienie oraz niezawodność, którą podzielono na cztery klasy w zależności od wystąpienia błędów w strumieniu i utraty pakietów.

System HeiRAT realizuje cztery podstawowe funkcje dotyczące zasobów: test przepustowości, kalkulacja QoS, rezerwacja i szeregowanie.

Funkcje te przeznaczone są do zapewnienia jakości usług na dwóch poziomach: gwarantowany i statystyczny.

Na rysunku 5 przedstawiono model architektury QoS systemu końcowego opracowany w Uniwersytecie Washington. Model ten dostarcza mechanizmów QoS poprzez cztery podstawowe elementy: specyfikacja QoS, mapowanie QoS, realizacja QoS oraz protokół. Na podstawie specyfikacji jakości usługi, poprzez mapowanie, następuje przyporządkowanie odpowiednich zasobów na poziomie sesji współpracujących ze sobą aplikacji końcowych.



Rys. 5. Architektura QoS-UW
Fig. 5. Architecture of QoS-UW

Głównymi zasobami jest procesor oraz pamięć stacji końcowej. Realizacja QoS odbywa się poprzez algorytmy szeregowania dzielonych zasobów w czasie transmisji danych, tak aby spełnić wymagania dotyczące danego połączenia. Mechanizmy specyfikacji, mapowania oraz realizacji QoS zintegrowano na poziomie implementacji protokołów przesyłu danych w ramach sesji aplikacji. Przyjęto, że procesy (wątki) aplikacji i protokołu realizowane są w tej samej przestrzeni adresowej, a wymiana informacji odbywa się poprzez pamięć dzieloną. W ten sposób wszystkie elementy realizujące wymianę danych zgodnie ze specyfikacją jakości usług podlegają tym samym mechanizmom, co ułatwia zarządzanie zasobami spełniając wymagania QoS. W celu zwiększenia efektywności realizacji transmisji (w celu zmniejszenia częstotliwości przełączania kontekstu) dodatkowo wprowadzono metodę opóźnionego wywłaszczania procesów, to znaczy, że proces wywłaszczany jest dopiero po przetworzeniu zadanej liczby jednostek danych.

W celu uproszczenia specyfikacji QoS oraz klasyfikacji wprowadzono trzy cechy: izochroniczność, wybuchowość oraz małe opóźnienie. Każda z aplikacji charakteryzowana jest pięcioma parametrami: maksymalny rozmiar ramki, prędkość ramek, średni rozmiar ramki, opóźnienie oraz przepustowość.

4. Podsumowanie

Modelowe architektury informatycznych środowisk multimedialnych gwarantujących realizację usług na ustalonym, przez użytkownika-aplikację, poziomie jest jednym z głównych kierunków prac badawczych w dziedzinie rozwoju systemów multimedialnych. Dotychczas jednak nie wypracowano jednolitego modelu, który stanowiłby odniesienie dla opracowania systemu operacyjnego uwzględniającego specyfikę zastosowań multimedialnych. Pojawiające się propozycje modeli i architektur systemów z usługą QoS proponują różne mechanizmy specyfikacji i realizacji gwarancji usług. Propozycja QoS-A jest jednym z częściej przedstawianych w literaturze przykładów i może stanowić odniesienie dla architektur środowisk wymiany informacji multimedialnych z gwarancją usług.

LITERATURA

1. Gopalakrishna R., Parulkar G: Efficient Quality of Service Support in Multimedia Computer Operating Systems, Washington University, Technical report WUCS-TM-94-04, August,1994.
2. Campbell A.T.: A Quality of Service Architecture, PhD Thesis, Computing Depart-ment, Lancaster University, 1996.
3. Siqueira F.: The Design of a Generic QoS Architecture for Open Systems, PhD Project, Department of Computer Science, Trinity College Dublin, February 1998.
4. Mallikarjum Tatipamula, Bhumip Khasnabish, eds.: Multimedia Communications Networks, ARTECH HOUSE,INC.,Boston-London.1998.
5. Lin C., Chu H., Nahrstedt K.: A Soft Real-time Scheduling Server on the Windows NT, 2nd USENIX Windows NT Symposium, Seattle Washington, August 3-5, 1998.
6. Chu H., Nahrstedt K.: A Soft Real Time Scheduling Server in UNIX Operating Sys-tem, European Workshop IDMS'97, Darmstadt, September 1997.
7. Lee Ch., Rajkumar R., Mercer C.: Experiences with Procesor Reservation and Dy-namic QoS in Real-Time Mach, Proceedings of Multimedia Japan, March 1996.
8. Nahrstedt K., Steinmetz R.: Resource Management in Multimedia Networked Sys-tems, IEEE Computer, May 1995.
9. Tenenhouse D., Smith J., Sincoskie W., Wetherall D., Minden G.: A survey of Active Network Research, IEEE Communications Magazine, vol. 35, no. 1, January 1997.
10. M.J. Perez-Luque, T.D.C. Little: A Temporal Reference Framework for Multimedia Synchronization, MCI Techn. Rep., Boston University,1995.

11. Stallings W.: High-speed Networks: TCP/IP and ATM Design Principles, Printice Hall, New Jersey,1998.
12. S. Frolund, J. Koistinen: Quality os Service Aware Distributed Object Systems, 5th Conference on OOT and Systems, San Diego, May, 1999.
13. J.B Stefani: Computational Aspects of QoS in an Object Based Distributed Architec-ture, 3th International Workshop on Responsive Computer Systems, Lincoln, NH, USA, September 1994.
14. A. Lakas, G. Blair, A Chetwynd: A Formal Approach to Design of QoS Parameters in Multimedia Systems, Lancaster University, 1996
15. W. Fang: Differentiated Services: Architecture, Mechanism and an Evaluation, PhD dissertation, Princeton University, November 2000.

Recenzent: Dr inż. Henryk Małysiak

Wpłynęło do Redakcji 30 marca 2001 r.

Abstract

In this paper the terminology and selected QoS architectures issues that emerged is summarized and discussed. While the area of QoS research in multimedia networking is mature, work on architecture remains in its early stage development with no unified model of architecture having been published.

This work is a short presentation of key principles, services and mechanisms which are responsible for providing QoS in distributed multimedia systems .