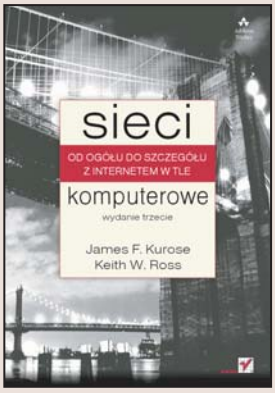
Sieci komputerowe. Od ogółu do szczegółu z internetem w tle. Wydanie III



Autorzy: James F. Kurose, Keith W. Ross
Tłumaczenie: Piotr Pilch, Grzegorz Werner
ISBN: 83-246-0353-0
Tytuł oryginału: Computer Networking: A Top-Down Approach Featuring the Internet (3rd Edition)
Format: B5, stron: 792

Kompleksowy przegląd wszystkich zagadnień związanych z sieciami komputerowymi

- Protokoły komunikacyjne
- Aplikacje sieciowe
- Sieci bezprzewodowe i mobilne
- Bezpieczeństwo sieci

Sieci komputerowe są już tak powszechne, że niemal nie zauważamy ich istnienia. Na co dzień używamy internetu, sieci bezprzewodowych, hot-spotów w hotelach i restauracjach, w zasadzie nie zastanawiając się, jak to wszystko działa. Jeśli jednak nie chcesz ograniczać się do biernego korzystania z tego, co wymyślili inni, ale masz ambicję, by samodzielnie opracowywać rozwiązania sieciowe, musisz poznać technologię, która pozwala na niezakłóconą wymianę informacji.

Książka „Sieci komputerowe. Od ogółu do szczegółu z internetem w tle. Wydanie III" umożliwia zrozumienie zasad działania systemów sieciowych. Czytając ją, zdobędziesz wiedzę, dzięki której sieci komputerowe odkryją przed Tobą wszystkie tajemnice. Poznasz warstwy sieci, dowiesz się, w jaki sposób realizowany jest przekaz informacji, jak działają serwery i aplikacje sieciowe oraz jakie metody zabezpieczeń są współcześnie stosowane.

- Historia sieci komputerowych
- Protokoły HTTP i FTP
- Poczta elektroniczna
- Usługi DNS
- Protokoły transportowe
- Algorytmy routingu
- Adresowanie w sieciach komputerowych
- Sieci bezprzewodowe
- Komunikacja mobilna
- Multimedia w sieci i telefonia internetowa
- Zabezpieczanie sieci i danych

Poznaj tajniki sieci komputerowych

Wydawnictwo Helion
ul. Chopina 6
44-100 Gliwice
tel. (32)230-98-63
e-mail: helion@helion.pl

Spis treści

1
Sieci komputerowe i internet

Gdy pomyśli się o przeglądarkach internetowych zainstalowanych w telefonach komórkowych, kawiarenkach z bezprzewodowym publicznym dostępem do internetu, sieciach domowych dysponujących szybkim szerokopasmowym dostępem, infrastrukturze informatycznej firmy, w przypadku której każdy komputer PC znajdujący się na biurku jest podłączony do sieci, samochodach z interfejsem sieciowym, czujnikach rozmieszczonych w środowisku naturalnym i łączących się z siecią, a także międzyplanetarnym internecie, można po prostu odnieść wrażenie, że właściwie sieci komputerowe są wszechobecne. Opracowywane są nowe ekscytujące zastosowania, które jeszcze bardziej rozszerzają zasięg obecnie istniejących sieci. Wydaje się, że sieci komputerowe są wszędzie! Książka w nowatorski sposób wprowadza w dynamicznie zmieniający się świat sieci komputerowych, przedstawiając zasady i praktyczne analizy, które są niezbędne do tego, aby zrozumieć nie tylko obecnie istniejące sieci, ale również te, które pojawią się w przyszłości.

W niniejszym rozdziale dokonano przeglądu sieci komputerowych i internetu. Celem rozdziału jest nakreślenie ogólnego obrazu, który pozwoli ujrzeć las spoza drzew. W rozdziale zamieściliśmy wiele podstawowych informacji i omówiliśmy mnóstwo elementów sieci komputerowej, mając jednocześnie na względzie sieć jako całość. Informacje znajdujące się w rozdziale stanowią przygotowanie do reszty książki.

Przegląd sieci komputerowych dokonany w tym rozdziale ma następujący układ. Po zaprezentowaniu kilku podstawowych terminów i zagadnień w pierwszej kolejności przyjrzymy się elementarnym składnikom sprzętowym i programowym, które tworzą sieć. Rozpoczniemy od obrzeża sieci, po czym omówimy systemy końcowe i aplikacje uruchamiane w sieci. Przyjrzymy się usługom transportowym zapewnianym tym aplikacjom. W dalszej kolejności objaśnimy rdzeń sieci komputerowej, omawiając łącza i przełączniki przesyłające dane, a także sieci dostępowe i fizyczne nośniki łączące systemy końcowe z rdzeniem sieci. Czytelnik dowie się, że internet jest siecią złożoną z wielu sieci, a ponadto omówimy, w jaki sposób są one ze sobą połączone.

Po zakończeniu przeglądu obrzeża sieci komputerowej i jej rdzenia w drugiej części rozdziału dokonamy szerszego i bardziej abstrakcyjnego omówienia. Przyjrzymy się przyczynom opóźnień w transmisji danych w sieciach i ich utracie. Zaprezentujemy proste modele ilościowe demonstrujące opóźnienie występujące między dwoma

węzłami końcowymi. Modele te będą uwzględniały opóźnienia transmisji, propagacji i kolejkowania. Dalej przedstawimy kilka kluczowych zasad dotyczących architektury sieci komputerowych, a dokładniej mówiąc, warstw protokołów i modeli usług. Na końcu rozdziału zamieszczono w skrócie historię sieci komputerowych.

1.1. Czym jest internet?

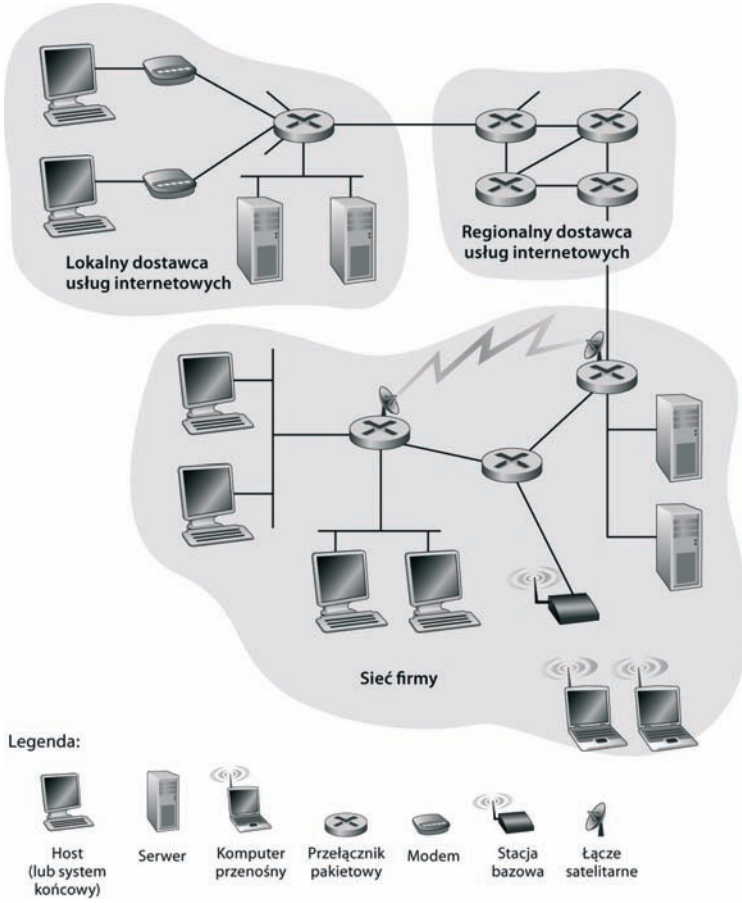
W książce publiczny internet będący specyficzną siecią komputerową traktujemy jako podstawowy fundament, na którym są oparte omówienia sieci i ich protokołów. Jednak czym jest internet? Chcielibyśmy przedstawić jednozdaniową definicję internetu, którą można podzielić się z rodziną i znajomymi. Niestety internet jest bardzo złożony i cały czas się zmienia pod względem składników sprzętowych i programowych, a także usług, które oferuje.

1.1.1. Opis podstawowych komponentów

Zamiast podawać jednozdaniową definicję internetu, spróbujmy skorzystać z bardziej opisowej metody. Można to zrobić na kilka sposobów. Jeden z nich polega na omówieniu podstawowych sprzętowych i programowych składników tworzących internet. Kolejnym sposobem jest scharakteryzowanie internetu w kategoriach infrastruktury sieciowej, która zapewnia usługi aplikacjom rozproszonym. Zacznijmy od opisu podstawowych składników, korzystając z rysunku 1.1 ilustrującego omówienie.

Publiczny internet jest globalną siecią komputerową łączącą miliony urządzeń komputerowych zlokalizowanych w różnych miejscach świata. Nie tak dawno temu tymi urządzeniami były przede wszystkim tradycyjne stacjonarne komputery PC, uniksowe stacje robocze i tak zwane serwery, które przechowują i przesyłają dane, takie jak strony internetowe i wiadomości pocztowe. Jednak wraz z upływem czasu do internetu zaczęto podłączać nowoczesne urządzenia końcowe, takie jak cyfrowe asystenty osobiste PDA (*Personal Digital Assistant*), telewizory, komputery przenośne, telefony komórkowe, samochody, czujniki środowiskowe, przeglądarki slajdów, domowe urządzenia elektryczne, systemy zabezpieczeń, kamery internetowe, a nawet tostery [BBC 2001]. Tak naprawdę termin *sieć komputerowa* zaczyna być trochę nieaktualny, gdy pod uwagę weźmie się to, że wiele nietradycyjnych urządzeń jest podłączanych do internetu. W żargonie internetowym wszystkie wymienione urządzenia są nazywane **hostami** lub **systemami końcowymi**. W styczniu 2003 r. z internetu korzystały ponad 233 miliony systemów końcowych. Wartość ta cały czas szybko się zwiększa [ISC 2004].

Systemy końcowe są połączone za pomocą **łączy komunikacyjnych**. W podrozdziale 1.4 będzie można się dowiedzieć, że istnieje wiele typów łączy komunikacyjnych, które są złożone z różnego rodzaju nośników fizycznych, takich jak kabel koncentryczny, kabel z miedzi, światłowód i widmo radiowe. Różne łącza mogą przesyłać dane z innymi szybkościami. **Szybkość transmisji** łącza jest wyrażana w bitach na sekundę.



RYSUNEK 1.1. Niektóre składniki internetu

Zwykle systemy końcowe nie są ze sobą bezpośrednio połączone przy użyciu pojedynczego łącza komunikacyjnego. Pośredniczą między nimi inne urządzenia przełączające, nazywane **przełącznikami pakietów**. Przełącznik pakietów pobiera porcję danych, którą odbierze za pomocą jednego ze swoich wejściowych łączy komunikacyjnych, i przekazuje ją do jednego z łączy wyjściowych. W żargonie związanym z sieciami komputerowymi porcja danych jest określana mianem **pakietu**. Choć dostępne są różne odmiany przełączników pakietów, aktualnie w internecie najczęściej wykorzystuje się dwa z nich — **routery** i **przełączniki warstwy łącza danych**. Oba typy przełączników przekazują pakiety w kierunku ich ostatecznego miejsca przeznaczenia. Routery szczegółowo zostaną omówione w rozdziale 4., natomiast przełączniki warstwy łącza w rozdziale 5.

Droga pokonywana przez pakiet wysłany przez nadawczy system końcowy, a następnie przechodzący przez kolejne łącza komunikacyjne i przełączniki pakietów, aby dotrzeć do odbiorczego systemu końcowego, jest nazywana **trasą** lub **ścieżką** sieciową. Zamiast tworzenia między komunikującymi się systemami końcowymi *dedykowanej* ścieżki, w internecie stosuje się technologię **przełączania pakietów**. Umożliwia ona jednoczesne współużytkowanie ścieżki lub jej fragmentów przez wiele systemów końcowych nawiązujących połączenia. Pierwsze sieci przełączające pakiety stworzone

w latach 70. są najstarszymi przodkami obecnego internetu. Choć dokładna ilość danych przesyłanych aktualnie w internecie jest kwestią podlegającą debacie [Odylsko 2003], według ostrożnych szacunków w ciągu miesiąca w amerykańskich sieciach rozległych przesyłanych jest około 100 000 terabajtów danych. Ponadto określa się, że każdego roku ilość transferowanych danych będzie w przybliżeniu dwukrotnie większa.

Systemy końcowe uzyskują dostęp do internetu za pośrednictwem **dostawcy usług internetowych ISP** (*Internet Service Provider*). Może to być lokalny dostawca lub operator telekomunikacyjny bądź telewizji kablowej. Może to być korporacyjny lub uniwersytecki dostawca ISP bądź taki jak firma T-Mobile, która zapewnia dostęp bezprzewodowy na lotniskach, w hotelach, kawiarniach i innych miejscach użyteczności publicznej. Sieć każdego dostawcy ISP jest złożona z przełączników pakietów i łączy komunikacyjnych. Dostawcy ISP oferują systemom końcowym różnego typu dostęp sieciowy, taki jak dostęp przy użyciu modemów telefonicznych o szybkości 56 kb/s, dostęp szerokopasmowy za pośrednictwem modemu kablowego lub DSL, dostęp za pomocą bardzo szybkiej sieci lokalnej i dostęp bezprzewodowy. Dostawcy ISP oferują też dostęp firmom zapewniającym informacje i korzystającym z serwerów WWW bezpośrednio podłączonych do internetu. Aby umożliwić użytkownikom internetowym komunikowanie się ze sobą i korzystanie z danych udostępnianych przez witryny WWW, wymienieni dostawcy ISP niższego poziomu są połączeni ze sobą przez krajowych lub międzynarodowych dostawców wyższego poziomu, takich jak AT&T i Sprint. Sieć takiego dostawcy zawiera bardzo szybkie routery połączone ze sobą przy użyciu szybkich łączy światłowodowych. Każda sieć dostawcy ISP, niezależnie od tego, czy niższego czy wyższego poziomu, jest oddzielnie zarządzana, a ponadto stosuje protokół IP (opisany niżej) i jest zgodna z określonymi konwencjami nazewniczymi i adresowymi. Dostawcom usług internetowych i ich powiązaniom przyjrzymy się bliżej w podrozdziale 1.5.

Ze względu na istotną rolę, jaką protokoły odgrywają w przypadku internetu, ważne jest jednoznaczne określenie przeznaczenia każdego protokołu. Właśnie to umożliwiają standardy. **Standardy internetowe** są opracowywane przez organizację IETF (*Internet Engineering Task Force*) [IETF 2004]. Dokumenty standardów tej organizacji są nazywane **dokumentami RFC** (*Requests for Comments*). Pierwotnie dokumenty RFC zawierały ogólne prośby o dołączanie własnych wniosków (stąd wywodzi się termin RFC — prośby o komentarze) umożliwiających rozwiązanie problemów związanych z architekturą sieci, z którymi borykali się prekursorzy internetu. Dokumenty RFC mają tendencję do bycia dość technicznymi i szczegółowymi. Definiują protokoły, takie jak TCP, IP, HTTP (obsługa witryn WWW) i SMTP (*Simple Mail Transfer Protocol*; obsługa poczty elektronicznej). Organizacja IETF dokonała również standaryzacji tego, jakie protokoły muszą być stosowane przez internetowy host (RFC 1122 i RFC 1123) i router (RFC 1812). Istnieje ponad 3500 dokumentów RFC. Inne organizacje też definiują standardy dotyczące komponentów sieciowych, a zwłaszcza łączy sieciowych. Przykładowo, organizacja IEEE 802 LAN/MAN Standards Committee [IEEE 802 2004] opracowuje standardy dotyczące Ethernetu i technologii bezprzewodowej Wi-Fi.

Systemy końcowe, przełączniki pakietów i inne składniki internetu wykorzystują **protokoły**, które kontrolują proces wysyłania i odbierania informacji transferowanych w internecie. **TCP** (*Transmission Control Protocol*) i **IP** (*Internet Protocol*) to dwa najważniejsze protokoły stosowane w internecie. Protokół IP określa format pakietów wysyłanych i odbieranych przez routery i systemy końcowe. W przypadku tych dwóch głównych protokołów internetowych używa się pojedynczego oznaczenia **TCP/IP**. Choć w tym rozdziale rozpoczniemy omawianie protokołów, jest to zaledwie początek. Spora część książki została poświęcona protokołom sieci komputerowych!

Publiczny internet, czyli globalna sieć złożona z wielu sieci, o czym wcześniej wspomniano, jest siecią, którą zwykle ma się na myśli, mówiąc o internecie. Istnieje też wiele prywatnych sieci, takich jak należące do korporacji i organizacji rządowych, których hosty nie mogą wymieniać się informacjami z innymi hostami znajdującymi się poza siecią prywatną (jeśli komunikaty nie przechodzą przez tak zwane firewalle, czyli zapory ogniowe, które ograniczają możliwości transferowania informacji do sieci i poza nią). Tego typu prywatne sieci często są nazywane **intranetami**, ponieważ znajdują się w nich takiego samego rodzaju hosty, routery, łącza i protokoły, jak w przypadku publicznego internetu.

1.1.2. Omówienie usług

Dotychczas zidentyfikowano wiele składników tworzących internet. Zrezygnujmy teraz z opisywania podstawowych elementów sieci i przyjrzyjmy się sieci z punktu widzenia usług.

- Internet pozwala na **zastosowania rozproszone**, w przypadku których systemy końcowe mogą przesyłać między sobą dane. Do zastosowań takich należy między innymi zaliczyć przeglądanie stron internetowych, komunikatory, transmisje strumieniowe danych audio i wideo, telefonię internetową, gry sieciowe, wymianę plików w sieciach równorzędnych (P2P — *Peer to Peer*), zdalne logowanie, pocztę elektroniczną. Warto podkreślić, że technologia WWW nie tworzy niezależnej sieci, a raczej jedynie stanowi jedno z wielu zastosowań rozproszonych korzystających z usług komunikacyjnych zapewnianych przez internet.
- Zastosowaniom rozproszonym internet zapewnia dwie usługi — **niezawodną usługę zorientowaną na połączenie** i **zawodną usługę bezpołączeniową**. Ogólnie mówiąc, pierwsza usługa gwarantuje, że dane przesyłane od nadawcy do odbiorcy ostatecznie zostaną do niego dostarczone w odpowiedniej kolejności i w całości. Z kolei druga usługa nie zapewnia, że dane zostaną dostarczone. Zwykle zastosowanie rozproszone korzysta tylko z jednej z tych dwóch usług.
- Aktualnie internet nie zapewnia usługi, która próbuje gwarantować *czas, jaki zajmie* dostarczenie danych od nadawcy do odbiorcy. Poza zwiększającą się szybkością przesyłania danych do dostawcy usług internetowych, obecnie nie można uzyskać lepszej usługi (na przykład ograniczającej opóźnienia), za którą się więcej zapłaci. Taki stan rzeczy niektórym osobom (zwłaszcza Amerykanom!) wydaje się być dziwny. W rozdziale 7. przyjrzymy się najnowszym badaniom internetowym, których celem jest zmiana tej sytuacji.

Drugi sposób opisania internetu, czyli w kategoriach usług, które zapewnia zastosowaniom rozproszonym, jest równie istotny. Cały czas postępy dokonywane w przypadku podstawowych komponentów internetu wynikają z potrzeb nowych zastosowań. W związku z tym ważne jest, aby wiedzieć, że internet stanowi infrastrukturę, w której nieustannie pojawiają się nowo projektowane i wdrażane zastosowania.

Właśnie zaprezentowaliśmy dwa opisy internetu. Pierwszy odnosi się do sprzętowych i programowych składników internetu, natomiast drugi do jego usług zapewnianych zastosowaniom rozproszonym. Jednak być może w dalszym ciągu Czytelnik

nie jest do końca pewien, czym jest internet. Co to jest przełączanie pakietów, protokoły TCP/IP i usługa zorientowana na połączenie? Czym są routery? Jakiego typu łącza komunikacyjne występują w internecie? Czym jest zastosowanie rozproszone? W jaki sposób do internetu można podłączyć toster lub czujnik meteorologiczny? Jeśli na chwilę obecną Czytelnik jest trochę przytłoczony tym wszystkim, nie ma powodu do obaw, ponieważ celem książki jest zaprezentowanie zarówno podstawowych składników internetu, jak i zasad, które decydują o jego funkcjonowaniu. W kolejnych podrozdziałach i rozdziałach zostaną omówione wszystkie wymienione zagadnienia i udzielone zostaną odpowiedzi na zadane pytania.

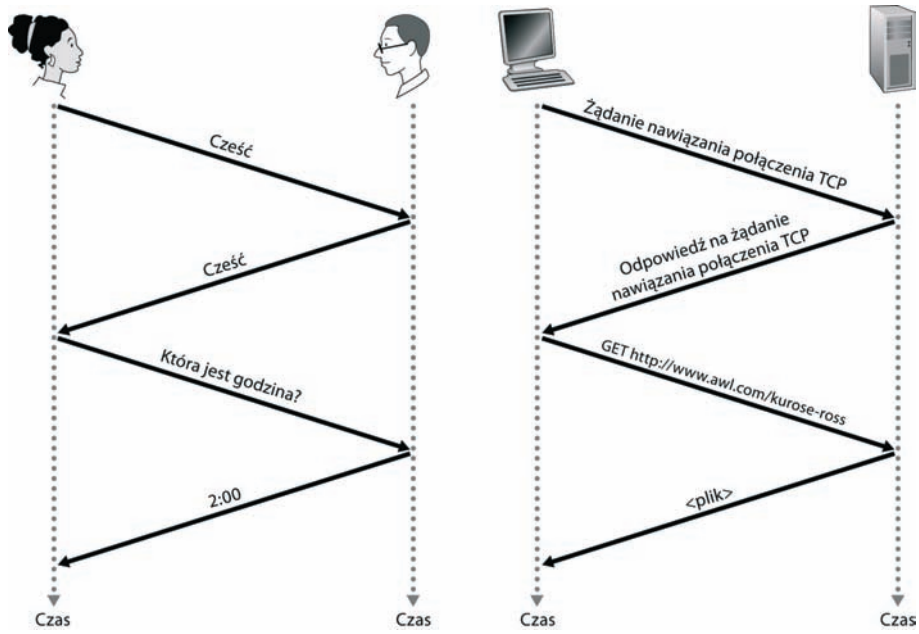
1.1.3. Czym jest protokół?

Gdy już trochę bliżej zapoznaliśmy się z internetem, pod uwagę weźmy kolejny istotny termin powiązany z sieciami komputerowymi, którym jest *protokół*. Czym jest protokół? Jaka jest jego rola? W jaki sposób stwierdzić, że ma się do czynienia z protokołem?

Analogia nawiązująca do komunikacji między ludźmi

Prawdopodobniej najprostszym sposobem pozwalającym zrozumieć protokół sieci komputerowej będzie zastosowanie w pierwszej kolejności analogii dotyczącej ludzi, którzy cały czas korzystają z protokołów. Zastanówmy się, co robimy, gdy chcemy zapytać kogoś o godzinę. Na rysunku 1.2 przedstawiono przebieg typowej wymiany. Ludzki protokół (lub przynajmniej dobre maniery) nakazuje, aby najpierw się przywitać (zwrot „Cześć" na rysunku 1.2), zanim rozpocznie się z kimś rozmowę. Typową odpowiedzią na zwrot „Cześć" jest takie samo słowo. Automatycznie przyjmuje się, że po usłyszeniu w odpowiedzi serdecznego zwrotu „Cześć" można kontynuować rozmowę i zapytać o godzinę. Inne różne odpowiedzi na początkowy zwrot „Cześć" (takie jak „Nie przeszkadzaj!", „Nie mówię po polsku" lub inne, których nie można tu przytoczyć) mogą wskazywać na to, że rozmówca nie ma ochoty na konwersację lub nie ma możliwości jej prowadzenia. W tym przypadku ludzki protokół nie umożliwi poznania godziny. Czasami na zadane pytanie nie otrzyma się żadnej odpowiedzi. W tego typu sytuacji zwykle rezygnuje się z pytania o godzinę. Warto zauważyć, że w przypadku ludzkiego protokołu *istnieją określone przekazywane komunikaty i działania podejmowane w odpowiedzi na otrzymane komunikaty lub wystąpienie innych zdarzeń (takich jak brak odpowiedzi w określonym przedziale czasu)*. Oczywiście przekazywane i otrzymywane komunikaty, a także czynności wykonywane w momencie transferowania komunikatów lub po wystąpieniu innych zdarzeń w przypadku ludzkiego protokołu odgrywają podstawową rolę. Jeśli ludzie posługują się różnymi protokołami, które nie są ze sobą zgodne (jeśli na przykład jedna osoba posiada maniery, a druga nie lub ktoś rozumie pojęcie czasu, natomiast ktoś inny nie), nie uzyska się żadnego pozytywnego efektu. To samo dotyczy sieci. W celu zrealizowania zadania dwie komunikujące się ze sobą jednostki muszą korzystać z tego samego protokołu.

Rozważmy drugą analogię dotyczącą komunikowania się ludzi. Załóżmy, że jesteśmy na zajęciach (na przykład z sieci komputerowych!). Nauczyciel w nieciekawy sposób mówi o protokołach, co powoduje, że traci się w tym orientację. Prowadzący zajęcia przerywa i pyta się, czy są jakieś pytania (komunikat jest wysyłany i odbierany przez wszystkich studentów, z wyjątkiem tych, którzy śpią). Jeden ze studentów podnosi



RYSUNEK 1.2. Porównanie ludzkiego protokołu i protokołu sieci komputerowej

rękę (przekazując automatycznie komunikat nauczycielowi). Prowadzący reaguje na to uśmiechem i mówi: „Tak, słucham" (przekazany komunikat zachęca studenta do zadania pytania). Nauczyciele *uwielbiają*, gdy zadaje im się pytania. Student zadaje pytanie (przekazuje komunikat nauczycielowi). Po usłyszeniu pytania (otrzymaniu komunikatu) nauczyciel udziela odpowiedzi na nie (przekazuje komunikat studentowi). Również w tym przypadku widać, że przekazywane i otrzymywane komunikaty, a także zestaw typowych działań podejmowanych w chwili wysyłania i odbierania wiadomości to rdzeń protokołu opartego na pytaniach i odpowiedziach.

Protokoły sieciowe

Protokół sieciowy przypomina ludzki protokół, z tą różnicą, że jednostki wymieniające się komunikatami i podejmujące działania są sprzętowymi lub programowymi składnikami określonego urządzenia (na przykład komputera, routera lub innego urządzenia sieciowego). Wszystkie operacje realizowane w internecie, które angażują dwie lub więcej zdalnych jednostek komunikacyjnych, są zarządzane przez protokół. Przykładowo, sprzętowe protokoły kart sieciowych dwóch połączonych ze sobą komputerów kontrolują przepływ bitów w kablu znajdującym się między dwoma interfejsami sieciowymi. Protokoły systemów końcowych kontrolujące przeciążenie decydują, z jaką szybkością pakiety będą transmitowane między nadawcą i odbiorcą. Protokoły routerów określają ścieżkę, jaką pakiet będzie podążał od miejsca źródłowego do docelowego. Protokoły są stosowane w każdym miejscu internetu. W związku z tym spora część książki jest poświęcona protokołom sieci komputerowych.

W ramach przykładu protokołu sieciowego, z którym Czytelnik prawdopodobnie jest zaznajomiony, zastanówmy się, co się stanie, gdy do serwera WWW wyśle się żądanie, czyli w oknie przeglądarki internetowej wprowadzi się adres URL witryny

WWW. Zostało to zilustrowane w prawej części rysunku 1.2. Najpierw komputer wysyła do serwera WWW żądanie nawiązania połączenia i czeka na odpowiedź. Po otrzymaniu żądania serwer WWW zwraca komunikat odpowiedzi na żądanie. Gdy komputer już „wie", że może żądać pobrania z serwera WWW strony internetowej, w komunikacie `GET` wysyła do niego jej nazwę. Na końcu serwer WWW zwraca komputerowi stronę internetową (plik).

Po przytoczeniu przykładów protokołów ludzkich i sieciowych można powiedzieć, że wymiana komunikatów i działania podejmowane w chwili wysyłania i odbierania komunikatów to kluczowe elementy definiujące protokół.

> ***Protokół*** *definiuje format i kolejność komunikatów wymienianych między dwoma lub większą liczbą komunikujących się jednostek, a także operacje wykonywane w momencie wysyłania i (lub) odbierania komunikatu bądź innego zdarzenia.*

Zasadniczo internet i sieci komputerowe intensywnie korzystają z protokołów. Różne protokoły są używane w celu zrealizowania odmiennych zadań komunikacyjnych. W trakcie lektury książki okaże się, że niektóre protokoły są proste i oczywiste, natomiast inne złożone i zaawansowane intelektualnie. Opanowanie sieci komputerowych jest równoznaczne ze zrozumieniem, jak i dlaczego są stosowane protokoły sieciowe.

1.2. Obrzeże sieci

W poprzednich punktach dokonaliśmy bardzo ogólnego przeglądu internetu i protokołów sieciowych. Teraz trochę bardziej szczegółowo przyjrzymy się składnikom sieci komputerowej, a w szczególności internetu. W tym podrozdziale zaczniemy omawiać obrzeże sieci i komponenty, z którymi jesteśmy najbardziej zaznajomieni, czyli codziennie używane komputery. W kolejnym podrozdziale przejdziemy z obrzeża sieci do jej rdzenia i zajmiemy się przełączaniem oraz routingiem wykorzystywanym w sieciach komputerowych. W podrozdziale 1.4 omówimy fizyczne łącza, które przenoszą sygnały przesyłane między komputerami i przełącznikami.

1.2.1. Systemy końcowe, klienty i serwery

W żargonie dotyczącym sieci komputerowych komputery podłączone do internetu często są nazywane **systemami końcowymi**. Wynika to stąd, że takie komputery są zlokalizowane na obrzeżu internetu (rysunek 1.3). Do internetowych systemów końcowych należy zaliczyć komputery stacjonarne (na przykład biurkowe komputery PC, a także stacje robocze z systemami Macintosh i UNIX), serwery (na przykład serwery WWW i pocztowe) i komputery mobilne (na przykład laptopy, cyfrowe asystenty osobiste PDA i telefony z bezprzewodowym połączeniem internetowym). Za systemy końcowe podłączane do internetu uważa się również coraz większą liczbę innych urządzeń, takich jak „cienkie" klienty, urządzenia domowe [Thinplanet 2002], przystawki internetowe do telewizora [Nesbitt 2002], cyfrowe aparaty, urządzenia stosowane w fabrykach i czujniki środowiskowe (poniższa ramka).

KĄCIK HISTORYCZNY

Oszałamiający zasób internetowych systemów końcowych

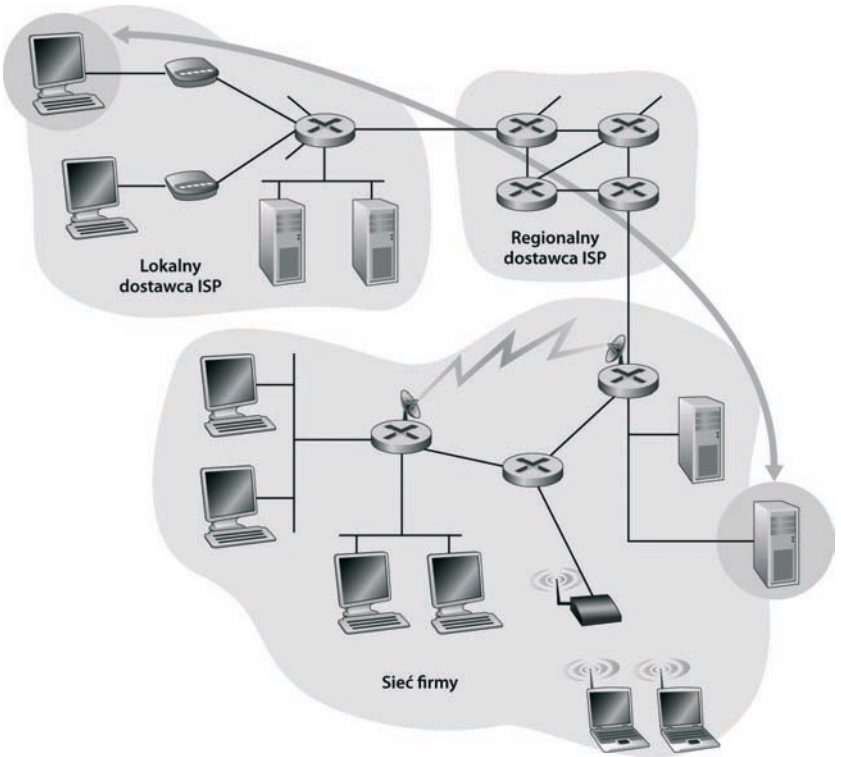
Nie tak dawno temu systemami końcowymi podłączonymi do internetu były przede wszystkim zwykłe komputery stacjonarne i wydajne serwery. Począwszy od końca lat 90., aż do chwili obecnej do internetu podłączono szeroką gamę interesujących urządzeń o coraz większej różnorodności. Urządzenia te cechują się wspólną potrzebą wysyłania cyfrowych danych do innych urządzeń i odbierania ich od nich. Biorąc pod uwagę wszechobecność internetu, jego dobrze zdefiniowane (poddane standaryzacji) protokoły i dostępność sprzętu powszechnego użytku przystosowanego do obsługi internetu, naturalnym jest zastosowanie technologii internetowych do połączenia ze sobą takich urządzeń.

Wydaje się, że część urządzeń powstała wyłącznie w celach rozrywkowych. Obsługująca protokół IP biurkowa przeglądarka slajdów [Ceiva 2004] pobiera cyfrowe obrazy ze zdalnego serwera i wyświetla je w czymś, co przypomina tradycyjną ramkę obrazu. Internetowy toster [BBC 2001] pobiera z serwera dane meteorologiczne i na porannym toście tworzy obraz prognozy pogody na aktualny dzień (na przykład kombinację chmur i słońca). Inne urządzenia oferują przydatne informacje. Kamery internetowe pokazują aktualną sytuację na drodze lub warunki pogodowe bądź monitorują interesujące nas miejsce. Zaprojektowano podłączone do internetu pralki, które mogą być zdalnie monitorowane przy użyciu przeglądarki internetowej. Tego typu pralki tworzą wiadomość pocztową, gdy pranie zostanie zakończone. Telefony komórkowe obsługujące protokół IP umożliwiają ich posiadaczom przeglądanie stron internetowych, wiadomości pocztowych i przesyłanych przez komunikatory. Nowej klasy sieciowe systemy pomiarowe mogą zrewolucjonizować to, w jaki sposób będziemy obserwować otoczenie i komunikować się z nim. Czujniki sieciowe umieszczone dookoła nas umożliwiają monitorowanie budynków, mostów i innych konstrukcji zbudowanych przez człowieka [Elgamal 2001], a także aktywności sejsmicznej [CISN 2004], środowiska naturalnego [Mainwaring 2002], ujścia rzek [Baptista 2003], funkcji biomedycznych [Schwiebert 2001] oraz wiatrów i zagrożeń meteorologicznych występujących w dolnej warstwie atmosfery [CASA 2004]. Zebrane informacje czujniki mogą udostępnić zdalnym użytkownikom. Jednostka Center for Embedded Networked Sensing podlegająca uniwersytetowi UCLA [CENS 2004] jest centrum naukowo-technologicznym organizacji NSF (*National Science Foundation*), którego celem jest wykorzystywanie sieci zintegrowanych czujników w krytycznych zastosowaniach naukowych i społecznych.

Systemy końcowe określa się też terminem *hosty* (*ang. gospodarze*), ponieważ goszczą (uruchamiają) aplikacje, takie jak przeglądarka internetowa, program serwera WWW, program pocztowy lub program serwera poczty. W książce zamiennie będą używane terminy host i system końcowy. Oznacza to, że są one równorzędne. Czasami hosty są dzielone na dwie kategorie — **klienty** i **serwery**. Nieoficjalnie za klienty uważa się stacjonarne i przenośne komputery PC, asystenty PDA itp. Z kolei serwerami są bardziej wydajne komputery, które przechowują i dystrybuują strony internetowe, strumieniowe dane wideo, wiadomości pocztowe itp.

W nawiązaniu do oprogramowania sieciowego należy wspomnieć o kolejnej definicji klienta i serwera, do której będziemy się odwoływać w różnych miejscach książki. **Aplikacja klienta** jest programem uruchomionym na jednym systemie końcowym, który żąda i korzysta z usługi świadczonej przez **aplikację serwera** uaktywnioną na drugim systemie końcowym. Taki model klient-serwer, który szczegółowo

omówiono w rozdziale 2., bez wątpienia jest najbardziej rozpowszechnioną strukturą aplikacji internetowych. Modelu tego używają serwery WWW, serwery pocztowe, serwery FTP, obsługujące zdalne logowania (na przykład serwer Telnet) i grupy dyskusyjne, a także wiele innych popularnych aplikacji. Ponieważ aplikacja klienta zwykle jest uruchamiana na jednym komputerze, natomiast aplikacja serwera na drugim, aplikacje internetowe klient-serwer z definicji są **aplikacjami rozproszonymi**. Aplikacje klienta i serwera komunikują się ze sobą przez wysyłanie komunikatów za pośrednictwem internetu. Na tym poziomie abstrakcji routery, łącza i inne podstawowe składniki internetu są jak „czarna skrzynka", która przesyła komunikaty między rozproszonymi i komunikującymi się składnikami aplikacji internetowej. Poziom ten przedstawiono na rysunku 1.3.



RYSUNEK 1.3. Interakcja systemów końcowych

Obecnie nie wszystkie aplikacje internetowe są złożone wyłącznie z aplikacji klienta, które komunikują się tylko z aplikacjami serwera. Przykładowo, w przypadku popularnych programów P2P służących do wymiany plików, takich jak KaZaA, aplikacja znajdująca się w systemie końcowym użytkownika jednocześnie pełni rolę aplikacji klienta i serwera. Program uaktywniony na węźle (komputer użytkownika) jest klientem, gdy żąda pliku od innego węzła. Z kolei program pełni rolę serwera, gdy do innego węzła wysyła plik. W telefonii internetowej dwie komunikujące się ze sobą strony są węzłami. W tym przypadku nie ma sensu określanie, która strona żąda usługi. W rozdziale 2. szczegółowo porównano model klient-serwer z architekturą aplikacji P2P.

1.2.2. Usługi zorientowane na połączenie i usługi bezpołączeniowe

Systemy końcowe wykorzystują internet do komunikowania się ze sobą. Dokładniej mówiąc, uruchomione na nich programy używają usług internetowych w celu przesyłania między sobą komunikatów. Łącza, routery i inne składniki internetu zapewniają środki umożliwiające transportowanie komunikatów między aplikacjami systemów końcowych. Jednak jakie są charakterystyki usług komunikacyjnych świadczonych przez internet podłączonym do niego systemom końcowym?

Sieci TCP/IP, a zwłaszcza internet, oferują aplikacjom systemów końcowych dwa typy usług — **usługę bezpołączeniową** i **usługę zorientowaną na połączenie**. Projektant piszący aplikację internetową (na przykład program pocztowy, aplikację przesyłającą pliki, aplikację dla serwera WWW lub program obsługujący telefonię internetową) musi utworzyć oprogramowanie, które będzie korzystało z jednego z dwóch wymienionych typów usług. Opiszemy teraz w skrócie oba rodzaje usług (bardziej szczegółowe omówienie usług można znaleźć w rozdziale 3., który poświęcono protokołom warstwy transportowej).

Usługa zorientowana na połączenie

Gdy aplikacja korzysta z usługi zorientowanej na połączenie, programy po stronie klienta i serwera (znajdujące się na różnych systemach końcowych) przed wysłaniem pakietów z faktycznymi danymi, które mają być przesłane, transferują między sobą pakiety sterujące. Ten tak zwany proces negocjowania (*ang. handshaking* to dosłownie uścisk dłoni) powiadamia klienta i serwer w celu umożliwienia im przygotowania się na konieczność nagłego odebrania pakietów danych. Po zakończeniu procesu negocjowania między dwoma systemami mówi się, że zostało nawiązane połączenie.

Godne uwagi jest to, że wstępnie wykonywany proces negocjowania przebiega podobnie jak w przypadku ludzkiego protokołu. Wymiana zwrotów „Cześć" widoczna na rysunku 1.2 jest przykładem procesu negocjowania stanowiącego część ludzkiego protokołu (nawet pomimo tego, że dwie osoby dosłownie nie ściskają sobie dłoni). W przypadku komunikacji z serwerem WWW zaprezentowanej również na rysunku 1.2 pierwsze dwa komunikaty są wymieniane w ramach procesu negocjacji. Kolejne dwa komunikaty — komunikat `GET` i komunikat odpowiedzi zawierający plik — uwzględniają faktyczne dane i są przesyłane tylko po nawiązaniu połączenia.

Dlaczego stosuje się termin *usługa zorientowana na połączenie*, a nie po prostu *usługa połączenia*? Wynika to z faktu, że systemy końcowe są połączone w bardzo swobodny sposób. W istocie tylko systemy końcowe są świadome nawiązanego połączenia. Internetowe przełączniki pakietów całkowicie nie zdają sobie sprawy z utworzonego połączenia. W rzeczywistości połączenie internetowe jest niczym więcej niż przydzielonymi buforami i zmiennymi stanu systemów końcowych. Pośredniczące przełączniki pakietów nie przechowują żadnych informacji na temat stanu połączenia.

Trzeba podkreślić, że choć internetowa usługa zorientowana na połączenie współpracuje z usługą niezawodnego transferu danych, kontroli przepływu i przeciążenia, w żadnym razie te trzy usługi nie stanową podstawowych składników usługi zorientowanej na połączenie. Różnego typu sieci komputerowe mogą oferować swoim aplikacjom usługę zorientowaną na połączenie bez uwzględniania jednej lub większej liczby wymienionych usług dodatkowych. Tak naprawdę, usługą zorientowaną na połączenie

jest dowolny protokół, który przed rozpoczęciem transferu danych między komunikującymi się jednostkami przeprowadzi proces negocjowania [Iren 1999].

Internetowa usługa zorientowana na połączenie współpracuje z kilkoma innymi usługami, takimi jak usługa niezawodnego transferu danych, kontroli przepływu i przeciążenia. Przez **niezawodny transfer danych** rozumiemy to, że aplikacja może liczyć, że połączenie umożliwi jej dostarczenie wszystkich danych bez błędów i w odpowiedniej kolejności. Niezawodność w internecie jest osiągana za pomocą potwierdzeń i retransmisji. Aby wstępnie zorientować się, w jaki sposób w internecie jest stosowana usługa niezawodnego transferu danych, weźmy pod uwagę aplikację, która nawiązała połączenie między dwoma systemami końcowymi A i B. Gdy system końcowy B odbierze pakiet od systemu A, odeśle potwierdzenie. Po otrzymaniu potwierdzenia system A będzie wiedział, że odpowiedni pakiet ostatecznie został odebrany. Jeśli system końcowy A nie dostanie potwierdzenia, przyjmie, że wysłany przez niego pakiet nie został odebrany przez system B i w związku z tym dokona retransmisji pakietu. **Kontrola przepływu** zapewnia, że żadna ze stron połączenia nie przeciąży drugiej strony przez zbyt szybkie wysłanie zbyt dużej liczby pakietów. W rozdziale 3. okaże się, że w internecie usługa kontroli przepływu jest implementowana przy użyciu nadawczych i odbiorczych buforów komunikujących się ze sobą systemów końcowych. Usługa **kontroli przeciążenia** pomaga uniknąć sytuacji, w której internet zostanie zupełnie zablokowany. Gdy przełącznik pakietów stanie się przeciążony, jego bufory mogą zostać przepełnione i może dojść do utraty pakietów. Jeżeli w takich okolicznościach każda para komunikujących się ze sobą systemów końcowych będzie nadal transferowała pakiety z maksymalną szybkością, dojdzie do zatoru i w efekcie niewiele pakietów zostanie dostarczonych do celu. W internecie unika się takiego problemu przez zmuszenie systemów końcowych do obniżenie szybkości, z jaką wysyłają pakiety w sieci podczas okresów przeciążenia. Gdy systemy końcowe przestają otrzymywać potwierdzenia powiązane z wysłanymi pakietami, wiedzą, że wystąpiło poważne przeciążenie.

Protokół **TCP** (*Transmission Control Protocol*) jest internetową usługą zorientowaną na połączenie. Pierwotna wersja protokołu została zdefiniowana w dokumencie RFC 793 [RFC 793]. Protokół TCP zapewnia aplikacji takie usługi jak niezawodny transport, kontrolę przepływu i przeciążenia. Protokół oferuje **poziom abstrakcji strumienia bajtów**, niezawodnie dostarczając strumień bajtów od nadawcy do odbiorcy. Godne uwagi jest to, że aplikacja musi jedynie korzystać z udostępnionych usług. Nie musi dbać o to, w *jaki* sposób protokół TCP zapewnia niezawodność, kontrolę przepływu lub przeciążenia. Oczywiście jesteśmy *bardzo* zainteresowani tym, jak protokół TCP implementuje te usługi. W związku z tym zostaną one szczegółowo omówione w rozdziale 3.

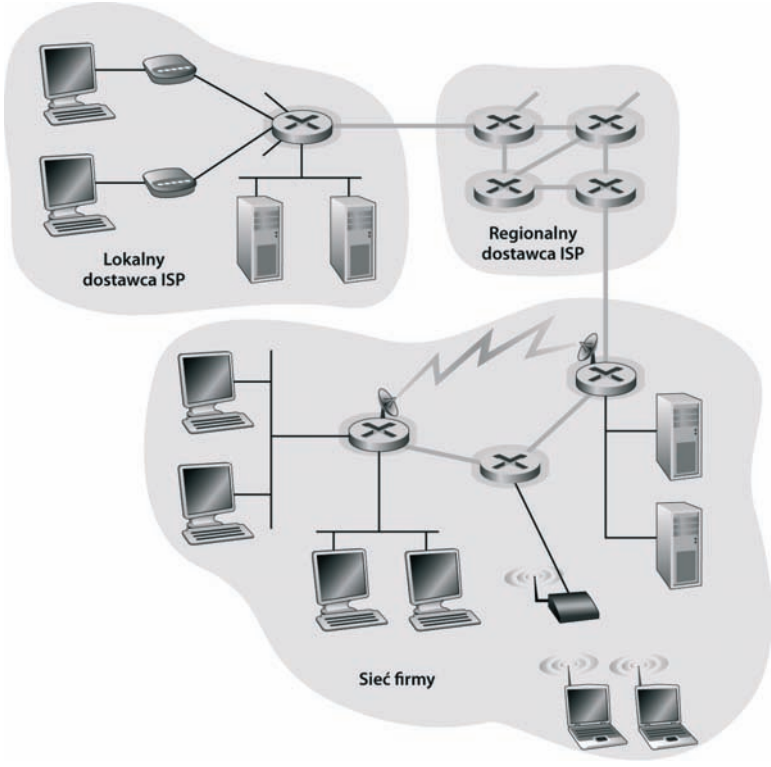
Usługa bezpołączeniowa

W przypadku internetowej usługi bezpołączeniowej nie jest przeprowadzany proces negocjowania. Gdy jedna część aplikacji chce wysłać pakiety do jej drugiej części, strona nadawcza po prostu rozpoczyna transmisję pakietów. Ponieważ przed zainicjowaniem transferu pakietów nie występuje proces negocjowania, dane mogą być szybciej dostarczone. To sprawia, że usługa bezpołączeniowa jest idealna w przypadku prostych zastosowań transakcyjnych. Jednak nie ma żadnej gwarancji niezawodnego transferu danych, dlatego nadawca nigdy nie ma całkowitej pewności, czy pakiety dotarły do miejsca przeznaczenia. Co więcej, internetowa usługa bezpołączeniowa nie zapewnia żadnej kontroli przepływu i przeciążenia. Tego typu usługą jest protokół **UDP** (*User Datagram Protocol*). Protokół zdefiniowano w dokumencie RFC 768.

Większość bardziej znanych internetowych zastosowań korzysta z protokołu TCP, będącego usługą zorientowaną na połączenie. Do zastosowań tych zalicza się usługa Telnet (obsługa zdalnego logowania), protokół SMTP (obsługa poczty elektronicznej), protokół FTP (przesyłanie plików) i protokół HTTP (witryny WWW). Niemniej jednak protokół UDP, czyli usługa bezpołączeniowa, jest wykorzystywany przez wiele zastosowań, takich jak spora liczba nowych technologii multimedialnych (na przykład telefonia internetowa i wideokonferencje).

1.3. Rdzeń sieci

Po omówieniu systemów końcowych i internetowego modelu usługi transportowej typu punkt-punkt w tym miejscu bardziej zagłębimy się w strukturę sieci. W kolejnym podrozdziale przyjrzymy się rdzeniowi sieci, czyli siatce routerów, które łączą ze sobą systemy końcowe podłączone do internetu. Na rysunku 1.4 grubymi cieniowanymi liniami wyróżniono rdzeń sieci.



RYSUNEK 1.4. Rdzeń sieci

1.3.1. Przełączanie obwodów i pakietów

W przypadku budowy rdzenia sieci obowiązują dwa podstawowe rozwiązania — **przełączanie obwodów** i **przełączanie pakietów**. W sieciach z przełączaniem obwodów zasoby wymagane przez ścieżkę (bufory, szybkość transmisji łącza) zapewniającą komunikację między systemami końcowymi są *rezerwowane* na czas trwania sesji. W sieciach z przełączaniem pakietów zasoby *nie są rezerwowane*. Komunikaty przesyłane w ramach sesji korzystają z zasobów na żądanie. W efekcie mogą być zmuszone do czekania w kolejce na uzyskanie dostępu do łącza komunikacyjnego. W celu przytoczenia prostej analogii rozważmy dwie restauracje, z których jedna wymaga rezerwacji, natomiast druga nie żąda ich, ani też ich nie przyjmuje. W przypadku pierwszej restauracji przed wyjściem z domu trzeba będzie zadzwonić. Jednak dzięki temu po pojawieniu się w restauracji w zasadzie będzie można od razu wezwać kelnera i zamówić danie. W przypadku drugiej restauracji nie musimy pamiętać o rezerwowaniu stolika. Jednak po przybyciu do niej możemy być zmuszeni do czekania na stolik, zanim będzie można przywołać kelnera.

Wszechobecne sieci telefoniczne są przykładem sieci z przełączaniem obwodów. Zastanówmy się nad tym, co się stanie, gdy jedna osoba będzie chciała wysłać dane (głos lub faks) do drugiej za pośrednictwem sieci telefonicznej. Zanim nadawca będzie w stanie wysłać informacje, w sieci musi zostać nawiązane połączenie między nim i odbiorcą. W przeciwieństwie do połączenia TCP, które omówiono w poprzednim punkcie, tu używane jest fizyczne połączenie, w przypadku którego przełączniki znajdujące się na ścieżce ustanowionej między nadawcą i odbiorcą przechowują dane dotyczące stanu połączenia. W żargonie związanym z sieciami telefonicznymi tego typu połączenie jest nazywane **obwodem**. Gdy w sieci zostanie utworzony obwód, na czas połączenia w łączach sieciowych jest rezerwowana stała szybkość transmisji. Ponieważ dla takiego połączenia między nadawcą i odbiorcą jest rezerwowane pasmo, nadawca może przesyłać dane do odbiorcy z *gwarantowaną* stałą szybkością.

Obecnie internet stanowi najważniejszą sieć z przełączaniem pakietów. Rozważmy, co się stanie, gdy jeden host będzie chciał za pośrednictwem internetu przesłać pakiet do innego hosta. Podobnie jak w przypadku przełączania obwodów pakiet jest przesyłany przy użyciu kilku łączy komunikacyjnych. Jednak w przypadku przełączania pakietów pakiet jest transferowany w sieci bez rezerwowania żadnego pasma. Jeśli jedno z łączy będzie przeciążone, ponieważ w tym samym czasie konieczne jest przesłanie za jego pośrednictwem innych pakietów, transmitowany pakiet będzie musiał poczekać w buforze znajdującym się po stronie nadawczej łącza transmisyjnego. W efekcie wystąpi opóźnienie. Choć w internecie są czynione *wszelkie starania*, aby pakiety dostarczać w odpowiednim czasie, nie ma na to żadnej gwarancji.

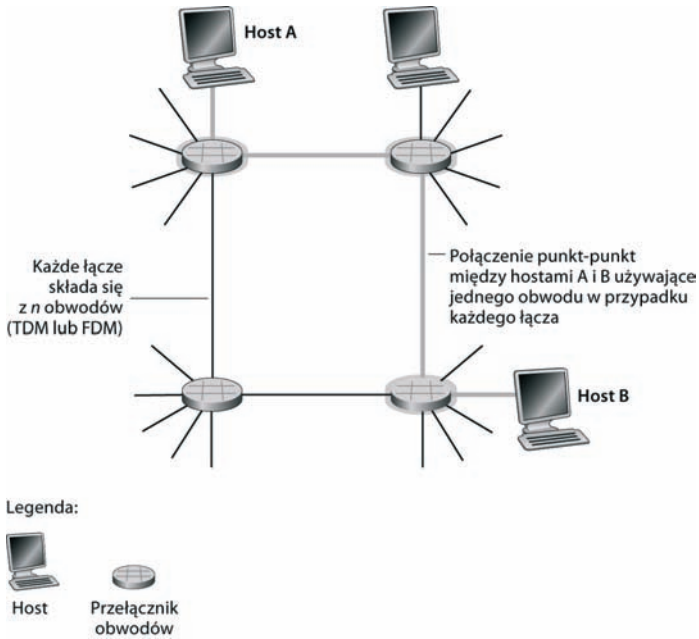
Nie wszystkie sieci telekomunikacyjne można w prosty sposób podzielić na tylko korzystające z przełączania obwodów i tylko używające przełączania pakietów. Niemniej jednak taka podstawowa klasyfikacja na sieci z przełączaniem obwodów i pakietów stanowi znakomity punkt wyjścia na drodze do zrozumienia technologii sieci telekomunikacyjnych.

Przełączanie obwodów

Książka jest poświęcona sieciom komputerowym, internetowi i przełączaniu pakietów, a nie sieciom telefonicznym i przełączaniu obwodów. Niemniej jednak ważne jest zrozumienie, dlaczego internet i inne sieci komputerowe korzystają z przełączania

pakietów, a nie z bardziej tradycyjnej technologii przełączania obwodów stosowanej w sieciach telefonicznych. Z tego powodu dokonamy teraz krótkiego przeglądu przełączania obwodów.

Na rysunku 1.5 zilustrowano sieć z przełączaniem obwodów. W przypadku tej sieci cztery przełączniki obwodów są ze sobą połączone za pomocą takiej samej liczby łączy. Każde łącze posiada *n* obwodów. W związku z tym każde łącze może obsługiwać *n* jednoczesnych połączeń. Hosty (na przykład komputery PC i stacje robocze) są bezpośrednio podłączone do jednego z przełączników. Gdy dwa hosty zamierzają się ze sobą komunikować, sieć ustanawia między nimi dedykowane **połączenie punkt-punkt**. Choć oczywiście możliwe jest realizowanie połączeń konferencyjnych między więcej niż dwoma urządzeniami, trzeba dążyć do prostoty. W tym celu załóżmy teraz, że w każdym połączeniu uczestniczą tylko dwa hosty. A zatem, aby host A mógł wysłać komunikaty do hosta B, sieć musi najpierw dla każdego z dwóch łączy zarezerwować jeden obwód. Ponieważ każde łącze posiada *n* obwodów, dla każdego łącza używanego przez połączenie punkt-punkt przypada część pasma łącza określona wartością *1/n* (dotyczy to czasu trwania połączenia).



RYSUNEK 1.5. Prosta sieć z przełączaniem obwodów złożona z czterech przełączników i czterech łączy

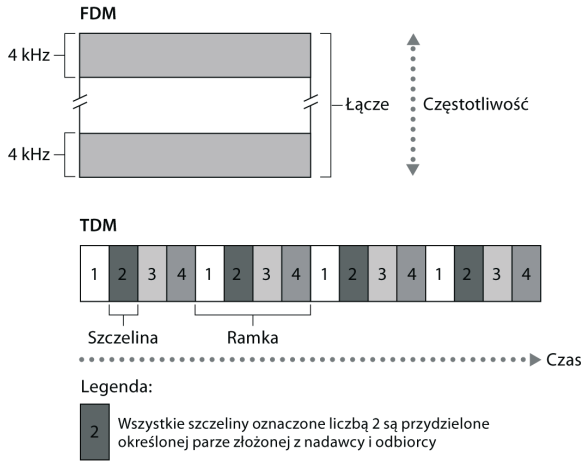
Multipleksowanie w sieciach z przełączaniem obwodów

Obwód w łączu jest implementowany za pomocą **multipleksowania z podziałem częstotliwości FDM** (*Frequency-Division Multiplexing*) lub **multipleksowania z podziałem czasu TDM** (*Time-Division Multiplexing*). W przypadku multipleksowania FDM widmo częstotliwości łącza jest dzielone między połączenia nawiązane za pośrednictwem tego łącza. Dokładniej mówiąc, łącze przydziela pasmo częstotliwości

każdemu połączeniu na czas jego trwania. W sieciach telefonicznych pasmo to zwykle ma szerokość wynoszącą 4 kHz (inaczej 4000 herców lub 4000 cykli na sekundę). Szerokość pasma jest też nazywana **przepustowością**. Stacje radiowe FM również korzystają z multipleksowania FDM w celu współużytkowania widma częstotliwości z przedziału od 88 do 108 MHz.

W przypadku łącza z multipleksowaniem TDM czas jest dzielony na ramki o stałej długości. Każda ramka składa się z niezmiennej liczby szczelin czasowych. Gdy w sieci jest ustanawiane połączenie za pośrednictwem łącza, połączeniu jest przydzielana jedna szczelina czasowa każdej ramki. Szczeliny te są przeznaczone wyłącznie na potrzeby tego konkretnego połączenia. Udostępniona szczelina czasowa każdej ramki służy do transmisji danych połączenia.

Na rysunku 1.6 zilustrowano użycie multipleksowania FDM i TDM dla określonego łącza sieciowego obsługującego maksymalnie cztery obwody. W przypadku multipleksowania FDM domena częstotliwości jest segmentowana na cztery pasma, z których każde ma szerokość 4 kHz. W przypadku multipleksowania TDM domena czasowa jest dzielona na ramki, z których każda posiada cztery szczeliny czasowe. Każdemu obwodowi jest przydzielana ta sama dedykowana szczelina w powtarzających się ramkach TDM. Gdy stosuje się multipleksowanie TDM, szybkość transmisji obwodu jest równa liczbie ramek przypadających na sekundę pomnożonej przez liczbę bitów w szczelinie. Jeśli na przykład łącze przesyła 8000 ramek w ciągu sekundy i każda szczelina zawiera 8 bitów, szybkość transmisji obwodu wyniesie 64 kb/s.



RYSUNEK 1.6. W przypadku multipleksowania FDM każdemu obwodowi nieprzerwanie jest przydzielana część przepustowości. W przypadku multipleksowania TDM każdy obwód okresowo podczas krótkich odcinków czasu (identyfikowanych przez szczeliny) dysponuje całą dostępną przepustowością

Zwolennicy przełączania pakietów zawsze uważali, że przełączanie obwodów jest marnotrawstwem, ponieważ podczas **okresów ciszy** dedykowane obwody znajdują się w stanie bezczynności. Przykładowo, gdy ktoś zakończy rozmawiać przez telefon, bezczynne zasoby sieciowe (pasma częstotliwości lub szczeliny w łączach tworzących ścieżkę połączenia) nie mogą być wykorzystane przez inne utworzone połączenia. Kolejnym przykładem tego, jak zasoby mogą być niedostatecznie użytkowane, jest

radiolog, który w celu uzyskania zdalnego dostępu do serii zdjęć rentgenowskich korzysta z sieci z przełączaniem obwodów. Radiolog nawiązuje połączenie, żąda pobrania zdjęcia, a następnie ogląda je i żąda kolejnego. Zasoby sieciowe są marnotrawione w czasie, gdy specjalista przegląda zdjęcia. Zwolennicy przełączania pakietów z przyjemnością zaznaczą, że definiowanie obwodów punkt-punkt i rezerwowanie dla nich przepustowości jest skomplikowane i wymaga zastosowania złożonego oprogramowania przetwarzającego sygnały, które koordynuje pracę przełączników znajdujących się na ścieżce biegnącej między dwoma węzłami.

Zanim zakończymy omawianie przełączania obwodów, przytoczymy liczbowy przykład, który powinien w większym stopniu przybliżyć to zagadnienie. Zastanówmy się, ile czasu zajmie wysłanie za pośrednictwem sieci z przełączaniem obwodów z hosta A do hosta B pliku liczącego 640 000 bitów. Załóżmy, że wszystkie łącza sieciowe korzystają z multipleksowania TDM oferującego 24 szczeliny i posiadają szybkość transmisji wynoszącą 1,536 Mb/s. Dodatkowo przyjmijmy, że 500 milisekund zajmuje utworzenie obwodu punkt-punkt, jeszcze zanim host A będzie mógł rozpocząć wysyłanie pliku. Ile czasu zajmie przesłanie pliku? Ponieważ szybkość transmisji każdego obwodu wynosi (1,536 Mb/s)/24 = 64 kb/s, transmisja pliku zajmie 10 sekund (640 000 bitów)/(64 kb/s)). Do 10 sekund należy dodać czas tworzenia obwodu, co powoduje, że czas wysyłania pliku zwiększa się do 10,5 sekundy. Warto zauważyć, że czas transmisji jest niezależny od liczby łączy. Czas ten wyniesie 10 sekund, zarówno gdy obwód punkt-punkt będzie korzystał z jednego, jak i stu łączy (rzeczywiste opóźnienie występujące między dwoma punktami uwzględnia też opóźnienie propagacji; więcej o tym można znaleźć w podrozdziale 1.6).

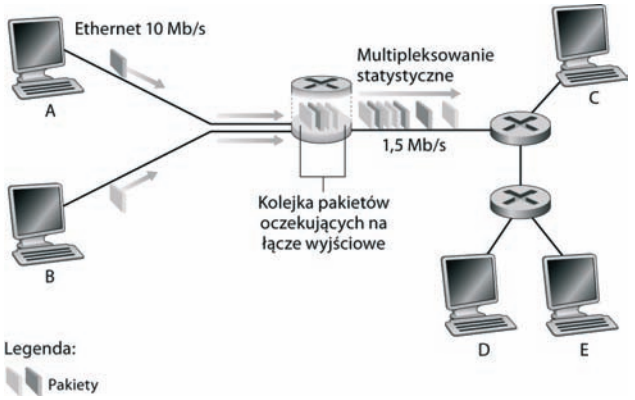
Przełączanie pakietów

W podrozdziale 1.1 wspomniano, że w celu zrealizowania zadania aplikacje wymieniają się **komunikatami**. Komunikaty mogą zawierać wszystko, czego zażąda twórca protokołu. Komunikaty mogą pełnić funkcję kontrolną (komunikaty „Cześć" zastosowane w przykładzie prezentującym komunikowanie się dwóch osób) lub przechowywać dane, takie jak wiadomość pocztowa, obraz JPEG lub plik audio MP3. W nowoczesnych sieciach komputerowych nadawca długie komunikaty dzieli na mniejsze porcje danych nazywane **pakietami**. Między nadawcą i odbiorcą każdy pakiet jest przesyłany przy użyciu łączy komunikacyjnych i **przełączników pakietów** (wśród nich dominują dwa typy — **routery** i przełączniki warstwy łącza). Pakiety są przesyłane przez każde łącze komunikacyjne z szybkością równą jego *maksymalnej* szybkości. Większość przełączników pakietów na wejściu łącza stosuje **transmisję buforowaną** (ang. *store-and-forward transmission*). Tego typu transmisja powoduje, że przełącznik musi odebrać cały pakiet, zanim będzie mógł rozpocząć wysyłanie jego pierwszego bitu do łącza wyjściowego. A zatem buforujące przełączniki pakietów na wejściu każdego łącza znajdującego się na trasie pokonywanej przez pakiet wprowadzają opóźnienie związane z transmisją buforowaną. Opóźnienie jest proporcjonalne do długości pakietu wyrażonej w bitach. Zwłaszcza gdy pakiet składa się z L bitów i zostanie przekazany do łącza wyjściowego o szybkości R b/s, opóźnienie buforowania występujące w przełączniku wyniesie L/R sekund.

Do każdego przełącznika pakietów jest podłączonych wiele łączy. Każdemu takiemu łączu przełącznik przydziela **bufor wyjściowy** (nazywany też **kolejką wyjściową**) przechowujący pakiety, które router prześle do łącza. W przypadku przełączania pakietów bufory wyjściowe odgrywają kluczową rolę. Jeśli odebrany pakiet musi zostać

przesłany łączem, które akurat jest zajęte transmisją innego pakietu, otrzymany pakiet będzie musiał poczekać w buforze wyjściowym. W związku z tym poza opóźnieniami związanymi z transmisją buforowaną na pakiet będą miały też wpływ **opóźnienia kolejkowania** w buforze wyjściowym. Tego typu opóźnienia są zmienne i zależą od poziomu przeciążenia sieci. Ponieważ pojemność bufora jest ograniczona, przy odbiorze pakietu może się okazać, że bufor jest całkowicie zapełniony innymi pakietami oczekującymi na transmisję. W tym przypadku dojdzie do **utraty pakietu**. Zostanie odrzucony otrzymany pakiet lub jeden z tych, które są już kolejkowane. Wracając do przytoczonej wcześniej analogii nawiązującej do restauracji, opóźnienie kolejkowania można porównać z czasem, przez jaki trzeba będzie oczekiwać przy barze na zwolnienie stolika. Utrata pakietu odpowiada stwierdzeniu przez kelnera, że trzeba będzie opuścić lokal, ponieważ już zbyt wiele osób oczekuje przy barze na stolik.

Na rysunku 1.7 przedstawiono prostą sieć z przełączaniem pakietów. Na tym i kolejnych rysunkach pakiety są reprezentowane przez trójwymiarowe płytki. Szerokość płytki identyfikuje liczbę bitów pakietu. Na rysunku 1.7 wszystkie pakiety są identycznej szerokości, a zatem również długości. Załóżmy, że hosty A i B wysyłają pakiety do hosta E. Hosty A i B najpierw swoje pakiety przesyłają łączami Ethernet o szybkości 10 Mb/s do pierwszego przełącznika pakietów. Przełącznik kieruje pakiety do łącza o szybkości 1,5 Mb/s. Jeśli szybkość, z jaką pakiety są transmitowane do przełącznika, przekracza jego szybkość przekazywania pakietów za pomocą łącza wyjściowego 1,5 Mb/s, dojdzie do przeciążenia, ponieważ przed przesłaniem pakietów łączem będą one kolejkowane w jego buforze wyjściowym.



RYSUNEK 1.7. Przełączanie pakietów

Zastanówmy się teraz, ile czasu zajmie wysłanie pakietu liczącego *L* bitów z jednego hosta do drugiego za pośrednictwem sieci z przełączaniem pakietów. Przyjmijmy, że między dwoma hostami znajduje się *Q* łączy, z których każde ma szybkość *R* b/s. Załóżmy, że opóźnienia kolejkowania i propagacji między dwoma punktami są nieznaczne, a ponadto, że nie jest przeprowadzane negocjowanie połączenia. Pakiet musi najpierw zostać przesłany pierwszym łączem wychodzącym z hosta A. Operacja zajmie *L/R* sekund. W dalszej kolejności pakiet musi zostać przesłany każdym pozostałym łączem, których jest *Q–1*. Oznacza to, że pakiet trzeba będzie buforować *Q–1* razy. W efekcie całkowite opóźnienie wyniesie *QL/R*.

Porównanie przełączania pakietów i obwodów oraz multipleksowanie statystyczne

Po omówieniu przełączania obwodów i pakietów dokonamy porównania obu rozwiązań. Krytycy przełączania pakietów często twierdzą, że ze względu na zmienne i nieprzewidywalne opóźnienia (chodzi głównie o opóźnienia kolejkowania) między dwoma punktami nie nadaje się ono do zastosowania w przypadku usług czasu rzeczywistego (na przykład rozmowy telefoniczne i wideokonferencje). Zwolennicy przełączania pakietów argumentują, że po pierwsze, w porównaniu z przełączaniem obwodów oferuje lepsze współużytkowanie przepustowości, a po drugie, przełączanie pakietów można wdrożyć prościej, efektywniej i taniej. Interesujące porównanie przełączania pakietów i przełączania obwodów znajduje się w zasobie [Molinero-Fernandez 2002]. Ogólnie mówiąc, osoby, które nie lubią dokonywać rezerwacji w restauracjach, nad przełączanie obwodów przedkładają przełączanie pakietów.

Dlaczego przełączanie pakietów jest efektywniejsze? Przytoczmy prosty przykład. Załóżmy, że użytkownicy korzystają ze wspólnego łącza 1 Mb/s. Ponadto przyjmijmy, że każdy użytkownik na zmianę ma do czynienia z okresem aktywności (generowane są dane ze stałą szybkością 100 kb/s) i nieaktywności (nie są generowane żadne dane). Załóżmy dodatkowo, że użytkownik jest aktywny tylko przez 10% czasu (przez pozostałe 90% tylko pije kawę). W przypadku przełączania obwodów przepustowość 100 kb/s musi być cały czas *zarezerwowana dla każdego* użytkownika. Przykładowo, gdy korzysta się z przełączania obwodów z multipleksowaniem TDM, po podzieleniu jednosekundowej ramki na 10 szczelin czasowych 100 ms, każdemu użytkownikowi zostanie przydzielona jedna szczelina ramki.

A zatem łącze może jednocześnie obsłużyć tylko 10 użytkowników (= 1Mb/s/ 100 kb/s). W przypadku przełączania pakietów prawdopodobieństwo, że określony użytkownik jest aktywny, wynosi 0,1, czyli 10%. Jeśli istnieje 35 użytkowników, prawdopodobieństwo, że w tym samym czasie aktywnych jest 11 lub więcej użytkowników w przybliżeniu wynosi 0,0004 (w problemie 8. zamieszczonym na końcu rozdziału wyjaśniono, w jaki sposób uzyskuje się wartość prawdopodobieństwa). Gdy jednocześnie aktywnych jest 10 lub mniej użytkowników (będzie tak przy prawdopodobieństwie 0,9996), całkowita szybkość, z jaką są wysyłane dane będzie mniejsza lub równa wyjściowej szybkości łącza wynoszącej 1 Mb/s. A zatem, jeśli aktywnych jest 10 lub mniej osób, wysyłane przez nie pakiety w zasadzie będą transmitowane przez łącze bez opóźnień, tak jak w przypadku przełączania obwodów. Gdy w danej chwili aktywnych będzie ponad 10 użytkowników, łączna szybkość, z jaką pakiety są wysyłane, przekroczy szybkość łącza. W efekcie zacznie się tworzyć kolejka wyjściowa (będzie to trwało do momentu, gdy całkowita szybkość wejściowa spadnie poniżej 1 Mb/s; wtedy kolejka zacznie się zmniejszać). Ponieważ w przytoczonym przykładzie prawdopodobieństwo tego, że jednocześnie aktywnych będzie ponad 10 użytkowników, jest znikome, przełączanie pakietów w zasadzie zapewnia taką samą wydajność, co przełączanie obwodów. *Jednak przełączanie pakietów dodatkowo umożliwia obsłużenie ponadtrzykrotnie większej liczby użytkowników*.

Omówmy teraz kolejny prosty przykład. Załóżmy, że istnieje 10 użytkowników i jeden z nich generuje nagle tysiąc pakietów 1000-bitowych, natomiast pozostałe osoby nie wykonują żadnej operacji powodującej tworzenie pakietów. W przypadku przełączania obwodów z multipleksowaniem TDM i ramki złożonej z 10 szczelin czasowych, z których każda zawiera 1000 bitów, w celu przesłania danych aktywny użytkownik będzie mógł zastosować tylko jedną szczelinę każdej ramki. Pozostałe

dziewięć szczelin czasowych znajdujących się w każdej ramce będzie bezczynnych. Upłynie 10 sekund, zanim zostaną przesłane wszystkie z miliona bitów wygenerowanych przez aktywnego użytkownika. Gdy użyje się przełączania pakietów, aktywny użytkownik może cały czas przesyłać swoje pakiety z pełną szybkością łącza wynoszącą 1 Mb/s. Wynika to stąd, że żaden inny użytkownik nie generuje pakietów, które muszą być multipleksowane z pakietami aktywnego użytkownika. W tym przypadku wszystkie jego dane zostaną przesłane w ciągu sekundy.

Powyższe przykłady demonstrują dwa przypadki, w których wydajność przełączania pakietów może być wyższa od oferowanej przez przełączanie obwodów. Ponadto przykłady uwidaczniają zasadniczą różnicę między dwoma odmianami współużytkowania przepustowości łącza przez wiele strumieni danych. Przełączanie obwodów odgórnie decyduje o wykorzystaniu łącza transmisyjnego niezależnie od zapotrzebowania. W efekcie marnowana jest część przydzielonego łącza, która w danej chwili nie jest używana. Z kolei przełączanie pakietów przydziela przepustowość łącza *na żądanie*. Pojemność łącza transmisyjnego w przypadku kolejnych pakietów będzie dzielona tylko przez użytkowników tworzących pakiety, które muszą zostać przesłane przy użyciu łącza. Tego typu przydzielanie zasobów na żądanie (zamiast odgórnie) czasami jest nazywane **multipleksowaniem statystycznym** zasobów.

Choć w obecnie istniejących sieciach telekomunikacyjnych rozpowszechnione jest zarówno przełączanie pakietów, jak i przełączanie obwodów, z pewnością obowiązuje tendencja zmierzania w stronę przełączania pakietów. Nawet wiele aktualnie stosowanych sieci telefonicznych z przełączaniem obwodów powoli jest migrowanych do technologii przełączania pakietów. Z przełączania pakietów sieci telefoniczne często korzystają zwłaszcza w celu obsługi kosztownej części połączenia międzykontynentalnego.

1.3.2. Sieci z przełączaniem pakietów — sieci datagramowe i sieci wirtualnych obwodów

Można wyróżnić dwie obszerne klasy sieci z przełączaniem pakietów — sieci datagramowe i sieci wirtualnych obwodów. Sieci różnią się tym, że ich przełączniki w celu przesłania pakietów w miejsce przeznaczenia korzystają z adresów docelowych lub z tak zwanych wartości wirtualnych obwodów. **Siecią datagramową** można nazwać każdą sieć, która przekazuje pakiety, używając adresów docelowych hostów. Ponieważ tak jest w przypadku routerów działających w internecie, jest on siecią datagramową. Za **sieć wirtualnych obwodów** można uznać każdą sieć, która pakiety przekazuje za pomocą wartości wirtualnych obwodów. X.25, Frame Relay i ATM (*Asynchronous Transfer Mode*) są przykładami technologii przełączania pakietów korzystających z wirtualnych obwodów. Choć różnica między stosowaniem adresów docelowych i wartości wirtualnych obwodów może wydawać się mniej istotna, wybór rozwiązania ma ogromny wpływ na to, w jaki sposób będą tworzone trasy i zarządzany routing (więcej o tym poniżej).

Sieci wirtualnych obwodów

Jak nazwa sugeruje, **wirtualny obwód** VC (*Virtual Circuit*) można potraktować jak wirtualne połączenie między hostem źródłowym i docelowym. Istotne jest to, że w utworzenie i utrzymanie takiego obwodu będą zaangażowane nie tylko dwa systemy końcowe,

ale też każdy przełącznik wchodzący w skład ścieżki obwodu VC zdefiniowanego między hostem źródłowym i docelowym. Gdy między hostami po raz pierwszy utworzy się obwód VC, zostanie mu przypisany **identyfikator wirtualnego obwodu** VC ID (*Virtual-Circuit Identifier*). Każdy pakiet przesyłany za pośrednictwem wirtualnego obwodu w swoim nagłówku będzie posiadał identyfikator. Każdy przełącznik pakietów dysponuje tabelą odwzorowującą identyfikatory VC ID na łącza wyjściowe. Po odebraniu pakietu przełącznik sprawdzi jego identyfikator, a następnie przejrzy indeks swojej tabeli i przekaże pakiet do wyznaczonego łącza wyjściowego. Warto zauważyć, że w przypadku wirtualnych obwodów ich punkty źródłowe i docelowe są jedynie pośrednio określane przez identyfikator VC ID. W celu wykonywania przełączania pakietów nie są potrzebne rzeczywiste adresy źródłowego i docelowego systemu końcowego. Oznacza to, że przełączanie może być szybko realizowane (przez wyszukiwanie identyfikatora VC ID otrzymanego pakietu w niewielkiej tabeli translacji wirtualnych obwodów zamiast szukania adresu docelowego w potencjalnie dużej przestrzeni adresowej).

Jak wyżej wspomniano, przełącznik sieci wirtualnych obwodów utrzymuje **informacje dotyczące stanu** jego połączeń wyjściowych. Dokładniej mówiąc, każdorazowo po nawiązaniu nowego połączenia za pośrednictwem przełącznika w jego tabeli translacji dla połączenia musi zostać umieszczony wpis. Ponadto każdorazowo po zakończeniu połączenia powiązany z nim wpis musi z tabeli zostać usunięty. Jeśli nawet nie występuje translacja identyfikatora VC ID, nadal jest konieczne utrzymywanie informacji na temat stanu, które kojarzą wartości wirtualnych obwodów z wartościami interfejsów wyjściowych. To, czy przełącznik pakietów będzie dla każdego nawiązanego połączenia utrzymywał takie informacje czy nie, jest bardzo istotną kwestią, do której niedługo powrócimy.

Sieci datagramowe

Pod wieloma względami sieci datagramowe można przyrównać do poczty. Gdy nadawca wysyła list w określone miejsce docelowe, umieszcza go w kopercie, na której podaje adres przeznaczenia. Adres ma strukturę hierarchiczną. Przykładowo, listy kierowane do miejsca znajdującego się w Stanach Zjednoczonych będą miały adres składający się z nazwy państwa (USA), stanu (np. Pennsylvania), miasta (np. Philadelphia), ulicy (np. Walnut Street) i numeru domu (np. 421). Aby list dostarczyć w miejsce przeznaczenia, poczta posłuży się adresem podanym na kopercie. Jeśli na przykład list został wysłany z Francji, miejscowy urząd pocztowy skieruje najpierw list do centrum pocztowego zlokalizowanego w Stanach Zjednoczonych. Stamtąd list zostanie przekazany do centrum pocztowego znajdującego się w Pensylwanii. Na końcu pracownik poczty stanowej dostarczy list w jego ostatecznie miejsce przeznaczenia.

Każdy pakiet przesyłany w sieci datagramowej w swoim nagłówku posiada adres docelowy. Podobnie jak w przypadku adresów pocztowych adres ten ma hierarchiczną strukturę. Gdy pakiet zostanie odebrany przez przełącznik pakietów znajdujący się w sieci, ten sprawdzi część pakietu zawierającą docelowy adres, a następnie przekaże pakiet do sąsiedniego przełącznika. Dokładniej mówiąc, każdy przełącznik pakietów korzysta z tabeli przekierowań, która odwzorowuje adresy docelowe (lub ich fragmenty) na łącze wyjściowe. Po otrzymaniu pakietu przełącznik sprawdza jego adres, a następnie za jego pomocą przeszukuje swoją tabelę, aby zlokalizować odpowiednie łącze wyjściowe. Gdy to nastąpi, przełącznik skieruje pakiet do tego łącza.

Proces routingu między dwoma punktami można przyrównać do kierowcy, który zamiast korzystać z mapy, woli zapytać kogoś o drogę. Dla przykładu załóżmy, że Joe jedzie z miasta Philadelphia do Orlando w stanie Florida, z zamiarem zlokalizowania

domu o numerze 156 znajdującego się przy ulicy Lakeside Drive. Joe najpierw udaje się na pobliską stację benzynową, aby dowiedzieć się, jak dojechać do wspomnianej ulicy. Pracownik stacji zwraca uwagę na część adresu dotyczącą stanu Florida i mówi Joemu, że musi wjechać na międzystanową autostradę I-95 South, od której prowadzi zjazd do kolejnej stacji. Ponadto radzi Joemu, aby po wjechaniu do stanu Florida zapytał kogoś o drogę. Joe jedzie autostradą I-95 South aż do chwili dotarcia do Jacksonville w stanie Florida. Tam pyta o drogę pracownika kolejnej stacji benzynowej. Ten przygląda się części adresu dotyczącego Orlando i mówi Joemu, że powinien dalej jechać autostradą I-95 aż do Daytona Beach i tam zapytać kogoś o drogę. W Daytona Beach kolejny pracownik stacji patrzy na część adresu związaną z Orlando i informuje Joego, że autostradą I-4 dojedzie bezpośrednio do tego miasta. Joe rusza autostradą I-4 i zjeżdża z niej do Orlando. Udaje się na następną stację benzynową. Tym razem jej pracownik zwraca uwagę na część adresu z ulicą Lakeside Drive i pokazuje Joemu drogę, którą musi pojechać, aby się tam dostać. Po dotarciu na ulicę Lakeside Drive Joe pyta napotkane dziecko na rowerze, gdzie znajduje się szukany dom. Dziecko przygląda się części adresu zawierającej numer 156 i wskazuje dom. Joe wreszcie lokalizuje cel swojej podróży.

Przekazywanie pakietów w sieciach datagramowych zostanie w książce bardzo dokładnie omówione. Jednak teraz wspomnimy tylko, że w porównaniu z sieciami wirtualnych obwodów *sieci datagramowe nie utrzymują w swoich przełącznikach informacji dotyczących stanu połączeń*. W rzeczywistości przełącznik pakietów znajdujący się w czystej sieci datagramowej jest zupełnie „nieświadom" wszelkich danych, które mogą być przez niego przesyłane. Przełącznik zlokalizowany w takiej sieci podejmuje decyzje dotyczące przekazywania na podstawie adresu docelowego pakietu, a nie połączenia, w ramach którego pakiet jest przesyłany. Ponieważ sieci wirtualnych obwodów muszą utrzymywać informacje (które muszą być zapisywane i usuwane odpowiednio po pojawieniu się i zniknięciu wirtualnego obwodu, a także czyszczone, gdy wirtualny obwód nieprawidłowo zakończy działanie) na temat stanu połączeń, wymagają w tym celu potencjalnie złożonych protokołów, których nie spotyka się w sieciach datagramowych.

Jak można sprawdzić, jaką trasę pokonują w internecie pakiety między dwoma punktami? Proponujemy poświęcić trochę czasu na zapoznanie się z narzędziem *Traceroute* dostępnym pod adresem *http://www.traceroute.org* (więcej na jego temat można znaleźć w podrozdziale 1.6).

Systematyka sieci

Do tej pory zapoznaliśmy się z kilkoma ważnymi zagadnieniami sieciowymi, takimi jak przełączanie obwodów, przełączanie pakietów, wirtualne obwody, usługa bezpołączeniowa i usługa zorientowana na połączenie. Jak to wszystko jest do siebie dopasowane?

Upraszczając wszystko, w sieci telekomunikacyjnej jest stosowane przełączanie obwodów lub pakietów (rysunek 1.8). Łącze sieci z przełączaniem obwodów może wykorzystywać multipleksowanie FDM lub TDM. Sieci z przełączaniem pakietów są sieciami wirtualnych obwodów lub sieciami datagramowymi. Przełączniki sieci wirtualnych obwodów przekazują pakiety, używając ich wartości wirtualnych obwodów, a także utrzymują informacje dotyczące stanu połączenia. Przełączniki znajdujące się w sieci datagramowej przekazują pakiety, korzystając z adresów docelowych pakietów, i nie przechowują informacji na temat stanu połączenia.

RYSUNEK 1.8. Systematyka sieci telekomunikacyjnych

1.4. Sieci dostępowe i fizyczne nośniki

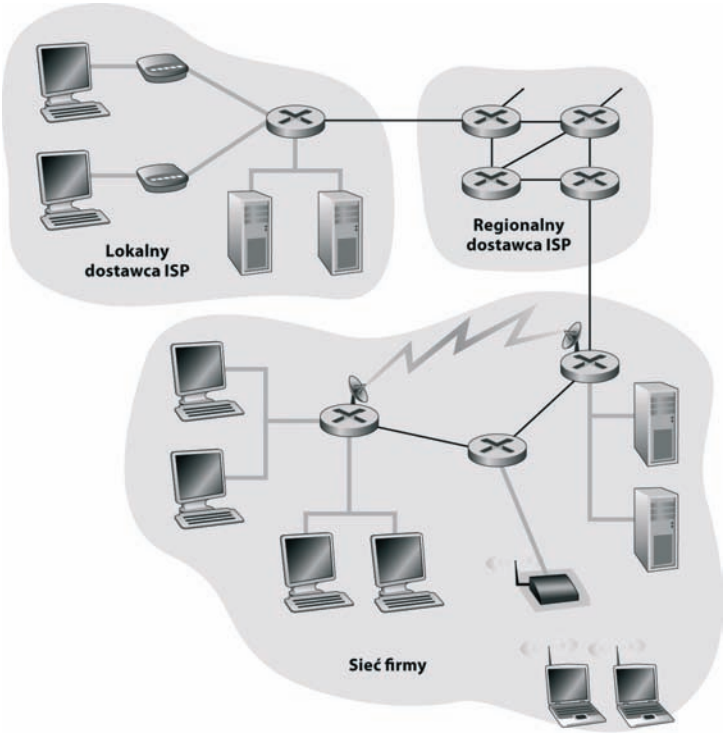
W podrozdziałach 1.2 i 1.3 omówiliśmy role pełnione przez systemy końcowe i przełączniki pakietów znajdujące się w sieci komputerowej. W niniejszym podrozdziale przyjrzymy się **sieciom dostępowym**, czyli fizycznym łączom łączącym system końcowy z jego **routerem brzegowym**. Jest to pierwszy router na ścieżce poprowadzonej od jednego systemu końcowego do dowolnego innego znacznie oddalonego. Sieć dostępowa zapewnia infrastrukturę łączącą tak zwane lokalizacje klientów z resztą sieci. Na rysunku 1.9 przedstawiono kilka typów łączy dostępowych poprowadzonych między systemem końcowym i routerem brzegowym. Łącza wyróżniono grubymi cieniowanymi liniami. Ponieważ technologie sieci dostępowych są ściśle powiązane z technologiami fizycznych nośników (światłowód, kabel koncentryczny, skrętka telefoniczna, widmo radiowe), oba zagadnienia omówiono w tym samym podrozdziale.

1.4.1. Sieci dostępowe

Ogólnie sieci dostępowe można podzielić na trzy następujące kategorie:

- **sieci dostępowe prywatnych użytkowników —** przyłączają do sieci domowe systemy końcowe;
- **sieci dostępowe firm —** przyłączają do sieci systemy końcowe znajdujące się w budynku firmy lub ośrodka edukacyjnego;
- **bezprzewodowe sieci dostępowe —** przyłączają do sieci systemy końcowe, które często są systemami mobilnymi.

Podane kategorie nie są ostateczne. Przykładowo, systemy końcowe określonej firmy mogą korzystać z technologii dostępowej, którą zaliczymy do pierwszej wymienionej kategorii. Poniższe omówienie dotyczy ogólnych przypadków zastosowań.



RYSUNEK 1.9. Sieci dostępowe

Sieci dostępowe prywatnych użytkowników

Sieci dostępowe prywatnych użytkowników łączą z routerem brzegowym system końcowy znajdujący się w domu (choć zwykle jest to komputer PC, coraz częściej podłącza się całą sieć domową) użytkownika. Jednym z wariantów dostępu uzyskiwanego z domu jest zastosowanie **modemu telefonicznego**, który za pośrednictwem zwykłej analogowej linii telefonicznej łączy się z lokalnym dostawcą ISP, takim jak America Online. Modem zamienia cyfrowe dane wysyłane przez komputer PC do formatu analogowego umożliwiającego transmisję za pomocą analogowej linii telefonicznej. Tego typu linia jest wykonana z miedzianej skrętki. Ta sama linia jest używana do przeprowadzania zwykłych rozmów telefonicznych (skrętka zostanie omówiona w dalszej części niniejszego podrozdziału). Po drugiej stronie linii telefonicznej znajduje się modem dostawcy ISP, który analogowe sygnały z powrotem zamienia do postaci cyfrowej, aby możliwe było przekazanie danych routerowi. A zatem sieć dostępowa po prostu składa się z pary modemów, między którymi znajduje się linia telefoniczna punkt-punkt. Obecnie spotykane modemy pozwalają na uzyskanie dostępu telefonicznego z maksymalną szybkością wynoszącą 56 kb/s. Jednak ze względu na kiepską jakość skrętki linii poprowadzonych między wieloma domami i dostawcami ISP spora część użytkowników może łączyć się z rzeczywistą szybkością znacznie mniejszą od 56 kb/s.

Wielu prywatnych użytkowników szybkość 56 kb/s oferowaną przez modemy telefoniczne uważa za wyjątkowo niską. Przykładowo, pobranie za pomocą modemu o takiej szybkości jednego trzyminutowego utworu zapisanego w formacie MP3 w przybliżeniu zajmie 8 minut. Ponadto dostęp modemowy wiąże się z zajmowaniem zwykłej

linii telefonicznej użytkownika. Gdy osoba prywatna przegląda strony internetowe, korzystając z modemu, nie może w tym czasie odbierać rozmów telefonicznych ani też sama dzwonić. Na szczęście nowe technologie dostępu szerokopasmowego zapewniają prywatnym użytkownikom wyższe szybkości transferu. Poza tym oferują środki umożliwiające jednoczesne łączenie się z internetem i rozmawianie przez telefon. Można wyróżnić dwa powszechnie stosowane typy dostępu szerokopasmowego przeznaczonego dla użytkowników domowych — **DSL** (*Digital Subscriber Line*) [DSL 2004] i **HFC** (*Hybrid Fiber-Coaxial Cable*) [Cable Labs 2004].

W 2003 r. dostęp szerokopasmowy był znacznie mniej rozpowszechniony od dostępu uzyskiwanego za pomocą modemu telefonicznego o szybkości 56 kb/s. Liczba łączy szerokopasmowych przypadających na 100-osobową populację w przybliżeniu wynosiła 23 osoby w Korei Południowej, 13 osób w Kanadzie i 7 osób w Stanach Zjednoczonych. W większości krajów europejskich liczba ta nie przekraczała 10 osób [Point Topic 2003]. Jednak technologie DSL i HFC są coraz szybciej wdrażane na całym świecie. Generalnie technologia HFC jest bardziej rozpowszechniona w Stanach Zjednoczonych, natomiast technologia DSL w Europie i Azji.

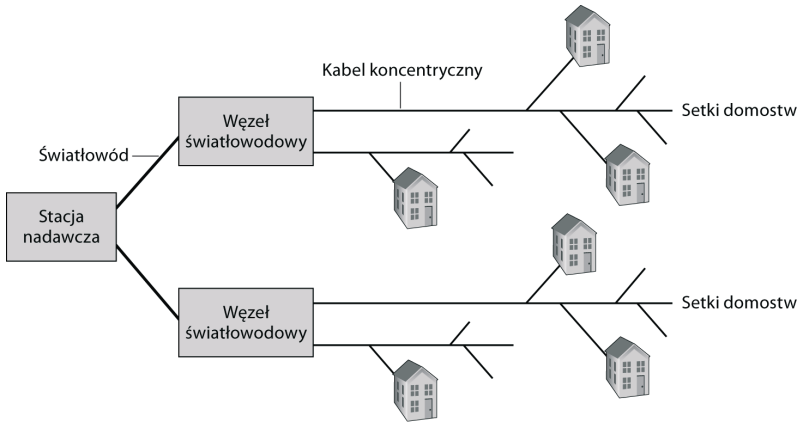
Dostęp za pośrednictwem urządzenia DSL zwykle jest oferowany przez operatora telekomunikacyjnego (na przykład Verizon lub France Telecom), a czasami we współpracy z niezależnym dostawcą usług internetowych. DSL, pod względem pojęciowym podobny do technologii modemów telefonicznych, jest nową technologią modemową, która ponownie wykorzystuje istniejące linie telefoniczne wykonane ze skrętki. Przez zmniejszenie odległości między użytkownikiem i modemem dostawcy ISP technologia DSL pozwala na wysyłanie i odbieranie danych ze znacznie większymi szybkościami. Szybkość transmisji danych w obu kierunkach jest zwykle asymetryczna. Dane są szybciej pobierane z routera dostawcy ISP do komputera użytkownika niż odwrotnie. Zastosowanie asymetrii w przypadku szybkości transferu danych wynika z przeświadczenia, że użytkownik domowy bardziej będzie odbiorcą danych niż ich nadawcą. W teorii technologia DSL jest w stanie zapewnić szybkości przekraczające 10 Mb/s (kierunek od dostawcy ISP do użytkownika) i 1 Mb/s (od użytkownika do dostawcy ISP). Jednak w praktyce szybkości oferowane przez dostawców usługi DSL są znacznie mniejsze. W 2004 r. typowa szybkość pobierania zawierała się w przedziale od 1 do 2 Mb/s, natomiast szybkość wysyłania wynosiła kilkaset kb/s.

Technologia DSL stosuje multipleksowanie z podziałem częstotliwości omówione w poprzednim podrozdziale. Technologia ta dzieli łącze komunikacyjne między domem i dostawcą ISP na trzy następujące nienachodzące na siebie pasma częstotliwości:

- pasmo o częstotliwości z przedziału od 50 kHz do 1 MHz — kanał pobierania o dużej szybkości;
- pasmo o częstotliwości z przedziału od 4 do 50 kHz — kanał wysyłania o średniej szybkości;
- pasmo o częstotliwości z przedziału od 0 do 4 kHz — zwykły dwukierunkowy kanał telefoniczny.

Rzeczywista szybkość pobierania i wysyłania dostępna dla użytkownika jest między innymi funkcją odległości między modemem domowym i modemem dostawcy ISP, grubości skrętki linii i poziomu zakłóceń elektrycznych. W przeciwieństwie do modemów telefonicznych modemy DSL zostały przez inżynierów zaprojektowane wyłącznie z myślą o niewielkich dystansach dzielących modemy domowe i modemy dostawcy ISP. Dzięki temu możliwe jest osiągnięcie znacznie większych szybkości transmisji niż w przypadku dostępu telefonicznego.

Modemy DSL i telefoniczne używają zwykłych linii telefonicznych, natomiast sieci dostępowe HFC stanowią rozszerzenie istniejącej sieci obsługiwanej przez operatorów telewizji kablowej. W tradycyjnym systemie telewizji kablowej stacja nadawcza przesyła sygnał do użytkowników za pośrednictwem sieci dystrybucyjnej złożonej z okablowania koncentrycznego i wzmacniaczy. Jak widać na rysunku 1.10, światłowody łączą stację nadawczą z sąsiadującymi z nią węzłami, od których do poszczególnych domów i mieszkań jest poprowadzony zwykły kabel koncentryczny. Każdy węzeł zwykle obsługuje od 500 do 5000 domostw.



RYSUNEK 1.10. Hybrydowa światłowodowo-koncentryczna sieć dostępowa

Podobnie jak w przypadku technologii DSL, technologia HFC wymaga użycia specjalnych modemów nazywanych **modemami kablowymi**. Firmy oferujące dostęp do internetu za pośrednictwem sieci kablowej wymagają od swoich klientów nabycia lub wydzierżawienia modemu. Zwykle modem kablowy jest zewnętrznym urządzeniem połączonym z domowym komputerem PC za pomocą portu Ethernet 10-BaseT (Ethernet bardziej szczegółowo omówiono w rozdziale 5.). Tego typu modemy dzielą sieć HFC na dwa kanały — pobierania i wysyłania. Tak jak w przypadku technologii DSL, kanał pobierania zwykle oferuje większą szybkość transmisji niż kanał wysyłania.

Ważną cechą technologii HFC jest to, że jest to wspólny nośnik transmisyjny. Dokładniej mówiąc, każdy pakiet wysłany przez stację nadawczą jest przesyłany do każdego domu kanałem pobierania wszystkich łączy. Ponadto każdy pakiet wysłany przez domowy komputer jest transmitowany kanałem wysyłania do stacji nadawczej. Z tego powodu, gdy jednocześnie kilku użytkowników będzie pobierało pliki MP3 za pośrednictwem kanału pobierania, rzeczywista szybkość przesyłania plików u każdego z nich będzie znacznie mniejsza od oferowanej. Z kolei gdy aktywnych będzie tylko kilku użytkowników, którzy będą przeglądali strony internetowe, każdy z nich może wyświetlać dane z pełną szybkością pobierania. Wynika to stąd, że użytkownicy rzadko w tym samym czasie będą żądali pobrania stron internetowych. Ponieważ kanał wysyłania też jest wspólny, niezbędne jest użycie protokołu rozproszonego wielodźródłowego dostępu, który będzie koordynował transmisje i zapobiegał kolizjom (zagadnieniu kolizji bardziej dokładnie przyjrzymy się w rozdziale 5. podczas omawiania Ethernetu). Zwolennicy technologii DSL szybko zauważą, że oferuje ona połączenie punkt-punkt nawiązywane między domowym komputerem i dostawcą ISP. W związku z tym w przypadku tej technologii cała dostępna szybkość transmisji jest dedykowana, a nie

współużytkowana. Z kolei obrońcy technologii HFC argumentują, że sieć kablowa o rozsądnej wielkości zapewni większe szybkości transmisji niż technologia DSL. Nie ulega wątpliwości, że rozpoczęła się rywalizacja między firmami świadczącymi prywatnym użytkownikom bardzo szybki dostęp do sieci oparty na technologii DSL lub HFC.

Jedną z atrakcyjnych funkcji technologii DSL i HFC jest **ciągła dostępność** usług. Oznacza to, że użytkownik może pozostawić włączony komputer, który cały czas będzie połączony z dostawcą ISP, i jednocześnie odbierać, a także wykonywać zwykłe połączenia telefoniczne.

Sieci dostępowe firm

Sieć lokalna LAN (*Local Area Network*) znajdująca się w budynkach firm i kampusów uniwersyteckich zwykle służy do łączenia systemów końcowych z routerem brzegowym. Jak się okaże w rozdziale 5., istnieje wiele typów sieci lokalnych. Jednak jak dotąd w sieciach firm najbardziej rozpowszechnionym rozwiązaniem dostępowym jest technologia Ethernet. Ethernet oferuje szybkość 10 lub 100 Mb/s (obecnie nawet 1 lub 10 Gb/s). W celu połączenia określonej liczby systemów końcowych ze sobą i z routerem brzegowym technologia ta korzysta z miedzianej skrętki lub kabla koncentrycznego. Router jest odpowiedzialny za trasowanie pakietów, których adres docelowy jest zlokalizowany poza siecią lokalną. Podobnie jak technologia HFC Ethernet korzysta ze wspólnego nośnika. Oznacza to, że użytkownicy końcowi dzielą między sobą szybkość transmisji oferowaną przez sieć lokalną. Nie tak dawno temu technologię Ethernet ze wspólnym nośnikiem zastąpiono jej wersją przełączaną. Ethernet z przełączaniem stosuje wiele segmentów wykonanych przy użyciu skrętki, które są podłączone do przełącznika umożliwiającego jednoczesne zaoferowanie różnym użytkownikom tej samej sieci lokalnej jej pełnej szybkości transmisji. Technologia Ethernet ze wspólnym nośnikiem i w wersji przełączanej zostanie szczegółowo omówiona w rozdziale 5.

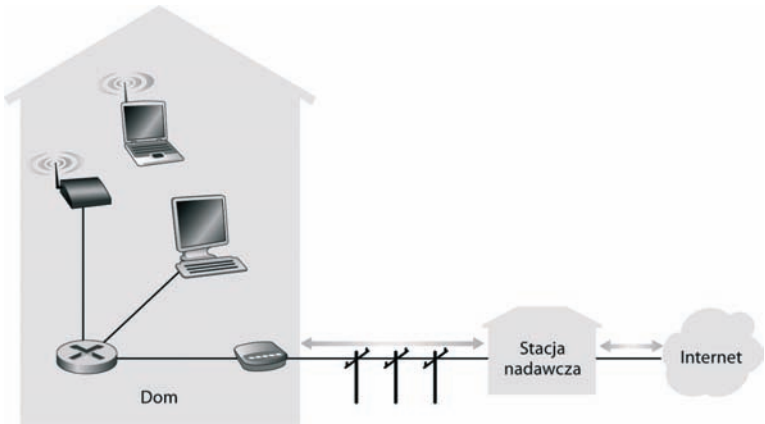
Bezprzewodowe sieci dostępowe

Towarzysząca trwającej właśnie rewolucji internetowej rewolucja technologii bezprzewodowej też ma znaczący wpływ na to, w jaki sposób ludzie pracują i żyją. Obecnie w Europie więcej osób posiada telefon komórkowy niż komputer PC lub samochód. Technologia bezprzewodowa cały czas zyskuje na popularności. Wielu analityków przewiduje, że bezprzewodowe (i często mobilne) urządzenia kieszonkowe, takie jak telefony komórkowe i cyfrowe asystenty osobiste w miejsce tradycyjnych komputerów będą uznawane za dominujące na świecie urządzenia z dostępem do internetu. Aktualnie istnieją dwa ogólne typy bezprzewodowego dostępu do internetu. W przypadku **bezprzewodowej sieci lokalnej** korzystający z niej użytkownicy wysyłają pakiety do stacji bazowej (nazywanej też bezprzewodowym punktem dostępowym) i odbierają je z niej (gdy znajdują się w promieniu kilkudziesięciu metrów od stacji). Stacja bazowa jest zwykle podłączona do tradycyjnego internetu. Oznacza to, że stacja pełni rolę pośrednika między bezprzewodowymi użytkownikami i siecią z okablowaniem. W przypadku **rozległej bezprzewodowej sieci dostępowej** stacja bazowa jest zarządzana przez operatora telekomunikacyjnego i obsługuje użytkowników znajdujących się w promieniu dziesiątek kilometrów.

Bezprzewodowe sieci lokalne oparte na technologii IEEE 802.11 (stosuje się również terminy takie jak bezprzewodowy Ethernet i Wi-Fi) są obecnie na szeroką skalę wdrażane w jednostkach uczelni, biurach firm, kawiarniach i domach. Przykładowo,

na obszarze uniwersytetów obu autorów książki zainstalowano stacje bazowe IEEE 802.11. Korzystając z infrastruktury takiej bezprzewodowej sieci lokalnej, studenci mogą wysyłać i odbierać wiadomości pocztowe bądź przeglądać strony internetowe z dowolnego miejsca kampusu (na przykład z biblioteki, pokoju w akademiku, sali wykładowej lub ławki znajdującej się na obszarze kampusu). Technologia 802.11, którą dokładnie omówiono w rozdziale 6., oferuje wspólną szybkość transmisji wynoszącą 11 Mb/s.

Obecnie w wielu domach możliwy jest zarówno dostęp szerokopasmowy (modemy kablowe lub DSL), jak i tania technologia bezprzewodowej sieci lokalnej pozwalająca na tworzenie sieci domowych o dużych możliwościach. Na rysunku 1.11 pokazano schemat typowej sieci domowej (w rzeczywistości jest to sieć wykonana przez autorów książki). Przykładowa sieć składa się z laptopa i trzech stacjonarnych komputerów PC (dwa z okablowaniem sieciowym, a jeden bez), a także ze stacji bazowej (bezprzewodowy punkt dostępowy), która komunikuje się z bezprzewodowymi komputerami PC. Ponadto w sieci znajduje się modem kablowy zapewniający szerokopasmowy dostęp do internetu i router, który pośredniczy między modemem oraz stacją bazową i stacjonarnymi komputerami PC. Sieć umożliwia domownikom dysponowanie szerokopasmowym dostępem do internetu. Jeden z nich może korzystać z laptopa, będąc w kuchni, na podwórku lub w sypialni. Całkowity koszt wykonania takiej sieci nie przekracza 750 zł (z uwzględnieniem ceny modemu kablowego lub DSL).



RYSUNEK 1.11. Schemat typowej sieci domowej

Gdy zamierza się uzyskać dostęp do internetu za pośrednictwem bezprzewodowej sieci lokalnej, zwykle trzeba się znajdować w odległości kilkudziesięciu metrów od stacji bazowej. Jest to możliwe w przypadku uzyskiwania dostępu do sieci z domu, kawiarni lub bardziej ogólnie, na obszarze budynku lub wokół niego. Jednak co zrobić, aby możliwe było skorzystanie z internetu, będąc na plaży lub w samochodzie? W przypadku tego typu rozległego dostępu mobilni użytkownicy internetowi korzystają z infrastruktury telefonii komórkowej, łącząc się ze stacjami bazowymi oddalonymi nawet o dziesiątki kilometrów od miejsca pobytu.

Protokół WAP (*Wireless Access Protocol*; wersja 2) [WAP 2004] ogólnie dostępny w Europie i protokół i-mode rozpowszechniony w Japonii to dwie technologie umożliwiające uzyskanie dostępu do internetu za pośrednictwem mobilnej infrastruktury telefonicznej. Przypominające zwykłe bezprzewodowe aparaty telefoniczne, lecz z trochę

większymi wyświetlaczami, telefony obsługujące protokół WAP oferują dostęp internetowy o niskiej lub średniej szybkości, a także usługę telefonii mobilnej. Zamiast z protokołu HTML telefony WAP korzystają ze specjalnego języka znaczników WML (*WAP Markup Language*), który zoptymalizowano pod kątem niewielkich wyświetlaczy i niskich szybkości dostępu. W Europie protokół WAP jest stosowany w infrastrukturze telefonii komórkowej GSM, która cieszy się największym powodzeniem. Protokół WAP 2.0 funkcjonuje za pomocą protokołów TCP/IP. Niestandardowy protokół i-mode, który pod względem działania przypomina protokół WAP, odniósł ogromny sukces w Japonii.

Aktualnie firmy telekomunikacyjne na ogromną skalę inwestują w technologię bezprzewodową 3G (*Third Generation*), która zapewnia rozległy dostęp internetowy oparty na przełączaniu pakietów z szybkością przekraczającą 384 kb/s [Kaaranen 2001] [Korhonen 2003]. Rozwiązania 3G oferują bardzo szybki dostęp do stron internetowych i interaktywnego wideo, a także powinny zapewnić jakość głosu lepszą od zapewnianej przez zwykły tradycyjny aparat telefoniczny. Pierwsze systemy 3G wdrożono w Japonii. Biorąc pod uwagę poziom inwestycji związanych z technologią 3G, infrastrukturą i licencjami, wielu analityków (i inwestorów!) zastanawia się, czy technologia odniesie duży sukces, o którym cały czas się mówi. A może technologia 3G przegra rywalizację z rozwiązaniami takimi jak IEEE 802.11? Czy technologie 3G i 802.11 zostaną ze sobą połączone, aby zapewnić rozpowszechniony, lecz heterogeniczny dostęp? Tego nadal nie wiadomo (zapoznaj się ze źródłem [Weinstein 2002] i zawartością ramki „KĄCIK HISTORYCZNY" zamieszczonej w podrozdziale 6.2). Obie technologie zostaną szczegółowo omówione w rozdziale 6.

1.4.2. Fizyczny nośnik

W poprzednim punkcie dokonaliśmy przeglądu kilku najważniejszych technologii dostępu do sieci stosowanych w internecie. Omawiając je, wspomnieliśmy również o używaniu fizycznego nośnika. Przykładowo, stwierdziliśmy, że technologia HFC korzysta z rozwiązania łączącego światłowód z kablem koncentrycznym. Napisaliśmy, że modemy telefoniczne 56 kb/s i technologia DSL używają miedzianej skrętki, jak również, że mobilne sieci dostępowe wykorzystują widmo radiowe. W tej części rozdziału dokonamy krótkiego przeglądu tych i innych nośników transmisyjnych powszechnie spotykanych w internecie.

Aby określić, co rozumie się przez fizyczny nośnik, opiszemy krótki żywot bita. Pod uwagę weźmy bit przemieszczający się z jednego systemu końcowego, przez serię łączy i routerów do innego systemu końcowego. Ten nieszczęsny bit jest transmitowany wiele, ale to wiele razy! Najpierw przesyła go źródłowy system końcowy. Chwilę potem bit jest odbierany przez pierwszy router z kilku, który transmituje go do kolejnego routera. Ten odbiera bit i powtarza operację. A zatem bit podróżujący od miejsca źródłowego do docelowego przechodzi przez serię par złożonych z nadajnika i odbiornika. W przypadku każdej takiej pary bit jest przesyłany **fizycznym nośnikiem** przy użyciu propagacji fal elektromagnetycznych lub impulsów optycznych. Fizyczny nośnik może przyjmować wiele kształtów i form, a ponadto nie musi być identycznego typu w przypadku poszczególnych par nadajnik-odbiornik, występujących na drodze pokonywanej przez bit. Przykładami fizycznych nośników są: miedziana skrętka, kabel koncentryczny, światłowód wielomodowy, a także naziemne i satelitarne widmo radiowe. Fizyczne

nośniki zalicza się do dwóch kategorii — **przewodowe** i **bezprzewodowe**. W przypadku nośników przewodowych fale przemieszczają się wzdłuż ciągłego nośnika, takiego jak światłowód, miedziana skrętka lub kabel koncentryczny. W przypadku nośnika bezprzewodowego fale rozchodzą się w atmosferze i przestrzeni kosmicznej (dotyczy to na przykład bezprzewodowej sieci lokalnej lub cyfrowego kanału satelitarnego).

Zanim zajmiemy się charakterystykami różnego typu nośników, należy w skrócie wspomnieć o ich cenie. Rzeczywisty koszt fizycznego łącza (kabel z miedzi, światłowód itp.) często jest dość nieznaczny w porównaniu z innymi kosztami związanymi z siecią. W rzeczywistości koszty robocizny dotyczące instalacji fizycznego łącza mogą być o rzędy wielkości większe od kosztów materiałów. Z tego powodu wielu wykonawców instaluje skrętkę, światłowód i kabel koncentryczny w każdym pomieszczeniu budynku. Jeśli nawet początkowo będzie używany tylko jeden nośnik, są spore szanse na to, że inny nośnik może okazać się przydatny w najbliższej przyszłości. W takiej sytuacji zaoszczędzi się na braku konieczności ponownego prowadzenia dodatkowego okablowania.

Skrętka miedziana

Najtańszym i najczęściej używanym przewodowym nośnikiem transmisyjnym jest skrętka miedziana. Przez ponad sto lat skrętka była używana w sieciach telefonicznych. W rzeczywistości ponad 99% połączeń kablowych poprowadzonych od aparatu telefonicznego do lokalnego przełącznika jest wykonanych ze skrętki miedzianej. Większość z nas widziała skrętkę w domach i miejscu pracy. Skrętka składa się z dwóch izolowanych przewodów miedzianych (każdy o grubości około 1 mm), które są ułożone w regularny spiralny wzór. Przewody są ze sobą skręcone, aby zredukować zakłócenia elektryczne wywoływane przez bliskość podobnych par. Zwykle kilka par przewodów jest umieszczanych w jednym kablu i otaczanych ochronnym ekranem. Para przewodów tworzy pojedyncze łącze komunikacyjne. **Skrętka nieekranowana UTP** (*Unshielded Twisted Pair*) jest powszechnie stosowana w sieciach komputerowych znajdujących się w budynku, czyli sieciach lokalnych. Obecnie szybkości transferu danych w przypadku sieci lokalnych używających skrętki zawierają się w przedziale od 10 Mb/s do 1 Gb/s. Szybkości, które mogą być uzyskiwane, zależą od grubości przewodu i odległości między nadajnikiem i odbiornikiem.

Gdy w latach 80. pojawiła się technologia światłowodowa, wiele osób zlekceważyło skrętkę, ponieważ oferuje stosunkowo niewielkie szybkości transmisji. Niektórzy uznali nawet, że światłowody powinny całkowicie zastąpić skrętkę. Jednak skrętka nie tak łatwo ustępowała pola. Nowsze technologie skrętki, takie jak kategoria UTP 5, pozwalają osiągnąć szybkość transmisji 100 Mb/s przy maksymalnej odległości wynoszącej kilkaset metrów. W przypadku krótszych dystansów są nawet możliwe większe szybkości. Ostatecznie skrętka zaczęła pełnić rolę dominującego rozwiązania dla bardzo szybkich sieci lokalnych.

Jak wspomniano w punkcie dotyczącym sieci dostępowych, skrętka jest też powszechnie wykorzystywana na potrzeby oferowania dostępu do internetu prywatnym użytkownikom. Stwierdziliśmy, że modem telefoniczny za pośrednictwem skrętki pozwala uzyskać maksymalną szybkość 56 kb/s. Ponadto technologia DSL wykorzystująca skrętkę umożliwiła prywatnym użytkownikom łączenie się z internetem z szybkościami przekraczającymi 6 Mb/s (gdy domostwa znajdują się blisko modemu dostawcy ISP).

Kabel koncentryczny

Kabel koncentryczny podobnie jak skrętka jest złożony z dwóch miedzianych przewodów. Jednak są one prowadzone koncentrycznie, a nie równolegle. Dzięki takiemu rozwiązaniu, a także specjalnemu izolowaniu i ekranowaniu kabel koncentryczny może oferować duże szybkości transmisji. Kabel koncentryczny jest dość rozpowszechniony w systemach telewizji kablowych. Jak wcześniej wspomniano, w ostatnim czasie w tego typu systemach zastosowano modemy kablowe, które prywatnym użytkownikom zapewniają dostęp do internetu z szybkościami rzędu 1 Mb/s lub wyższymi. W przypadku telewizji kablowej i kablowego dostępu internetowego nadajnik przesuwa sygnał cyfrowy na określone pasmo częstotliwości, a następnie uzyskany sygnał analogowy jest wysyłany do jednego lub kilku odbiorników. Kabel koncentryczny może pełnić rolę przewodowego **wspólnego nośnika**. Dokładniej mówiąc, kilka systemów końcowych może zostać bezpośrednio podłączonych do kabla. Każdy taki system odbierze dowolne dane przesyłane przez pozostałe systemy.

Światłowód

Światłowód jest cienkim i elastycznym nośnikiem, który przenosi impulsy świetlne. Każdy impuls reprezentuje bit. Pojedynczy światłowód jest w stanie oferować niebywałe szybkości transmisji, które maksymalnie mogą wynosić dziesiątki, a nawet setki gigabitów na sekundę. Światłowody są odporne na zakłócenia elektromagnetyczne. Na odległości do 100 kilometrów cechują się bardzo niewielkim tłumieniem sygnału, a ponadto bardzo trudno się do nich „podpiąć". Wymienione cechy sprawiły, że światłowody stały się preferowanym przewodowym nośnikiem transmisyjnym, zwłaszcza w przypadku łączy międzykontynentalnych. Aktualnie wiele długodystansowych sieci telefonicznych istniejących w Stanach Zjednoczonych i innych państwach wykorzystuje wyłącznie światłowody. Są one też rozpowszechnione w sieci szkieletowej internetu. Jednak wysoki koszt urządzeń optycznych, takich jak transmitery, odbiorniki i przełączniki, zmniejsza tempo wdrażania technologii światłowodowej w przypadku mniejszych odległości (dotyczy to sieci lokalnych lub dostępu sieciowego uzyskiwanego z domostw prywatnych użytkowników). W źródłach [IEC Optical 2003], [Goralski 2001], [Ramaswami 1998] i [Mukherjee 1997] można znaleźć omówienie różnych aspektów dotyczących sieci optycznych. Szybkości łączy optycznych mogą maksymalnie wynieść dziesiątki gigabitów na sekundę.

Naziemne kanały radiowe

Kanały radiowe przenoszą sygnały za pośrednictwem widma elektromagnetycznego. Jest to atrakcyjny nośnik, ponieważ nie wymaga prowadzenia żadnego okablowania, radzi sobie ze ścianami, zapewnia łączność użytkownikom mobilnym i ewentualnie może być użyty do przesyłania sygnału na duże odległości. Charakterystyki kanału radiowego w dużym stopniu zależą od środowiska propagacji i odległości, na jaką sygnał będzie przenoszony. Ze środowiskiem są związane takie kwestie jak utrata ścieżki i „zacienianie" (osłabianie sygnału w trakcie jego przesyłania spowodowane przez przeszkody znajdujące się na drodze sygnału lub w jego pobliżu), zanikanie wielodrożne (spowodowane przez odbicia sygnału od przeszkód) i zakłócenia (wywołane przez inne kanały radiowe lub sygnały elektromagnetyczne).

Naziemne kanały radiowe można ogólnie zaklasyfikować do dwóch grup. Pierwszą stanowią kanały wykorzystywane lokalnie, które zwykle swoim zasięgiem obejmują od 10 do kilkuset metrów. Z kolei do drugiej grupy zaliczają się kanały używane na większych odległościach liczących dziesiątki kilometrów. Z pierwszej grupy kanałów radiowych korzystają urządzenia bezprzewodowej sieci lokalnej omówione w punkcie 1.4.1. W tej samym punkcie wspomniano o technologiach WAP, i-mode i 3G, które stosują kanały radiowe zaliczane do drugiej grupy. Źródło [Dornan 2001] zawiera przegląd i omówienie technologii i produktów. Kanały radiowe zostaną szczegółowo opisane w rozdziale 6.

Satelitarne kanały radiowe

Satelita komunikacyjny łączy ze sobą dwa lub więcej znajdujących się na ziemi mikrofalowych transmiterów lub odbiorników nazywanych też stacjami naziemnymi. Satelita odbiera transmisje na jednym paśmie częstotliwości, a następnie regeneruje sygnał za pomocą wzmacniacza (omówiony niżej) i przesyła sygnał, korzystając z innej częstotliwości. Satelity są w stanie zapewnić szybkości transmisji rzędu gigabitu na sekundę. W komunikacji są stosowane dwa typy satelitów — **geostacjonarne** i **niskoorbitalne**.

Satelity geostacjonarne cały czas pozostają nad tym samym miejscem Ziemi. Taką stacjonarność osiąga się przez umieszczenie satelity na orbicie położonej na wysokości 36 000 kilometrów nad powierzchnią Ziemi. Tak duża odległość, która musi być pokonana przez sygnał (od stacji naziemnej do satelity i z powrotem), powoduje znaczne opóźnienie jego propagacji wynoszące 250 ms. Niemniej jednak łącza satelitarne, które mogą działać z szybkościami równymi setkom megabitów na sekundę, są często wykorzystywane w sieciach telefonicznych i sieci szkieletowej internetu.

Satelity niskoorbitalne są zlokalizowane znacznie bliżej Ziemi i nie znajdują się cały czas nad jednym jej punktem. Tego typu satelity obracają się wokół Ziemi, podobnie jak Księżyc. Aby zapewnić niezmienne obejmowanie przez satelitę określonego obszaru, na orbicie musi być umieszczonych wiele satelitów. Aktualnie jest wdrażanych wiele systemów komunikacyjnych używających satelitów niskoorbitalnych. Na stronie internetowej Lloyda Wooda [Wood 2004] zamieszczono informacje na temat konstelacji satelitów używanych przez systemy komunikacyjne. Być może w przyszłości satelity niskoorbitalne zostaną zastosowane na potrzeby oferowania dostępu internetowego.

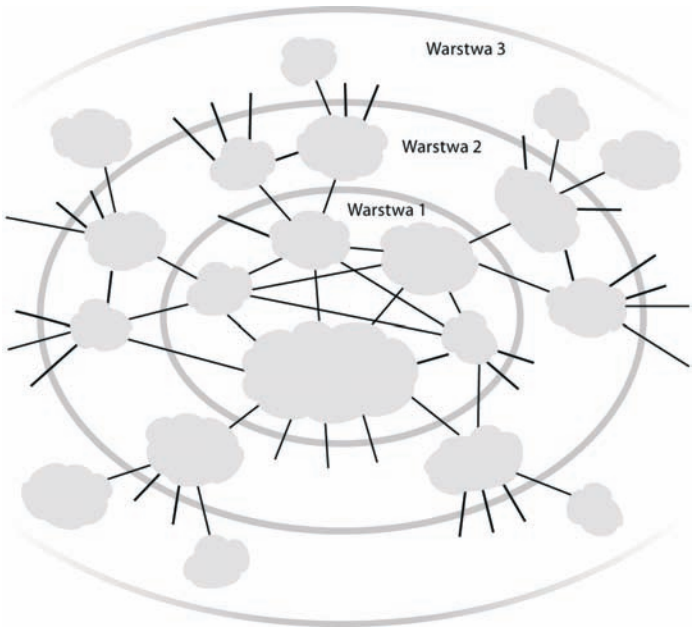
1.5. Dostawcy ISP i sieci szkieletowe internetu

Wcześniej wspomnieliśmy, że systemy końcowe (komputery PC, cyfrowe asystenty osobiste, serwery WWW, serwery poczty itp.) są łączone z internetem za pośrednictwem sieci dostępowej. Siecią dostępową może być przewodowa lub bezprzewodowa sieć lokalna (znajdująca się na przykład w firmie, szkole lub bibliotece). Taką sieć może też oferować regionalny dostawca ISP (na przykład AOL lub MSN). Można się z nią połączyć za pośrednictwem modemu telefonicznego, kablowego lub DSL. Jednak podłączanie użytkowników końcowych i dostawców treści do sieci dostępowych

stanowi tylko niewielki element układanki łączącej ze sobą setki milionów użytkowników i setki tysięcy sieci tworzących internet. Internet jest *siecią sieci*. Zrozumienie tego pojęcia jest kluczem do poradzenia sobie z tą układanką.

Sieci dostępowe zlokalizowane na obrzeżu publicznego internetu są połączone z jego pozostałą częścią za pośrednictwem warstwowej hierarchii dostawców ISP (rysunek 1.12). Na spodzie hierarchii znajdują się dostawcy ISP oferujący sieci dostępowe (na przykład dostawcy obsługujący prywatnych użytkowników, tacy jak AOL, i świadczący usługi firmom, korzystając z sieci lokalnych). Na samym szczycie hierarchii znajduje się stosunkowo niewielka liczba tak zwanych **dostawców ISP pierwszej warstwy**. Pod wieloma względami tego typu dostawca korzysta z sieci takiej samej jak każda inna. W sieci, która jest połączona z innymi sieciami, znajdują się łącza i routery. Jednak sieci dostawców ISP pierwszego poziomu w kilku elementach są wyjątkowe. Szybkości ich łączy często wynoszą 622 Mb/s lub więcej. W przypadku większych dostawców ISP pierwszej warstwy szybkość łączy zawiera się w przedziale od 2,5 do 10 Gb/s. Oznacza to, że routery znajdujące się w sieciach takich dostawców muszą być w stanie przekazywać pakiety z wyjątkowo dużymi szybkościami. Dostawcy ISP pierwszej warstwy cechują się następującymi dodatkowymi elementami:

- dysponowanie bezpośrednim połączeniem z *wszystkimi* innymi dostawcami pierwszej warstwy;
- połączenie z dużą liczbą dostawców ISP drugiej warstwy i innymi sieciami abonenckimi;
- międzynarodowy zasięg.



RYSUNEK 1.12. Schemat połączeń dostawców ISP

Sieci dostawców ISP pierwszej warstwy są też nazywane sieciami **szkieletowymi internetu**. Do tego typu dostawców ISP należy zaliczyć takie firmy jak Sprint, MCI (wcześniej UUNet/WorldCom), AT&T, Level3 (przejęła firmę Genuity), Qwest i Cable & Wireless. W połowie 2002 r. firma WorldCom była największym dostawcą ISP pierwszej warstwy, ponaddwukrotnie większym od następnego w kolejności pod względem kilku wskaźników wielkości [Teleography 2002]. Interesujące jest to, że oficjalnie żadna grupa nie sankcjonuje statusu pierwszej warstwy. Jest takie powiedzenie: „Jeśli trzeba się dowiedzieć, czy należy się do określonej grupy, prawdopodobnie się do niej nie należy".

Dostawcy ISP drugiej warstwy zwykle mają zasięg regionalny lub krajowy, a ponadto, co jest istotne, są połączeni tylko z kilkoma dostawcami ISP pierwszej warstwy (rysunek 1.12). A zatem, aby uzyskać dostęp do sporej części globalnego internetu, dostawca ISP drugiej warstwy będzie musiał przesyłać dane za pośrednictwem dostawcy pierwszej warstwy, z którym jest połączony. Dostawca ISP drugiej warstwy jest nazywany **klientem** dostawcy pierwszej warstwy, z którym jest połączony. Z kolei dostawca pierwszej warstwy dla swojego klienta jest **dostawcą**. Wiele dużych firm i instytucji podłącza swoje sieci bezpośrednio do sieci dostawcy ISP pierwszej lub drugiej warstwy. W ten sposób stają się klientem takiego dostawcy ISP. Nadrzędny dostawca ISP od dostawcy ISP będącego klientem pobiera opłatę, która zwykle zależy od szybkości transmisji łącza poprowadzonego między nimi. Sieć dostawcy ISP drugiej warstwy może też być bezpośrednio połączona z sieciami innych dostawców tej warstwy. W tym przypadku dane między dwoma dostawcami drugiej warstwy mogą być przesyłane bez pośrednictwa sieci dostawcy ISP pierwszej warstwy. Poniżej warstwy drugiej znajdują się dostawcy ISP niższej warstwy, którzy z większym obszarem internetu łączą się za pośrednictwem jednego lub większej liczby dostawców ISP drugiej warstwy. Na dole hierarchii są zlokalizowani dostawcy ISP oferujący sieci dostępowe. Aby wszystko jeszcze bardziej skomplikować, niektórzy dostawcy ISP pierwszej warstwy znajdują się jednocześnie w drugiej warstwie (oznacza to, że są zintegrowani w sposób pionowy). Tego typu dostawcy oferują dostęp do internetu bezpośrednio użytkownikom końcowym i dostawcom treści, a także dostawcom ISP niższych warstw. Gdy dwaj dostawcy ISP są ze sobą bezpośrednio połączeni, są dla siebie **węzłami**. W interesującej pracy [Subramanian 2002] podjęto się próby bardziej precyzyjnego zdefiniowania warstwowej struktury internetu przez przeanalizowanie jego topologii pod względem powiązań klient-dostawca i węzeł-węzeł.

Punkty zlokalizowane w sieci dostawcy ISP, w których jest ona połączona z sieciami innych dostawców ISP (niezależnie od tego, czy dostawca znajduje się w tej samej warstwie hierarchii, czy poniżej jej), określa się mianem **punktów POP** (*Points of Presence*). Punkt POP jest po prostu grupą liczącą jeden lub kilka routerów sieci dostawcy ISP, z którymi mogą się łączyć routery znajdujące się w sieciach jego klientów lub sieciach innych dostawców. Dostawca pierwszej warstwy zazwyczaj posiada w swojej sieci wiele punktów POP rozmieszczonych w różnych geograficznych lokalizacjach. Do każdego punktu POP jest podłączonych wiele sieci klientów i innych dostawców ISP. Aby klient mógł do punktu POP dostawcy podłączyć swoją sieć, zwykle od innego operatora telekomunikacyjnego dzierżawi łącze o dużej szybkości i jeden ze swoich routerów łączy bezpośrednio z routerem dostawcy ISP, będącego w posiadaniu punktu POP. Dwaj dostawcy ISP pierwszej warstwy mogą też być dla siebie węzłami, gdy każdy z nich udostępni punkt POP, a następnie punkty te zostaną ze sobą połączone. Ponadto dwaj dostawcy ISP mogą dysponować wieloma połączeniami wykonanymi między dwoma lub większą liczbą par punktów POP.

Poza łączeniem się ze sobą przy użyciu własnych punktów dostawcy ISP często wykonują między sobą połączenia przy użyciu **punktów NAP** (*Network Access Point*). Każdy taki punkt może być w posiadaniu niezależnego operatora telekomunikacyjnego lub dostawcy sieci szkieletowych internetu. Za pośrednictwem punktów NAP między wieloma dostawcami ISP jest wymieniana ogromna ilość danych. Jednak coraz częściej dostawcy ISP pierwszej warstwy pomijają punkty NAP i łączą się ze sobą bezpośrednio za pomocą własnych węzłów [Kende 2000]. W przypadku dostawców ISP pierwszej warstwy występuje tendencja do wzajemnego bezpośredniego łączenia się przy użyciu własnych węzłów. Z kolei dostawcy ISP drugiej warstwy łączą się ze sobą i z dostawcami pierwszej warstwy za pomocą punktów NAP. Ponieważ punkty NAP przekazują i przełączają ogromne ilości danych, same są złożonymi bardzo szybkimi sieciami z przełączaniem, często zlokalizowanymi w pojedynczym budynku.

Podsumowując, topologia internetu jest złożona. Składa się z dziesiątków dostawców ISP pierwszej i drugiej warstwy, a także tysięcy dostawców niższych warstw. Dostawcy ISP różnią się skalą działania. Niektórzy z nich swoim zakresem obejmują wiele kontynentów i oceanów, natomiast inni ograniczają się do niewielkich regionów świata. Dostawcy ISP niższych warstw łączą się z dostawcami wyższych warstw, natomiast te są ze sobą połączone (zazwyczaj) za pomocą prywatnych węzłów i punktów NAP. Użytkownicy i dostawcy treści są klientami dostawców ISP niższych warstw, a te klientami dostawców wyższych warstw.

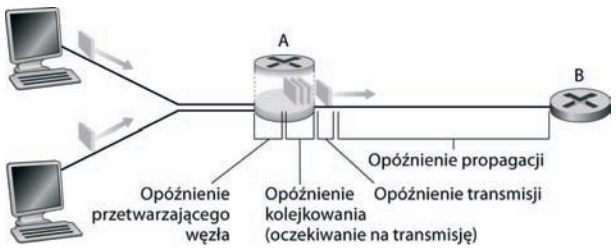
W ramach wniosku z tego podrozdziału wspomnimy, że każdy z nas może stać się dostawcą ISP, gdy tylko będzie dysponował połączeniem internetowym. Aby to było możliwe, trzeba jedynie nabyć sprzęt (na przykład router i pulę modemów) niezbędny do przyłączenia innych użytkowników. A zatem nowe warstwy i gałęzie mogą być uwzględniane w topologii internetu, tak jak nowe klocki Lego można dodawać do już istniejącej konstrukcji.

1.6. Opóźnienie i utrata pakietów w sieciach z przełączaniem pakietów

Gdy już pobieżnie omówiono podstawowe elementy architektury internetu, takie jak aplikacje, systemy końcowe, protokoły transportowe punkt-punkt, routery i łącza, przyjrzyjmy się, co się dzieje z pakietami przesyłanymi od miejsca źródłowego do docelowego. Jak wcześniej wspomniano, pakiet jest najpierw wysyłany przez host źródłowy, a następnie przechodzi przez serię routerów i swoją podróż kończy, docierając do hosta docelowego. Po drodze pakiet jest transmitowany od jednego węzła (host lub router) do kolejnego. W tym czasie po trafieniu do *każdego* kolejnego węzła na pakiet ma wpływ kilka typów opóźnień. Do najważniejszych z nich należy zaliczyć **opóźnienie przetwarzającego węzła**, **kolejkowania**, **transmisji** i **propagacji**. Razem wymienione opóźnienia tworzą **całkowite opóźnienie węzła**. Aby dogłębnie zrozumieć przełączanie pakietów i sieci komputerowe, trzeba poznać naturę tych opóźnień i ich istotną rolę.

1.6.1. Typy opóźnień

Opóźnienia objaśnijmy, korzystając z rysunku 1.13. W trakcie pokonywania trasy od jednego punktu (źródłowego) do drugiego (docelowego) pakiet jest transmitowany kolejno przez routery A i B. Naszym celem jest scharakteryzowanie opóźnienia węzła występującego w routerze A. Warto zauważyć, że router A posiada łącze wyjściowe prowadzące do routera B. Przed łączem znajduje się kolejka (inaczej bufor). Gdy pakiet wysłany przez węzeł dotrze do routera A, ten sprawdzi nagłówek pakietu, aby identyfikować dla niego odpowiednie łącze wyjściowe i skierować go do niego. W omawianym przykładzie łączem wyjściowym pakietu jest łącze prowadzące do routera B. Pakiet może być transmitowany łączem, tylko gdy w danej chwili nie przesyła ono żadnego innego pakietu, a ponadto w kolejce nie ma pakietów. Jeśli łącze akurat jest zajęte lub są już kolejkowane inne pakiety przeznaczone dla tego łącza, nowy pakiet, który odebrano, zostanie umieszczony w kolejce.



RYSUNEK 1.13. Opóźnienie węzła występujące w routerze A

Opóźnienie przetwarzającego węzła

Czas wymagany do sprawdzenia nagłówka pakietu i stwierdzenia, gdzie należy go skierować, jest elementem **opóźnienia przetwarzającego węzła**. Opóźnienie to może też uwzględniać inne czynniki, takie jak czas niezbędny do sprawdzenia błędów na poziomie bitów, które wystąpiły podczas przesyłania bitów pakietu od węzła nadawczego do routera A. Opóźnienie przetwarzającego węzła w przypadku bardzo szybkich routerów jest określane w mikrosekundach lub mniejszych jednostkach czasu. Po przetworzeniu pakietu router umieszcza go w kolejce znajdującej się przed łączem prowadzącym do routera B (w rozdziale 4. szczegółowo wyjaśniono zasady działania routera).

Opóźnienie kolejkowania

Opóźnienie kolejkowania dotyczy pakietu oczekującego w kolejce na przesłanie łączem. Wielkość opóźnienia w przypadku określonego pakietu będzie zależała od liczby pakietów wcześniej umieszczonych w kolejce i oczekujących na transmisję przy użyciu łącza. Jeśli kolejka jest pusta i w danej chwili nie jest przesyłany żaden inny pakiet, opóźnienie kolejkowania dla rozważanego pakietu będzie zerowe. Z kolei gdy obciążenie sieci jest duże i wiele innych pakietów oczekuje na przesłanie, opóźnienie będzie znaczne. Wkrótce okaże się, że liczba pakietów, które mogą się znaleźć w kolejce przed odebranym pakietem, jest funkcją natężenia i charakteru danych umieszczanych w kolejce. W praktyce wartość opóźnienia kolejkowania może być wyrażona w mikrosekundach lub milisekundach.

Opóźnienie transmisji

Zakładając, że pakiety są transmitowane zgodnie z zasadą „pierwszy na wejściu, pierwszy na wyjściu", powszechnie wykorzystywaną w sieciach z przełączaniem pakietów, rozważany pakiet będzie mógł być przesłany dopiero, gdy zostaną wysłane wszystkie pakiety, które wcześniej odebrano. Określmy długość pakietu jako L bitów, natomiast szybkość łącza między routerem A i B jako R bitów/s. Szybkość R jest uzależniona od szybkości transmisji łącza prowadzącego do routera B. Przykładowo, w przypadku łącza Ethernet 10 Mb/s R wyniesie 10 Mb/s. W przypadku łącza Ethernet 100 Mb/s R = 100 Mb/s. **Opóźnienie transmisji** (nazywane też opóźnieniem buforowanym, omówione w podrozdziale 1.3) wynosi L/R. Opóźnienie to określa czas niezbędny do umieszczenia (wysłania) w łączu wszystkich bitów pakietu. W praktyce wartość opóźnienia transmisji zwykle jest wyrażana w mikrosekundach lub milisekundach.

Opóźnienie propagacji

Po umieszczeniu bitu w łączu trzeba go przesłać do routera B. Czas potrzebny na propagację bitu z początku łącza do routera B wyznacza **opóźnienie propagacji**. Bit jest przemieszczany z szybkością propagacji łącza. Zależy ona od fizycznego nośnika (światłowód, skrętka miedziana itp.) łącza i zawiera się w następującym przedziale:

$$2 \cdot 10^8 \text{ — } 3 \cdot 10^8 \, m/s$$

Szybkość jest równa lub trochę mniejsza od prędkości światła. Opóźnienie propagacji jest odległością między dwoma routerami podzieloną przez szybkość propagacji. Oznacza to, że opóźnienie to określa zależność d/s, w której d jest dystansem między routerem A i B, natomiast s szybkością propagacji łącza. Gdy zakończy się propagacja do routera B ostatniego bitu pakietu, wraz z wszystkimi wcześniejszymi bitami zostanie zapisany w routerze B. Gdy to nastąpi, router B będzie kontynuował proces, wykonując operację przekazywania pakietu. W sieciach rozległych opóźnienia propagacji są wyrażane w milisekundach.

Porównanie opóźnień transmisji i propagacji

Nowicjusze w dziedzinie sieci komputerowych czasami mają problemy ze zrozumieniem różnicy występującej między opóźnieniami transmisji i propagacji. Różnica jest subtelna, ale istotna. Opóźnienie transmisji określa czas, jakiego router potrzebuje na wysłanie pakietu. Opóźnienie to jest funkcją długości pakietu i szybkości transmisji łącza. Jednak nie ma nic wspólnego z odległością między dwoma routerami. Z kolei opóźnienie propagacji jest czasem, jaki jest wymagany na propagację bitu z jednego routera do drugiego. Opóźnienie jest funkcją odległości między dwoma routerami. Jednak nie ma związku z długością pakietu lub szybkością transmisji łącza.

Analogia powinna rozwiać wątpliwości dotyczące różnic między opóźnieniami transmisji i propagacji. Weźmy pod uwagę autostradę, na której co 100 kilometrów znajduje się bramka. Segmenty autostrady między kolejnymi bramkami można potraktować jak łącza, natomiast bramki jak routery. Załóżmy, że samochody przemieszczają się (propagują) po autostradzie z prędkością 100 km/h (oznacza to, że po minięciu bramki pojazd natychmiast przyspieszy do prędkości 100 km/h i utrzyma ją do kolejnej

bramki). Przyjmijmy, że autostradą przemieszcza się karawana złożona z 10 samochodów jadących w określonej niezmiennej kolejności. Każdy pojazd można potraktować jak bit, natomiast karawanę jak pakiet. Ponadto załóżmy, że każda bramka w ciągu 12 sekund obsługuje (transmituje) jeden samochód. Ponieważ jest późna pora, samochody tworzące karawanę są jedynymi znajdującymi się na autostradzie. Przyjmijmy jeszcze, że gdy pierwszy pojazd grupy dotrze do bramki, poczeka przy niej do momentu, aż pozostałe 9 samochodów ustawi się za nim (w związku z tym przed ruszeniem w dalszą drogę cała karawana musi być pomieszczona przy bramce). Czas potrzebny na przepuszczenie przez bramkę całej karawany wynosi (10 samochodów)/(5 pojazdów na minutę) = 2 minuty. Czas ten można przyrównać do opóźnienia transmisji występującego w routerze. Czas potrzebny na przemieszczenie się samochodu od wylotu jednej bramki do kolejnej bramki wynosi 100 km/(100 km/h) = 1 godzina. Czas ten odpowiada opóźnieniu propagacji. A zatem czas, jaki upływa od momentu zgromadzenia całej karawany przy wlocie bramki do chwili znalezienia się pojazdów przy wlocie kolejnej bramki, jest sumą opóźnienia transmisji i propagacji. W omawianym przykładzie czas ten wyniesie 62 minuty.

Rozwińmy bardziej przytoczoną analogię. Co by się stało, gdyby czas obsługi karawany przez bramkę był dłuższy niż czas pokonania przez samochód dystansu dzielącego dwie bramki? Dla przykładu załóżmy, że teraz pojazdy jadą z prędkością 1000 km/h, natomiast bramka obsługuje samochody z szybkością jednego na minutę. W tym przypadku opóźnienie związane z przejechaniem odcinka między dwoma bramkami wyniesie 6 minut, natomiast czas obsługi karawany 10 minut. Pierwsze samochody karawany pojawią się przy drugiej bramce, zanim ostatnie jej pojazdy miną pierwszą bramkę. Taka sytuacja występuje również w sieciach z przełączaniem pakietów. Pierwsze bity pakietu mogą zostać odebrane przez router, gdy wiele z pozostałych bitów w dalszym ciągu oczekuje na wysłane przez poprzedni router.

Jeśli opóźnienia przetwarzającego węzła, kolejkowania, transmisji i propagacji oznaczy się odpowiednio symbolami o_{przet}, o_{kolejk}, o_{trans} i o_{prop}, całkowite opóźnienie węzła określi następujące równanie:

$$o_{\text{węz}} = o_{\text{przet}} + o_{\text{kolejk}} + o_{\text{trans}} + o_{\text{prop}}$$

Udział poszczególnych opóźnień może się bardzo znacząco różnić. Przykładowo, opóźnienie o_{prop} może być nieznaczne (kilka mikrosekund) w przypadku łącza poprowadzonego między dwoma routerami znajdującymi się na tym samym kampusie uniwersyteckim. Jednak to samo opóźnienie będzie miało wartość setek milisekund, gdy dwa routery będą ze sobą połączone przy użyciu łącza obsługiwanego przez satelitę geostacjonarnego. W efekcie opóźnienie to może odgrywać dominującą rolę w równaniu określającym wartość opóźnienia $o_{\text{węz}}$. Podobnie opóźnienie o_{trans} może być nieznaczne, ale też dość istotne. Wkład wnoszony przez to opóźnienie zwykle jest pomijalny w przypadku szybkości transmisji rzędu 10 Mb/s i wyższych (dotyczy to na przykład sieci lokalnych). Jednak wartość opóźnienia o_{trans} może wynieść setki milisekund, gdy za pośrednictwem wolnego łącza obsługiwanego przez modem telefoniczny przesyła się duże pakiety internetowe. Choć opóźnienie o_{przet} jest często nieznaczne, ma duży wpływ na maksymalną przepustowość routera, która jest największą szybkością, z jaką urządzenie to może przekazywać pakiety.

1.6.2. Opóźnienie kolejkowania i utrata pakietów

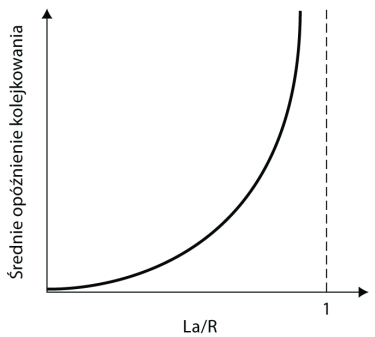
Najbardziej skomplikowanym i interesującym składnikiem opóźnienia węzła jest opóźnienie kolejkowania o_{kolejk}. W rzeczywistości opóźnienie kolejkowania występujące w sieci komputerowej jest tak ważne i ciekawe, że poświęcono mu tysiące artykułów i kilka książek [Bertsekas 1991] [Daigle 1991] [Kleinrock 1975, 1976] [Ross 1995]. W tym miejscu dokonamy jedynie bardzo ogólnego omówienia opóźnienia kolejkowania. Bardziej zainteresowani tym zagadnieniem mogą sprawdzić kilka książek, a ostatecznie nawet napisać pracę doktorską na ten temat! W przeciwieństwie do pozostałych trzech opóźnień (o_{przet}, o_{trans} i o_{prop}) opóźnienie kolejkowania może być inne dla każdego pakietu. Jeśli na przykład jednocześnie w pustej kolejce pojawi się 10 pakietów, te, które będą transmitowane w pierwszej kolejności, nie doświadczą żadnego opóźnienia. Z kolei w przypadku pakietu wysłanego jako ostatni opóźnienie kolejkowania będzie stosunkowo duże (wynika to z faktu czekania pakietu na przesłanie pozostałych dziewięciu). A zatem opisując opóźnienie kolejkowania, zwykle korzysta się ze wskaźników statystycznych, takich jak średnie opóźnienie, wariancja opóźnienia i prawdopodobieństwo tego, że opóźnienie przekroczy określoną wartość.

Kiedy opóźnienie kolejkowania jest duże, a kiedy nieznaczne? Odpowiedź na to pytanie zależy od szybkości, z jaką dane są umieszczane w kolejce, szybkości transmisji łącza i sposobu odbierania danych (chodzi o to, czy pojawiają się w kolejce cyklicznie czy impulsowo). Aby trochę bardziej to przybliżyć, literą a oznaczmy średnią szybkość (wyrażoną w pakietach na sekundę), z jaką pakiety pojawiają się w kolejce. Jak pamiętamy, R jest szybkością transmisji (w bitach/s), z jaką bity są pobierane z kolejki. Dla uproszczenia załóżmy, że wszystkie pakiety składają się z L bitów. W związku z tym średnia szybkość, z jaką bity będą umieszczane w kolejce, wyniesie La bitów/s. Przyjmijmy jeszcze, że kolejka jest bardzo duża, na tyle, że w zasadzie może pomieścić nieskończoną liczbę bitów. Współczynnik La/R nazywany **natężeniem ruchu** często odgrywa istotną rolę w szacowaniu zakresu opóźnienia kolejkowania. Jeśli $La/R > 1$, średnia szybkość, z jaką bity pojawiają się w kolejce, przekracza szybkość pobierania z niej bitów. W takiej niekorzystnej sytuacji kolejka będzie miała tendencję do powiększania się bez ograniczeń i opóźnienie kolejkowania będzie zbliżało się do nieskończoności! W związku z tym jedna z podstawowych zasad obowiązujących w przypadku inżynierii ruchu sieciowego brzmi: „*System należy tak projektować, aby natężenie ruchu nie przekroczyło wartości 1*".

Rozważmy teraz sytuację, gdy $La/R \leq 1$. W tym przypadku charakterystyka ruchu wejściowego ma wpływ na opóźnienie kolejkowania. Jeśli na przykład pakiety są odbierane cyklicznie, czyli co L/R sekund pojawia się jeden pakiet, każdy z nich będzie umieszczany w pustej kolejce, dzięki czemu nie wystąpi opóźnienie kolejkowania. Z kolei gdy pakiety są odbierane w sposób okresowy, ale też impulsowy, może pojawić się średnie opóźnienie o znacznej wartości. Dla przykładu załóżmy, że co $(L/R)N$ sekund jednocześnie odbieranych jest N pakietów. Pierwszy transmitowany pakiet nie posiada żadnego opóźnienia kolejkowania. Drugi przesyłany pakiet ma opóźnienie wynoszące L/R sekund. Mówiąc bardziej ogólniej, z n-tym transmitowanym pakietem jest związane opóźnienie kolejkowania równe $(n–1)L/R$ sekund. W ramach ćwiczenia Czytelnikowi pozostawiamy obliczenie dla przytoczonego przykładu średniego opóźnienia kolejkowania.

Powyższe dwa przykłady dotyczące okresowego odbierania danych są trochę akademickie. Zwykle proces umieszczania bitów w kolejce ma przebieg *losowy*. Oznacza to, że bity nie są odbierane według żadnego schematu i pakiety pojawiają się w przypadkowych odstępach czasu. W tym bardziej realistycznym wariancie zależność La/R

zwykle nie jest wystarczająca do pełnego scharakteryzowania statystyk dotyczących opóźnienia. Niemniej jednak zależność jest pomocna w zrozumieniu zakresu oddziaływania opóźnienia kolejkowania. W szczególności wtedy, gdy natężenie ruchu jest bliskie zeru, w kolejce pojawia się niewiele pakietów i jest między nimi duży odstęp. Ponadto mało prawdopodobne jest, że odebrany pakiet napotka w kolejce na inny pakiet. A zatem średnie opóźnienie kolejkowania będzie niemal zerowe. Z kolei jeśli natężenie ruchu jest zbliżone do wartości 1, wystąpią okresy, w których szybkość pojawiania się bitów przekroczy możliwości transmisyjne (na skutek impulsowego charakteru odbierania danych), co w efekcie spowoduje utworzenie kolejki. W miarę zbliżania się wartości natężenia ruchu do 1 średnia długość kolejki będzie coraz większa. Na rysunku 1.14 przedstawiono zależność średniego opóźnienia kolejkowania od natężenia ruchu.



RYSUNEK 1.14. Zależność średniego opóźnienia kolejkowania od natężenia ruchu

Ważną kwestą uwidocznioną na rysunku 1.14 jest to, że wraz ze zbliżaniem się wartości natężenia ruchu do 1 szybko zwiększa się średnie opóźnienie kolejkowania. Niewielki procentowy wzrost natężenia powoduje znacznie większy procentowy wzrost wartości opóźnienia. Być może z takim zjawiskiem miało się do czynienia na autostradzie. Jeśli regularnie jeździ się drogą, która zazwyczaj jest zatłoczona, oznacza to, że w jej przypadku wartość natężenia ruchu jest bliska 1. Jeśli jakieś zdarzenie spowoduje nawet jeszcze większy ruch na drodze niż zwykle, opóźnienia mogą być ogromne.

Utrata pakietów

W powyższej analizie założyliśmy, że kolejka jest w stanie przechowywać nieskończoną liczbę pakietów. W rzeczywistości kolejka zlokalizowana przed łączem posiada ograniczoną pojemność, która w znacznym stopniu zależy od konstrukcji przełącznika i jego ceny. Ponieważ pojemność kolejki jest ograniczona, tak naprawdę przy zbliżaniu się wartości natężenia ruchu do 1 opóźnienia pakietów nie zmierzają do nieskończoności. Może się okazać, że po odebraniu pakietu kolejka będzie już pełna. Gdy nie będzie możliwe przechowanie pakietu, router **odrzuci** go, co jest równoznaczne z **utratą** pakietu. Z punktu widzenia systemu końcowego wygląda to tak, jakby pakiet został przesłany do rdzenia sieci, lecz nigdy nie dotarł do miejsca przeznaczenia. Liczba utraconych pakietów zwiększa się wraz ze wzrostem natężenia ruchu. A zatem wydajność węzła jest często określana nie tylko w kategoriach opóźnienia, ale też pod względem prawdopodobieństwa utraty pakietu. W kolejnych rozdziałach Czytelnik dowie się, że utracony pakiet może być ponownie przesłany między poszczególnymi węzłami przez aplikację lub protokół warstwy transportowej.

Opóźnienie międzywęzłowe

Do tego momentu skoncentrowaliśmy się na opóźnieniu węzła występującego w pojedynczym routerze. W ramach podsumowania pobieżnie przyjrzymy się opóźnieniu między węzłem źródłowym i docelowym. Aby przybliżyć to zagadnienie, załóżmy, że między źródłowym i docelowym hostem znajduje się N–1 routerów. Przyjmijmy również, że sieć nie jest przeciążona (opóźnienia kolejkowania są pomijalne), opóźnienie każdego przetwarzającego routera i źródłowego hosta wynosi o_{przet}, szybkość wysyłania danych przez routery i źródłowy host jest równa R bitów/s, natomiast opóźnienie propagacji dla każdego łącza ma wartość o_{prop}. Po zsumowaniu opóźnień węzła uzyska się opóźnienie międzywęzłowe wyrażone następującą zależnością:

$$o_{\text{węz-węz}} = N(o_{\text{przet}}+o_{\text{trans}}+o_{\text{prop}})$$

W równaniu $o_{\text{trans}} = L/R$, gdzie L jest wielkością pakietu. Czytelnikowi pozostawiamy uogólnienie powyższego równania pod kątem różnorodnych opóźnień występujących w węzłach i istnienia w każdym z nich średniego opóźnienia kolejkowania.

1.6.3. Opóźnienia i trasy występujące w internecie

Aby od strony praktycznej przedstawić opóźnienie występujące w sieci komputerowej, skorzystamy z programu diagnostycznego *Traceroute*. Jest to proste narzędzie, które można uruchomić na dowolnym komputerze podłączonym do internetu. Gdy użytkownik zidentyfikuje docelowy host, program uaktywniony na źródłowym komputerze wyśle w jego kierunku wiele specjalnych pakietów. W trakcie zmierzania do docelowego hosta pakiety przechodzą przez serię routerów. Gdy router odbierze jeden z takich pakietów, do źródłowego hosta odeśle krótki komunikat zawierający nazwę i adres routera.

Załóżmy, że między źródłowym i docelowym hostem znajduje się N–1 routerów. Źródłowy host wyśle do sieci N specjalnych pakietów, z których każdy posiada adres hosta docelowego. Pakiety te są ponumerowane od 1 do N (pierwszy pakiet ma numer 1, natomiast ostatni N). Gdy n-ty router odbierze n-ty pakiet o numerze n, nie przekaże go w kierunku jego miejsca przeznaczenia, lecz wyśle komunikat do źródłowego hosta. Po odebraniu n-tego pakietu docelowy host również prześle komunikat źródłowemu hostowi. Źródłowy host rejestruje czas, jaki upłynie od chwili wysłania pakietu do momentu otrzymania powiązanego z nim komunikatu zwrotnego. Ponadto host zapisuje nazwę i adres routera lub docelowego hosta, który zwrócił komunikat. W ten sposób źródłowy host może odtworzyć trasę pokonaną przez pakiety transmitowane z miejsca źródłowego do docelowego. Poza tym źródłowy host jest w stanie określić opóźnienia całkowitej trasy generowane przez pośredniczące routery. Program *Traceroute* trzykrotnie wykonuje wyżej opisany proces. Oznacza to, że źródłowy host wysyła do docelowego $3 * N$ pakietów. Narzędzie zostało szczegółowo omówiono w dokumencie RFC 1393.

Przedstawimy teraz przykład wyników zwróconych przez program *Traceroute*. W tym przypadku trasa wiodła od źródłowego hosta *gaia.cs.umass.edu* (zlokalizowany na uniwersytecie w Massachusetts) do hosta *cis.poly.edu* (znajduje się na uniwersytecie w Brooklynie). Wynik ma postać sześciu kolumn. W pierwszej jest podana wyżej opisana wartość n określająca liczbę routerów znajdujących się na trasie pakietów. W drugiej kolumnie jest widoczna nazwa routera. W trzeciej kolumnie jest wstawiony

adres routera (formatu *xxx.xxx.xxx.xxx*), natomiast w trzech ostatnich opóźnienia trasy dla trzech prób. Jeśli źródłowy host od dowolnego routera otrzyma mniej niż trzy komunikaty (na skutek utraty pakietu w sieci), program *Traceroute* umieści znak gwiazdki tuż za numerem routera i nie poda dla niego trzech wartości opóźnień całkowitej trasy.

```
1 cs-gw (128.119.240.254) 1.009 ms  0.899 ms  0.993 ms
2 128.119.3.154 (128.119.3.154)  0.931 ms 0.441 ms  0.651 ms
3 border4-rt-gi-1-3.gw.umass.edu (128.119.2.194) 1.032 ms  0.484 ms  0.451 ms
4 acr1-ge-2-1-0.Boston.cw.net (208.172.51.129) 10.006 ms  8.150 ms  8.460 ms
5 agr4-loopback.NewYork.cw.net (206.24.194.104) 12.272 ms  14.344 ms  13.267 ms
6 acr2-loopback.NewYork.cw.net (206.24.194.62) 13.225 ms  12.292 ms  12.148 ms
7 pos10-2.core2.NewYork1.Level3.net (209.244.160.133) 12.218 ms  11.823 ms
    11.793 ms
8 gige9-1-52.hsipaccess1.NewYork1.Level3.net (64.159.17.39) 13.081 ms  11.556 ms
    13.297 ms
9 p0-0.polyu.bbnplanet.net (4.25.109.122) 12.716 ms  13.052 ms  12.786 ms
10 cis.poly.edu (128.238.32.126) 14.080 ms  13.035 ms  12.082 ms
```

W powyższym wyniku znajduje się dziewięć routerów zlokalizowanych między hostem źródłowym i docelowym. Wszystkie routery posiadają adres, natomiast część z nich nazwę. Przykładowo, nazwą routera 3. jest `border4-rt-gi-1-3.gw.umass.edu`, a jego adresem `128.119.2.194`. Po przyjrzeniu się wynikom podanym dla tego routera stwierdzi się, że pierwsze z trzech opóźnień całkowitej trasy związanych z przesłaniem pakietu między źródłowym hostem i routerem wynosi 1,03 ms. Pozostałe dwa opóźnienia mają wartości 0,48 i 0,45 ms. Wszystkie trzy opóźnienia całkowitej trasy uwzględniają wszystkie wcześniej omówione opóźnienia, takie jak opóźnienia transmisji, propagacji, przetwarzania przez router i kolejkowania. Ponieważ opóźnienie kolejkowania zmienia się w czasie, opóźnienie całkowitej trasy pakietu *n* wysłanego do routera *n* w rzeczywistości może być większe od opóźnienia całkowitej trasy pakietu *n*+1 przesłanego do routera *n*+1. W powyższym wyniku należy zauważyć, że w przypadku routera 6. opóźnienia okazują się być większe od opóźnień routera 7. Trzeba zauważyć, że jest to artefakt procesu pomiarowego. Każdy pakiet zmierzający do routera 7. musi koniecznie przejść przez router 6.

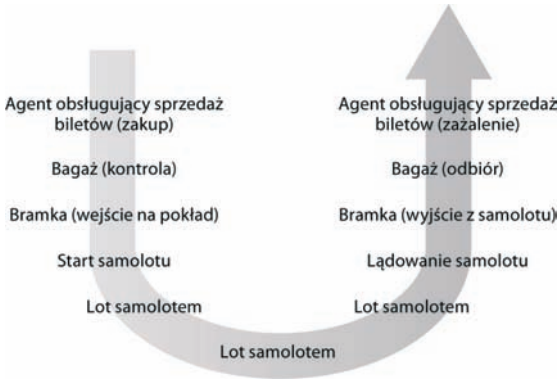
Czy Czytelnik chciałby sam wypróbować program *Traceroute*? *Gorąco* namawiamy do odwiedzenia strony internetowej (*http://www.traceroute.org*), na której zamieszczono interfejs udostępniający rozbudowaną listę źródeł umożliwiających śledzenie trasy pokonywanej przez pakiety. Po wybraniu źródła podaje się nazwę dowolnego hosta docelowego. Program *Traceroute* zajmie się całą resztą.

1.7. Warstwy protokołów i modele ich usług

Z naszej dotychczasowej analizy wynika, że *internet* jest wyjątkowo złożonym systemem. Przekonaliśmy się, że internet składa się z wielu elementów, takich jak wiele aplikacji i protokołów, różnego typu systemy końcowe i połączenia między nimi, routery i różnorodne odmiany nośników łączy. Biorąc pod uwagę tę złożoność, czy istnieje jakakolwiek szansa na zorganizowanie architektury sieci lub przynajmniej na jej omówienie? Na szczęście odpowiedź na obie części pytania jest twierdząca.

1.7.1. Architektura warstwowa

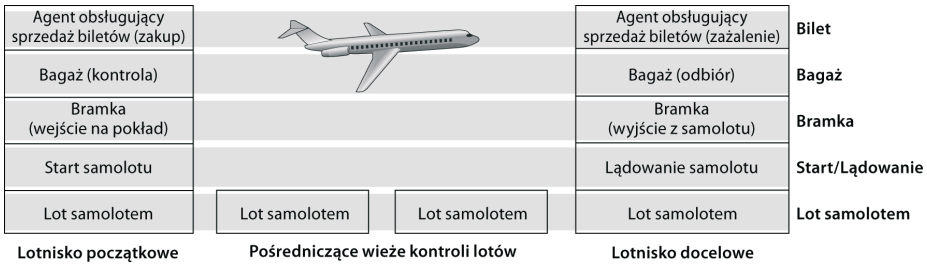
Zanim podejmiemy się próby uporządkowania przemyśleń dotyczących architektury internetu, przeanalizujmy analogię ze świata ludzi. W rzeczywistości nieustannie w codziennym życiu mamy do czynienia ze złożonymi systemami. Wyobraźmy sobie, że na przykład ktoś poprosi nas o opisanie systemu linii lotniczych. W jaki sposób określiłoby się strukturę charakteryzującą tak skomplikowany twór złożony z agentów obsługujących bilety, kontrolerów bagażu, personelu bramek, pilotów, samolotów, wieży kontroli ruchu i systemu obsługującego w skali globalnej trasy lotu samolotów? Jeden ze sposobów może polegać na opisaniu serii działań podejmowanych samemu lub przez innych, gdy zamierza się polecieć samolotem. W tym celu kupuje się bilet, oddaje bagaż do kontroli, udaje do bramki i ostatecznie wchodzi na pokład samolotu. Samolot startuje i jest kierowany do docelowego lotniska. Po wylądowaniu pasażer przemieszcza się do bramki i odbiera bagaż. Jeśli podróż nie była dla nas zadowalająca, agentowi obsługującemu bilety zgłasza się zażalenie (nic nie załatwiając). Zostało to zilustrowane na rysunku 1.15.



RYSUNEK 1.15. Działania podejmowane, gdy planuje się polecieć samolotem

Już można zauważyć, występują pewne analogie z siecią komputerową. Pasażer jest przewożony samolotem z miejsca początkowego do docelowego. Pakiet jest przesyłany z hosta źródłowego do docelowego za pośrednictwem internetu. Jednak nie do końca chodzi tu o taką analogię. Na rysunku 1.15 szukamy pewnej *struktury*. Przyglądając się mu, zauważymy, że na obu końcach strzałki znajduje się obsługa biletów. Z obu stron widoczny jest też punkt obsługi bagażu pasażerów posiadających bilet i bramki, do których podchodzą osoby mające bilet i będące już po kontroli bagażu. Pasażerowie, którzy znajdą się za bramką (posiadający bilet i sprawdzony bagaż oraz mający za sobą kontrolę przy bramce), będą mogli znaleźć się w samolocie. Samolot wystartuje, a następnie wyląduje. Dodatkowo osoby te będą uczestniczyły w locie samolotem. Można z tego wywnioskować, że funkcje przedstawione na rysunku 1.15 mogą być postrzegane *horyzontalnie* (rysunek 1.16).

Na rysunku 1.16 funkcje systemu linii lotniczych podzielono na warstwy, tworząc strukturę pozwalającą omówić podróż samolotem. Warto zauważyć, że każda warstwa połączona ze znajdującymi się niżej oferuje określoną funkcjonalność lub *usługę*. W warstwie agentów obsługujących bilety i niżej jest realizowany transfer pasażera na poziomie kas sprzedaży należących do linii lotniczej. W warstwie bagażowej i niżej



RYSUNEK 1.16. Poziome warstwy odpowiadające funkcjom systemu linii lotniczych

ma miejsce transfer pasażera i bagażu na poziomie punktu kontroli i odbioru bagażu. Warto zauważyć, że warstwa bagażowa świadczy usługę tylko tym pasażerom, którzy posiadają już bilet. W warstwie bramki jest wykonywana operacja transferu pasażera i bagażu na poziomie bramek zlokalizowanych w sali odlotów i przylotów. W przypadku warstwy startu i lądowania samolotu ma miejsce przemieszczanie ludzi i bagażu na poziomie pasa startowego. Każda warstwa świadczy usługę przez wykonanie w swoim obrębie określonych operacji (na przykład w warstwie bramki jest to wchodzenie pasażerów do samolotu i wychodzenie z niego) i skorzystanie z usług warstwy znajdującej się bezpośrednio poniżej (na przykład w przypadku warstwy bramki zostanie użyta usługa transferu pasażerów po pasie startowym oferowana przez warstwę startu i lądowania samolotu).

Architektura warstwowa umożliwia nam omówienie dobrze zdefiniowanego wybranego elementu dużego i złożonego systemu. Ponadto takie rozwiązanie przedstawia znaczną wartość, ponieważ zapewnia modularność. Dzięki temu znacznie łatwiej zmodyfikować sposób realizowania usługi oferowanej przez warstwę. Dopóki warstwa świadczy taką samą usługę warstwie znajdującej się nad nią i korzysta z identycznych usług oferowanych przez warstwę położoną niżej, reszta systemu nie ulegnie zmianie, gdy wprowadzi się w warstwie modyfikacje dotyczące sposobu realizowania usługi (warto zauważyć, że jest to coś zupełnie innego niż zmiana samej usługi!). Jeśli na przykład zmodyfikowano funkcje bramki (pasażerowie wchodzą do samolotu i wychodzą z niego w kolejności określanej przez ich wzrost), pozostała część systemu linii lotniczych będzie taka sama, ponieważ warstwa bramki nadal pełni identyczną rolę, czyli wpuszczanie ludzi do samolotu i wypuszczanie ich z niego. Po prostu po dokonaniu zmiany warstwa realizuje funkcję w inny sposób. W przypadku dużych i skomplikowanych systemów, które są nieustannie uaktualniane, możliwość zmiany sposobu świadczenia usługi bez wpływu na inne składniki systemu jest kolejną istotną zaletą architektury warstwowej.

Warstwy protokołów

Na razie wystarczy o liniach lotniczych. Skoncentrujmy się teraz na protokołach sieciowych. Aby zapewnić strukturę wykorzystywaną przy tworzeniu protokołów sieciowych, projektanci umieścili je w **warstwach** wraz z urządzeniami i oprogramowaniem sieciowym obsługującym protokoły. Każdy protokół należy do jednej z warstw, podobnie jak funkcje systemu linii lotniczej widoczne na rysunku 1.16. Ponownie interesują nas **usługi**, które warstwa świadczy warstwie znajdującej się ponad nią. Jest to tak zwany **model usług** warstwy. Tak jak w przypadku systemu linii lotniczej każda warstwa sieci oferuje swoje usługi przez wykonanie w swoim obrębie określonych operacji i skorzystanie z usług warstwy znajdującej się bezpośrednio poniżej. Przykładowo, usługi

świadczone przez warstwę *n* mogą uwzględniać niezawodne dostarczanie komunikatów z obrzeża jednej sieci do drugiej. Usługa może być realizowana przez zawodną usługę warstwy $n-1$ polegającą na dostarczaniu komunikatów między sieciami, a także przez funkcję warstwy *n*, która wykrywa utracone komunikaty i ponownie je przesyła.

Warstwa protokołów może być obsługiwana na poziomie programowym, sprzętowym lub tak i tak. Protokoły warstwy aplikacji, takie jak HTTP i SMTP, prawie zawsze są umieszczane w oprogramowaniu systemów końcowych. Tak samo jest w przypadku protokołów warstwy transportowej. Ponieważ warstwy fizyczna i łącza danych są odpowiedzialne za obsługę komunikacji realizowanej przy użyciu określonego łącza, zwykle są oferowane przez kartę sieciową (na przykład kartę Ethernet lub Wi-Fi) powiązaną z łączem. Warstwa sieci często jest obsługiwana przez kombinację sprzętu i oprogramowania. Warto zauważyć, że podobnie do funkcji systemu linii lotniczej o architekturze warstwowej, które były rozmieszczone na różnych lotniskach i centrach kontroli lotów tworzących system, protokół warstwy *n* jest *rozproszony* między systemami końcowymi, przełącznikami pakietów i innymi składnikami, z których jest złożona sieć. Oznacza to, że często w każdym z wymienionych komponentów sieciowych występuje element protokołu warstwy *n*.

Warstwy protokołów oferują korzyści pod względem pojęciowym i strukturalnym. Jak pokazaliśmy, warstwy pozwalają w strukturalny sposób omawiać składniki systemu. Modularność ułatwia uaktualnianie składników. Jednak trzeba wspomnieć, że niektórzy naukowcy i inżynierowie sieciowi zdecydowanie są przeciwni stosowaniu warstw [Wakeman 1992]. Potencjalną wadą warstw jest to, że jedna warstwa może powielać funkcje innej położonej niżej. Przykładowo, wiele stosów protokołów zapewnia funkcję usuwania błędów zarówno na poziomie łącza, jak i węzłów końcowych. Kolejną potencjalną wadą jest to, że funkcja jednej warstwy może wymagać danych (na przykład wartości znacznika czasu) dostępnych tylko w innej warstwie. Coś takiego przeczy idei oddzielania warstw.

Kombinacja protokołów różnych warstw jest nazywana **stosem protokołów**. Stos protokołów internetowych jest złożony z protokołów pięciu warstw — fizycznej, łącza danych, sieci, transportowej i aplikacji (rysunek 1.17).

RYSUNEK 1.17. Stos protokołów internetowych

Warstwa aplikacji

W warstwie aplikacji znajdują się aplikacje sieciowe i ich protokoły. Internetowa warstwa aplikacji uwzględnia wiele protokołów, takich jak HTTP (obsługuje żądania pobrania stron internetowych), SMTP (umożliwia przesyłanie wiadomości pocztowych) i FTP (pozwala na transfer plików między dwoma systemami końcowymi). Okaże się, że niektóre funkcje sieciowe, takie jak translacja przyjaznych dla użytkownika nazw internetowych systemów końcowych (np. *gaia.cs.umass.edu*) na 32-bitowe adresy

sieciowe, są realizowane przy użyciu protokołu warstwy aplikacji oferującego usługę DNS (*Domain Name System*). W rozdziale 2. Czytelnik przekona się, że w bardzo prosty sposób można tworzyć własne nowe protokoły warstwy aplikacji.

W definicji protokołu zamieszczonej w podrozdziale 1.1 podano, że komunikaty są wymieniane między rozproszonymi jednostkami stosującymi protokół. W książce komunikaty te będą nazywane komunikatami warstwy aplikacji.

Warstwa transportowa

Internetowa warstwa transportowa przesyła komunikaty warstwy aplikacji między klientem i serwerem tworzącym aplikację. Można wyróżnić dwa protokoły internetowej warstwy transportowej — TCP i UDP. Oba są w stanie transportować komunikaty warstwy aplikacji. Protokół TCP zapewnia swoim aplikacjom usługę zorientowaną na połączenie. Tego typu usługa gwarantuje dostarczenie komunikatów warstwy aplikacji do miejsca ich przeznaczenia, a także oferuje kontrolę przepływu (zapewnia zgodność szybkości transmisji nadawcy i odbiorcy). Ponadto protokół TCP dzieli długie komunikaty na krótsze segmenty i zapewnia mechanizm kontroli przeciążenia sieci, dzięki czemu w momencie jego wystąpienia źródłowy węzeł zmniejszy szybkość transmisji. Protokół UDP świadczy swoim aplikacjom usługę bezpołączeniową (omawiana w podrozdziale 1.2), która jest bardzo ograniczona. W książce pakiet warstwy transportowej określa się mianem **segmentu**.

Warstwa sieci

Internetowa warstwa sieci jest odpowiedzialna za przesyłanie swoich pakietów nazywanych *datagramami* od jednego hosta do drugiego. Protokół internetowej warstwy transportowej (TCP lub UDP) źródłowego hosta przekazuje segment i adres docelowy warstwie sieci, tak jak nadawca zostawiający na poczcie list z adresem odbiorcy. Warstwa sieci oferuje następnie segmentowi usługę polegającą na dostarczeniu go do warstwy transportowej docelowego hosta.

Internetowa warstwa sieci jest złożona z dwóch podstawowych składników. W warstwie znajduje się protokół definiujący pola datagramu, a także to, jak mogą one być przetwarzane przez systemy końcowe i routery. Protokołem tym jest słynny protokół IP. Istnieje tylko jeden taki protokół, który musi być używany przez wszystkie składniki internetu posiadające warstwę sieci. Internetowa warstwa sieci zawiera również protokoły routingu określające trasy, którymi podążają datagramy między źródłowymi i docelowymi węzłami. W internecie jest stosowanych wiele protokołów routingu. Jak wspomniano w podrozdziale 1.5, internet jest siecią sieci. W obrębie sieci jej administrator może uaktywnić dowolny żądany protokół routingu. Choć warstwa sieci zawiera zarówno protokół IP, jak i kilka protokołów routingu, często jest po prostu nazywana warstwą IP, co odzwierciedla fakt, że protokół IP pełni rolę spoiwa wiążącego internet.

Warstwa łącza danych

Internetowa warstwa sieci przesyła datagram za pośrednictwem serii przełączników pakietów (w przypadku internetu nazywanych *routerami*) znajdujących się między źródłowym i docelowym węzłem. Aby przemieścić pakiet z jednego węzła (host lub przełącznik pakietów) do kolejnego zlokalizowanego na trasie, warstwa sieci korzysta

z usług warstwy łącza danych. W szczególności warstwa sieci każdego węzła przekazuje datagram niżej położonej warstwie łącza danych, która dostarcza go do kolejnego węzła znajdującego się na trasie. Warstwa łącza danych tego węzła przemieszcza datagram do warstwy sieci.

To, jakie usługi są zapewniane przez warstwę łącza danych, zależy od jej określonego protokołu, który obsługuje łącze. Przykładowo, niektóre protokoły oferują niezawodne dostarczanie danych na poziomie łącza, które łączy ze sobą węzeł nadawczy i odbiorczy. Warto zauważyć, że taka usługa różni się od usługi niezawodnego dostarczania świadczonej przez protokół TCP, która dotyczy przesyłania danych między dwoma systemami końcowymi. Ethernet i PPP (*Point-to-Point Protocol*) są przykładami protokołów warstwy łącza danych. Ponieważ zwykle datagramy w celu przemieszczenia się od źródłowego do docelowego węzła muszą pokonać kilka łączy, w różnych łączach wchodzących w skład trasy mogą być obsługiwane przez inne protokoły warstwy łącza. Przykładowo, w jednym łączu datagram może być obsługiwany przez protokół Ethernet, natomiast w kolejnym przez protokół PPP. Przez każdy odmienny protokół warstwy łącza danych warstwie sieci zostanie zaoferowana inna usługa. W książce pakiety warstwy łącza danych są nazywane **ramkami**.

Warstwa fizyczna

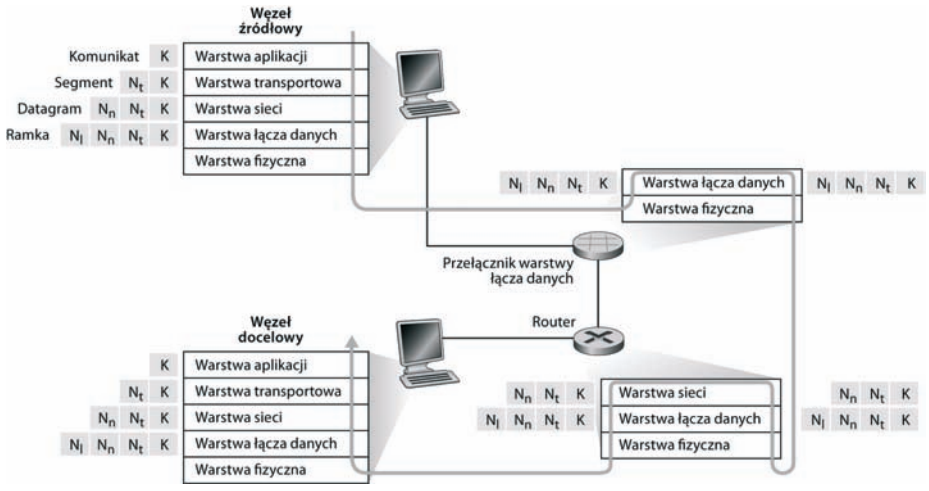
Zadaniem warstwy łącza danych jest przemieszczanie całych ramek od jednego elementu sieci do kolejnego z nim sąsiadującego, natomiast rolą warstwy fizycznej jest przesyłanie *poszczególnych bitów* ramki pomiędzy dwoma węzłami. Protokoły warstwy fizycznej są zależne od łącza, a także od rzeczywistej szybkości transmisji oferowanej przez nośnik łącza (na przykład skrętka lub światłowód jednomodowy). Przykładowo, standard Ethernet korzysta z wielu protokołów warstwy fizycznej. Poszczególne protokoły są powiązane ze skrętką, kablem koncentrycznym, światłowodem i innymi nośnikami. W każdym przypadku bit jest w inny sposób przesyłany łączem.

Wystarczy zapoznać się ze spisem treści książki, aby zauważyć, że z grubsza została ona zorganizowana w oparciu o warstwy stosu protokołów internetowych. Omawiając warstwy, skorzystamy z **metody „od góry do dołu”**, co oznacza, że najpierw przyjrzymy się warstwie aplikacji, a następnie kolejnym niższym warstwom.

1.7.2. Warstwy, komunikaty, segmenty, datagramy i ramki

Na rysunku 1.18 przedstawiono fizyczną ścieżkę, którą zmierzają dane w dół stosu protokołów nadawczego systemu końcowego, w górę i w dół stosów protokołów pośredniczących przełączników warstwy łącza danych i routerów, a następnie w górę stosu protokołów odbiorczego systemu końcowego. Jak się okaże w dalszej części książki, zarówno routery, jak i przełączniki warstwy łącza danych są przełącznikami pakietów. Podobnie jak systemy końcowe routery i przełączniki warstwy łącza danych swoje składniki programowe i sprzętowe organizują przy użyciu warstw. Jednak routery i przełączniki warstwy łącza danych nie wykorzystują *wszystkich* warstw stosu protokołów. Zwykle stosują jedynie jego dolne warstwy. Jak widać na rysunku 1.18, przełączniki warstwy łącza danych używają warstw 1. i 2. Routery wykorzystują warstwy od 1. do 3.

Oznacza to na przykład, że routery internetowe są w stanie zastosować protokół IP (należący do warstwy 3.), natomiast przełączniki warstwy łącza danych nie. Później okaże się, że tego typu przełączniki nie rozpoznają adresów IP, ale obsługują adresy warstwy drugiej, takie jak adresy ethernetowe. Warto zauważyć, że hosty wykorzystują wszystkie pięć warstw. Jest to zgodne ze stwierdzeniem, że architektura internetu większą część swojej złożoności lokalizuje na obrzeżach sieci.



RYSUNEK 1.18. Hosty, routery i przełączniki warstwy łącza danych. Każdy z tych węzłów posiada inny zestaw warstw, które są odzwierciedleniem ich różnego przeznaczenia

Na rysunku 1.18 zilustrowano też ważne zagadnienie, jakim jest **kapsułkowanie**. **Komunikat warstwy aplikacji** (litera K na rysunku 1.18) nadawczego hosta jest przekazywany warstwie transportowej. W najprostszym wariancie warstwa transportowa pobiera komunikat i dodaje do niego dodatkowe informacje (jest to zawartość tak zwanego nagłówka warstwy transportowej, oznaczonego na rysunku 1.18 symbolem N_t), które zostaną wykorzystane przez warstwę transportową węzła odbiorczego. Komunikat warstwy aplikacji i nagłówek warstwy transportowej razem tworzą **segment warstwy transportowej**. Segment kapsułkuje komunikat warstwy aplikacji. Dodatkowe informacje zawarte w nagłówku mogą uwzględniać dane umożliwiające warstwie transportowej strony odbiorczej dostarczenie komunikatu do odpowiedniej aplikacji, a także bity wykrywania błędów, które pozwalają węzłowi odbiorczemu stwierdzić, czy bity komunikatu zostały zmodyfikowane podczas przesyłania. Warstwa transportowa przekazuje następnie segment warstwie sieci, która dodaje własny nagłówek (na rysunku 1.18 symbol N_s) zawierający takie informacje jak adresy źródłowego i docelowego systemu końcowego. W efekcie jest tworzony **datagram warstwy sieci**. Datagram jest następnie przekazywany warstwie łącza danych, która (oczywiście!) dołączy swój nagłówek i utworzy **ramkę warstwy łącza danych**.

W tym miejscu przydatną analogią będzie wysłanie służbowej notatki za pośrednictwem ogólnie dostępnej usługi pocztowej. Sama notatka jest komunikatem warstwy aplikacji. Notatka jest umieszczana w biurowej kopercie. Na jej przedzie są podawane personalia odbiorcy notatki i dział, w którym pracuje. Informacje te ułatwią pracownikowi docelowego oddziału firmy zajmującemu się korespondencją przekazanie notatki bezpośrednio jej odbiorcy. Biurową kopertę zawierającą informacje nagłówkowe

(personalia i dział odbiorcy) i kapsułkującą komunikat warstwy aplikacji (notatka) można przyrównać do segmentu warstwy transportowej. Dział oddziału firmy zajmujący się korespondencją otrzymuje notatkę służbową, umieszcza ją w kopercie przeznaczonej do wysyłania za pośrednictwem publicznej poczty, a następnie podaje na kopercie adresy oddziału nadawcy i odbiorcy oraz przykleja znaczek. Tę kopertę można porównać z datagramem, który kapsułkuje segment warstwy transportowej (biurowa koperta wraz z jej zawartością). Z kolei segment kapsułkuje oryginalny komunikat (notatka). Poczta dostarcza zwykłą kopertę działowi oddziału firmy odbiorcy zajmującego się korespondencją. Pracownicy działu otwierają kopertę i wyjmują biurową kopertę z notatkę służbową. Koperta jest przekazywana właściwej osobie, która ją otwiera i wyciąga notatkę.

Proces kapsułkowania może być bardziej złożony od właśnie omówionego. Przykładowo, duży komunikat może zostać podzielony na wiele segmentów warstwy transportowej (one same mogą być podzielone na wiele datagramów warstwy sieci). Po stronie odbiorczej taki segment musi być odtworzony z tworzących go datagramów.

1.8. Historia sieci komputerowych i internetu

W podrozdziałach od 1.1 do 1.7 dokonano przeglądu technologii sieci komputerowych i internetu. Czytelnik powinien teraz dysponować wiedzą wystarczającą do tego, aby zrobić wrażenie na rodzinie i znajomych! Jeśli jednak chciałoby się naprawdę zabłysnąć na kolejnym przyjęciu, powinno się wzbogacić planowane przemówienie o ciekawostki dotyczące fascynującej historii internetu [Segaller 1998].

1.8.1. Rozwój technologii przełączania pakietów: 1961 – 1972

Branżę związaną z sieciami komputerowymi i fundamenty obecnej postaci internetu zaczęto tworzyć z początkiem lat 60., gdy sieć telefoniczna była dominującą w skali światowej siecią komunikacyjną. W podrozdziale 1.3 wspomniano, że sieć telefoniczna korzysta z przełączania obwodów w celu przesłania informacji od nadawcy do odbiorcy (jest to właściwa opcja, biorąc pod uwagę to, że głos między dwoma węzłami jest transmitowany ze stałą szybkością). Uwzględniając coraz większe znaczenie (i znaczny koszt) komputerów na początku lat 60. i pojawienie się komputerów z podziałem czasu, być może oczywistym (lepiej późno niż wcale!) było zastanowienie się nad tym, w jaki sposób połączyć ze sobą komputery, aby mogły być wykorzystywane przez użytkowników znajdujących się w różnych lokalizacjach geograficznych. Ruch generowany przez takich użytkowników miał raczej charakter *impulsowy*. Polegało to na tym, że po okresach aktywności, takich jak przesłanie polecenia do zdalnego komputera, następowały okresy bezczynności, podczas których oczekiwano na odpowiedź lub ją sprawdzano.

Trzy grupy badawcze z różnych miejsc świata, które nie wiedziały, czym pozostałe grupy się zajmują [Leiner 1998] zaczęły prace nad przełączaniem pakietowym, które miało stanowić efektywną i pewną alternatywę dla przełączania obwodów. Pierwsza

opublikowana praca na temat metod przełączania pakietów została napisana przez Leonarda Kleinrocka [Kleinrock 1961] [Kleinrock 1964], który wówczas ukończył studia na uczelni MIT. Wykorzystująca teorię kolejkowania praca Kleinrocka znakomicie demonstrowała efektywność technologii przełączania pakietów w przypadku źródeł danych o charakterze impulsowym. W 1964 r. Paul Baran [Baran 1964] rozpoczął w instytucie Rand Institute prace nad zastosowaniem przełączania pakietów do bezpiecznej transmisji głosu w sieciach wojskowych. Donald Davies i Roger Scantlebury, pracujący w angielskiej organizacji NPL (*National Physical Laboratory*), rozwijali swoje pomysły dotyczące przełączania pakietów.

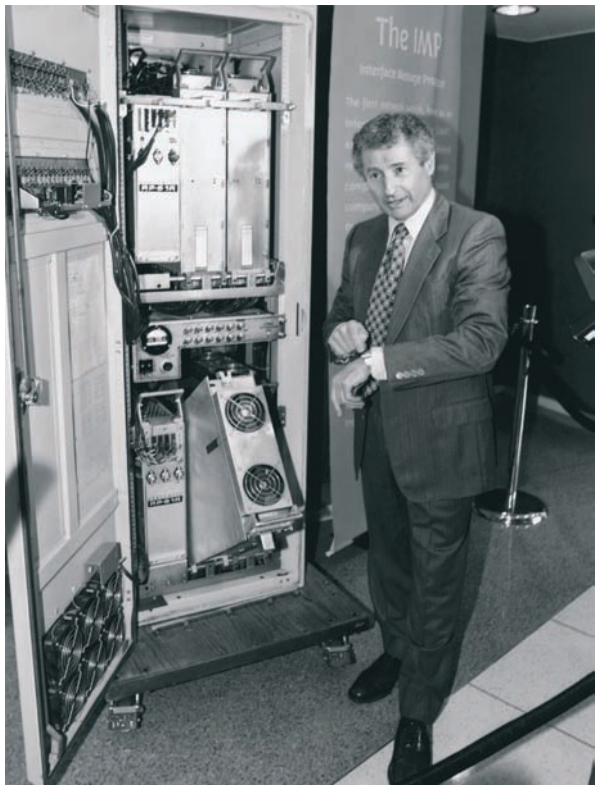
Wyniki prac zrealizowanych na uczelni MIT, w instytucie Rand i organizacji NPL stanowią fundament obecnego internetu. Z internetem jest też związana długa historia działań typu „zbudujmy to i zademonstrujmy", której początki sięgają wczesnych lat 60. J.C.R. Licklider [DEC 1990] i Lawrence Roberts, będący kolegami Kleinrocka z czasów studiów na uczelni MIT, kierowali projektem informatycznym realizowanym przez amerykańską agencję ARPA (*Advanced Research Projects Agency*). Roberts udostępnił ogólny plan sieci ARPAnet [Roberts 1967], która była pierwszą siecią komputerową z przełączaniem pakietów i bezpośrednim protoplastą obecnego internetu. Pierwsze przełączniki pakietów nazywano również **procesorami IMP** (*Interface Message Processor*). Kontrakt na wykonanie takich przełączników zdobyła firma BBN. Pierwszego maja 1969 r. na uczelni UCLA pod kierownictwem Kleinrocka zainstalowano pierwszy procesor IMP. Wkrótce kolejne trzy takie urządzenia umieszczono w instytucie SRI (*Stanford Research Institute*) uniwersytetu UC położonego w Santa Barbara, a także na Uniwersytecie Utah (rysunek 1.19). Nowo powstały prekursor internetu pod koniec 1969 r. liczył cztery węzły. Kleinrock wspomina, że pierwsze użycie sieci w celu zdalnego zalogowania się na komputerze znajdującym się w instytucie SRI przy użyciu komputera zlokalizowanego na uczelni UCLA zakończyło się zawieszeniem systemu [Kleinrock 2004].

W 1972 r. sieć ARPAnet w przybliżeniu liczyła 15 węzłów. W tym samym roku sieć została po raz pierwszy publicznie zademonstrowana przez Roberta Kahna podczas konferencji International Conference on Computer Communications. Stworzono pierwszy protokół NCP (*Network-Control Protocol*) [RFC 001], którego użyto w sieci ARPAnet do obsługi komunikacji między systemami końcowymi. Po udostępnieniu protokołu międzywęzłowego można było pisać aplikacje. W 1972 r. Ray Tomlinson z firmy BBN stworzył pierwszy program pocztowy.

1.8.2. Sieci zastrzeżone i łączenie sieci: 1972 – 1980

Pierwotnie sieć ARPAnet była pojedynczą zamkniętą siecią. Aby można było się skomunikować z hostem sieci ARPAnet, trzeba było go podłączyć do kolejnego procesora IMP. W pierwszej połowie lat 70. poza siecią ARPAnet pojawiły się następujące sieci z przełączaniem pakietów:

- ALOHAnet, sieć mikrofalowa łącząca uniwersytety wysp hawajskich [Abramson 1970], a także satelitarną sieć organizacji DARPA (*Defense Advanced Research Projects Agency*) [RFC 829] i sieci radiowe [Kahn 1978].
- Telenet, komercyjna sieć firmy BBN z przełączaniem pakietów, oparta na technologii sieci ARPAnet.



RYSUNEK 1.19. Pierwszy procesor IMP, obok którego stoi L. Kleinrock (Mark J. Terrill, AP/Wide World Photos)

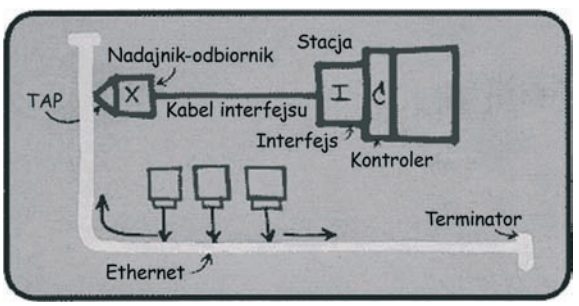
- Cyclades, francuska sieć z przełączaniem pakietów stworzona przez Louisa Pouzina [Think 2002].
- Sieci z podziałem czasu, takie jak Tymnet i GE Information Services, które istniały od końca lat 60. do początku lat 70. [Schwartz 1977].
- Sieć SNA firmy IBM (1969 – 1974), która przypominała sieć ARPAnet [Schwartz 1977].

Liczba sieci zwiększała się. Z perspektywy czasu można stwierdzić, że wówczas był odpowiedni moment na to, aby zacząć projektować ogólną architekturę umożliwiającą połączenie sieci. Pionierskie prace w tym zakresie (sponsorowane przez organizację DARPA), w zasadzie polegające na utworzeniu *sieci sieci*, zostały wykonane przez Vintona Cerfa i Roberta Kahna [Cerf 1974]. W celu opisania rezultatu tych prac wymyślono termin *internetting*.

Te architektoniczne podstawy zostały uwzględnione w protokole TCP. Jednak pierwsze wersje protokołu TCP dość znacznie różniły się od jego obecnej postaci. W pierwszych wersjach protokołu połączono mechanizm niezawodnego sekwencyjnego dostarczania danych przy użyciu retransmisji realizowanej przez system końcowy (nadal stanowi część obecnej wersji protokołu) z funkcjami przekazywania (aktualnie są wykonywane przez protokół IP). Wczesne doświadczenia związane z protokołem

TCP, a także stwierdzenie potrzeby użycia zawodnej usługi transportu między węzłami bez kontroli przepływu w przypadku zastosowań takich jak pakietowe przesyłanie głosu, spowodowało wydzielenie protokołu IP z protokołu TCP i stworzenie protokołu UDP. Pod koniec lat 70. istniały już koncepcje dotyczące trzech podstawowych protokołów internetowych, które są stosowane obecnie (TCP, UDP i IP).

Poza badaniami związanymi z internetem prowadzonymi przez organizację DARPA realizowanych było wiele innych istotnych działań związanych z sieciami. Na Hawajach Norman Abramson stworzył pakietową sieć radiową ALOHAnet, która umożliwiała komunikowanie się wielu zdalnym jednostkom znajdującym się na wyspach hawajskich. Protokół ALOHA [Abramson 1970] był pierwszym protokołem wielodostępowym zezwalającym użytkownikom rozproszonym w różnych geograficznych lokalizacjach na współużytkowanie pojedynczego nośnika transmisyjnego (częstotliwość radiowa). Projektując protokół Ethernet [Metcalfe 1976] przeznaczony dla kablowych sieci o wspólnym nośniku transmisyjnym, Metcalfe i Boggs korzystali z wyników prac Abramsona dotyczących jego protokołu wielodostępowego (rysunek 1.20). Interesujące jest to, że protokół Ethernet został stworzony przez Metcalfe'a i Boggsa, ponieważ zależało im na połączeniu ze sobą wielu komputerów PC, drukarek i współużytkowanych dysków [Perkins 1994]. Dwadzieścia pięć lat temu, jeszcze sporo przed rewolucją związaną z komputerami PC i eksplozją sieci, Metcalfe i Boggs stworzyli podwaliny pod obecnie stosowane sieci lokalne. Technologia Ethernet stanowiła ważny krok również w przypadku łączenia sieci. Każda lokalna sieć ethernetowa była niezależną siecią. Wraz z rozpowszechnianiem się sieci lokalnych coraz bardziej istotna stawała się potrzeba połączenia ich. Ethernet, protokół ALOHA i inne technologie sieci lokalnych zostaną szczegółowo omówione w rozdziale 5.



RYSUNEK 1.20. Oryginalna koncepcja Ethernetu opracowana przez Metcalfe'a

1.8.3. Popularyzacja sieci: 1980 – 1990

Pod koniec lat 70. do sieci ARPAnet było podłączonych w przybliżeniu dwieście hostów. Z końcem lat 80. liczba hostów podłączonych do publicznego internetu (konfederacja sieci bardzo przypominająca obecny internet) osiągnęła sto tysięcy. Na lata 80. przypadł okres niebywałego rozwoju sieci.

W znacznej części rozwój wynikał z kilku różnych działań mających na celu stworzenie sieci komputerowych łączących ze sobą uniwersytety. Sieć BITNET kilku wyższym uczelniom znajdującym się w północno-wschodniej części Stanów Zjednoczonych zapewniała usługę poczty elektronicznej i transferu plików. Sieć CSNET

(*Computer Science Network*) została zbudowana w celu umożliwienia komunikacji naukowcom uniwersyteckim, którzy nie dysponowali dostępem do sieci ARPAnet. W 1986 r. stworzono sieć NSFNET, aby udostępnić centra superkomputerowe sponsorowane przez organizację NSF. Początkowo szkielet sieci NSFNET oferował szybkość 56 kb/s. Jednak pod koniec dekady szybkość sieci NSFNET wynosiła już 1,5 Mb/s, dzięki czemu pełniła ona rolę szkieletu łączącego sieci regionalne.

W sieci ARPAnet stosowano wiele ostatecznych elementów obecnej architektury internetu. Pierwszego stycznia 1983 r. w sieci ARPAnet oficjalnie wdrożono nowy standard protokołów TCP/IP przeznaczonych dla hostów, który zastąpił protokół NCP. Przejście z protokołu NCP na protokoły TCP/IP [RFC 801] miało miejsce 14 czerwca 1983 r. Tego dnia wszystkie hosty musiały zacząć korzystać z protokołów TCP/IP. Pod koniec lat 80. w protokole TCP wprowadzono istotne rozszerzenia uwzględniające kontrolę przeciążenia realizowaną na poziomie hosta [Jacobson 1988]. Opracowano również system DNS [RFC 1034] służący do translacji przyjaznych dla użytkownika nazw internetowych (na przykład *gaia.cs.umass.edu*) na odpowiadające im 32-bitowe adresy IP.

Równolegle z rozwojem sieci ARPAnet (w zdecydowanej części znajdującej się w Stanach Zjednoczonych) na początku lat 80. Francja rozpoczęła realizację ambitnego projektu Minitel, który miał umożliwić uzyskanie dostępu do sieci z każdego domu. System tworzony w ramach projektu Minitel sponsorowanego przez francuski rząd składał się z publicznej sieci z przełączaniem pakietów (oparta na zestawie protokołów X.25 korzystających z wirtualnych obwodów), serwerów i tanich terminali z wbudowanymi modemami o niewielkiej szybkości. W 1984 r. system ten okazał się wielkim sukcesem, gdy rząd francuski za darmo przekazywał terminal każdemu obywatelowi, który chciał go użyć. Wśród witryn Minitela istniały witryny darmowe świadczące usługi, takie jak książka telefoniczna, a także prywatne witryny, które od każdego użytkownika pobierały opłatę uzależnioną od stopnia korzystania z usług. W szczytowej fazie rozwoju, który przypadł na połowę lat 90., system Minitel oferował ponad 20 000 usług, począwszy od domowej bankowości, a skończywszy na specjalistycznych bazach wyszukujących. System był używany przez ponad 20% mieszkańców Francji i każdego roku generował przychód przekraczający 1 miliard dolarów. Ponadto system Minitel pozwolił utworzyć 10 000 miejsc pracy. Dziesięć lat przed tym, zanim większość Amerykanów w ogóle usłyszała o internecie, system Minitel był dostępny w sporej części francuskich domostw.

1.8.4. Eksplozja internetu: lata 90.

Lata 90. przyniosły kilka wydarzeń, które symbolizują ciągły rozwój i bliską już komercjalizację internetu. Sieć ARPAnet, będąca przodkiem internetu, przestała istnieć. W latach 80. sieci MILNET i Defense Data Network zostały powiększone w celu przesyłania większości danych powiązanych z amerykańskim departamentem obrony. Sieć NSFNET zaczęła pełnić rolę sieci szkieletowej łączącej amerykańskie sieci regionalne i sieci państw znajdujących się na różnych kontynentach. W 1991 r. sieć NSFNET zwiększyła ograniczenia dotyczące zastosowania jej do celów komercyjnych. Sieć NSFNET zlikwidowano w 1995 r., gdy szkielet internetu zaczął być obsługiwany przez komercyjnych dostawców usług internetowych.

Jednak głównym wydarzeniem lat 90. było pojawienie się technologii WWW (*World Wide Web*), która sprawiła, że internet zagościł w domach i firmach milionów osób z całego świata. Technologia ta pełniła też rolę platformy umożliwiającej realizowanie i wdrażanie setek nowych zastosowań, takich jak handel akcjami i bankowość elektroniczna, usługi strumieniowej transmisji multimediów i pobierania informacji. W celu zapoznania się z początkami technologii WWW należy skorzystać ze źródła [W3C 1995].

Technologia WWW została opracowana przez Tima Berners-Lee (pracownik CERN-u) między 1989 i 1991 r. [Berners-Lee 1989] w oparciu o pomysły zawarte we wcześniejszych pracach dotyczących hipertekstu, które były realizowane w latach 40. przez Busha [Bush 1945] i Teda Nelsona [Ziff-Davis 1998]. Począwszy od lat 60., Berners-Lee i jego współpracownicy stworzyli pierwotne wersje języka HTML, protokołu HTTP, serwera WWW i przeglądarki, czyli czterech kluczowych składników technologii WWW. Oryginalne przeglądarki używane w CERN-ie oferowały jedynie interfejs wiersza poleceń. Mniej więcej z końcem 1992 r. czynnych było około 200 serwerów WWW. Taka kolekcja serwerów stanowiła jedynie zwiastun tego, co miało nastąpić. W tym czasie kilku naukowców projektowało przeglądarki internetowe z graficznym interfejsem użytkownika. Wśród nich był Marc Andreesen, który kierował pracami nad popularną graficzną przeglądarką Mosaic. W 1994 r. Marc Andreesen i Jim Clark stworzyli firmę Mosaic Communications, która później przyjęła nazwę Netscape Communications Corporation [Cusumano 1998] [Quittner 1998]. W 1995 r. studenci uniwersytetów w codziennej pracy używali przeglądarek Mosaic i Netscape do wyświetlania stron internetowych. Mniej więcej w tym czasie duże i małe firmy zaczęły udostępniać serwery WWW i za ich pośrednictwem prowadzić działalność handlową. W 1996 r. Microsoft zaczął prace nad własnymi przeglądarkami. Zapoczątkowało to rywalizację między przeglądarkami firm Netscape i Microsoft, którą kilka lat później wygrała ta druga [Cusumano 1998].

W drugiej połowie lat 90. nastąpił okres niebywałego rozwoju internetu i wprowadzania związanych z nim innowacji. W tym czasie duże korporacje i tysiące nowych firm tworzyło produkty i usługi internetowe. Internetowa poczta elektroniczna oferowała bogate w możliwości przeglądarki wiadomości wyposażone w książki adresowe, załączniki, odnośniki i mechanizmy przesyłania multimediów. Pod koniec drugiego tysiąclecia internet obsługiwał setki popularnych zastosowań, uwzględniających następujące cztery najważniejsze:

- poczta elektroniczna umożliwiająca wysyłanie załączników i udostępniająca wiadomości z poziomu przeglądarki internetowej;
- technologia WWW pozwalająca na przeglądanie stron internetowych i prowadzenie handlu elektronicznego;
- komunikatory zawierające listy kontaktowe (pionierem było narzędzie ICQ);
- wymiana plików (na przykład MP3) typu sieci równoległych (pionierem było narzędzie Napster).

Interesujące jest to, że pierwsze dwa z wyżej wymienionych zastosowań wywodzą się ze społeczności naukowców, natomiast dwa pozostałe zostały zapoczątkowane przez kilku młodych przedsiębiorców.

Okres od 1995 do 2001 r. stoi pod znakiem bardzo dynamicznego wykorzystania internetu przez rynki finansowe. Zanim jeszcze zaczęły przynosić zyski, setki nowych firm internetowych robiły publiczne oferty i ich akcje pojawiały się na giełdach papierów

wartościowych. Wiele spółek, które nie miały żadnych znaczących przychodów, wyceniano na miliardy dolarów. „Bańka internetowa" pękła w latach 2000 – 2001 i w efekcie wiele nowych firm zbankrutowało. Niemniej jednak kilka spółek, takich jak Microsoft, Cisco, AOL, Yahoo, e-Bay i Amazon, odniosło sukces w branży internetowej (nawet jeśli ceny ich akcji ucierpiały podczas krachu).

W latach 90. osoby zajmujące się badaniami i pracami rozwojowymi związanymi z siecią poczyniły spore postępy w zakresie bardzo szybkich routerów i routingu (rozdział 4.), a także sieci lokalnych (rozdział 5.). Przedstawiciele branży technicznej borykali się z problemami polegającymi na zdefiniowaniu i wdrożeniu modelu usług internetowych na potrzeby danych wymagających zapewnienia przetwarzania w czasie rzeczywistym, tak jak na przykład w przypadku strumieniowych aplikacji multimedialnych (rozdział 7.). Bardzo istotna stała się również potrzeba zabezpieczenia i zarządzania infrastrukturą internetową (rozdziały 8. i 9.), gdy rozpowszechniły się zastosowania związane z handlem elektronicznym, a ponadto internet zaczął pełnić rolę podstawowego składnika światowej infrastruktury telekomunikacyjnej.

1.8.5. Ostatnie dokonania

W dalszym ciągu jest utrzymywane duże tempo wprowadzania innowacji w sieciach komputerowych. Postępy są czynione we wszystkich dziedzinach uwzględniających projektowanie nowych aplikacji, dystrybucję treści, telefonię internetową, zwiększanie szybkości transmisji w sieciach lokalnych i szybsze routery. Jednak na szczególną uwagę zasługują trzy rzeczy — rozpowszechnienie bardzo szybkich sieci dostępowych (w tym sieci bezprzewodowych), zabezpieczenia i technologie P2P.

Jak wspomniano w podrozdziale 1.4, zwiększająca się liczba prywatnych użytkowników dysponujących szerokopasmowym dostępem internetowym uzyskiwanym za pośrednictwem modemu kablowego lub DSL umożliwia wprowadzenie wielu nowych zastosowań multimedialnych, takich jak strumieniowa transmisja na żądanie wysokiej jakości wideo i interaktywne wideokonferencje. Coraz większa wszechobecność bardzo szybkich (11 Mb/s i więcej) publicznych sieci Wi-Fi i średniej szybkości (kilkaset kilobitów) sieci telefonii komórkowej udzielającej dostępu do internetu powoduje, że nie tylko można cały czas być podłączonym do sieci, ale również korzystać z ekscytującego nowego zestawu usług powiązanych z określoną lokalizacją. Sieci mobilne i bezprzewodowe zostaną omówione w rozdziale 6.

Po serii ataków typu DoS (*Denial of Service*), których ofiarą pod koniec 1990 r. padły znane serwery WWW, i rozpowszechnieniu się ataków przeprowadzanych za pomocą robaków (na przykład robaka o nazwie Blaster) infekujących systemy końcowe i przeciążających sieć bezpieczeństwo sieci stało się ogromnie istotną kwestią. Tego typu ataki spowodowały powstanie systemów wykrywających włamania, które zapewniają wczesne ostrzeganie o ataku. Ponadto zaczęto stosować firewalle, które przed wpuszczeniem danych do sieci dokonywały ich filtrowania, a także metodę śledzenia adresów IP pozwalającą zlokalizować źródło ataku. W rozdziale 8. omówiono kilka ważnych zagadnień związanych z zabezpieczeniami.

Ostatnią innowacją, na którą zwracamy szczególną uwagę, jest technologia P2P. Wykorzystuje ona zasoby komputerów użytkowników, takie jak dyski, dane, cykle procesora. Aplikacje P2P są w dużej mierze niezależne od centralnych serwerów. Zwykle komputery użytkowników (czyli węzły) dysponują stałym połączeniem. W czasie pisania tego rozdziału program KaZaA był najpopularniejszą aplikacją P2P służącą

do wymiany plików. Do sieci obsługiwanej przez ten program zazwyczaj jest podłączonych ponad 4 miliony systemów końcowych, natomiast generowany przez nią ruch stanowi od 20 do 50% całego ruchu w internecie [Sariou 2002].

1.9. Podsumowanie

W niniejszym rozdziale zawarliśmy ogromną ilość materiału! Przyjrzeliśmy się różnym sprzętowym i programowym składnikom tworzącym internet w szczególności i ogólnie sieci komputerowe. Zaczęliśmy od omówienia obrzeża sieci, omawiając systemy końcowe i aplikacje, a także usługę transportową oferowaną aplikacjom uruchamianym na systemach końcowych. Posługując się jako przykładem sieciowymi zastosowaniami rozproszonymi, zaprezentowaliśmy pojęcie protokołu, czyli kluczowego zagadnienia związanego z sieciami. W dalszej kolejności zagłębiliśmy się bardziej w strukturę sieci, analizując jej rdzeń, identyfikując przełączanie pakietów i obwodów jako dwie podstawowe metody transportowania danych w sieciach telekomunikacyjnych. Dalej przedstawiliśmy zalety i wady każdej metody. Później przyjrzeliśmy się najniżej położonym elementom sieci (z punktu widzenia architektury), czyli technologiom warstwy łącza danych i fizycznym nośnikom spotykanym zwykle w sieciach dostępowych. Zaprezentowaliśmy też strukturę globalnego internetu, stwierdzając, że jest to sieć sieci. Czytelnicy dowiedzieli się, że hierarchiczna struktura internetu składająca się z dostawców ISP niższych i wyższych warstw może być skalowana, aby możliwe było uwzględnienie w niej tysięcy sieci.

W drugiej części tego wprowadzającego rozdziału omówiliśmy kilka zagadnień odgrywających ważną rolę w przypadku sieci komputerowych. Najpierw przeanalizowaliśmy przyczyny opóźnień i utraty pakietów w sieci z przełączaniem pakietów. Utworzyliśmy proste modele ilościowe dotyczące opóźnienia transmisji, propagacji i kolejkowania. Będziemy z nich intensywnie korzystać, rozwiązując problemy zamieszczone w różnych miejscach książki. W dalszej kolejności przyjrzeliśmy się warstwom protokołów i modelom usług, będącym kluczowymi elementami architektury sieci, do których będziemy się w książce odwoływali. Na zakończenie wprowadzenia do sieci przedstawiliśmy w skrócie historię sieci komputerowych. Pierwszy rozdział sam w sobie stanowi minikurs z zakresu sieci komputerowych.

W pierwszym rozdziale zawarliśmy naprawdę niebywałą ilość podstawowych informacji! Jeśli Czytelnik czuje się trochę przytłoczony, nie ma powodu do obaw. W kolejnych rozdziałach ponownie znacznie bardziej szczegółowo zostaną omówione wszystkie zagadnienia (to jest obietnica, a nie groźba!). Mamy nadzieję, że po przeczytaniu tego rozdziału Czytelnik orientuje się w składnikach tworzących sieć i terminologii z nią związanej (należy śmiało powracać do niniejszego rozdziału), a także ma ochotę na dalsze pogłębianie wiedzy na temat sieci. Temu mają służyć pozostałe rozdziały książki.

Struktura książki

Przed rozpoczęciem każdej podróży zawsze powinno się spojrzeć na mapę, aby zapoznać się z głównymi drogami i pobliskimi połączeniami. W przypadku podróży, w którą my się wybierzemy, ostatecznym celem jest pełne zrozumienie sieci komputerowych. Nasza mapa drogowa jest złożona z następujących rozdziałów książki:

1. Sieci komputerowe i internet.
2. Warstwa aplikacji.
3. Warstwa transportowa.
4. Warstwa sieci.
5. Warstwa łącza danych i sieci lokalne.
6. Technologie bezprzewodowe i mobilne.
7. Multimedia i sieci.
8. Bezpieczeństwo sieci komputerowych.
9. Zarządzanie siecią.

Rozdziały od 2. do 5. są podstawowymi rozdziałami książki. Należy zwrócić uwagę na to, że rozdziały te poświęcono czterem górnym warstwom 5-warstwowego stosu protokołów internetowych. Każdy rozdział dotyczy jednej warstwy. Dodatkowo należy zauważyć, że nasza podróż rozpocznie się od najwyżej położonej warstwy stosu protokołów, czyli od warstwy aplikacji, a następnie będzie przebiegała przez kolejne niższe warstwy. Powodem obrania takiego kierunku podróży (od góry do dołu) jest to, że po zrozumieniu aplikacji będzie łatwiej zrozumieć usługi sieciowe wymagane przez aplikacje. Przyjrzymy się różnym metodom, które można zastosować do wdrażania takich usług przy użyciu architektury sieciowej. Omówienie aplikacji w pierwszej kolejności będzie stanowiło motywację do zapoznania się z resztą książki.

W drugiej części książki (rozdziały od 6. do 9.) skupiono się na czterech wyjątkowo istotnych (i w pewnym stopniu niezależnych) zagadnieniach dotyczących nowoczesnych sieci komputerowych. W rozdziale 6. omówimy technologie bezprzewodowe i mobilne, takie jak sieci lokalne Wi-Fi, GSM i Mobile IP. W rozdziale 7. („Multimedia i sieci") przyjrzymy się zastosowaniom audio-wideo, takim jak telefonia internetowa, wideokonferencje i transmisja strumieniowa magazynowanych multimediów. Zastanowimy się też, w jaki sposób mogą być projektowane sieci z przełączaniem pakietów, aby były w stanie zapewnić zastosowaniom audio-wideo stałą jakość usług. W rozdziale 8. („Bezpieczeństwo sieci komputerowych") najpierw przeanalizujemy podstawy szyfrowania i zabezpieczeń sieciowych, a następnie sprawdzimy, jak teoria jest wykorzystywana w praktyce w wielu różnych zastosowaniach internetowych. W ostatnim rozdziale („Zarządzanie siecią") zaprezentujemy kluczowe zagadnienia dotyczące zarządzania siecią, a także podstawowe protokoły internetowe, które to umożliwiają.

Problemy do rozwiązania i pytania

Rozdział 1. Pytania kontrolne

PODROZDZIAŁY 1.1 – 1.5

1. Jaka jest różnica między hostem i systemem końcowym? Podaj typy systemów końcowych. Czy serwer WWW jest systemem końcowym?
2. Termin *protokół* jest często używany do opisania stosunków dyplomatycznych. Podaj przykład takiego protokołu.

3. Czym jest aplikacja klienta, a czym aplikacja serwera? Czy aplikacja serwera żąda usług od aplikacji klienta i korzysta z nich?
4. Jakie są dwa typy usług transportowych zapewnianych przez internet swoim aplikacjom? Jakie są cechy charakterystyczne każdej z usług?
5. Stwierdzono, że kontrola przepływu i kontrola przeciążenia są sobie równorzędne. Czy jest to prawda w przypadku internetowej usługi zorientowanej na połączenie? Czy cele postawione kontroli przepływu i kontroli przeciążenia są takie same?
6. Podaj krótki ogólny opis tego, w jaki sposób internetowa usługa zorientowana na połączenie zapewnia niezawodny transport.
7. Jakie w porównaniu z siecią z przełączaniem pakietów są zalety sieci z przełączaniem obwodów? Jakie w porównaniu z multipleksowaniem FDM są zalety multipleksowania TDM sieci z przełączaniem obwodów?
8. Dlaczego mówi się, że przełączanie pakietów wykorzystuje multipleksowanie statystyczne? Porównaj je z multipleksowaniem TDM.
9. Załóżmy, że między hostem nadawczym i odbiorczym znajduje się dokładnie jeden przełącznik pakietów. Szybkości transmisji między hostem nadawczym i przełącznikiem oraz między przełącznikiem i hostem odbiorczym wynoszą odpowiednio R_1 i R_2. Przyjmując, że przełącznik stosuje przełączanie pakietów z buforowaniem, jakie będzie całkowite opóźnienie międzywęzłowe w przypadku wysyłania pakietu o długości L (należy zignorować opóźnienia kolejkowania, propagacji i przetwarzania)?
10. Co rozumie się przez informacje o stanie połączenia w przypadku sieci wirtualnych obwodów? Jeśli w przełączniku znajdującym się w takiej sieci połączenia są nawiązywane i przerywane ze średnią szybkością jednego połączenia w ciągu milisekundy, z jaką szybkością musi być modyfikowana tabela przekierowań przełącznika?
11. Załóżmy, że tworzymy standard dla nowego typu sieci z przełączaniem pakietów. Trzeba zdecydować, czy na potrzeby routingu sieć będzie używała wirtualnych obwodów czy datagramów. Jakie są plusy i minusy zastosowania wirtualnych obwodów?
12. Wymień sześć technologii sieci dostępowych. Każdą z nich zaklasyfikuj do jednej z kategorii: sieci dostępowe prywatnych użytkowników, sieci dostępowe firm i bezprzewodowe sieci dostępowe.
13. Jaka jest kluczowa różnica między dostawcą ISP pierwszej i drugiej warstwy?
14. Czym się różnią punkty POP i NAP?
15. Czy szybkość transmisji oferowana przez technologię HFC jest dedykowana czy dzielona między wszystkimi użytkownikami? Czy w kanale pobierania urządzenia HFC mogą wystąpić kolizje? Dlaczego do tego może dojść lub dlaczego nie?
16. Jaka jest szybkość transmisji ethernetowych sieci lokalnych? Biorąc pod uwagę określoną szybkość transmisji, czy każdy użytkownik sieci lokalnej może cały czas przesyłać dane z taką szybkością?
17. Jakie fizyczne nośniki mogą być wykorzystane w przypadku technologii Ethernet?
18. Na potrzeby sieci dostępowych prywatnych użytkowników używa się modemów telefonicznych, a także modemów HFC i ADSL. Dla każdej z tych technologii dostępowej określ zakres szybkości transmisji i stwierdź, czy szybkość jest dedykowana, czy współużytkowana.

PODROZDZIAŁY 1.6 – 1.8

19. Pod uwagę weźmy przesłanie pakietu z hosta nadawczego do odbiorczego za pośrednictwem ustalonej trasy. Wymień składniki opóźnienia międzywęzłowego. Które ze składowych opóźnień są stałe, a które zmienne?
20. Podaj pięć zadań, które może zrealizować warstwa. Czy możliwe jest, aby jedno lub więcej zadań mogło zostać wykonanych przez dwie lub więcej warstw?
21. Jakich pięć warstw wchodzi w skład internetowego stosu protokołów? Jakie są podstawowe funkcje każdej z warstw?
22. Czym jest komunikat warstwy aplikacji? Co to jest segment warstwy transportowej? Czym jest datagram warstwy sieci, a czym ramka warstwy łącza danych?
23. Które warstwy internetowego stosu protokołów są używane przez router? Które warstwy są wykorzystywane przez przełącznik warstwy łącza danych? Które warstwy są stosowane przez host?

Problemy

1. Utwórz i opisz protokół warstwy aplikacji, który będzie pośredniczył między zautomatyzowanym bankomatem i komputerem znajdującym się w centrali banku. Protokół powinien umożliwiać weryfikację karty i hasła użytkownika, sprawdzanie salda konta (przechowywane na komputerze w centrali) i realizowanie operacji na koncie (wypłacanie pieniędzy użytkownikowi). Elementy protokołu powinny obsługiwać częsty przypadek, w którym na koncie nie ma wystarczającej kwoty, aby zrealizować wypłatę. Zdefiniuj protokół przez podanie wymienianych komunikatów i działań podejmowanych przez bankomat lub komputer w centrali banku podczas wysyłania i odbierania komunikatów. Narysuj schemat przedstawiający funkcjonowanie protokołu w przypadku zwykłej bezbłędnej operacji wypłaty (skorzystaj z diagramu pokazanego na rysunku 1.2). Wyraźnie określ założenia protokołu dotyczące używanej przez niego międzywęzłowej usługi transportowej.
2. Weźmy pod uwagę aplikację, która transmituje dane ze stałą szybkością (przykładowo, nadawca generuje N-bitową jednostkę danych co k jednostek czasu, gdzie k jest wartością niewielką i niezmienną). Ponadto, gdy aplikacja zostanie uruchomiona, będzie aktywna przez stosunkowo długi okres. Udziel odpowiedzi na następujące pytania i krótko je uzasadnij:
    a. Czy w przypadku tej aplikacji bardziej odpowiednia byłaby sieć z przełączaniem pakietów czy z przełączaniem obwodów? Dlaczego jedna z sieci będzie lepsza?
    b. Załóżmy, że jest używana sieć z przełączaniem pakietów, w której jedyne przesyłane dane są generowane przez wyżej opisaną aplikację. Dodatkowo przyjmijmy, że suma szybkości przesyłania danych aplikacji jest mniejsza od oferowanej przez każde łącze. Czy będzie potrzebna jakiegoś typu kontrola przeciążenia? Jeśli tak, to dlaczego?

3. Weźmy pod uwagę sieć z przełączaniem obwodów pokazaną na rysunku 1.5. W jej przypadku na każde łącze przypadało n obwodów.
   a. Jaka jest maksymalna liczba jednoczesnych połączeń, które w danej chwili mogą być aktywne w sieci?
   b. Przyjmijmy, że wszystkie połączenia są nawiązywane między przełącznikiem widocznym w górnym lewym narożniku rysunku 1.5 i przełącznikiem zlokalizowanym w dolnym prawym narożniku. Jaka jest maksymalna liczba jednoczesnych aktywnych połączeń?
4. Zapoznaj się z analogią karawany samochodów przedstawionej w podrozdziale 1.6. Ponownie przyjmijmy, że prędkość propagacji (przemieszczania) wynosi 100 km/h.
   a. Załóżmy, że karawana pokonuje dystans 200 kilometrów. Trasa podróży zaczyna się od wejścia jednej bramki, przebiega przez drugą bramkę i kończy tuż przed trzecią bramką. Jakie jest opóźnienie międzywęzłowe?
   b. Powtórnie rozpatrz przypadek przedstawiony w punkcie (a), ale przyjmij, że zamiast dziesięciu samochodów karawana liczy siedem.
5. Pod uwagę weźmy transmisję ścieżką złożoną z Q łączy pakietu zawierającego F bitów. Każde łącze przesyła dane z szybkością R b/s. Ponieważ sieć jest nieznacznie obciążona, nie występują opóźnienia kolejkowania. Opóźnienie propagacji jest pomijalne.
   a. Załóżmy, że sieć jest siecią z przełączaniem pakietów korzystającą z wirtualnych obwodów. Czas tworzenia wirtualnego obwodu oznaczmy jako t_s (w sekundach). Przyjmijmy, że warstwy zajmujące się wysyłaniem dodają do pakietu nagłówek liczący w sumie n bitów. Ile czasu zajmie przesłanie pliku z węzła źródłowego do docelowego?
   b. Załóżmy, że sieć jest datagramową siecią z przełączaniem pakietów i korzysta z usługi bez połączenia. Dodatkowo przyjmijmy, że nagłówek każdego pakietu liczy $2n$ bitów. Ile czasu zajmie wysłanie pakietu?
   c. Załóżmy, że jest stosowana sieć z przełączaniem obwodów, a ponadto, że szybkość transmisji obwodu między węzłem źródłowym i docelowym wynosi R b/s. Przyjmując, że czas tworzenia obwodu wynosi t_s, a ponadto, że do pakietu jest dodawany nagłówek liczący n bitów, ile potrwa wysyłanie pakietu?
6. Przedstawiony w tym punkcie elementarny problem rozpoczyna rozpatrywanie opóźnień propagacji i transmisji, będących dwoma głównymi zagadnieniami związanymi z przesyłaniem danych w sieci. Weźmy pod uwagę hosty A i B połączone za pomocą pojedynczego łącza o szybkości R b/s. Załóżmy, że dwa hosty są oddalone od siebie o m metrów, a ponadto, że szybkość propagacji w przypadku łącza wynosi s m/s. Host A zamierza wysłać do hosta B pakiet liczący L bitów.
   a. Posługując się wielkościami m i s, określ opóźnienie propagacji o_{prop}.
   b. Używając wielkości L i R, wyznacz opóźnienie transmisji pakietu o_{trans}.
   c. Ignorując opóźnienia przetwarzania i kolejkowania, sformułuj wyrażenie określające opóźnienie międzywęzłowe.
   d. Zakładając, że host A rozpocznie transmisję pakietu w chwili czasu $t = 0$, gdzie znajdzie się ostatni bit pakietu w chwili czasu $t = o_{trans}$?
   e. Przyjmijmy, że o_{prop} jest większe od o_{trans}. Gdzie znajdzie się pierwszy bit pakietu w chwili czasu $t = o_{trans}$?

f. Przyjmijmy, że o_{prop} jest mniejsze od o_{trans}. Gdzie znajdzie się pierwszy bit pakietu w chwili czasu $t = o_{trans}$?

g. Załóżmy, że $s = 2{,}5 \cdot 10^8$, $L = 100$ bitów i $R = 28$ kb/s. Wyznacz odległość m, tak aby o_{prop} było równe o_{trans}.

7. W przypadku tego problemu pod uwagę weźmiemy przesyłanie głosu z hosta A do hosta B za pośrednictwem sieci z przełączaniem pakietów (na przykład używanej przez telefonię internetową). Host A zamienia „w locie" głos z postaci analogowej na cyfrowy strumień bitów transmitowany z szybkością 64 kb/s. Host A grupuje następnie bity w 48-bajtowe pakiety. Między hostami znajduje się tylko jedno łącze. Jego szybkość transmisji wynosi 1 Mb/s, natomiast opóźnienie propagacji 2 ms. Od razu po utworzeniu pakietu host A wysyła go do hosta B. Natychmiast po odebraniu całego pakietu host B dokonuje konwersji jego bitów na sygnał analogowy. Ile czasu upłynie od utworzenia bitu przez host A (przy użyciu oryginalnego sygnału analogowego) do chwili poddania bitu dekodowaniu przez host B (w ramach konwersji na sygnał analogowy)?

8. Załóżmy, że użytkownicy korzystają ze wspólnego łącza o szybkości 1 Mb/s, a ponadto, że każdy z nich do transmisji wymaga szybkości 100 kb/s. Jednak każdy użytkownik przesyła dane tylko przez 10% czasu pracy (zapoznaj się z porównaniem przełączania pakietów z przełączaniem obwodów, zamieszczonym w podrozdziale 1.3).

   a. Ilu użytkowników może być obsługiwanych, gdy używa się przełączania obwodów?

   b. W przypadku pozostałej części tego problemu przyjmijmy, że jest wykorzystywane przełączanie pakietów. Określ prawdopodobieństwo tego, że określony użytkownik transmituje dane.

   c. Załóżmy, że istnieje 40 użytkowników. Określ prawdopodobieństwo tego, że w dowolnej chwili czasu dokładnie n użytkowników jednocześnie przesyła dane (*wskazówka*: należy skorzystać z rozkładu dwumianowego).

   d. Wyznacz prawdopodobieństwo tego, że 11 lub więcej użytkowników jednocześnie transmituje dane.

9. Pod uwagę weźmy omówienie zamieszczone w podpunkcie „Porównanie przełączania pakietów i obwodów oraz multipleksowanie statystyczne" znajdującym się w podrozdziale 1.3. Zawarto w nim przykład łącza o szybkości 1 Mb/s. Gdy użytkownicy są zajęci, generują dane z szybkością 100 kb/s. Jednak prawdopodobieństwo tego, że tak jest, wynosi zaledwie $p = 0{,}1$. Załóżmy, że łącze 1 Mb/s zostanie zastąpione łączem o szybkości 1 Gb/s.

   a. Jaka jest maksymalna liczba N użytkowników, którzy jednocześnie mogą być obsługiwani przez sieć z przełączaniem obwodów?

   b. Weźmy pod uwagę przełączanie pakietów i populację liczącą M użytkowników. Podaj wzór (używając wielkości p, M i N) określający prawdopodobieństwo tego, że więcej niż N użytkowników przesyła dane.

10. Weźmy pod uwagę opóźnienie kolejkowania bufora routera (znajdującego się przed łączem wyjściowym). Załóżmy, że wszystkie pakiety liczą L bitów, szybkość transmisji wynosi R b/s, natomiast co LN/R sekund w buforze pojawia się jednocześnie N pakietów. Wyznacz dla pakietu średnie opóźnienie kolejkowania (*wskazówka*: w przypadku pierwszego pakietu opóźnienie jest zerowe; dla drugiego

pakietu wynosi L/R, natomiast dla trzeciego $2L/R$; N-ty pakiet będzie już wysłany, gdy w buforze pojawi się druga partia pakietów).

11. Uwzględnijmy opóźnienie kolejkowania bufora routera. Niech I określa natężenie ruchu wynoszące La/R. Przyjmijmy, że dla $I < 1$ opóźnienie kolejkowania przyjmuje postać IL/R $(1–I)$.
    a. Podaj wzór dla całkowitego opóźnienia, które jest sumą opóźnienia kolejkowania i transmisji.
    b. Sporządź wykres całkowitego opóźnienia jako funkcję L/R.
12. a. Uogólnij wzór zamieszczony w podrozdziale 1.6 i określający opóźnienie międzywęzłowe, aby uwzględniał różne szybkości przetwarzania, szybkości transmisji i opóźnienia propagacji.
    b. Powtórnie rozpatrz przypadek przedstawiony w punkcie (a), z tym że przyjmij dodatkowo, że w każdym węźle występuje średnie opóźnienie kolejkowania o_{kolejk}.
13. Za pomocą narzędzia *Traceroute* w trzech różnych porach dnia prześledź trasę pokonywaną przez pakiety między węzłem źródłowym i docelowym, które znajdują się na tym samym kontynencie.
    a. W przypadku każdej z trzech pór dnia dla opóźnień całkowitej trasy określ średnie i standardowe odchylenie.
    b. W przypadku każdej z trzech pór dnia wyznacz liczbę routerów znajdujących się na drodze pakietów. Czy poszczególne ścieżki różnią się od siebie?
    c. Spróbuj określić liczbę sieci dostawców ISP, przez które narzędzie *Traceroute* przesyła pakiety od węzła źródłowego do docelowego. Routery o podobnych nazwach i (lub) adresach IP powinny być uznane za należące do tych samych dostawców ISP. Czy w przeprowadzonych doświadczeniach największe opóźnienia występują w przypadku interfejsów węzłów znajdujących się między sąsiadującymi ze sobą sieciami dostawców ISP?
    d. Powyższe trzy punkty powtórnie wykonaj dla węzłów źródłowego i docelowego zlokalizowanych na różnych kontynentach. Porównaj wyniki uzyskane dla węzłów położonych na tym samym i na innych kontynentach.
14. Załóżmy, że dwa hosty A i B są od siebie oddalone o 10 000 kilometrów i połączone bezpośrednim łączem o szybkości $R = 1$ Mb/s. Przyjmijmy też, że w przypadku łącza szybkość propagacji wynosi $2{,}5 \cdot 10^8$ m/s.
    a. Oblicz iloczyn przepustowości i opóźnienia propagacji $R \cdot o_{\text{prop}}$.
    b. Weźmy pod uwagę przesłanie pliku liczącego 400 000 bitów z hosta A do hosta B. Załóżmy, że plik zostanie wysłany jako jeden wielki komunikat. Jaka jest maksymalna liczba bitów, które w danej chwili znajdą się w łączu?
    c. Przedstaw interpretację iloczynu przepustowości i opóźnienia propagacji.
    d. Jaka jest długość bita (w metrach) znajdującego się w łączu? Czy jest dłuższy od boiska piłkarskiego?
    e. Uwzględniając szybkość propagacji s, szybkość transmisji R i długość łącza m, utwórz zależność określającą długość bitu.
15. Nawiązując do problemu 14., załóżmy, że można zmodyfikować wartość R. Dla jakiej wartości R długość bitu będzie równa długości łącza?

16. Weźmy pod uwagę problem 14., z tym że niech szybkość łącza wynosi $R = 1$ Gb/s.
    a. Oblicz iloczyn przepustowości i opóźnienia propagacji $R \cdot o_{\text{prop}}$.
    b. Weźmy pod uwagę przesłanie pliku liczącego 400 000 bitów z hosta A do hosta B. Załóżmy, że plik zostanie wysłany jako jeden wielki komunikat. Jaka jest maksymalna liczba bitów, które w danej chwili znajdą się w łączu?
    c. Jaka jest długość bitu (w metrach) znajdującego się w łączu?
17. Ponownie odnieśmy się do problemu 14.
    a. Ile czasu zajmie wysłanie pliku przy założeniu, że operacja jest wykonywana bez żadnych przerw?
    b. Przyjmijmy, że plik podzielono na 10 pakietów, z których każdy liczy 40 000 bitów. Ponadto załóżmy, że dla każdego pakietu węzeł odbiorczy odsyła potwierdzenie i czas jego transmisji jest nieznaczny. Przyjmijmy wreszcie, że nadawca nie może wysłać kolejnego pakietu, dopóki dla poprzedniego nie otrzyma potwierdzenia dostarczenia. Ile czasu zajmie przesłanie pliku?
    c. Porównaj ze sobą wyniki uzyskane w punktach (a) i (b).
18. Załóżmy, że między geostacjonarnym satelitą i jego naziemną stacją bazową znajduje się łącze mikrofalowe o szybkości 10 Mb/s. Co minutę satelita wykonuje cyfrowe zdjęcie i przesyła je do stacji. Przyjmijmy, że szybkość propagacji wynosi $2{,}4 \cdot 10^8$ m/s.
    a. Jakie jest dla łącza opóźnienie propagacji?
    b. Jaka jest wartość iloczynu przepustowości i opóźnienia propagacji $R \cdot o_{\text{prop}}$?
    c. Niech x oznacza wielkość zdjęcia. Jaka musi być dla łącza mikrofalowego minimalna wartość x, aby transmisja była ciągła?
19. Pod uwagę weźmy analogię dotyczącą podróży samolotem przytoczoną podczas omawiania warstw w podrozdziale 1.7, a także dodawanie nagłówków do zawartości pakietów przemieszczających się przez kolejne warstwy stosu protokołów. Czy istnieją informacje odpowiadające danym znajdującym się w nagłówkach, które są kojarzone z bagażem i pasażerami podczas pokonywania przez nich kolejnych poziomów stosu protokołów systemu linii lotniczej?
20. W nowoczesnych sieciach z przełączaniem pakietów źródłowy host segmentuje długie komunikaty warstwy aplikacji (na przykład obraz lub plik audio) na mniejsze pakiety i umieszcza je w sieci. Host odbiorczy łączy pakiety z powrotem, aby uzyskać oryginalny komunikat. Proces ten określa się mianem *segmentacji komunikatów*. Na rysunku 1.21 zilustrowano transport komunikatu między węzłami bez stosowania segmentacji i z jej wykorzystaniem. Pod uwagę weźmy komunikat liczący $7{,}5 \cdot 10^6$ bitów, który zostanie przesłany między źródłowym i docelowym hostem (rysunek 1.21). Załóżmy, że każde łącze widoczne na rysunku ma szybkość 1,5 Mb/s. Opóźnienia propagacji, kolejkowania i przetwarzania należy zignorować.
    a. Rozważmy przesłanie komunikatu między źródłowym i docelowym węzłem *bez* zastosowania segmentacji. Ile czasu zajmie przemieszczenie komunikatu ze źródłowego hosta do pierwszego przełącznika pakietów? Pamiętając o tym, że każdy przełącznik korzysta z przełączania pakietów z buforowaniem, jaki będzie całkowity czas potrzebny na przesłanie komunikatu między hostem źródłowym i docelowym?

RYSUNEK 1.21. Transport komunikatu między węzłami: (a) — bez zastosowania segmentacji; (b) — z wykorzystaniem segmentacji

b. Przyjmijmy teraz, że komunikat jest segmentowany na 5000 pakietów, z których każdy liczy 1500 bitów. Ile potrwa przemieszczenie pierwszego pakietu ze źródłowego hosta do pierwszego przełącznika? Gdy pierwszy pakiet jest przesyłany między pierwszym i drugim przełącznikiem, drugi pakiet jest transmitowany od źródłowego hosta do pierwszego przełącznika. W jakim czasie drugi pakiet zostanie całkowicie odebrany przez pierwszy przełącznik?

c. Ile czasu zajmie przesłanie pliku między źródłowym i docelowym hostem, gdy zostanie przeprowadzona segmentacja komunikatu? Uzyskany wynik porównaj z odpowiedzią udzieloną w przypadku punktu (a) i skomentuj to.

d. Omów wady segmentacji komunikatów.

21. Poeksperymentuj z apletem Java segmentującym komunikaty, który znajduje się na stronie internetowej poświęconej książce. Czy opóźnienia występujące w przypadku apletu odpowiadają tym, które pojawiły się w poprzednim problemie? Jak opóźnienia propagacji powiązane z łączem wpływają na ogólne opóźnienie międzywęzłowe w przypadku przełączania pakietów (z segmentacją komunikatów) i przełączania komunikatów?

22. Pod uwagę weźmy wysłanie między hostami A i B dużego pliku liczącego F bitów. Między hostami znajdują się dwa łącza i jeden przełącznik. Łącza nie są przeciążone (oznacza to, że nie występują opóźnienia kolejkowania). Host A dzieli plik na segmenty, z których każdy liczy S bitów i do każdego dodaje 40-bitowy nagłówek. W efekcie powstają pakiety o długości $L = 40+S$ bitów. Każde łącze oferuje szybkość transmisji wynoszącą R b/s. Określ wartość S, która zminimalizuje opóźnienie związane z przesyłaniem pliku między hostami A i B. Opóźnienie propagacji należy zignorować.

Dodatkowe pytania

1. Jakiego typu usługi bezprzewodowych sieci komórkowych są dostępne w miejscu zamieszkania?
2. Używając bezprzewodowej technologii sieci lokalnych 802.11, zaprojektuj sieć w domu swoim lub rodziców. Wykonaj zestawienie modeli produktów znajdujących się w sieci domowej, uwzględniające ich cenę.

3. Czym jest telefonia internetowa? Zlokalizuj kilka witryn WWW firm, które się tym zajmują.
4. Czym jest usługa SMS (*Short Message Service*)? Czy jest to usługa popularna w dowolnej części świata? Jeśli tak jest, gdzie na przykład cieszy się powodzeniem i jak bardzo? Czy jest możliwe wysłanie SMS-a do mobilnego telefonu za pośrednictwem witryny WWW?
5. Co to jest strumieniowa transmisja magazynowanych danych audio? Opisz kilka istniejących produktów służących do strumieniowej transmisji danych audio w internecie. Znajdź kilka witryn WWW firm, które się tym zajmują.
6. Czym jest wideokonferencja internetowa? Opisz kilka istniejących produktów służących do przeprowadzania tego typu wideokonferencji. Znajdź kilka witryn WWW firm, które się tym zajmują.
7. Poszukaj pięć firm świadczących usługi typu P2P umożliwiające wymianę plików. Jakiego rodzaju pliki są obsługiwane przez aplikacje każdej z firm?
8. Czym są komunikatory? Czy istnieją produkty umożliwiające skorzystanie z usługi przesyłania wiadomości błyskawicznych za pomocą urządzenia kieszonkowego?
9. Kto stworzył ICQ, czyli pierwszą usługę przesyłania wiadomości błyskawicznych? Kiedy to było i ile lat mieli twórcy usługi? Kto zaprojektował narzędzie Napster? Kiedy to miało miejsce i ile lat mieli wynalazcy?
10. Porównaj dwie technologie bezprzewodowego dostępu do internetu — Wi-Fi i 3G. Jakie są szybkości oferowane przez obie technologie? Jaka jest ich cena? Omów mobilność i dostępność technologii.
11. Dlaczego usługa Napster przestała istnieć? Czym się zajmuje organizacja RIAA i jakie są podejmowane kroki w celu ograniczenia wymiany za pośrednictwem aplikacji P2P danych chronionych prawami autorskimi? Jaka jest różnica między bezpośrednim i pośrednim naruszaniem praw autorskich?
12. Czy uważasz, że za 10 lat użytkownicy za pośrednictwem sieci komputerowych nadal powszechnie będą się wymieniali plikami chronionymi prawami autorskimi? Szczegółowo uzasadnij swoją odpowiedź.

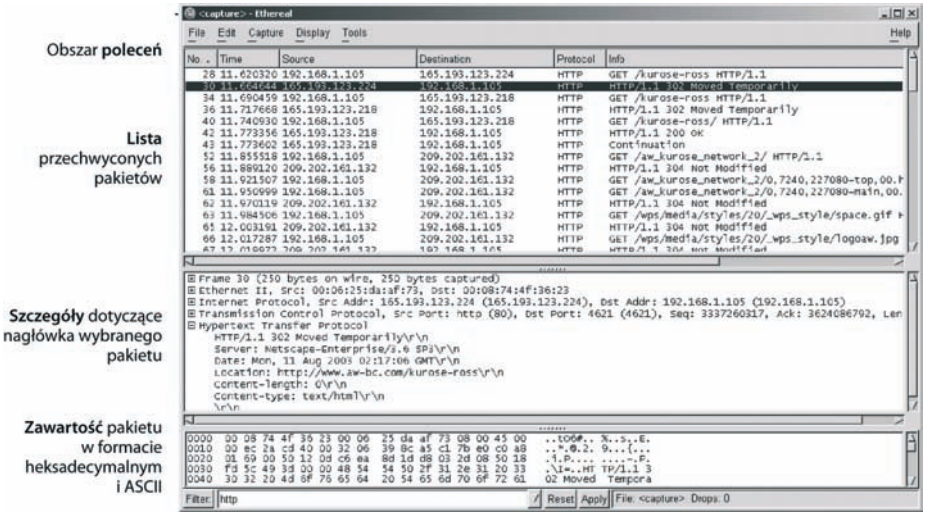
Ćwiczenie 1. realizowane za pomocą narzędzia Ethereal

„*Powiedz mi, a zapomnę. Pokaż mi, a zapamiętam. Zaangażuj mnie, a zrozumiem*” — chińskie przysłowie.

Poziom zrozumienia protokołów sieciowych często można znacznie zwiększyć przez zobaczenie ich w akcji i poeksperymentowaniu z nimi (obserwacja sekwencji komunikatów wymienianych między dwoma składnikami protokołu, zagłębianie się w szczegóły dotyczące działania protokołu, powodowanie wykonywania przez protokół określonych operacji, a także monitorowanie ich efektów). Można to zrealizować przy użyciu symulacji lub rzeczywistego środowiska sieciowego, jakim jest internet. W przypadku pierwszego wariantu zostaną użyte aplety Java zamieszczone na stronie internetowej

poświęconej książce. Drugi wariant będzie realizowany w ramach ćwiczeń wykorzystujących narzędzie *Ethereal*. W przypadku różnych wariantów aplikacje sieciowe będą uruchamiane na stacjonarnym komputerze znajdującym się w domu lub laboratorium. Czytelnik będzie obserwował protokoły komputera współpracujące i wymieniające się komunikatami ze składnikami protokołów funkcjonujących w innym miejscu internetu. Oznacza to, że posiadany komputer będzie integralną częścią rzeczywistych doświadczeń. Przez obserwację będzie się zdobywało wiedzę.

Podstawowym narzędziem umożliwiającym obserwację komunikatów wymienianych między aktywnymi składnikami protokołu jest **program analizujący pakiety** (ang. *packet sniffer*). Tego typu narzędzie w pasywny sposób kopiuje komunikaty, które są wysyłane i odbierane przez komputer. Ponadto program wyświetla zawartość różnych pól protokołu znajdujących się w przechwyconych komunikatach. Na rysunku 1.22 przedstawiono zrzut ekranu okna programu *Ethereal* służącego do analizowania pakietów. Jest to darmowe narzędzie dostępne w wersji dla systemów Windows, Linux/Unix i Mac. W rozdziałach od 1. do 5. zamieszczono ćwiczenia wykorzystujące program *Ethereal*, które umożliwią Czytelnikowi zapoznanie się z kilkoma protokołami omawianymi w danym rozdziale. Niniejsze pierwsze ćwiczenie polega na pobraniu i zainstalowaniu kopii narzędzia *Ethereal*, wyświetleniu strony internetowej, a następnie przechwyceniu i przeanalizowaniu komunikatów protokołu wymienianych między przeglądarką i serwerem WWW.



RYSUNEK 1.22. Okno programu Ethereal

WYWIAD

LEONARD KLEINROCK

Leonard Kleinrock jest profesorem informatyki na Uniwersytecie Kalifornijskim znajdującym się w Los Angeles. W 1969 r. jego uczelniany komputer stał się pierwszym węzłem internetu. Stworzone przez niego w 1961 r. podstawy technologii przełączania pakietów stanowiły fundament internetu. Leonard jest też prezesem i założycielem firmy Nomadix Inc., której technologia zapewnia większą dostępność usługi internetu szerokopasmowego. Leonard ukończył inżynierię elektryczną na uczelni CCNY (*City College of New York*), a na uczelni MIT zdobył tytuł magistra i doktora z tej samej dziedziny.

Co sprawiło, że postanowił Pan specjalizować się w technologiach związanych z sieciami i internetem?

Po uzyskaniu w 1959 r. na uczelni MIT tytułu doktora zorientowałem się, że większość znajomych ze studiów zajmuje się badaniami z zakresu teorii informacji i kodowania. Na uczelni MIT pracował Claude Shannon, znakomity naukowiec, który był pionierem w obu dziedzinach i poradził sobie już z większością istotnych problemów. Problemy badawcze, które nadal nie były rozwiązane, zaliczały się do trudnych, ale też mniej ważnych. W związku z tym postanowiłem zainicjować badania w nowej dziedzinie, o której nikt jeszcze nie pomyślał. Trzeba pamiętać, że na uczelni MIT wokół mnie znajdowało się mnóstwo komputerów. Jasne było dla mnie, że wkrótce będą musiały się ze sobą komunikować. W tamtym czasie nie istniała skuteczna metoda, która by to umożliwiała, dlatego postanowiłem opracować technologię pozwalającą na tworzenie w efektywny sposób sieci danych.

Jaka była Pana pierwsza praca w branży komputerowej? Na czym polegała?

Od 1951 r. do 1957 r. uczęszczałem na wieczorowe zajęcia prowadzone na uczelni CCNY, których zwieńczeniem było uzyskanie przeze mnie wykształcenia w zakresie inżynierii elektrycznej. W ciągu dnia pracowałem najpierw jako technik, a później jako inżynier w niewielkiej firmie Photobell, zajmującej się elektroniką przemysłową. W tamtym czasie do linii produkcyjnej firmy wprowadziłem technologię cyfrową. Przede wszystkim korzystaliśmy z fotoelektrycznych urządzeń wykrywających obecność określonych obiektów (skrzynek, ludzi itp.). Układ nazywany *multiwibratorem bistabilnym* był tą technologią, której potrzebowaliśmy, aby w dziedzinie detekcji wykorzystać cyfrowe przetwarzanie. Tego typu układy okazały się elementami tworzącymi komputery. W obecnie używanej terminologii układy te nazywa się *przerzutnikami* lub *przełącznikami*.

Co Panu przyszło na myśl po przesłaniu pierwszego komunikatu między dwoma hostami zlokalizowanymi na uczelni UCLA i w instytucie Stanford Research Institute?

Pierwszy komunikat przesłany między węzłami trochę mnie rozczarował. Jak pamiętam, większe wrażenie zrobiło na mnie zdarzenie, które miało miejsce 2 września 1969 r., gdy pierwsze urządzenie sieciowe połączyło się z pierwszym działającym systemem znajdującym się w innym miejscu świata (był to mój komputer na uczelni UCLA). Właśnie wtedy narodził się internet. Wcześniej w tym samym roku na konferencji prasowej odbywającej się na uczelni UCLA powtórzono moje słowa, w których stwierdziłem, że gdy sieć powstanie i zacznie funkcjonować, uzyskanie dostępu z naszych domów i biur do zasobów komputerowych będzie tak proste, jak w przypadku eksploatacji urządzeń elektrycznych i telefonicznych. W tamtym czasie moja wizja była taka, że internet stanie się wszechobecny, niedostrzegalny, zawsze aktywny i dostępny, a ponadto umożliwi podłączenie się do niego każdemu z dowolnego miejsca przy użyciu każdego urządzenia. Jednak nie przewidziałem, że moja 94-letnia mama będzie dzisiaj korzystać z internetu. Naprawdę to robi!

Jaka jest Pana wizja dotycząca przyszłości sieci?

Najbardziej wyrazista część mojej wizji dotyczy mobilnych technologii komputerowych i inteligentnych miejsc. Dostępność prostych, tanich, przenośnych i bardzo wydajnych urządzeń obliczeniowych, a także wszechobecność internetu pozwoliły nam stać się nomadami. *Mobilne technologie komputerowe* umożliwiają użytkownikom przemieszczającym się z miejsca w miejsce korzystanie z usług internetowych w bezproblemowy sposób, niezależnie od tego, dokąd się udadzą. Jednak mobilne technologie komputerowe są tylko jednym krokiem. Kolejny pozwoli nam przenieść się ze świata cyberprzestrzeni do fizycznego świata inteligentnych miejsc. Nasze otoczenie (biurka, ściany, pojazdy, zegarki, pasy itp.) ożyje dzięki technologii reprezentowanej przez urządzenia uruchamiające, czujniki, układy logiczne, procesory, magazyny danych, kamery, mikrofony, głośniki, wyświetlacze i sprzęt komunikacyjny. Taka wbudowana technologia umożliwi otaczającemu nas środowisku zaoferowanie usług opartych na protokole IP, których zażądamy. Gdy wejdę do pokoju, ten będzie o tym „wiedział". Będę w stanie w naturalny sposób komunikować się z otoczeniem, posługując się naszym językiem. Moje żądania spowodują wygenerowanie odpowiedzi przekazujących zawartość stron internetowych za pośrednictwem ściennych wyświetlaczach, szkieł kontaktowych, mowy, hologramów itp.

Sięgając w trochę dalszą przyszłość, wyobrażam sobie sieć uwzględniającą dodatkowe kluczowe komponenty. Widzę inteligentnych agentów programowych zastosowanych w sieci, których celem jest analizowanie i przetwarzanie danych, obserwowanie trendów, a także realizowanie zadań w sposób dynamiczny i adaptacyjny. Ponadto zdecydowanie większy ruch sieciowy będzie generowany nie przez ludzi, a przez wbudowane urządzenia i ich inteligentne agenty. Mogę sobie wyobrazić, że duży zbiór systemów, które same sobą zarządzają, będzie sprawować kontrolę nad taką rozległą i szybką siecią. Ogromna ilość informacji natychmiast będzie przesyłana w sieci. Będzie z tym związane intensywne przetwarzanie i filtrowanie. W zasadzie internet będzie rozprzestrzeniającym się globalnym układem nerwowym. Według mnie wszystko to i nie tylko nastanie już w XXI wieku.

Kto był dla Pana inspiracją podczas kariery zawodowej?

Jak na razie taką osobą był Claude Shannon pracujący na uczelni MIT. Był to znakomity naukowiec, który w bardzo intuicyjny sposób potrafił połączyć swoje matematyczne pomysły ze światem fizycznym. Claude wchodził w skład komisji związanej z moim doktoratem.

Czy ma Pan jakieś rady dla studentów zaczynających zajmować się sieciami komputerowymi i internetem?

Internet i wszystko, co pozwala z niego korzystać, jest nowym, rozległym obszarem pełnym niesamowitych wyzwań. Jest w nim miejsce na sporo innowacji. Nie należy być ograniczonym przez obecną technologię. Warto sięgać w przyszłość i wyobrażać sobie, co mogłoby w niej być, a następnie dążyć do tego, aby tak było.