



**INSTYTUT BADAŃ SYSTEMOWYCH
POLSKIEJ AKADEMII NAUK**

**ANALIZA SYSTEMOWA W FINANSACH
I ZARZĄDZANIU**

Wybrane problemy
Tom 4

Pod redakcją
Jerzego HOŁUBCA

Warszawa 2002



**INSTYTUT BADAŃ SYSTEMOWYCH
POLSKIEJ AKADEMII NAUK**

ANALIZA SYSTEMOWA W FINANSACH I ZARZĄDZANIU

**Wybrane problemy
Tom 4**

**Pod redakcją
Jerzego HOŁUBCA**

Warszawa 2002

Wykaz opiniodawców artykułów zamieszczonych w tomie:

doc. dr hab. Mieczysław KŁOPOTEK

prof. dr hab. Stanisław PIASECKI

prof. dr Elżbieta RAKUS-ANDERSON

prof. dr hab. Andrzej STRASZAK

doc. dr hab. Sławomir WIERZCHOŃ

dr Sławomir ZADROŻNY

Publikacja dofinansowana przez
Agencję Wydawniczo-Poligraficzną "ARGRAF", Warszawa

© Instytut Badań Systemowych PAN, Warszawa 2002

ISBN 83-85847-74-X

Wydawca: INSTYTUT BADAŃ SYSTEMOWYCH PAN
ul. Nowelska 6 01-447 Warszawa

Redakcja: Dział Informacji Naukowej i Wydawnictw

Barbara Katuszewska, Joanna Runowska, tel. 837-68-22

Druk: Agencja Wydawniczo-Poligraficzna "ARGRAF", Warszawa

Nakład 200 egz., 15 ark.wyd.; 12,8 .ark. druk.

PROBLEMY OPÓŹNIEŃ TRANSMISJI I ZARZĄDZANIA OPÓŹNIENIAMI W SIECIACH TELEINFORMATYCZNYCH

Maciej Łopaciński

Zaoczne Studia Doktoranckie IBS PAN

M.Lopacinski@ibspan.waw.pl

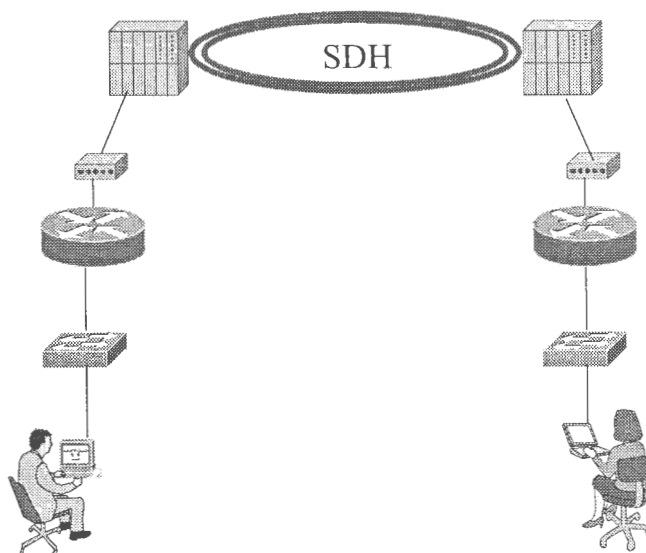
Słowa kluczowe: systemy informatyczne, sieci komputerowe, voip.

1. Wprowadzenie

Czas przesyłania danych w sieciach komputerowych jest istotnym czynnikiem decydującym o działaniu systemów informatycznych w środowiskach rozproszonych. Czas ten, zwany opóźnieniem, jest nierozdzielnie związany z zagadnieniem przesyłania danych.

2. Źródła opóźnień

Typowa sieć komputerowa jest przedstawiona na poniższym rysunku.



Rys. 1 Struktura typowej sieci telekomunikacyjnej

Dane przesyłane w sieci są w postaci pakietów, zwykle o rozmiarze od 64 do 1500 bajtów. Pakiet danych stacji A do stacji B wędruje przez szereg urządzeń i łączy. Każde z łączy wnosi czas związany z przekazywaniem pakietu przez łącze. Każde urządzenie wnosi czas związany z przetwarzaniem pakietu.

Głównymi źródłami opóźnień są: serializacja, kolejkowanie i czas transmisji.

Czas serializacji jest czasem związanym z przesłaniem pakietu bit po bicie przez łącze szeregowe. Czas przesłania każdego bitu jest ściśle określony i zależy od przepustowości interfejsy przez który dane są przesyłane. Tabela poniżej (źródło www.cisco.com) przedstawia czasy przesłania pakietu danych przez interfejsy o różnej szybkości.

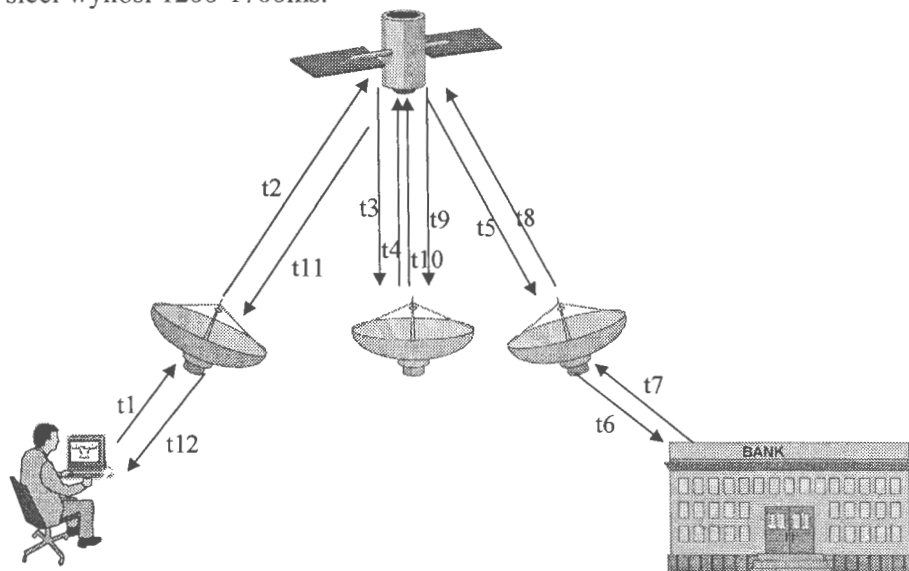
| | 1 Byte | 64 Bytes | 128 Bytes | 256 Bytes | 512 Bytes | 1024 Bytes | 1500 Bytes |
|---------|--------|----------|-----------|-----------|-----------|------------|------------|
| 56kbps | 143us | 9ms | 18ms | 36ms | 72ms | 144ms | 214ms |
| 64kbps | 125us | 8ms | 16ms | 32ms | 64ms | 128ms | 187ms |
| 128kbps | 62.5us | 4ms | 8ms | 16ms | 32ms | 64ms | 93ms |
| 256kbps | 31us | 2ms | 4ms | 8ms | 16ms | 32ms | 46ms |
| 512kbps | 15.5us | 1ms | 2ms | 4ms | 8ms | 16ms | 23ms |
| 768kbps | 10us | 640us | 1.28ms | 2.56ms | 5.12ms | 10.24ms | 15ms |
| 1536kbs | 5us | 320us | 640us | 1.28ms | 2.56ms | 5.12ms | 7.5ms |

Jeśli łącze ma jedno zakończenie o większej przepustowości, a drugie o mniejszej to czas serializacji pakietu jest związany z łączem o mniejszej przepustowości.

Czas kolejkowania. Urządzenia sieciowe sterują ruchem pakietów. W chwili, gdy nowy pakiet trafia do urządzenia sieciowego musi ono określić do jakiego interfejsu wyjściowego ma go skierować. Określenie tego i ew. przebudowa pakietu zajmuje pewien czas. Następnie jeśli interfejs jest zajęty przez poprzedni pakiet, nowy pakiet musi poczekać, przez czas związany z serializacją pakietu poprzedniego. Pakiety oczekujące są umieszczane w kolejce wyjściowej. Opróżnianie kolejki wyjściowej może odbywać się zgodnie ze strategią FIFO (pierwszy wchodzi, pierwszy wychodzi) lub według strategii uwzględniających znaczenie pakietu dla systemu

informatycznego. Niektóre strategie są bardzo wyrafinowane i mogą uwzględniać np. strumienie danych pomiędzy poszczególnymi aplikacjami, wielkości pakietów, interaktywność pracy itp.

Czas transmisji jest związany z czasem wędrówki sygnału przez medium transmisyjne. Czas transmisji zauważany przez użytkownika jest sumarycznym czasem transmisji sygnału na wszystkich odcinkach i w obu kierunkach. Całkowity czas transmisji t_c dla sieci zrealizowanej w technologii VSAT jest zilustrowany na rysunku poniżej. Typowo t_c w takiej sieci wynosi 1200-1700ms.



$$t_c = t_1 + t_2 + t_3 + t_4 + t_5 + t_6 + t_7 + t_8 + t_9 + t_{10} + t_{11} + t_{12}$$

Rys. 2 Opóźnienie w sieci VSAT

W sieciach pakietowych takich jak frame-relay, czy IPVPN opóźnienie związane jest dodatkowo z kolejkowaniem w urządzeniach operatora. Na to kolejkowanie użytkownik ma wpływ jedynie przez wykupienie odpowiednio droższej usługi

3. Opóźnienie a przepustowość

Dla analizy skutków opóźnień bardzo ważne jest zrozumienie różnicy pomiędzy przepustowością, a opóźnieniem. Przepustowość jest to liczba bitów jakie możemy przesłać w jednostce czasu. Opóźnienie jest to czas jaki zajmuje nam przesłanie pakietu danych o określonej długości przez sieć telekomunikacyjną. Prostą analogią z życia jest przykład autostrady.

Wyobraźmy sobie dwie autostrady. Jedną o dwóch pasach ruchu w każdą stronę bezpośrednio z Warszawy do Łodzi. Drugą, o czterech pasach ruchu w każdą stronę, ale wiodącą z Warszawy do Łodzi przez Gdańsk. Czas przejazdu pierwszą wynosi godzinę, a drugą sześć godzin. Która ma większą przepustowość? Ta czteropasmowa wiodąca przez Gdańsk. Jeśli wyślemy nią ciągły strumień ciężarówek to możemy dowieźć dwa razy więcej towaru, mimo, że podróż potrwa sześć razy dłużej.

4. Skutki opóźnień

Różne systemy informatyczne mają różną wrażliwość na opóźnienia.

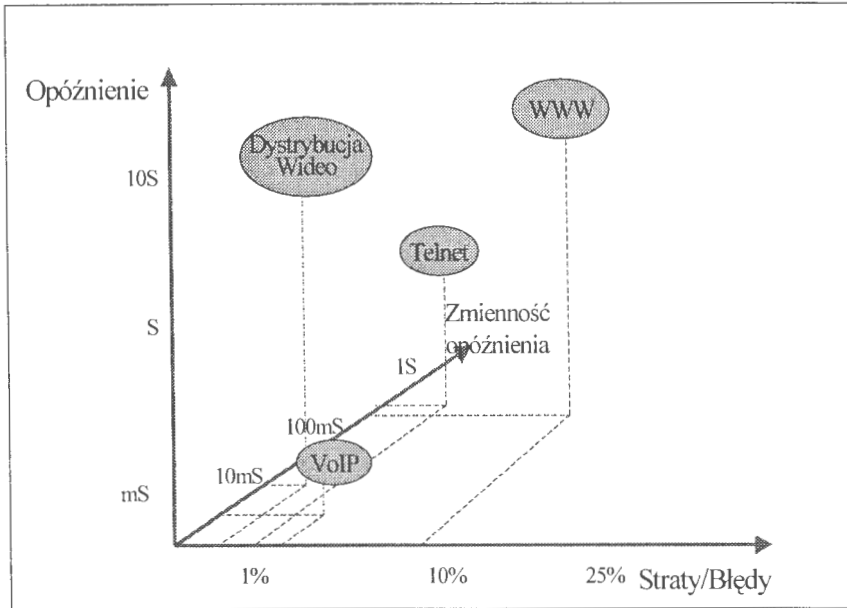
$$T = \frac{wp * 8}{pl} + tc = \frac{100MB * 8}{1Mb / s} + tc = 800,01s$$

Są systemy zupełnie nie wrażliwe na opóźnienie. Przykładem takiego systemu może być transmisja dużych plików protokołem FTP w systemach z dobrym stosem TCP/IP. Przesłanie pliku o wielkości $wp=100MB$, łączem o przepustowości $pl=1Mb/s$ trwa i opóźnieniu $tc=10ms$ trwa T .

Jeśli opóźnienie tc zwiększy się nawet do 1 sekundy to T wyniesie $T=801$ sekund. Proporcjonalnie do T wzrost jest praktycznie nie istotny.

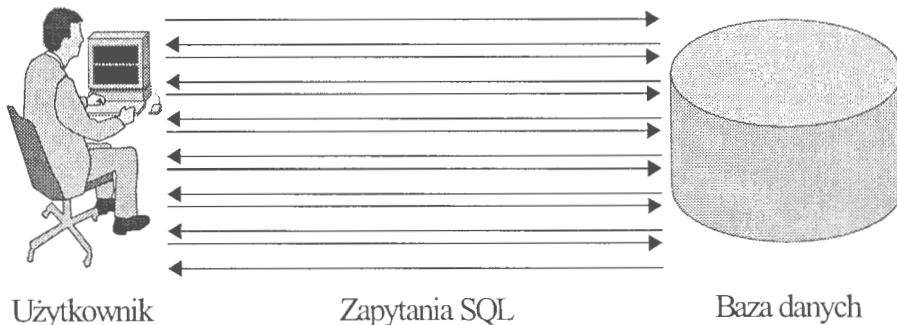
Systemy terminalowe są znacznie bardziej wrażliwe na opóźnienia. W systemie terminalowym opóźnienie objawia się oczekiwaniem operatora na pojawienie się na ekranie znaku naciśniętego na klawiaturze. Opóźnienie jednosekundowe oznacza że po naciśnięciu klawisza operator musi odczekać sekundę na pojawienie się odpowiedniego znaku na ekranie. Dopuszczalne, opóźnienie zależy od indywidualnej percepcji użytkownika.

Systemy telefonii w sieci (VoIP, VoFR, VoATM) - odczuwana przez użytkownika jakość rozmowy zależy nie tylko od poziomu szumów, ale również od opóźnień. Rozmowa w której opóźnienia są zbyt duże staje się nienaturalna. Rozmówcy czują dyskomfort czekając na reakcję drugiej osoby po wypowiedzianej kwestii. Jak wykazują badania [Steinmetz] dopuszczalne opóźnienie dla rozmowy telefonicznej, łącznie z czasami kodowania i dekodowania sygnału mowy, wynosi ok. 200ms.



Rys.3 Wrażliwość aplikacji multimedialnych na opóźnienie, straty i zmienność opóźnień – źródło [Steinmetz]

Systemy klient-serwer. W architekturze klient serwer działają często systemy bazodanowe. Przetwarzanie jest realizowane na stacji klienckiej, a zapytania wysyłane do serwera. Dobrze podręcznikowo napisany program wykonuje sekwencje zapytań do bazy i w zależności od wyniku zadaje kolejne pytania. Dla programisty napisanie całego zapytania do bazy w jednym wielkim zapytaniu połączonym warunkami logicznymi AND i OR jest co najmniej nieeleganckie. Prześledźmy jak działa poprawnie i elegancko napisany program. Rysunek poniżej przedstawia schemat komunikacji.



Rys. 4 Działanie aplikacji klient serwer

Przyjmijmy, że aplikacja wykonuje 100 zapytań do serwera w celu wykonania operacji sprzedaży. Czas wykonania operacji sprzedaży wynosi

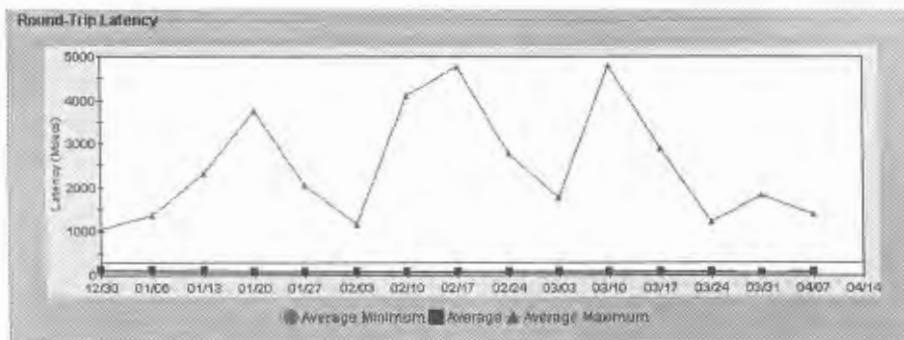
$$Ts = n * (t_{SQL} + t_c)$$

gdzie T_s czas transakcji sprzedaży, t_{SQL} – czas wykonania operacji na serwerze, t_c – czas przesłania danych przez sieć. W sieci lokalnej przy $n=100$, $t_c=1ms$ i $t_{SQL} = 1ms$, czas T_s wynosi 200ms. Jest to akceptowalny czas zawarcia transakcji sprzedaży. Jeśli przeniesiemy aplikację w środowisko sieci rozległej z $t_c=40$ ms, to czas sprzedaży wzrośnie do ponad 4 sekund. Ten czas jest prawdopodobnie na granicy akceptowalności. Jeśli jednak nasza sieć rozległa korzysta z usług operatora sieci frame-relay, w której opóźnienia potrafią dochodzić do 600ms to czas sprzedaży wyniesie

$$T_s = n * (t_{SQL} + t_c) = 100 * (1ms + 600ms) = 60100ms = 60,1s.$$

ponad minutę. Taki czas jest nie do zaakceptowania. Prawdopodobnie nasz klient następnym razem pójdzie do konkurencji.

Rzeczywiste opóźnienia w sieci frame-relay Polpak-T w relacji Górzów Zielona Góra w postaci średnich tygodniowych, w okresie 30 grudnia 2001 – 7 kwietnia 2002, przedstawione są w tabeli i na wykresie.



Rys. 5 Opóźnienia w sieci Polpak-T wykres

5. Opóźnienie w sieciach operatorów telekomunikacyjnych

Na dzień dzisiejszy – kwiecień 2002 gwarancję nie przekraczania maksymalnego założonego opóźnienia można uzyskać jedynie od operatorów sieci TDM sprzedających cyfrowe łącza dzierżawione. Wynika to jednak z technologii sieci TDM, a nie z zarządzania jakością usługi.

Również niektórzy dostawcy interentu gwarantują maksymalne opóźnienie np. Ipartners, Telia, Teleglobe. Jest to opóźnienie tylko i wyłącznie liczone w obrębie sieci tego operatora, a nie w całym internecie.

| Round-Trip Latency | | |
|--------------------|-------|--|
| Average Maximum: | 2,324 | |
| Average: | 78 | |
| Average Minimum: | 57 | |
| Minimum Latency: | 0 | |
| Maximum Latency: | 4,789 | |

| Errors | | |
|--------------------------|--------|-------|
| Total Pkt Loss Src-Dest: | 3,628 | 2.69% |
| Total Pkt Loss Dest-Src: | 398 | 0.30% |
| Total Buses: | 0 | 0.00% |
| Total Sequence Errors: | 0 | 0.00% |
| Total Drops: | 10,196 | 7.56% |
| Total Packets MA: | 1,773 | 1.31% |
| Total Packets Late: | 0 | 0.00% |

Rys. 6 opóźnienia i straty pakietów w sieci Polpak-T

Analizując takie oferty trzeba pamiętać, że opóźnienie to jeden z elementów SLA (service level agreement). Oferowany poziom SLA jest natomiast narzędziem walki konkurencyjnej działów marketingu i sprzedaży, które zrobią wszystko, aby ich SLA wyglądało lepiej od konkurencyjnych. Najczęściej stosowanymi, zwłaszcza w zakresie dostępności, ale również opóźnień sztuczkami statystycznymi w liczeniu SLA są:

- uśrednianie w ciągu roku;
- uśrednianie wszystkich punktów w których klient kupuje usługę;
- wyłączenie problemów krótszych niż 5 minut;
- liczenie czasu trwania problemu dopiero od momentu zgłoszenia przez klienta problemu faxem.

Główny dostawca usług pakietowej transmisji danych w Polsce – Polpak-T nie jest zainteresowany gwarantowaniem maksymalnych opóźnień.

Zgodnie z wypowiedzią dyrektora Polpaku Euzebiusza Sowy mało klientów jest zainteresowanych zapłaceniem za usługę z gwarantowanym opóźnieniem. Dla tej grupy klientów nie opłaca się wdrażać systemu monitorującego opóźnienia, a parametr który nie jest monitorowany nie może być elementem oferty.

6. Sposoby zapobiegania opóźnieniom

Do skutecznego zapobiegania opóźnieniom potrzebna jest prawidłowa diagnoza przyczyny opóźnień. W przypadku opóźnień wynikających z natury medium transmisyjnego np. łącza satelitarne lub z działania operatora telekomunikacyjnego jedyną metodą eliminacji opóźnień jest zmiana operatora lub medium. Na słabo rozwiniętych rynkach telekomunikacyjnych nie zawsze jest to możliwe.

Opóźnienia wynikające z dużego natłoku danych w sieci eliminuje się przez zastosowanie zaawansowanych technik kolejkowania na urządzeniach sieciowych. Mechanizmy kolejkowania mogą same próbować wykryć dane które mają priorytet lub mogą polegać na znakowaniu ważnych pakietów przez aplikacje. Temat tej jest szeroko opisywany w materiałach producentów urządzeń sieciowych [cisco].

Opóźnienia spowodowane serializacją można wyeliminować przez fragmentację pakietów na routerach obsługujących wolne łącza. W sieciach IP jest to realizowane przez technologię „Link Fragmentation and Interleaving” [cisco3], a w sieciach Frame-relay jest to technologia opisana normą FRF-12.

W przypadku korzystania z usług operatorów telekomunikacyjnych oferujących usługi frame-relay na łączach o małej wykupionej przepustowości sieci możemy zastosować dużą przepustowość portu. Przykładowo przy wykupionej przepustowości łącza 64kb/s możemy wykupić przepustowość portu 2Mb/s. Mamy prawo wówczas wprowadzać pakiet danych do sieci z prędkością 2Mb/s, ale nie możemy wprowadzić więcej niż 64kb danych w ciągu jednej sekundy.

Literatura

[Steinmetz] Ralf Steinmetz, Clemens Engler, "*Human perception of Media Synchronization*", Technical Report, No. 43.9310, IBM European Networking Center, Heidelberg, Germany, 1993

[Pashby] Ron Pashby Testing IP QoS Workshop
<http://smartbits.spirentcom.com/technology/smartvoipqos/presentation/QoS-Workshop5.ppt>

[frf1] Frame Relay Forum „Basic Guide to Frame Relay Networking”

[cisco1] Designing Internetworks for Multimedia
<http://www.cisco.com/univercd/cc/td/doc/cisintwk/idg4/nd2013.htm>

[walsh] Mikel Walsh „Integrating Voice and Data VPN: Voice over Frame vs. Voice over ATM/IP” Prezentacja z NetWorld+Interop Paris 2000

[ftf2] www.frforum.com

[cisco2] http://www.cisco.com/warp/public/788/voip/voice_rfc.html

[cisco3] http://www.cisco.com/univercd/cc/td/doc/product/software/ios120/12cgcr/qos_c/qcpart6/qclfi.htm

ISBN 83-85847-74-X

)