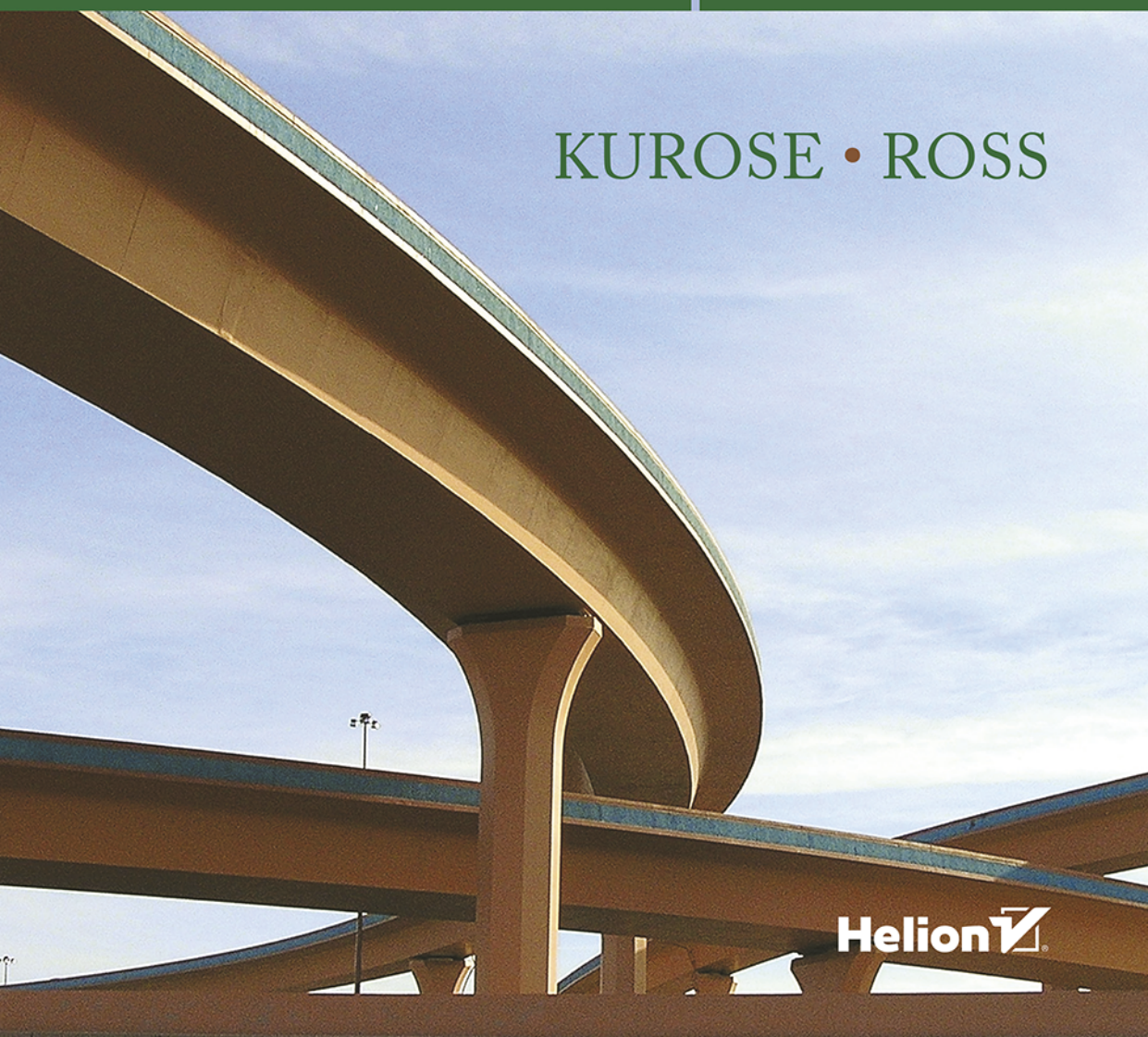
Sieci komputerowe

UJĘCIE CAŁOŚCIOWE | WYDANIE VII

KUROSE • ROSS



Helion

Tytuł oryginału: Computer Networking: A Top-Down Approach (7th Edition)

Tłumaczenie: Tomasz Walczak

ISBN: 978-83-289-2935-7

Authorized translation from the English language edition, entitled: COMPUTER NETWORKING: A TOP-DOWN APPROACH, Seventh Edition; ISBN 0133594149; by James F. Kurose; and by Keith W. Ross; published by Pearson Education, Inc.

Helion S.A.
ul. Kościuszki 1c, 44-100 Gliwice
tel. 32 230 98 63
e-mail: *helion@helion.pl*
WWW: *helion.pl* (księgarnia internetowa, katalog książek)

Drogi Czytelniku!
Jeżeli chcesz ocenić tę książkę, zajrzyj pod adres
helion.pl/user/opinie/sie7vv
Możesz tam wpisać swoje uwagi, spostrzeżenia, recenzję.

Dodatkowe materiały do książki można znaleźć pod adresem:
https://ftp.helion.pl/przyklady/sieu7v.zip

Printed in Poland.

Spis treści

ROZDZIAŁ 7

Sieci bezprzewodowe i mobilne

W świecie telekomunikacji ostatnich 20 lat było czasem telefonii komórkowej. Liczba jej abonentów na całym świecie zwiększyła się z 34 milionów w roku 1993 do prawie siedmiu miliardów w 2014, przewyższając liczbę stałych linii telefonicznych. Obecnie na świecie liczba używanych telefonów komórkowych jest większa niż ludzi. Zalety telefonów komórkowych są oczywiste — nieskrępowany dostęp do globalnej sieci telefonicznej za pomocą przenośnego, lekkiego urządzenia. Od niedawna laptopy, smartfony i tablety łączą się bezprzewodowo z internetem za pomocą sieci komórkowych lub Wi-Fi. Coraz częściej także urządzenia takie jak konsole, termostaty, systemy ochrony domu, sprzęt AGD, zegarki, okulary, samochody, systemy kontroli ruchu drogowego i inne są bezprzewodowo podłączone do internetu.

Problemy, które występują w tych sieciach — zwłaszcza w warstwie łącza danych i warstwie sieciowej — są na tyle inne niż w tradycyjnych, przewodowych sieciach komputerowych, że wymagają oddzielnego rozdziału poświęconego łączności bezprzewodowej i mobilnej (tzn. *niniejszego* rozdziału).

Rozdział zaczniemy od omówienia użytkowników mobilnych, łączy i sieci bezprzewodowych oraz ich związków z większymi (zwykle przewodowymi) sieciami, do których są podłączone. Rozróżnimy problemy stwarzane przez *bezprzewodową* naturę łączy komunikacyjnych oraz przez *mobilność*, która z niej wynika. To ważne rozróżnienie między bezprzewodowością a mobilnością pozwoli nam lepiej wyizolować, zidentyfikować i opanować kluczowe pojęcia w każdej dziedzinie. Zwróćmy uwagę, że istnieją środowiska, w których węzły sieci są bezprzewodowe, ale nie mobilne (na przykład bezprzewodowe sieci domowe lub biurowe z komputerami stacjonarnymi

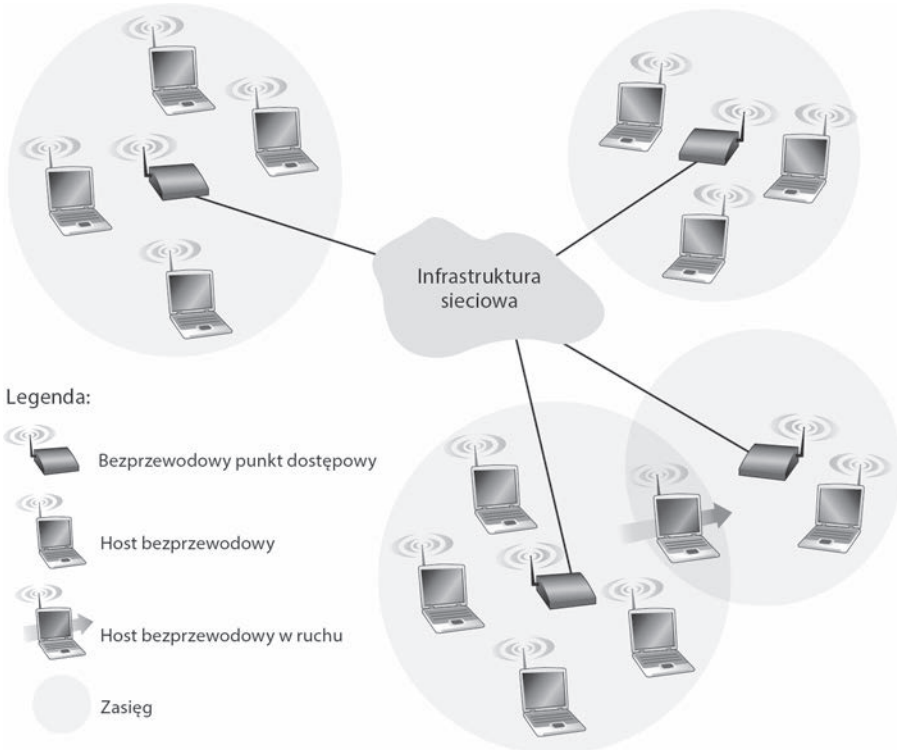
wyposażonymi w duże monitory), a także ograniczone formy mobilności, które nie wymagają łączy bezprzewodowych (na przykład pracownik, który używa w domu przewodowego laptopa, wyłącza go, jedzie do pracy i podłącza laptop do przewodowej sieci firmy). Oczywiście najbardziej ekscytujące środowiska sieciowe są zarówno bezprzewodowe, *jak i* mobilne — wyobraźmy sobie scenariusz, w którym użytkownik mobilny (na tylnym siedzeniu samochodu) prowadzi rozmowę VoIP i utrzymuje wiele połączeń TCP, podczas gdy samochód pędzi autostradą z szybkością 160 kilometrów na godzinę. Właśnie tutaj, na styku bezprzewodowości i mobilności, znajdziemy najciekawsze wyzwania techniczne!

Zaczniemy od zilustrowania środowiska, w którym będziemy badać bezprzewodową komunikację i mobilność — sieci, w której urządzenia (niektóre mobilne) są połączone bezprzewodowo z większą infrastrukturą sieciową przez łącza bezprzewodowe na obrzeżach danej sieci. Następnie w podrozdziale 7.2 rozważymy właściwości łącza bezprzewodowego i krótko opiszemy protokół dostępu do wspólnego nośnika Code Division Multiple Access (CDMA), używany w wielu sieciach bezprzewodowych. W podrozdziale 7.3 zbadamy aspekty warstwy łącza w bezprzewodowych sieciach lokalnych IEEE 802.11 (Wi-Fi); powiemy także kilka słów o protokole Bluetooth i sieciach WPAN. W podrozdziale 7.4 omówimy komórkowy dostęp do internetu, w tym nowe technologie 3G i 4G, które zapewniają zarówno przesyłanie głosu, jak i szybką transmisję danych. W podrozdziale 7.5 zajmiemy się mobilnością, skupiając się na problemach lokalizowania użytkownika mobilnego, routingu oraz „transferu" użytkownika, który dynamicznie porusza się między punktami przyłączeniowymi. W podrozdziałach 7.6 i 7.7 Czytelnicy dowiedzą się, jak te usługi są zrealizowane w standardzie Mobile IP w firmowych sieciach 802.11 i w komórkowych sieciach LTE. Wreszcie w podrozdziale 7.8 rozważymy wpływ łączy bezprzewodowych i mobilności na protokoły warstwy transportowej oraz aplikacje sieciowe.

7.1. Wprowadzenie

Na rysunku 7.1 przedstawiono środowisko, w którym będziemy rozważać zagadnienia mobilności oraz bezprzewodowej transmisji danych. Na początku przedstawimy informacje ogólne, dotyczące szerokiej gamy sieci, w tym bezprzewodowych sieci lokalnych takich jak IEEE 802.11 oraz sieci komórkowych (na przykład 4G); konkretnymi architekturami bezprzewodowymi zajmiemy się w dalszych podrozdziałach. W sieci bezprzewodowej możemy zidentyfikować następujące elementy:

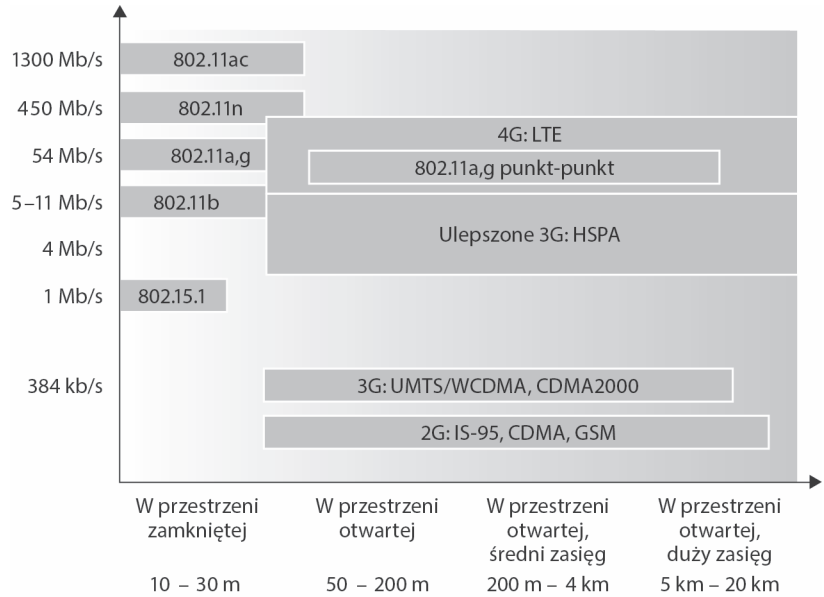
- *Hosty bezprzewodowe.* Podobnie jak w sieciach przewodowych, hosty to urządzenia końcowe, w których działają aplikacje. **Host bezprzewodowy** może być laptopem, tabletem, smartfonem albo komputerem stacjonarnym. Same hosty mogą, ale nie muszą być mobilne.



Rysunek 7.1. Elementy sieci bezprzewodowej

- *Łącza bezprzewodowe.* Host łączy się ze stacją bazową (definicja poniżej) albo z innym hostem bezprzewodowym poprzez **bezprzewodowe łącze komunikacyjne**. Różne technologie bezprzewodowe mają różne szybkości i zasięgi transmisji. Na rysunku 7.2 przedstawiono dwie kluczowe cechy (zasięg i szybkość) najpopularniejszych łączy bezprzewodowych. Rysunek ma jedynie ogólnie obrazować te właściwości. Pewne typy sieci są dopiero wdrażane, a szybkość niektórych łączy może się zwiększyć lub zmniejszyć w porównaniu do przedstawionych wartości (zależy to od odległości, warunków działania kanału i liczby użytkowników sieci bezprzewodowej). Opiszemy te standardy dalej w pierwszej połowie rozdziału; w podrozdziale 7.2 rozważymy też ich inne właściwości (na przykład współczynniki i przyczyny błędów).

  Na rysunku 7.1 łącza bezprzewodowe wiążą hosty położone na obrzeżach sieci w większą infrastrukturę sieciową. Warto zauważyć, że łączy bezprzewodowych używa się czasem *wewnątrz* sieci oraz do łączenia routerów, przełączników i innego sprzętu sieciowego. Jednak w tym rozdziale skupimy się na komunikacji bezprzewodowej na obrzeżach sieci, ponieważ właśnie tutaj pojawiają się najbardziej interesujące problemy techniczne i tu występuje najszybszy wzrost.



Rysunek 7.2. Charakterystyka łączy w wybranych standardach komunikacji bezprzewodowej

- *Stacja bazowa.* Jest to kluczowy składnik infrastruktury sieci bezprzewodowej. W przeciwieństwie do hosta i łącza bezprzewodowego **stacja bazowa** nie ma oczywistego odpowiednika w sieci przewodowej. Zadaniem stacji bazowej jest przesyłanie i odbieranie danych (tzn. pakietów) od związanego z nią hosta bezprzewodowego. Często jest również odpowiedzialna za koordynowanie transmisji wielu związanych z nią hostów. Kiedy mówimy, że host bezprzewodowy jest „związany" ze stacją bazową, mamy na myśli, iż (1) host znajduje się w zasięgu komunikacyjnym stacji bazowej oraz (2) host wykorzystuje tę stację bazową do przekazywania danych między sobą a większą siecią. Przykładami stacji bazowych mogą być **stacje transmisyjne** w sieciach komórkowych oraz **punkty dostępowe** w bezprzewodowych sieciach lokalnych 802.11.

  Na rysunku 7.1 stacja bazowa jest podłączona do większej sieci (na przykład internetu, sieci firmowej, domowej albo telefonicznej), a zatem funkcjonuje jako przekaźnik łącza danych między bezprzewodowym hostem a resztą świata.

  O hostach związanych ze stacją bazową często mówi się, że działają w **trybie infrastrukturalnym**, ponieważ wszystkie tradycyjne usługi (na przykład przypisywanie adresów i routing) są świadczone przez sieć, do której host jest podłączony za pośrednictwem stacji bazowej. W **sieciach doraźnych** (**ad hoc**) hosty bezprzewodowe nie dysponują taką infrastrukturą, a zatem same muszą zająć się routingiem, przypisywaniem adresów, tłumaczeniem nazw w stylu DNS itd.

KĄCIK HISTORYCZNY

PUBLICZNY DOSTĘP WI-FI: JUŻ WKRÓTCE W LAMPACH W TWOJEJ OKOLICY?

Hotspoty Wi-Fi (czyli publiczne miejsca, w których użytkownicy mogą uzyskać bezprzewodowy dostęp do sieci 802.11) coraz częściej pojawiają się w hotelach, kawiarniach i na lotniskach na całym świecie. Większość miasteczek uniwersyteckich zapewnia na całym terenie dostęp bezprzewodowy do sieci. Trudno jest też znaleźć hotel, który nie oferuje bezprzewodowego dostępu do internetu.

W ostatnim dziesięcioleciu wiele miast zaprojektowało, wdrożyło i udostępniło miejskie sieci Wi-Fi. Atrakcyjna jest wizja zaoferowania społeczności wszechobecnej sieci Wi-Fi jako usługi publicznej (podobnej do oświetlenia ulic), co pomaga wyeliminować cyfrowe podziały przez zapewnienie wszystkim mieszkańcom dostępu do internetu i przyspiesza rozwój ekonomiczny. W wielu miastach, takich jak Filadelfia, Toronto, Hongkong, Minneapolis, Londyn i Auckland, ogłoszono plany zapewnienia powszechnego bezprzewodowego dostępu do internetu w całej miejscowości. W Filadelfii za cel postawiono „przekształcenie miasta w największy w kraju hotspot Wi-Fi oraz usprawnienie edukacji, zniesienie cyfrowych podziałów, wsparcie nawiązywania więzi między sąsiadami i zmniejszenie kosztów zarządzania". Ten ambitny program realizowany w ramach współpracy między miastem, organizacją non profit Wireless Philadelpia i dostawcą ISP Earthlink doprowadził do powstania działającej sieci hotspotów 802.11b na lampach ulicznych i urządzeniach sterowania ruchem drogowym. Sieć ta pokrywała 80% miasta. Jednak problemy finansowe i operacyjne spowodowały w 2008 roku sprzedaż sieci grupie inwestorów prywatnych, którzy później (w 2010 roku) odsprzedali ją jednak miastu. W innych miastach, takich jak Minneapolis, Toronto, Hongkong i Auckland, z powodzeniem realizowano mniejsze projekty.

Ponieważ sieci 802.11 działają w pasmach nielicencjonowanych (dzięki czemu można je instalować bez wykupywania kosztownych praw do określonych pasm), wydają się atrakcyjne finansowo. Jednak punkty dostępowe 802.11 (zob. podrozdział 7.3) mają znacznie mniejszy zasięg niż komórkowe stacje bazowe 4G (podrozdział 7.4). Dlatego do pokrycia tego samego obszaru wymagają większej liczby punktów końcowych. Z kolei sieci komórkowe zapewniające dostęp do internetu działają w pasmach licencjonowanych. Operatorzy sieci komórkowych płacą miliardy złotych za prawa do pasm, przez co sieci komórkowe są traktowane jak biznes, a nie jak przedsięwzięcie miejskie.

Kiedy host mobilny przemieszcza się z zasięgu jednej stacji bazowej do zasięgu drugiej, zmienia swój punkt podłączenia do większej sieci (tzn. stację, z którą jest związany). Proces ten określamy mianem **transferu** (*handoff*). Taka mobilność wiąże się z wieloma problemami. Jeśli host może się przemieszczać, jak znaleźć jego bieżącą lokację w sieci i przesłać mu dane? Jak przypisywać adresy, zważywszy, że host może znajdować się w wielu możliwych lokacjach? Jeśli host przemieści się *w czasie* połączenia TCP albo rozmowy telefonicznej, jak przekierować dane, aby nie zakłócić połączenia? Te kwestie (i wiele, wiele innych) sprawiają, że łączność bezprzewodowa i mobilna jest ekscytującą dziedziną badań naukowych.

- *Infrastruktura sieciowa.* Jest to większa sieć, z którą komunikuje się host bezprzewodowy.

Po omówieniu „części" sieci bezprzewodowych warto zauważyć, że elementy te można połączyć na wiele różnych sposobów, aby zbudować różne wersje takich sieci. Systematyka tych sieci może okazać się przydatna w trakcie lektury tego rozdziału lub czytania o sieciach bezprzewodowych i poznawania ich ze źródeł innych niż ta książka. Na najbardziej ogólnym poziomie sieci bezprzewodowe można sklasyfikować według dwóch kryteriów: *(i)* czy pakiety w sieci są przekazywane przez *dokładnie jeden punkt przeskoku, czy przez wiele takich punktów* i *(ii)* czy sieć ma *infrastrukturę* (na przykład stację bazową).

- *Sieci infrastrukturalne typu single-hop.* Takie sieci obejmują stację bazową podłączoną do większej sieci przewodowej (na przykład do internetu). Ponadto cała komunikacja między stacją bazową i bezprzewodowym hostem odbywa się w jednym przeskoku w kanale bezprzewodowym. Do tej kategorii należą omówione dalej sieci 802.11 używane w salach zajęć, kawiarniach lub bibliotekach, i sieci 4G LTE, z którymi się wkrótce zapoznasz. Zdecydowana większość codziennych interakcji odbywa się właśnie z sieciami infrastrukturalnymi typu single-hop.
- *Pozbawione infrastruktury sieci typu single-hop.* W tych sieciach nie ma stacji bazowej podłączonej do sieci bezprzewodowej. Jednak — jak wyjaśnimy — w tej odmianie sieci typu single-hop jeden z węzłów może koordynować transmisję danych dla wszystkich pozostałych węzłów. Sieci Bluetooth (omawiamy je w punkcie 7.3.6) i sieci 802.11 działające w trybie ad hoc to właśnie pozbawione infrastruktury sieci typu single-hop.
- *Sieci infrastrukturalne typu multi-hop.* W takich sieciach działa stacja bazowa podłączona do większej sieci. Jednak niektóre węzły bezprzewodowe muszą przekazywać dane przez inne węzły, aby móc komunikować się z siecią za pośrednictwem stacji bazowej. Do tej kategorii należą niektóre bezprzewodowe sieci sensorowe i tak zwane **bezprzewodowe sieci kratowe** (ang. *wireless mesh network*).
- *Pozbawione infrastruktury sieci typu multi-hop.* W tych sieciach nie ma stacji bazowej, a węzły czasem muszą przekazywać dane przez inne hosty, aby móc przesłać informacje do celu. Węzły w tych sieciach mogą być mobilne, a połączenia między węzłami zmieniają się (tak działają sieci klasy **MANET**; ang. *Mobile Ad Hoc Network*). Jeśli te mobilne węzły to pojazdy, mamy do czynienia z siecią klasy **VANET** (ang. *Vehicular Ad Hoc Network*). Można się domyślić, że projektowanie protokołów na potrzeby takich sieci jest trudne i stanowi temat wielu badań naukowych.

W tym rozdziale zajmiemy się przede wszystkim sieciami typu single-hop, a wśród nich — głównie sieciami infrastrukturalnymi.

Zbadamy teraz dokładniej problemy techniczne, które występują w sieciach bezprzewodowych i mobilnych. Najpierw rozważymy pojedyncze łącze bezprzewodowe; mobilność zostanie omówiona w dalszej części rozdziału.

7.2. Cechy łączy i sieci bezprzewodowych

Zacznijmy od rozważenia prostej sieci przewodowej, na przykład domowej, z przewodowym przełącznikiem ethernetowym łączącym hosty (podrozdział 6.4). Jeśli postanowimy zastąpić przewodowy Ethernet bezprzewodową siecią 802.11, w hostach trzeba będzie wymienić ethernetowe karty sieciowe na karty bezprzewodowe, a zamiast przełącznika zastosować punkt dostępowy, ale w warstwie sieci i wyżej nie trzeba będzie dokonywać praktycznie żadnych zmian. Sugeruje to, że szukając istotnych różnic między sieciami przewodowymi a bezprzewodowymi, powinniśmy skupić się na warstwie łącza. I rzeczywiście, możemy znaleźć kilka ważnych różnic między łączem przewodowym a bezprzewodowym:

- *Malejąca siła sygnału.* Natężenie promieniowania elektromagnetycznego zmniejsza się, kiedy przechodzi ono przez materię (na przykład sygnał radiowy przenika przez ścianę). Nawet w wolnej przestrzeni sygnał ulega rozproszeniu, co powoduje zmniejszanie siły sygnału (czasem określane mianem **utraty ścieżki**) w miarę wzrostu odległości między nadajnikiem a odbiornikiem.
- *Interferencja z innymi źródłami.* Źródła radiowe nadające na tej samej częstotliwości wzajemnie się zakłócają. Przykładowo telefony bezprzewodowe 2,4 GHz oraz sieci lokalne 802.11b nadają na tym samym paśmie, zatem użytkownik sieci lokalnej 802.11b rozmawiający przez telefon 2,4 GHz nie może oczekiwać, że zarówno sieć, jak i telefon będą działać idealnie. Oprócz źródeł nadawczych interferencję może powodować elektromagnetyczny szum w otoczeniu (na przykład pobliski silnik albo kuchenka mikrofalowa).
- *Propagacja wielościeżkowa.* **Propagacja wielościeżkowa** występuje wtedy, kiedy części fali elektromagnetycznej odbijają się od przeszkód i od podłoża, wędrując od nadajnika do odbiornika ścieżkami o różnej długości. Powoduje to rozmycie sygnału przy odbiorniku. Poruszające się obiekty między nadajnikiem a odbiornikiem mogą zmieniać charakter propagacji wielościeżkowej.

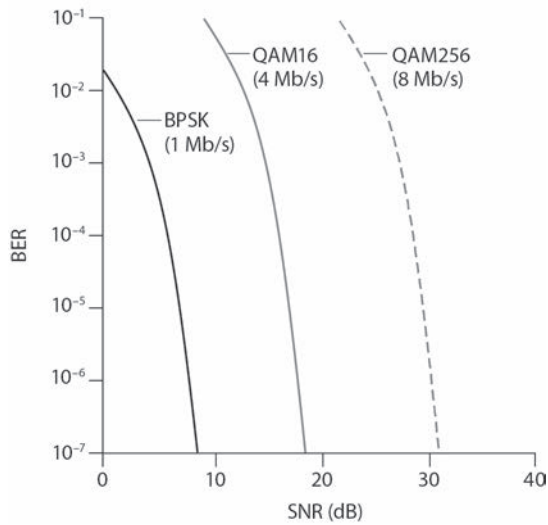
Szczegółowe omówienie właściwości, modeli i pomiarów kanałów bezprzewodowych przedstawiono w źródle [Anderson 1995].

Powyższy opis sugeruje, że błędy bitowe występują częściej w łączach bezprzewodowych niż w przewodowych. Nie dziwi zatem, że protokoły łącza bezprzewodowego (na przykład protokół 802.11, który omówimy w następnym podrozdziale) używają nie tylko zaawansowanych kodów detekcji błędów CRC, ale również protokołów niezawodnej transmisji danych funkcjonujących na poziomie łącza, retransmitujących uszkodzone ramki.

Po omówieniu problemów, jakie mogą wystąpić w kanale bezprzewodowym, skoncentrujmy się na hoście odbierającym sygnały bezprzewodowe. Taki host otrzymuje sygnał elektromagnetyczny, który jest połączeniem osłabionego pierwotnego sygnału wysłanego przez nadawcę (degradacja sygnału wynika ze zmniejszenia natężenia promieniowania elektromagnetycznego i propagacji wielościeżkowej; zjawiska te opisaliśmy wcześniej) i szumu tła obecnego w środowisku. **Stosunek sygnału do szumu** (ang. *Signal to Noise Ratio* — **SNR**) to względna miara mocy odbieranego sygnału

(czyli przesyłanych informacji) i wspomnianego szumu. Wskaźnik ten jest zwykle wyrażany w decybelach (dB). Jest to jednostka miary, która zdaniem niektórych osób jest stosowana przez inżynierów elektryków głównie do utrudniania pracy informatykom. SNR (mierzony w dB) to dwudziestokrotność stosunku logarytmu o podstawie 10 z amplitudy otrzymanego sygnału do amplitudy szumu. Na potrzeby tego rozdziału wystarczy wiedzieć, że wyższy SNR ułatwia odbiorcy wyodrębnienie sygnału z szumu tła.

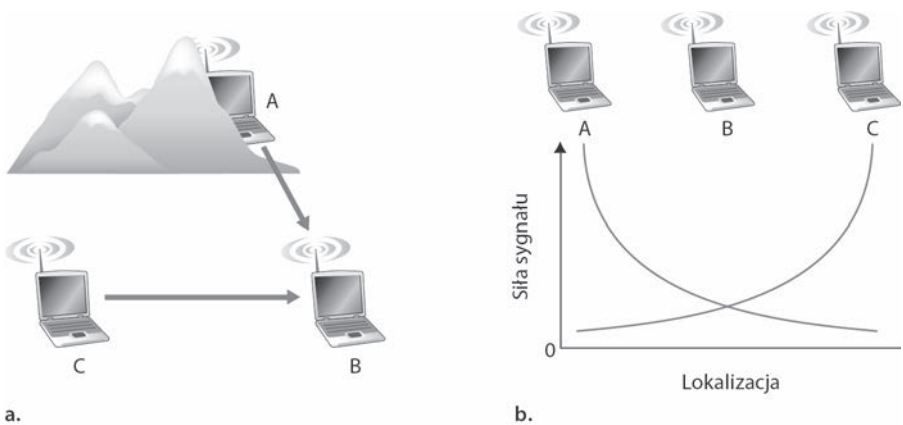
Rysunek 7.3 (zaadaptowany z pracy [Holland 2001]) przedstawia zależność współczynnika błędów bitowych (ang. *Bit Error Rate* — **BER**; można ogólnie powiedzieć, że jest to prawdopodobieństwo tego, iż przesłany bit będzie błędny po stronie odbiorcy) od miary SNR dla trzech różnych technik modulacji stosowanych do kodowania informacji na potrzeby transmisji w wyidealizowanym kanale bezprzewodowym. Omawianie teorii modulacji i kodowania oraz wyodrębniania sygnałów i współczynnika BER znacznie wykracza poza zakres tej książki (opis tych zagadnień zawiera praca [Schwartz 1980]). Tym niemniej rysunek 7.3 ilustruje kilka cech warstwy fizycznej ważnych do zrozumienia protokołów komunikacji bezprzewodowej stosowanych w wyższych warstwach.



Rysunek 7.3. Współczynnik BER, szybkość transmisji i SNR

- *Dla danego systemu modulacji im wyższy SNR, tym niższy współczynnik BER.* Ponieważ nadawca może zwiększyć SNR przez podniesienie mocy transmisji, może w ten sposób zmniejszyć prawdopodobieństwo wystąpienia błędów w odebranej ramce. Warto jednak zauważyć, że powyżej określonego progu wzrost mocy daje małe praktyczne zyski i pozwala zmniejszyć współczynnik BER na przykład z 10^{-12} do 10^{-13}. Istnieją też *wady* związane ze wzrostem mocy transmisji — nadawca musi zużyć więcej energii (jest to ważna kwestia dla użytkowników urządzeń mobilnych z zasilaniem bateryjnym), a dane wysyłane przez nadawcę z większym prawdopodobieństwem będą zakłócać transmisję z innych źródeł (rysunek 7.4(b)).

- *Dla danej wartości SNR technika zapewniająca szybszą transmisję bitów (z błędami i bez) da wyższy współczynnik BER.* Na przykład na rysunku 7.3 przy SNR na poziomie 10 dB modulacja BPSK z szybkością transmisji 1 Mb/s daje współczynnik BER na poziomie poniżej 10^{-7}, natomiast dla modulacji QAM16 (prędkość transmisji 4 Mb/s) współczynnik ten wynosi 10^{-1} i jest zdecydowanie za wysoki, aby ta technika była w praktyce użyteczna. Jednak przy SNR równym 20 dB modulacja QAM16 zapewnia szybkość transmisji na poziomie 4 Mb/s przy współczynniku BER 10^{-7}, natomiast modulacja BPSK ma prędkość transmisji tylko 1 Mb/s, a współczynnik BER jest tak niski, że — dosłownie — nie mieści się na wykresie. Jeśli wartość 10^{-7} dla współczynnika BER jest akceptowalna, wyższa szybkość transmisji zapewniana przez metodę QAM16 sprawia, że w tych warunkach będzie to preferowana technika modulacji. Te rozważania prowadzą do ostatniej, opisanej poniżej cechy.
- *Dynamiczny wybór techniki modulacji w warstwie fizycznej można wykorzystać do dostosowania tej techniki do warunków panujących w kanale.* SNR, a tym samym i współczynnik BER mogą się zmieniać w wyniku przemieszczania się użytkownika lub zmian środowiskowych. W systemach transmisji danych w sieciach komórkowych, w sieciach 802.11 Wi-Fi i w sieciach 4G (omawiamy je w podrozdziałach 7.3 i 7.4) wykorzystywane jest adaptacyjne modulowanie i kodowanie. Rozwiązanie to umożliwia na przykład zastosowanie na podstawie cech kanału techniki modulacji zapewniającej najwyższą możliwą szybkość transmisji przy określonym współczynniku BER.



Rysunek 7.4. Problem ukrytego terminalu (a) i zanikanie (b)

Większa i zmieniająca się w czasie liczba błędów to nie jedyna różnica między łączem przewodowym a bezprzewodowym. Jak wiemy, w przypadku rozgłoszeniowego łącza przewodowego każdy węzeł odbiera transmisje od wszystkich innych węzłów. W przypadku łącza bezprzewodowego sytuacja nie jest taka prosta (rysunek 7.4). Przypuśćmy, że zarówno stacja A, jak i C nadają do stacji B. Tak zwany **problem ukrytego terminalu** występuje wtedy, gdy fizyczne przeszkody w środowisku (na przykład

wzgórze albo budynek) uniemożliwiają stacjom A i C wzajemny odbiór swoich transmisji, mimo że ulegają one interferencji u celu (stacji B). Pokazano to na rysunku 7.4(a). Drugi scenariusz prowadzący do niewykrywalnych kolizji przy odbiorniku wynika z **zanikania** sygnału w miarę propagacji w nośniku bezprzewodowym. Na rysunku 7.4(b) zilustrowano sytuację, w której stacje A i C są umieszczone tak, że jedna nie może wykryć transmisji drugiej, a jednak ich sygnały *są* na tyle silne, że zakłócają się wzajemnie przy stacji B. Jak Czytelnicy zobaczą w podrozdziale 7.3, problem ukrytego terminalu oraz zanikanie sprawiają, że wspólny dostęp w sieci bezprzewodowej jest znacznie bardziej skomplikowany niż w przewodowej.

7.2.1. CDMA

W rozdziale 6. wyjaśniono, że kiedy hosty porozumiewają się przez współużytkowany nośnik, potrzebny jest jakiś protokół, aby sygnały wysyłane przez wiele nadajników nie zakłócały się nawzajem przy odbiorniku. Opisano też trzy klasy protokołów dostępu do nośnika: podział kanału, dostęp losowy i dostęp cykliczny. **CDMA** (ang. *Code Division Multiple Access*) to czwarty typ protokołu dostępu do współużytkowanego nośnika, dominujący w bezprzewodowych sieciach lokalnych i technologiach komórkowych. Ponieważ protokół CDMA jest tak istotny w świecie bezprzewodowym, przyjrzymy mu się teraz, zanim w następnych podrozdziałach zajmiemy się konkretnymi technikami dostępowymi.

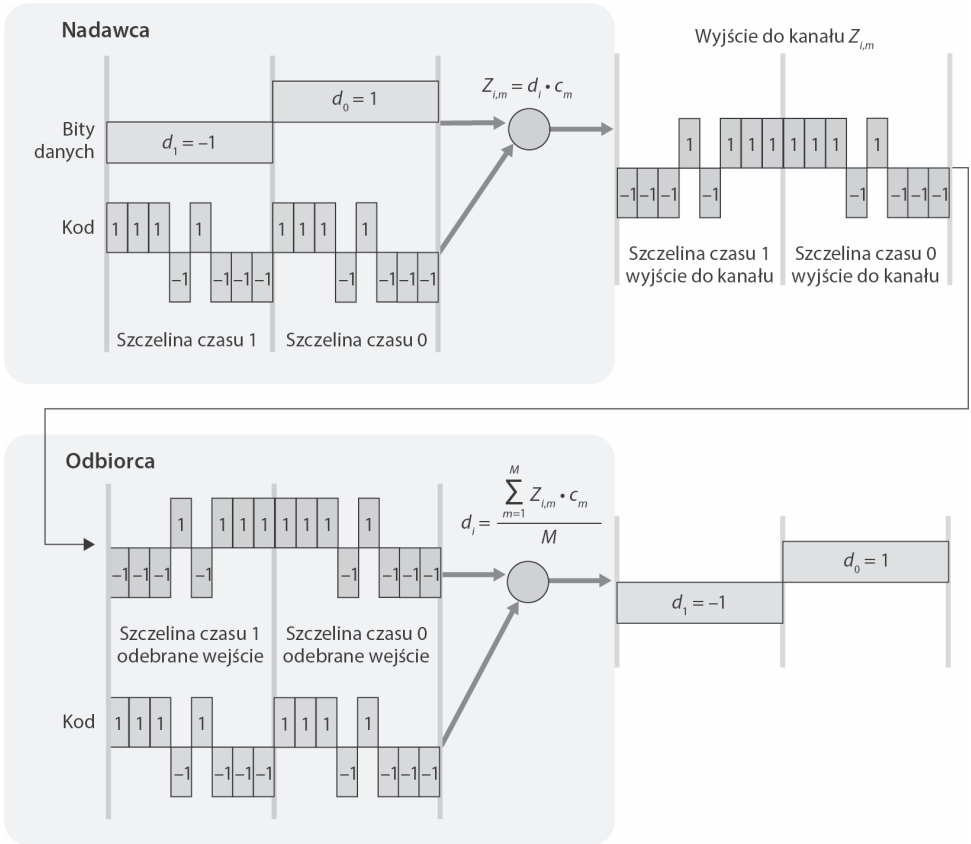
W protokole CDMA każdy wysyłany bit jest kodowany przez pomnożenie go przez sygnał (kod), który zmienia się ze znacznie większą częstotliwością (zwaną **częstotliwością kodowania**, ang. *chipping rate*) niż pierwotna sekwencja bitów danych. Na rysunku 7.5 pokazano prosty, wyidealizowany scenariusz kodowania-dekodowania CDMA. Przypuśćmy, że tempo, w którym pierwotne bity danych docierają do kodera CDMA, definiuje jednostkę czasu, tzn. każdy przeznaczony do wysłania bit zajmuje jednobitową szczelinę czasu. Niech d_i będzie wartością bitu danych w i-tej szczelinie. Dla wygody będziemy reprezentować bit danych o wartości 0 jako –1. Każda szczelina bitu jest podzielona na M miniszczelin; na rysunku 7.5 $M = 8$, choć w praktyce wartość ta jest znacznie większa. Kod CDMA używany przez nadajnik składa się z sekwencji M wartości c_m (gdzie $m = 1, \ldots, M$), każdej równej +1 albo –1. W przykładzie z rysunku 7.5 M-bitowy kod CDMA używany przez nadajnik to (1, 1, 1, –1, 1, –1, –1, –1).

Aby zrozumieć, jak działa CDMA, skupmy się na i-tym bicie danych, d_i. Przez m-tą miniszczelinę czasu transmisji bitu d_i na wyjściu kodera CDMA, $Z_{i,m}$, pojawia się wartość d_i pomnożona przez m-ty bit kodu CDMA, c_m:

$$Z_{i,m} = d_i \cdot c_m \tag{7.1}$$

W prostym świecie, bez interferencji innych nadajników, odbiornik otrzymałby zakodowane bity, $Z_{i,m}$, i odtworzył pierwotny bit danych według poniższego wzoru:

$$d_i = \frac{1}{M} \sum_{m=1}^{M} Z_{i,m} \cdot c_m \tag{7.2}$$



Rysunek 7.5. Prosty przykład CDMA: nadajnik koduje, odbiornik dekoduje

Warto szczegółowo prześledzić przykład z rysunku 7.5, aby upewnić się, że pierwotne bity danych rzeczywiście zostaną prawidłowo odtworzone przez odbiornik z wykorzystaniem równania 7.2.

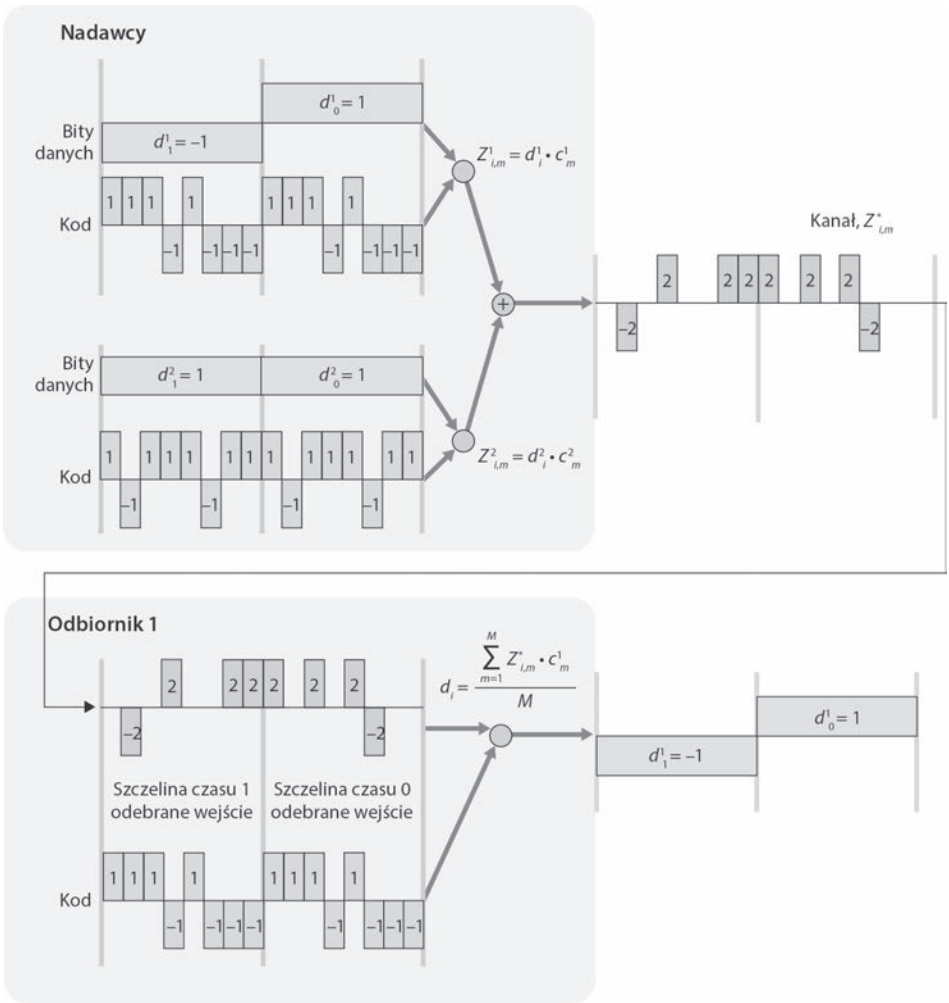
Świat nie jest jednak idealny — jak już wspomniano, CDMA musi działać w środowisku z wieloma nadajnikami, które kodują i transmitują dane przy użyciu różnych kodów. Jak odbiornik CDMA ma odtworzyć pierwotne bity danych, skoro nakładają się one na bity wysyłane przez inne nadajniki? CDMA opiera się na założeniu, że interferujące sygnały bitowe są addytywne. Oznacza to, że jeśli w tej samej miniszczelinie czasu trzy nadajniki wysyłają wartość 1, a jeden nadajnik wysyła wartość –1, to wszystkie odbiorniki otrzymują wartość 2 (1+1+1–1 = 2). W środowisku z wieloma nadajnikami każdy nadajnik *s* oblicza kodowaną transmisję, $Z^s_{i,m}$, w dokładnie taki sam sposób jak w równaniu 7.1. Jednak wartość otrzymywana przez odbiornik w *m*-tej miniszczelinie *i*-tej szczeliny bitowej jest teraz *sumą* bitów wysłanych przez *N* nadajników w tej miniszczelinie:

$$Z^*_{i,m} = \sum_{s=1}^{N} Z^s_{i,m}$$

Zdumiewające jest to, że jeśli kody nadajników zostaną odpowiednio dobrane, każdy odbiornik może wyodrębnić ze zbiorczego sygnału dane wysłane przez konkretny nadajnik. W tym celu musi użyć kodu nadajnika w dokładnie ten sam sposób, który pokazano w równaniu 7.2:

$$d_i = \frac{1}{M}\sum_{m=1}^{M} Z_{i,m}^{*} \cdot c_m \tag{7.3}$$

Rysunek 7.6 przedstawia przykład CDMA z dwoma nadajnikami. *M*-bitowy kod CDMA używany przez górny nadajnik to (1, 1, 1, –1, 1, –1, –1, –1); dolny nadajnik używa kodu (1, –1, 1, 1, 1, –1, 1, 1). Na rysunku pokazano, jak odbiornik odtwarza pierwotne bity danych otrzymane od górnego nadajnika. Zwróćmy uwagę, że nadajnik może wyodrębnić dane nadajnika 1 mimo interferencji z transmisją nadajnika 2.



Rysunek 7.6. Przykład z dwoma nadajnikami CDMA

Przypomnijmy sobie analogię przyjęcia z rozdziału 6. Protokół CDMA przypomina uczestników przyjęcia mówiących w różnych językach; w takich okolicznościach ludzie potrafią skupić się na konwersacjach w języku, który rozumieją, odfiltrowując pozostałe rozmowy. Jak widać, protokół CDMA jest protokołem dzielącym w tym sensie, że dzieli przestrzeń kodową (a nie czas lub częstotliwość) i przypisuje każdemu węzłowi część tej przestrzeni.

Powyższy opis CDMA był z konieczności skrótowy; w praktyce protokół ten wymaga rozwiązania kilku trudnych problemów. Po pierwsze, aby nadajniki CDMA mogły wyodrębniać sygnały poszczególnych nadajników, kody muszą być starannie dobrane. Po drugie, założyliśmy tu, że siła sygnałów odbieranych od poszczególnych nadajników jest taka sama; w rzeczywistości trudno jest to osiągnąć. Napisano wiele książek poświęconych tym i innym aspektom CDMA; szczegółowe informacje można znaleźć w [Pickholz 1982; Viterbi 1995].

7.3. Wi-Fi — bezprzewodowe sieci lokalne 802.11

Coraz częściej spotykane w biurach, domach, instytucjach edukacyjnych, kawiarniach, na lotniskach i rogach ulic bezprzewodowe sieci lokalne są obecnie jedną z najważniejszych technologii dostępu do internetu. Choć w latach 1990. opracowano wiele technologii i standardów komunikacji bezprzewodowej, jedna z klas tych standardów zdobyła zdecydowaną przewagę: **bezprzewodowa sieć lokalna IEEE 802.11**, znana również jako **Wi-Fi**. W tym podrozdziale przyjrzymy się bliżej bezprzewodowym sieciom lokalnym 802.11. Zbadamy strukturę ramki 802.11, protokół dostępu do nośnika oraz kwestie łączenia sieci 802.11 z przewodowymi, ethernetowymi sieciami lokalnymi.

Istnieje kilka standardów komunikacji w rodzinie bezprzewodowych sieciach lokalnych 802.11 (Wi-Fi). Przedstawiono je w tabeli 7.1. Różne standardy 802.11 mają wiele cech wspólnych. Używają tego samego protokołu dostępu do nośnika, CSMA/CA, który omówimy niebawem. Mają też jednakową strukturę ramki w warstwie łącza. Mogą zmniejszać szybkość transmisji, aby zwiększyć jej zasięg. Ważne jest też to, że produkty obsługujące standardy 802.11 są zgodne wstecz. Oznacza to na przykład, że urządzenie obsługujące tylko standard 802.11g może komunikować się z nowszą stacją bazową 802.11 ac.

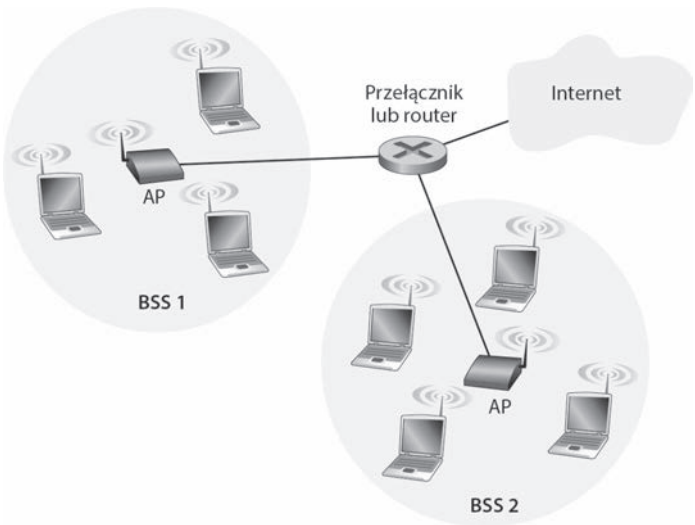
Standard	Zakres częstotliwości	Szybkość transmisji danych
802.11b	2,4 GHz	do 11 Mb/s
802.11a	5 GHz	do 54 Mb/s
802.11g	2,4 GHz	do 54 Mb/s
802.11n	2,5 i 5 GHz	do 450 Mb/s
802.11ac	5 GHz	do 1300 Mb/s

Tabela 7.1. Podsumowanie standardów IEEE 802.11

Jednak, jak pokazano w tabeli 7.1, standardy różnią się znacznie w warstwie fizycznej. Urządzenia 802.11 działają w dwóch różnych zakresach częstotliwości: 2,4 – 2,485 GHz (określanym jako zakres 2,4 Ghz) i 5,1 – 5,8 GHz (określanym jako zakres 5 Ghz). Zakres 2,4 Ghz jest nielicencjonowany, a posługujące się nim urządzenia współzawodniczą o pasmo z telefonami 2,4 GHz i kuchenkami mikrofalowymi. Sieci lokalne 802.11 używające zakresu 5 Ghz przy tym samym poziomie mocy mają mniejszy zasięg transmisji i są bardziej podatne na propagację wielościeżkową. Dwa najnowsze standardy, 802.11n [IEEE 802.11n 2012] i 802.11ac [IEEE 802.11ac 2013; Cisco 802.11ac 2015], korzystają z anten **MIMO** (ang. *multiple-input, multiple-output*). Oznacza to, że do nadawania i odbioru różnych sygnałów po stronie nadawcy i odbiorcy używane są przynajmniej po dwie anteny [Diggavi 2004]. Stacje bazowe 802.11ac mogą jednocześnie przesyłać dane do wielu stacji i wykorzystują „inteligentne" anteny do adaptacyjnego kształtowania wiązki sygnału w celu kierowania transmisji do odbiorcy. Zmniejsza to interferencje i zwiększa zasięg przy tej samej szybkości transmisji danych. Szybkości podane w tabeli 7.1 dotyczą idealnych warunków — odbiorca jest umieszczony metr od stacji bazowej, a interferencje nie występują. W praktyce taki scenariusz jest nieprawdopodobny. Dlatego wszystko zależy od sytuacji.

7.3.1. Architektura sieci 802.11

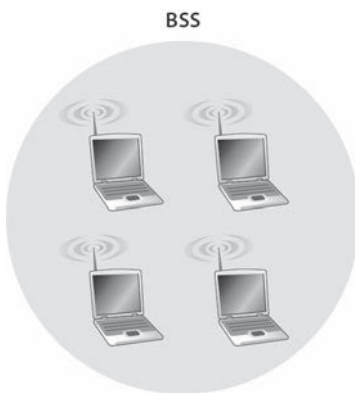
Na rysunku 7.7 przedstawiono podstawowe komponenty bezprzewodowej sieci lokalnej 802.11. Fundamentalnym elementem architektury 802.11 jest **grupa usługowa** (ang. *Basic Service Set* — **BSS**). BSS składa się z jednej lub wielu stacji bezprzewodowych oraz centralnej **stacji bazowej**, w terminologii 802.11 określanej mianem **punktu dostępowego** (ang. *access point* — **AP**). Na rysunku pokazano dwie grupy BSS, każdą z punktem dostępowym podłączonym do urządzenia takiego jak przełącznik lub router, które z kolei łączy się z internetem. W typowej sieci domowej znajduje się jeden punkt dostępowy i jeden router (często zintegrowane w jednej obudowie), który łączy BSS z internetem.



Rysunek 7.7. Architektura sieci lokalnej 802.11

Podobnie jak urządzenia ethernetowe każda stacja bezprzewodowa 802.11 ma 6-bajtowy adres MAC przechowywany w oprogramowaniu firmowym adaptera (to znaczy karty sieciowej 802.11). Każdy punkt dostępowy również ma adres MAC przypisany interfejsowi bezprzewodowemu. Tak jak w przypadku Ethernetu, adresy MAC są zarządzane przez IEEE i (teoretycznie) globalnie unikatowe.

Jak wspomniano w podrozdziale 7.1, bezprzewodowe sieci lokalne z punktami dostępowymi często określa się mianem sieci **infrastrukturalnych**, przy czym „infrastrukturą" są punkty dostępowe wraz z przewodowym Ethernetem, który łączy je z routerem. Na rysunku 7.8 pokazano, że stacje IEEE 802.11 mogą również łączyć się w sieć doraźną (ad hoc) — bez centralnej kontroli i bez połączenia z zewnętrznym światem. Sieć taka jest tworzona „w locie" przez urządzenia mobilne, które znalazły się blisko siebie, muszą nawiązać komunikację, a nie mogą odszukać istniejącej infrastruktury sieciowej. Sieć doraźna może powstać na przykład wtedy, kiedy spotkają się dwie osoby z laptopami (w sali konferencyjnej, w pociągu, w samochodzie) i zechcą wymienić dane pod nieobecność scentralizowanego punktu dostępowego. Sieci doraźne cieszą się coraz większym zainteresowaniem w miarę wzrostu liczby przenośnych urządzeń komunikacyjnych. W tym podrozdziale skoncentrujemy się jednak na sieciach infrastrukturalnych.



Rysunek 7.8. Doraźna sieć IEEE 802.11

Kanały i powiązania

W sieci 802.11 każda stacja bezprzewodowa musi związać się z punktem dostępowym, zanim będzie mogła wysyłać lub odbierać dane warstwy sieci. Choć powiązania stosuje się we wszystkich standardach 802.11, omówimy to zagadnienie w kontekście sieci IEEE 802.11b/g.

Kiedy administrator sieci instaluje punkt dostępowy, przypisuje mu jedno- albo dwuwyrazowy **identyfikator grupy usługowej** (ang. *Service Set Identifier* — **SSID**). Przykładowo podczas przeglądania dostępnych sieci w iPhonie wyświetlana jest lista pokazująca SSID każdego punktu dostępowego w zasięgu urządzenia. Administrator musi również przypisać punktowi dostępowemu numer kanału. Jak wiemy, sieci 802.11

operują w zakresie częstotliwości od 2,4 GHz do 2,485 GHz. W tym paśmie o szerokości 85 MHz standard 802.11 definiuje 11 częściowo nakładających się kanałów. Dowolne dwa kanały nie nakładają się na siebie, jeśli są oddzielone czterema lub większą liczbą kanałów. W szczególności kanały 1., 6. oraz 11. tworzą jedyny zbiór trzech nienakładających się kanałów. Oznacza to, że administrator może utworzyć bezprzewodową sieć lokalną z maksymalną sumaryczną szybkością transmisji równą 33 Mb/s, instalując w jednym miejscu trzy punkty dostępowe 802.11b, przypisując im kanały 1., 6. oraz 11. i podłączając każdy z nich do przełącznika.

Skoro rozumiemy już pojęcie kanałów 802.11, opiszmy interesującą (i wcale nierzadką) sytuację: dżunglę Wi-Fi. **Dżunglą Wi-Fi** nazywamy każdą fizyczną lokację, w której stacja bezprzewodowa odbiera wystarczająco silny sygnał od dwóch lub wielu punktów dostępowych. Na przykład w wielu nowojorskich kawiarniach stacja bezprzewodowa może odebrać sygnał od punktu zarządzanego przez kawiarnię oraz od punktów działających w okolicznych mieszkaniach. Każdy z tych punktów prawdopodobnie będzie znajdował się w innej podsieci i miał niezależnie przypisany kanał.

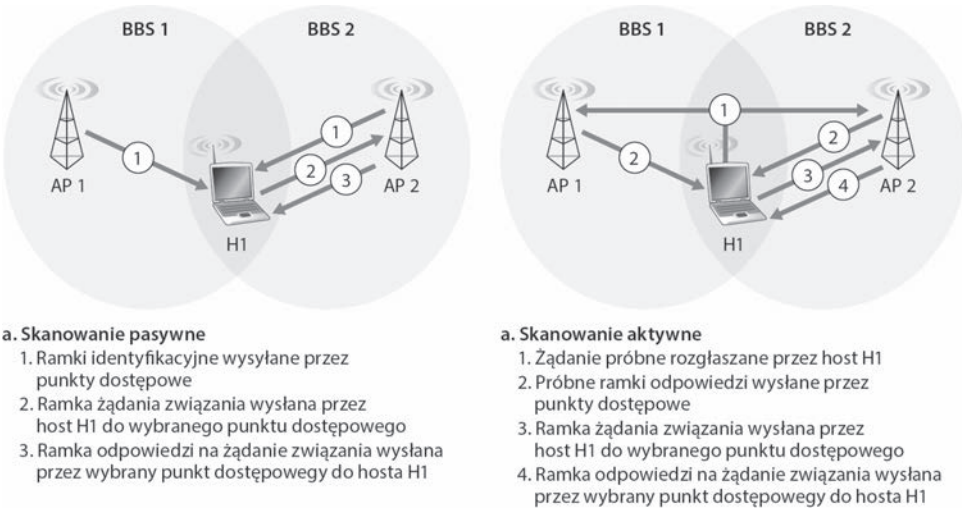
Przypuśćmy teraz, że wchodzimy do dżungli Wi-Fi z telefonem, tabletem lub laptopem, szukając dostępu do internetu i bułki z jagodami. Załóżmy, że w dżungli znajduje się pięć punktów dostępowych. Aby uzyskać dostęp do internetu, stacja bezprzewodowa musi podłączyć się do dokładnie jednej podsieci, a zatem **związać się** z dokładnie jednym punktem dostępowym. Wiązanie jest równoznaczne z utworzeniem wirtualnego „przewodu" miedzy stacją a punktem dostępowym. Mówiąc ściślej, tylko powiązany punkt dostępowy będzie wysyłał ramki danych (tzn. ramki zawierające dane, takie jak datagramy) do stacji bezprzewodowej, a stacja będzie wysyłać ramki danych do internetu tylko za pośrednictwem powiązanego punktu. Jednak w jaki sposób stacja bezprzewodowa wiąże się z określonym punktem dostępowym? I bardziej ogólnie — skąd stacja bezprzewodowa „wie", jakie punkty dostępowe — jeśli w ogóle — kryją się w dżungli?

Standard 802.11 wymaga, aby punkt dostępowy okresowo wysyłał **ramki identyfikacyjne** (ang. *beacon frames*), z których każda zawiera jego identyfikator SSID i adres MAC. Stacja bezprzewodowa skanuje wszystkich 11 kanałów, szukając ramek identyfikacyjnych jakiegoś punktu dostępowego (niektóre mogą nadawać na tym samym kanale — jesteśmy w dżungli!). Po wykryciu dostępnych punktów dostępowych stacja (albo jej użytkownik) wybiera ten, z którym się zwiąże.

W standardzie 802.11 nie określono algorytmu wyboru jednego z punktów dostępowych, z którym stacja się zwiąże. Za opracowanie takiego algorytmu odpowiadają projektanci oprogramowania firmowego urządzeń 802.11 i oprogramowania bezprzewodowego hosta. Zwykle host wybiera punkt, który ma najsilniejszy sygnał na etapie wysyłania ramki identyfikacyjnej. Choć mocny sygnał jest korzystny (zobacz na przykład rysunek 7.3), nie jest to jedyna cecha punktu dostępowego wpływająca na poziom usług świadczonych hostowi. Punkt dostępowy może generować silny sygnał, ale obsługiwać zbyt wiele powiązanych hostów (muszą one współdzielić przepustowość bezprzewodowego kanału tego punktu), jednak nieprzeciążony punkt dostępowy nie zostanie wybrany ze względu na nieco słabszy sygnał. Dlatego w ostatnim czasie

zaproponowano wiele nowych sposobów wyboru punktu dostępowego [Vasudevan 2005; Nicholson 2006; Sudaresan 2006]. Ciekawe i przystępne omówienie pomiaru siły sygnału przedstawiono w pracy [Bardwell 2004].

Proces skanowania kanałów i oczekiwania na ramki sygnalizacyjne to tak zwane **skanowanie pasywne** (rysunek 7.9(a)). Bezprzewodowy host może też przeprowadzić **skanowanie aktywne** przez rozgłaszanie ramki próbnej (ang. *probe frame*). Odbiorą ją wszystkie punkty dostępowe w zasięgu danego hosta (rysunek 7.9(b)). Punkt dostępowy po odebraniu próbnej ramki żądania przesyła próbną ramką odpowiedzi. Następnie bezprzewodowy host może spośród odpowiadających punktów dostępowych wybrać ten, z którym się zwiąże.



a. Skanowanie pasywne
1. Ramki identyfikacyjne wysyłane przez punkty dostępowe
2. Ramka żądania związania wysłana przez host H1 do wybranego punktu dostępowego
3. Ramka odpowiedzi na żądanie związania wysłana przez wybrany punkt dostępowegy do hosta H1

a. Skanowanie aktywne
1. Żądanie próbne rozgłaszane przez host H1
2. Próbne ramki odpowiedzi wysłane przez punkty dostępowe
3. Ramka żądania związania wysłana przez host H1 do wybranego punktu dostępowego
4. Ramka odpowiedzi na żądanie związania wysłana przez wybrany punkt dostępowegy do hosta H1

Rysunek 7.9. Aktywne i pasywne skanowanie w poszukiwaniu punktów dostępowych

Bezprzewodowy host po wybraniu punktu dostępowego, z którym się zwiąże, wysyła do tego punktu ramkę żądania związania, a punkt odsyła ramkę odpowiedzi na żądanie związania. Warto zauważyć, że drugie negocjacje żądanie-odpowiedź przy aktywnym skanowaniu są potrzebne, ponieważ punkt dostępowy reagujący na początkową próbną ramkę żądania nie wie, który z reagujących (potencjalnie wielu) punktów host wybierze. Podobnie klient DHCP może wybrać jeden z wielu serwerów DHCP (rysunek 4.21). Po związaniu się z punktem dostępowym host próbuje połączyć się z podsiecią (w sensie adresowania IP opisanym w punkcie 4.3.3), do której należy ten punkt. Dlatego zwykle wysyła poprzez punkt dostępowy komunikat wykrywania serwera DHCP (rysunek 4.21) do podsieci, aby pozyskać adres IP z tej podsieci. Kiedy już go otrzyma, świat zewnętrzny będzie postrzegał dany host jako jeszcze jedną stację o adresie IP z tej podsieci.

Aby związać się z danym punktem dostępowym, stacja bezprzewodowa czasem musi się uwierzytelnić. Bezprzewodowe sieci lokalne 802.11 zapewniają kilka sposobów uwierzytelniania i dostępu. Jedną z metod, używaną w wielu firmach, jest zezwalanie na dostęp do sieci bezprzewodowej na podstawie adresu MAC stacji. Inna metoda,

często wykorzystywana w kawiarenkach internetowych, wymaga podania nazwy użytkownika i hasła. W obu przypadkach punkt dostępowy zwykle komunikuje się z serwerem uwierzytelniania, przekazując informacje między stacją bezprzewodową a serwerem przy użyciu protokołu takiego jak RADIUS [RFC 2865] albo DIAMETER [RFC 3588]. Oddzielenie serwera uwierzytelniania od punktu dostępowego pozwala na obsługę wielu punktów przez jeden serwer, dzięki czemu decyzje dotyczące uwierzytelniania i dostępu (często poufne) są podejmowane centralnie, a punkty dostępowe mogą być względnie tanie i nieskomplikowane. W rozdziale 8 przekonamy się, że w nowym protokole IEEE 802.11i definiującym aspekty bezpieczeństwa rodziny protokołów 802.11 przyjęto dokładnie takie podejście.

7.3.2. Protokół kontroli dostępu do nośnika 802.11

Kiedy stacja bezprzewodowa jest już związana z punktem dostępowym, może rozpocząć wysyłanie i odbieranie ramek danych. Ponieważ jednak może się zdarzyć, że wiele stacji będzie jednocześnie dążyć do przesłania ramek danych tym samym kanałem, potrzebny jest protokół kontroli dostępu, który będzie koordynował transmisje. W tym kontekście termin stacja odnosi się zarówno do stacji bezprzewodowej, jak i do punktu dostępowego. Jak wyjaśniono w rozdziale 6. i w punkcie 7.2.1, istnieją trzy ogólne klasy protokołów wielokrotnego dostępu: dzielenie kanału (w tym CDMA), dostęp losowy i dostęp cykliczny. Zainspirowani sukcesem Ethernetu i protokołów dostępu losowego projektanci standardu 802.11 postanowili wykorzystać taki protokół również w sieciach bezprzewodowych. Protokół ten określany jest mianem **CSMA z unikaniem kolizji** (ang. *CSMA with collison avoidance*), w skrócie **CSMA/CA**. Podobnie jak w przypadku ethernetowego protokołu CSMA/CD „CSMA" to skrót od „carrier sense multiple access", co oznacza, że każda stacja bada kanał przed wysłaniem danych i wstrzymuje transmisję, kiedy kanał jest zajęty. Choć zarówno Ethernet, jak i 802.11 używają dostępu swobodnego z badaniem nośnika, istnieją między nimi ważne różnice. Po pierwsze, zamiast wykrywania kolizji protokół 802.11 używa technik unikania kolizji. Po drugie, ze względu na względnie wysoki współczynnik błędów bitowych w kanałach bezprzewodowych protokół 802.11 (w przeciwieństwie do Ethernetu) wykorzystuje technikę potwierdzeń i retransmisji (ARQ) w warstwie łącza. Unikanie kolizji i potwierdzanie transmisji w warstwie łącza opiszemy poniżej.

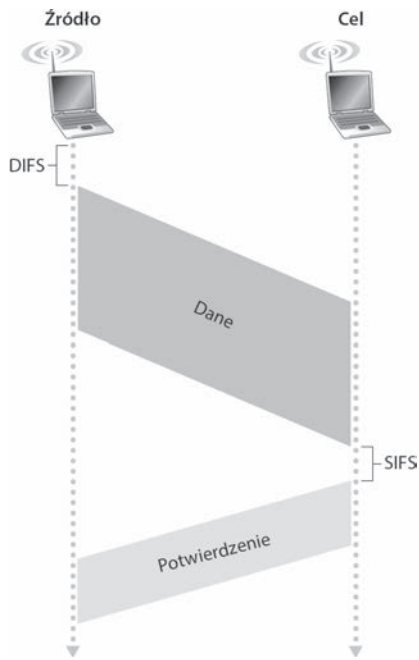
W podrozdziałach 6.3.2 i 6.4.2 wyjaśniono, że w ethernetowym algorytmie wykrywania kolizji stacja prowadzi nasłuch kanału transmisji. Jeśli podczas wysyłania danych wykryje transmisję innej stacji, przerywa na chwilę własną transmisję i ponawia próbę po krótkim, wybranym losowo okresie oczekiwania. W przeciwieństwie do ethernetowego protokołu 802.3 protokół MAC 802.11 *nie stosuje* wykrywania kolizji. Zdecydowano się na to z dwóch ważnych powodów:

- Wykrywanie kolizji wymaga zdolności do jednoczesnego wysyłania (własnego sygnału) oraz odbierania danych (w celu ustalenia, czy inna stacja również prowadzi transmisję). Ponieważ moc odbieranego sygnału jest zwykle znacznie niższa od mocy sygnału nadawanego przez adapter 802.11, sprzęt do wykrywania kolizji byłby kosztowny.

- Co ważniejsze, gdyby nawet adapter mógł jednocześnie nadawać i odbierać dane (i rezygnować z transmisji w razie wykrycia zajętego kanału), nadal nie mógłby wykrywać wszystkich kolizji ze względu na problem ukrytego terminalu oraz zanikanie (podrozdział 7.2).

Ponieważ bezprzewodowe sieci lokalne 802.11 nie wykrywają kolizji, kiedy stacja rozpocznie transmisję ramki, to *wysyła ją całą*; innymi słowy, transmisje nigdy nie są przerywane. Łatwo się domyślić, że wysyłanie całych ramek (zwłaszcza długich) w środowisku z wieloma kolizjami może znacznie zmniejszyć wydajność protokołu dostępowego. Aby ograniczyć ten niepożądany efekt, protokół 802.11 stosuje kilka technik unikania kolizji, które niebawem omówimy.

Zanim jednak zajmiemy się unikaniem kolizji, najpierw musimy zbadać metodę **potwierdzania transmisji w warstwie łącza** 802.11. W podrozdziale 7.2 stwierdzono, że kiedy stacja w bezprzewodowej sieci lokalnej wysyła ramkę, ramka ta nie zawsze trafia do stacji odbiorczej w nienaruszonym stanie. Aby uwzględnić względnie wysokie prawdopodobieństwo niepowodzenia, protokół MAC 802.11 stosuje potwierdzenia w warstwie łącza. Jak pokazano na rysunku 7.10, kiedy docelowa stacja odbiera ramkę, która pomyślnie przechodzi test CRC, odczekuje ona przez okres zwany **krótkim opóźnieniem międzyramkowym** (ang. *Short Inter-frame Spacing* — **SIFS**), a następnie odsyła ramkę potwierdzenia. Jeśli stacja nadawcza nie otrzyma potwierdzenia w określonej szczelinie czasu, to zakłada, że wystąpił błąd i retransmituje ramkę, ponownie uzyskując dostęp do kanału za pomocą protokołu CSMA/CA. Jeśli po określonej z góry liczbie retransmisji nie nadejdzie potwierdzenie, stacja nadawcza rezygnuje z dalszych prób i odrzuca ramkę.



Rysunek 7.10. Protokół 802.11 i potwierdzenia w warstwie łącza

Wyjaśniwszy, jak protokół 802.11 wykorzystuje potwierdzenia w warstwie łącza, możemy omówić protokół CSMA/CA. Przypuśćmy, że stacja (stacja bezprzewodowa albo punkt dostępowy) ma ramkę do przesłania.

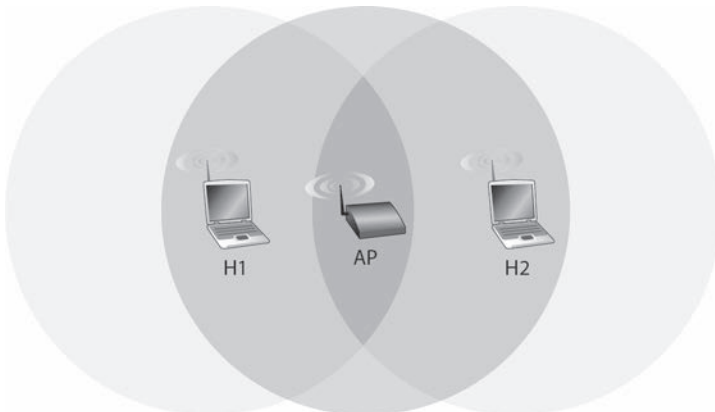
1. Jeśli początkowo stacja ustali, że kanał jest wolny, przesyła ramkę po okresie zwanym **opóźnieniem międzyramkowym** (ang. *Distributed Inter-frame Space* — **DIFS**), jak pokazano na rysunku 7.10.
2. W przeciwnym razie stacja wybiera losowy interwał wstrzymania z użyciem odczekiwania wykładniczego (opisanego w punkcie 6.3.2) i odlicza czas po upływie opóźnienia DIFS, pod warunkiem jednak, że kanał jest wolny. Jeśli kanał jest zajęty, licznik pozostaje zamrożony.
3. Kiedy licznik zmniejszy się do zera (zauważmy, że może tak się stać tylko wtedy, gdy kanał jest wolny), stacja transmituje całą ramkę i czeka na potwierdzenie.
4. Jeśli nadejdzie potwierdzenie, stacja nadawcza „wie", że jej ramka została prawidłowo odebrana przez stację docelową. Jeśli jest kolejna ramka do wysłania, stacja wznawia protokół CSMA/CA od etapu 2. Jeśli potwierdzenie nie nadejdzie, stacja nadawcza ponownie wchodzi w fazę wstrzymania w etapie 2., wybierając losową wartość z większego zakresu.

Warto przypomnieć, że w ethernetowym protokole wielodostępu, CSMA/CD (punkt 6.3.2), stacja rozpoczyna nadawanie, kiedy tylko wykryje wolny kanał. Jednak w protokole CSMA/CA stacja wstrzymuje transmisję na czas odliczania nawet wtedy, gdy stwierdzi, że kanał jest wolny. Dlaczego w protokołach CSMA/CA i CSMA/CD zastosowano odmienne podejścia?

Aby odpowiedzieć na to pytanie, rozważmy scenariusz, w którym dwie stacje mają ramkę danych do przesłania, ale nie robią tego natychmiast, ponieważ wykrywają transmisję trzeciej stacji. W ethernetowym protokole CSMA/CD obie stacje rozpoczęłyby transmisję, kiedy tylko wykryłyby, że trzecia stacja zakończyła nadawanie. Spowodowałoby to kolizję, która w protokole CSMA/CD nie stanowi poważnego problemu, ponieważ obie stacje przerwałyby transmisje i uniknęły niepotrzebnego wysyłania reszty uszkodzonych ramek. W protokole 802.11 sytuacja wygląda jednak zupełnie inaczej. Protokół ten nie wykrywa kolizji i nie przerywa transmisji, więc obie ramki zostałyby przesłane w całości. Celem protokołu 802.11 jest zatem unikanie kolizji, jeśli tylko jest to możliwe. Jeśli dwie stacje wykryją, że kanał jest zajęty, obie wchodzą w tryb oczekiwania wykładniczego, zazwyczaj wybierając inne okresy. Jeśli te losowe okresy rzeczywiście się różnią, po zwolnieniu kanału jedna stacja zacznie nadawać przed drugą. Stacja „przegrana" odbierze sygnał „zwycięskiej" (pod warunkiem że stacje nie są przed sobą ukryte), zamrozi swój licznik i nie zacznie nadawać, dopóki tamta nie zakończy swojej transmisji. Pozwala to uniknąć kosztownych kolizji. Oczywiście w protokole 802.11 w opisanym scenariuszu nadal pozostaje pewne prawdopodobieństwo kolizji: stacje mogą być przed sobą ukryte albo wybrać na tyle zbliżone okresy wstrzymania, że transmisja z pierwszej stacji nie zdąży dotrzeć do drugiej. Na problem ten natrafiliśmy już w czasie analizy algorytmów dostępu losowego w kontekście sytuacji z rysunku 6.12.

Ukryte terminale: RTS i CTS

Protokół MAC 802.11 oferuje również pomysłową (choć opcjonalną) metodę rezerwacji, która pomaga unikać kolizji w środowisku z ukrytymi terminalami. Zbadajmy tę metodę w kontekście rysunku 7.11, na którym pokazano dwie stacje bezprzewodowe i jeden punkt dostępowy. Obie stacje znajdują się w zasięgu punktu dostępowego (zasięgi są przedstawione jako cieniowane koła) i obie są z nim związane. Jednakże ze względu na zanikanie sygnał stacji bezprzewodowych jest ograniczony do wnętrza kół. Zatem każda stacja jest ukryta przed drugą, choć obie są widoczne dla punktu dostępowego.

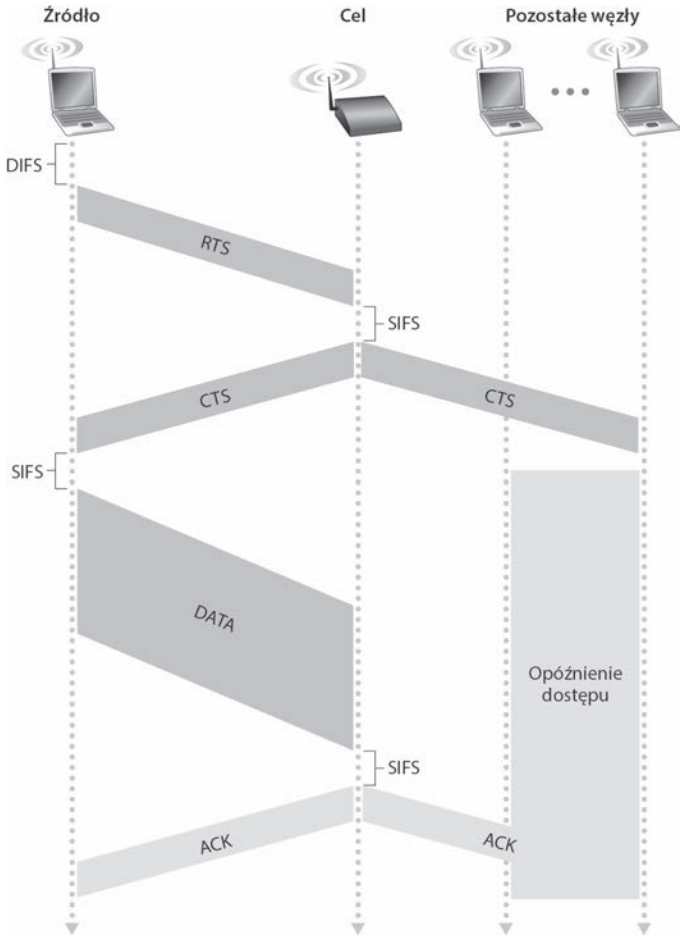


Rysunek 7.11. Przykład ukrytego terminalu: stacja H1 jest ukryta przed H2 i odwrotnie

Zastanówmy się, dlaczego ukryte terminale mogą powodować problemy. Przypuśćmy, że stacja H1 wysyła ramkę, a w połowie jej transmisji stacja H2 chce zacząć przekazywanie ramki do punktu dostępowego. Stacja H2 nie odbiera sygnału stacji H1, więc najpierw odczeka przez krótki, losowy okres DIFS, a następnie wyśle ramkę, tym samym powodując kolizję. Kanał będzie więc bezużyteczny przez cały okres transmisji stacji H1, a także H2.

Aby uniknąć tego problemu, protokół IEEE 802.11 zezwala na użycie krótkich ramek **Request to Send** (**RTS**) oraz **Clear To Send** (**CTS**) w celu *zarezerwowania* dostępu do kanału. Kiedy nadawca chce wysłać ramkę DATA, najpierw wysyła do punktu dostępowego ramkę RTS, określając łączny czas transmisji ramki DATA oraz ramki potwierdzenia (ACK). Kiedy punkt dostępowy odbierze ramkę RTS, odpowiada ramką CTS. Pełni ona dwie funkcje: daje nadawcy jawne zezwolenie na transmisję, a jednocześnie zabrania innym stacjom wysyłania danych przez zarezerwowany okres.

Tak więc w przykładzie z rysunku 7.12 przed wysłaniem ramki DATA stacja H1 najpierw rozgłasza ramkę RTS, którą odbierają wszystkie stacje w jej zasięgu, w tym punkt dostępowy. Punkt dostępowy odpowiada ramką CTS, którą odbierają wszystkie stacje w jego zasięgu, w tym stacje H1 i H2. Stacja H2 po odebraniu ramki CTS opóźnia transmisję o podany w niej czas. Ramki RTS, CTS, DATA i ACK są pokazane na rysunku 7.12.



Rysunek 7.12. Unikanie kolizji z wykorzystaniem ramek RTS i CTS

Zastosowanie ramek RTS i CTS pozwala zwiększyć wydajność sieci na dwa sposoby:

- problem ukrytego terminalu zostaje złagodzony, ponieważ ramka DATA jest transmitowana dopiero po zarezerwowaniu kanału;
- ponieważ ramki RTS i CTS są niewielkie, związane z nimi kolizje są krótkotrwałe. Kiedy ramki RTS i CTS zostaną przesłane poprawnie, transmisja ramek DATA i ACK powinna być bezkolizyjna.

W witrynie internetowej opracowanej na potrzeby niniejszej książki znajduje się interaktywna animacja związana ze standardem 802.11, która ilustruje działanie protokołu CSMA/CA, w tym wymianę sygnałów RTS-CTS.

Choć wymiana RTS-CTS pomaga ograniczyć kolizje, to wprowadza opóźnienia i pochłania zasoby kanału. Z tej przyczyny wykorzystuje się ją (jeśli w ogóle) tylko do rezerwowania kanału na transmisję długiej ramki DATA. W praktyce każda stacja

bezprzewodowa może ustawić próg RTS w taki sposób, aby sekwencja RTS-CTS była stosowana tylko w przypadku ramek przekraczających ten próg. W wielu stacjach bezprzewodowych domyślna wartość progu RTS jest większa od maksymalnej długości ramki, więc wszystkie ramki są wysyłane z pominięciem sekwencji RTS-CTS.

Używanie protokołu 802.11 jako łącza punkt-punkt

Dotychczas skupialiśmy się na zastosowaniu protokołu 802.11 w środowisku wielodostępnym. Warto jednak wspomnieć, że jeśli dwa węzły mają anteny kierunkowe, mogą one wykorzystać protokół 802.11 do komunikacji przez łącze typu punkt-punkt. Zważywszy na niski koszt sprzętu 802.11, zastosowanie anten kierunkowych i zwiększonej mocy transmisji pozwala używać protokołu 802.11 jako niedrogiej metody bezprzewodowej komunikacji na odległości liczone w dziesiątkach kilometrów. [Raman 2007] opisuje sieć bezprzewodową typu multi-hop działającą na obszarach wiejskich w Indiach (w dolinie Gangesu), która zawiera łącza punkt-punkt oparte na protokole 802.11.

7.3.3. Ramka IEEE 802.11

Choć ramka 802.11 jest pod wieloma względami podobna do ramki ethernetowej, zawiera kilka pól specyficznych dla łączności bezprzewodowej. Ramkę 802.11 przedstawiono na rysunku 7.13. Liczby umieszczone nad poszczególnymi polami ramki reprezentują ich długość w *bajtach*; liczby ponad poszczególnymi podpolami pola kontroli ramki reprezentują ich długość w *bitach*. Zbadajmy teraz pola ramki, a także najważniejsze podpola pola kontroli ramki.

Rysunek 7.13. Ramka 802.11

Pola treści i CRC

Zasadniczą częścią ramki jest pole treści, które zwykle zawiera datagram IP albo pakiet ARP. Choć jego maksymalna długość wynosi 2312 bajtów, zwykle nie przekracza ono 1500 bajtów. Podobnie jak ramka ethernetowa ramka 802.11 zawiera 32-bitowe pole kontroli błędów (ang. *cyclic redundancy check* — **CRC**), dzięki któremu odbiorca może wykryć błędy w przesłanej ramce. Jak widzieliśmy, błędy bitowe zdarzają się znacznie częściej w sieciach bezprzewodowych niż w przewodowych, więc pole CRC jest tu jeszcze bardziej użyteczne.

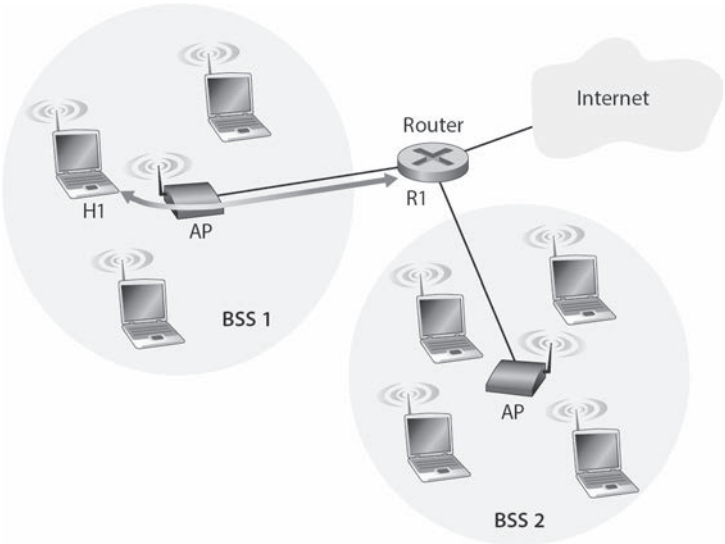
Pola adresu

Być może najbardziej uderzającą różnicą jest to, że ramka 802.11 zawiera *cztery* pola adresu, z których każde może pomieścić 6-bajtowy adres MAC. Po co jednak aż tyle pól adresu? Czy nie wystarczyłyby pola źródłowego i docelowego adresu MAC, jak w Ethernecie? Okazuje się, że trzy pola adresu są niezbędne do łączenia sieci — mówiąc ściślej, do przekazywania datagramu warstwy sieci od stacji bezprzewodowej przez punkt dostępowy do interfejsu routera. Czwarte pole adresu jest używane tylko w sieciach doraźnych. Ponieważ rozważamy tutaj wyłącznie sieci infrastrukturalne, skupmy się na pierwszych trzech polach adresu. Standard 802.11 definiuje je w następujący sposób:

- Adres 2 to adres MAC stacji nadawczej. Jeśli zatem ramka została wysłana przez stację bezprzewodową, to w polu adresu 2 znajduje się jej adres MAC. Podobnie jeśli ramka została wysłana przez punkt dostępowy, to w polu adresu 2 znajduje się jego adres MAC.
- Adres 1 to adres MAC stacji odbiorczej. Jeśli zatem ramka została wysłana przez stację bezprzewodową, to w polu adresu 1 znajduje się jej adres MAC punktu dostępowego. Podobnie jeśli ramka została wysłana przez punkt dostępowy, to w polu adresu 1 znajduje się adres MAC stacji bezprzewodowej.
- Aby zrozumieć, do czego służy adres 3, należy sobie przypomnieć, że grupa BSS (składająca się z punktu dostępowego i stacji bezprzewodowych) stanowi część podsieci, która łączy się z innymi podsieciami za pośrednictwem interfejsu routera. Pole adresu 3 zawiera adres MAC tego interfejsu.

Aby dokładniej wyjaśnić przeznaczenie adresu 3, przyjrzyjmy się sieciom na rysunku 7.14. Znajdują się w nich dwa punkty dostępowe obsługujące po kilka stacji bezprzewodowych. Każdy punkt dostępowy dysponuje bezpośrednim połączeniem z routerem, który z kolei jest podłączony do internetu. Powinniśmy zdawać sobie sprawę, że punkt dostępowy jest urządzeniem warstwy łącza, a zatem nie „mówi" protokołem IP i nie rozpoznaje jego adresów. Zastanówmy się teraz, jak przekazywany jest datagram od interfejsu routera R1 do stacji bezprzewodowej H1. Router nie „wie", że między nim a stacją H1 znajduje się punkt dostępowy; z jego perspektywy stacja H1 jest po prostu hostem w jednej spośród podsieci routera.

- Router, który zna adres IP stacji H1 (pobrany z pola docelowego adresu datagramu), używa protokołu ARP, aby ustalić adres MAC stacji H1, dokładnie tak samo jak w zwykłej ethernetowej sieci lokalnej. Po uzyskaniu adresu MAC stacji H1 interfejs routera umieszcza datagram w ramce ethernetowej. Pole adresu źródłowego tej ramki zawiera adres MAC routera R1, a pole adresu docelowego — adres MAC stacji H1.
- Po odebraniu ethernetowej ramki 802.3 punkt dostępowy przekształca ją w ramkę 802.11, która zostanie przesłana kanałem bezprzewodowym. Punkt wypełnia pola adresu 1 i adresu 2 adresem MAC stacji H1 oraz własnym adresem MAC, jak opisano powyżej. W polu adresu 3 punkt umieszcza adres MAC routera R1. Dzięki temu stacja H1 może ustalić (na podstawie adresu 3) adres MAC interfejsu routera, który wysłał datagram do podsieci.



Rysunek 7.14. Zastosowanie pól adresu w ramkach 802.11: przesyłanie ramki między hostem H1 a routerem R1

Teraz zbadajmy, co się stanie, kiedy stacja H1 udzieli odpowiedzi, przesyłając datagram do routera R1.

- Stacja H1 tworzy ramkę 802.11, wypełniając pola adresu 1 i adresu 2 adresem MAC punktu dostępowego oraz własnym adresem MAC, jak opisano powyżej. W polu adresu 3 stacja umieszcza adres MAC routera R1.
- Po odebraniu ramki 802.11 punkt dostępowy przekształca ją w ramkę ethernetową. W tej ramce pole adresu źródłowego zawiera adres MAC stacji H1, a pole adresu docelowego — adres MAC routera R1. Zatem pole adresu 3 pozwala punktowi dostępowemu ustalić odpowiedni docelowy adres MAC podczas konstruowania ramki ethernetowej.

Podsumowując — pole adresu 3 odgrywa kluczową rolę w łączeniu grupy BSS z przewodową siecią lokalną.

Pola numeru sekwencyjnego, okresu oraz kontroli ramki

Jak już wspomniano, po prawidłowym odebraniu ramki stacja 802.11 odsyła potwierdzenie. Ponieważ potwierdzenia mogą zostać utracone, zdarza się, że stacja nadawcza wysyła wiele kopii danej ramki. Omawiając protokół rdt2.1 (punkt 3.4.1), stwierdziliśmy, że zastosowanie numerów sekwencyjnych pozwala odbiorcy odróżnić transmisję nowej ramki od retransmisji poprzedniej. Pole numeru sekwencyjnego w ramce 802.11 pełni więc w warstwie łącza dokładnie tę samą funkcję, którą pełniło w warstwie transportu opisanej w rozdziale 3.

Pamiętajmy, że protokół 802.11 pozwala stacji nadawczej zarezerwować kanał na czas potrzebny do przesłania ramki danych oraz odebrania ramki potwierdzenia. Wartość ta jest umieszczana w polu okresu (zarówno w ramkach danych, jak i w ramkach RTS oraz CTS).

Jak pokazano na rysunku 7.13, pole kontroli ramki składa się z wielu podpól. Opiszemy tu krótko najważniejsze podpola; dokładniejsze informacje można znaleźć w specyfikacji standardu 802.11 [Held 2001; Crow 1997; IEEE 802.11 1999]. Pola *typ* i *podtyp* służą do rozróżniania ramek wiązania, RTS, CTS, ACK i danych. Pola *od* i *do* definiują znaczenie poszczególnych pól adresu. Zmienia się ono w zależności od trybu (doraźnego lub infrastrukturalnego) oraz — w trybie infrastrukturalnym — od tego, czy ramka jest wysyłana przez stację bezprzewodową, czy też przez punkt dostępowy. Wreszcie pole WEP wskazuje, czy używane jest szyfrowanie (protokół WEP zostanie opisany w rozdziale 8.).

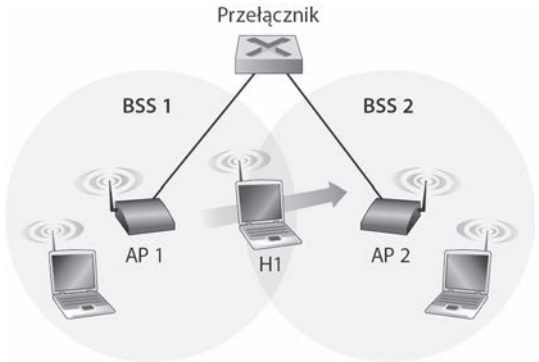
7.3.4. Mobilność w tej samej podsieci IP

Aby zwiększyć fizyczny zasięg bezprzewodowej sieci lokalnej, w firmach i na uniwersytetach często korzysta się z wielu grup BSS należących do tej samej podsieci IP. Oczywiście pojawia się tu kwestia mobilności między grupami BSS — jak stacje bezprzewodowe przechodzą od jednej grupy do drugiej, zachowując bieżące sesje TCP? Niebawem przekonamy się, że kiedy grupy BSS należą do tej samej podsieci, jest to względnie proste. Kiedy stacje poruszają się między podsieciami, potrzebne są bardziej zaawansowane protokoły zarządzania mobilnością, które przestudiujemy w podrozdziałach 7.5 i 7.6.

Przyjrzyjmy się teraz konkretnemu przykładowi mobilności między grupami BSS w tej samej podsieci. Na rysunku 7.15 przedstawiono dwie połączone grupy BSS oraz host H1, który przechodzi z grupy BSS1 do BSS2. Ponieważ w tym przykładzie urządzenie łączące dwie grupy BSS *nie jest* routerem, wszystkie stacje w obu grupach, w tym punkt dostępowy, należą do tej samej podsieci IP. Kiedy więc stacja H1 zmienia grupę, może zachować swój adres IP i bieżące połączenia TCP. Gdyby urządzeniem łączącym był router, stacja H1 musiałaby pozyskać nowy adres IP z podsieci, do której przechodzi. Ta zmiana wywołałaby wstrzymanie (i ostatecznie przerwanie) wszystkich aktywnych połączeń TCP tej stacji. W podrozdziale 7.6 pokazujemy, jak uniknąć tego problemu przez zastosowanie protokołu obsługującego mobilność w warstwie sieci, takiego jak Mobile IP.

Co się jednak dzieje, kiedy stacja H1 przechodzi z grupy BSS1 do BSS2? Kiedy stacja oddala się od punktu dostępowego AP1, wykrywa słabnący sygnał i zaczyna szukać silniejszego. Stacja H1 otrzymuje ramkę identyfikacyjną od punktu dostępowego AP2 (który w wielu sieciach firmowych i uniwersyteckich ma ten sam identyfikator SSID co punkt AP1). Następnie stacja przerywa powiązanie z punktem AP1 i wiąże się z punktem AP2, zachowując swój adres i bieżące sesje TCP.

To rozwiązuje problem z perspektywy hostów i punktów dostępowych. Co jednak z przełącznikiem z rysunku 7.15? W jaki sposób wykrywa on, że host przeszedł z jednego punktu dostępowego do drugiego? Jak Czytelnicy pamiętają z rozdziału 6., przełączniki



Rysunek 7.15. Mobilność w tej samej podsieci

są urządzeniami „samouczącymi", które automatycznie budują swoje tablice przekazywania. Funkcja ta dobrze sprawdza się w sporadycznych ruchach (na przykład kiedy pracownik zostaje przeniesiony do innego działu), ale przełączników nie projektowano z myślą o użytkownikach mobilnych, którzy chcą podtrzymywać połączenia TCP, kiedy poruszają się między grupami BSS. Aby zrozumieć, na czym polega problem, przypomnijmy sobie, że w tablicy przekazywania przełącznika znajduje się wpis kojarzący adres MAC stacji H1 z interfejsem przełącznika, przez który można do niej dotrzeć. Jeśli stacja H1 początkowo znajduje się w grupie BSS1, to zmierzające do niej datagramy są kierowane do punktu dostępowego AP1. Kiedy stacja zwiąże się z grupą BSS2, jej ramki powinny być kierowane do punktu AP2. Możliwe rozwiązanie (choć nieco prowizoryczne) polega na tym, że punkt AP2 może bezpośrednio po ustanowieniu powiązania przesłać do przełącznika rozgłoszeniową ramkę ethernetową z adresem źródłowym stacji H1. Kiedy przełącznik odbierze ramkę, zaktualizuje swoją tablicę przekazywania, umożliwiając dotarcie do stacji H1 przez punkt dostępowy AP2. Grupa standaryzacyjna 802.11f opracowuje protokół komunikacji między punktami dostępowymi, który rozwiąże tego rodzaju problemy.

Wcześniejsze omówienie dotyczyło mobilności w ramach tej samej podsieci w sieci lokalnej. Pamiętaj, że opisane w punkcie 6.4.4 sieci VLAN można wykorzystać do połączenia niezależnych sieci lokalnych w duże wirtualne sieci obejmujące spore obszary geograficzne. Przenoszenie użytkowników między stacjami bazowymi w tego rodzaju sieciach VLAN można obsługiwać w taki sam sposób jak opisany w tym miejscu [Yu 2011].

7.3.5. Zaawansowane funkcje protokołu 802.11

Omawianie protokołu 802.11 podsumowujemy krótką analizą dwóch zaawansowanych funkcji sieci 802.11. Jak się okaże, *nie* są one w pełni opisane w standardzie protokołu 802.11, ale można z nich skorzystać dzięki mechanizmom określonym w tym standardzie. Umożliwia to producentom realizowanie tych funkcji za pomocą własnych (zastrzeżonych) metod, potencjalnie zapewniających przewagę nad konkurencją.

Dostosowywanie szybkości w sieciach 802.11

Wcześniej pokazaliśmy (rysunek 7.3), że dla różnych poziomów wskaźnika SNR odpowiednie są różne techniki modulacji o odmiennych szybkościach transmisji. Rozważmy na przykład mobilnego użytkownika sieci 802.11, który początkowo znajduje się 20 metrów od stacji bazowej. Poziom SNR jest wtedy wysoki, dlatego odbiorca może komunikować się ze stacją bazową za pomocą tej techniki modulacji z warstwy fizycznej, która zapewni wysoką szybkość transmisji przy niskim współczynniku BER. Ten użytkownik ma szczęście! Załóżmy teraz, że osoba ta zaczyna się poruszać i oddalać od stacji bazowej. Wraz ze wzrostem odległości między odbiorcą i stacją poziom SNR zaczyna spadać. Jeśli nie zmieni się technika modulacji wykorzystywana w protokole 802.11 do obsługi komunikacji między stacją bazową i użytkownikiem, współczynnik BER wzrośnie do nieakceptowalnie wysokiego poziomu w wyniku spadku wartości SNR. Ostatecznie żadne z transmitowanych ramek nie będą poprawnie odbierane.

Dlatego niektóre wersje sieci 802.11 udostępniają funkcję dostosowywania szybkości. Polega ona na adaptacyjnym dobieraniu techniki modulacji stosowanej w warstwie fizycznej. Odbywa się to na podstawie obecnych (lub niedawnych) właściwości kanału. Jeśli węzeł nie otrzyma potwierdzenia dla dwóch kolejnych ramek (jest to pośredni wskaźnik wystąpienia błędów bitowych w kanale), zmniejszy prędkość transferu. Jeżeli nadejdzie potwierdzenie dla 10 kolejnych ramek lub upłynie określony czas od ostatniego błędu, szybkość transmisji wzrośnie. Ten mechanizm dostosowywania prędkości jest oparty na tej samej metodzie „testowania", co mechanizm kontroli przeciążenia w protokole TCP. Jeśli warunki są dobre (wskazują na to potwierdzenia ACK), szybkość transmisji jest podnoszona do momentu wystąpienia niepożądanego zdarzenia (braku potwierdzenia ACK). Kiedy wystąpi problem, prędkość transferu jest obniżana. Dostosowywanie szybkości w protokole 802.11 i kontrola przeciążenia w protokole TCP przypominają zachowanie dziecka, które nieustannie nagabuje rodziców o więcej (małe dziecko będzie prosić na przykład o cukierki, natomiast starsze — o późniejszą godzinę powrotu do domu), dopóki nie usłyszy odpowiedzi „Wystarczy!". Wtedy następuje wycofanie, a następnie kolejna próba przy — miejmy nadzieję — bardziej sprzyjających warunkach. Zaproponowano też wiele innych rozwiązań usprawniających podstawowy system automatycznego dostosowywania szybkości [Kamerman 1997; Holland 2001; Lacage 2004].

Zarządzanie poborem energii

Energia to cenny zasób w urządzeniach przenośnych. Dlatego standard 802.11 obejmuje funkcje zarządzania poborem energii, które umożliwiają węzłom w sieciach 802.11 zminimalizowanie czasu aktywności mechanizmów wykrywających, transmitujących i odbierających dane oraz innych układów. Zarządzanie energią w standardzie 802.11 działa w następujący sposób. Węzeł może przechodzić bezpośrednio między stanem uśpienia i aktywności (podobnie jak zaspany student na sali wykładowej!). Węzeł informuje punkt dostępowy o tym, że przechodzi w stan uśpienia, przez ustawienie bitu zarządzania energią w nagłówku ramki 802.11 na 1. Następnie zegar w węźle jest

ustawiany tak, aby wznowić działanie urządzenia tuż przed zaplanowanym wysłaniem ramki identyfikacyjnej przez punkt dostępowy (pamiętajmy, że takie punkty zwykle przesyłają ramki identyfikacyjne co 100 milisekund). Ponieważ punkt dostępowy po otrzymaniu bitu zarządzania energią rejestruje, że węzeł przechodzi w stan uśpienia, „wie", iż nie powinien wysyłać do danego urządzenia żadnych ramek. Dlatego jeśli otrzyma ramki skierowane do uśpionego węzła, zapisze je w buforze w celu ich późniejszego przesłania.

Węzeł wznawia działanie tuż przed przesłaniem ramki identyfikacyjnej przez punkt dostępowy i szybko przechodzi w stan pełnej aktywności (to przejście — inaczej niż w przypadku śpiącego studenta — zajmuje tylko 250 mikrosekund [Kamerman 1997]!). Ramka identyfikacyjna wysyłana przez punkt dostępowy obejmuje listę węzłów, dla których zapisano ramki w buforze. Jeśli w buforze nie ma ramek dla danego urządzenia, może ono ponownie przejść w stan uśpienia. W przeciwnym razie węzeł może bezpośrednio zażądać przesłania ramek z bufora. W tym celu musi wysłać komunikat odpytujący do punktu dostępowego. Przy 100 milisekundach odstępu między ramkami identyfikacyjnymi, 250 mikrosekundach wzbudzania i podobnym czasie odbioru ramki identyfikacyjnej oraz sprawdzania zawartości bufora węzeł bez ramek do wysłania lub odebrania może pozostawać w stanie uśpienia przez 99% czasu. Prowadzi to do znacznych oszczędności w zużyciu energii.

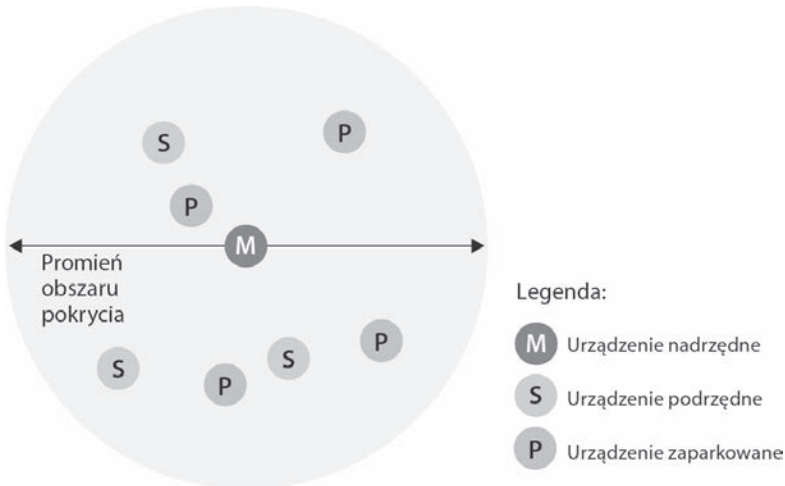
7.3.6. Sieci PAN — Bluetooth i Zigbee

Jak pokazano na rysunku 7.2, standard Wi-Fi IEEE 802.11 jest przeznaczony do komunikacji na odległość nieprzekraczającą 100 metrów (wyjątkiem jest wykorzystanie standardu 802.11 w konfiguracji typu punkt-punkt z antenami kierunkowymi). Dwa inne protokoły komunikacji bezprzewodowej z rodziny IEEE 802 to Bluetooth i Zigbee (standardy IEEE 802.15.1 i IEEE 802.15.4 [IEEE 802.15 2012]).

Bluetooth

Sieć IEEE 802.15.1 charakteryzuje się krótkim zasięgiem, niską mocą i niewielkim kosztem. Można ją traktować jako „zamiennik kabla" do łączenia komputerów z klawiaturą, myszą lub innym bezprzewodowym sprzętem peryferyjnym, telefonów komórkowych, głośników, słuchawek i wielu innych urządzeń. Natomiast 802.11 to technologia „dostępowa" o wyższej mocy, średnim zasięgu i większej szybkości. Dlatego sieci 802.15.1 czasem określa się mianem bezprzewodowych sieci osobistych (ang. *wireless personal area network* — **WPAN**). Warstwa łącza i warstwa fizyczna 802.15.1 opierają się na wcześniejszej specyfikacji sieci osobistych **Bluetooth** [Held 2001; Bisdikian 2001]. Sieci 802.15.1 operują w nielicencjonowanym paśmie 2,4 GHz i wykorzystują technikę TDM ze szczeliną czasu równą 625 mikrosekundom. W każdej szczelinie czasu nadawca transmituje dane na jednym z 79 kanałów, zmieniając je w znanej, ale pseudolosowej kolejności. Ta postać przełączania kanałów, znana jako **FHSS** (ang. *Frequency-Hopping Spread Spectrum*), rozprasza transmisje w całym zakresie częstotliwości. Standard 802.15.1 pozwala osiągnąć szybkość transmisji do 4 Mb/s.

Sieci 802.15.1 są sieciami doraźnymi: do łączenia urządzeń 802.15.1 niepotrzebna jest żadna infrastruktura sieciowa (na przykład punkty dostępowe). Urządzenia 802.15.1 muszą zatem organizować się same. Najpierw tworzą one **pikosieć** (rysunek 7.16) złożoną z najwyżej ośmiu aktywnych urządzeń. Jedno z nich pełni funkcję węzła nadrzędnego, a pozostałe — węzłów podrzędnych. Węzeł nadrzędny jest prawdziwym władcą pikosieci — jego zegar wyznacza czas, tylko on może nadawać w nieparzystych ramkach, a węzeł podrzędny może rozpocząć transmisję, ale tylko pod warunkiem że węzeł nadrzędny skontaktował się w nim w poprzedniej ramce, zresztą nawet w takim przypadku można nadawać tylko do węzła nadrzędnego. Oprócz urządzeń podrzędnych w sieci może znajdować się do 255 urządzeń zaparkowanych. Urządzenia te nie mogą nawiązać komunikacji, dopóki węzeł nadrzędny nie zmieni ich stanu z zaparkowanego na aktywny.



Rysunek 7.16. Pikosieć w standardzie Bluetooth

Więcej informacji o sieciach WPAN można znaleźć w materiałach poświęconych standardowi Bluetooth [Held 2001; Bisdikian 2001] albo w oficjalnej witrynie standardu IEEE 802.15 [IEEE 802.15 2009].

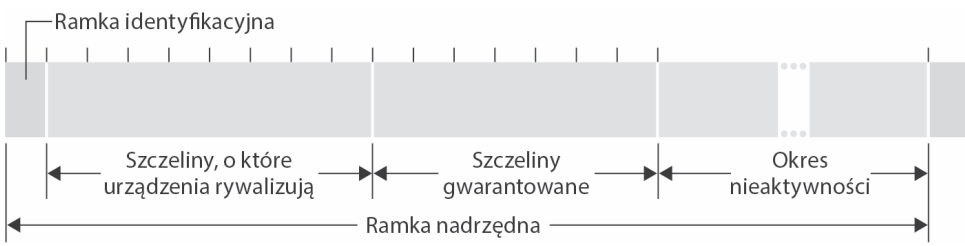
Zigbee

Drugi typ sieci PAN ustandaryzowany przez IEEE oparty jest na standardzie 802.15.4 [IEEE 802.15 2012] nazywanym Zigbee. Sieci Bluetooth stanowią zastępnik sieci kablowych i oferują szybkość transferu na poziomie Mb/s, natomiast sieci Zigbee są bardziej energooszczędne i przeznaczone do zastosowań z niższą szybkością transferu i krótszymi okresami aktywności. Choć zwykle wydaje nam się, że „większe i szybsze jest lepsze", nie wszystkie aplikacje sieciowe wymagają wysokiej przepustowości i wyższych kosztów (zakupu oraz energii). Na przykład domowe czujniki temperatury i światła

oraz włączniki oświetlenia na ścianach to bardzo proste, tanie i wymagające niewielkiej ilości energii urządzenia o krótkich okresach aktywności. Standard Zigbee dobrze się sprawdza w takich zastosowaniach. W zależności od częstotliwości kanału w tym standardzie dostępne są szybkości transferu rzędu 20, 40, 100 i 250 kb/s.

W sieciach Zigbee występują węzły dwóch rodzajów. Tak zwane urządzenia o ograniczonej funkcjonalności przypominają urządzenia podrzędne ze standardu Bluetooth i pracują jako urządzenia podrzędne kontrolowane przez jedno urządzenie o pełnej funkcjonalności. To ostatnie działa podobnie jak urządzenie nadrzędne w sieciach Bluetooth i kontroluje wiele urządzeń podrzędnych. Grupę urządzeń o pełnej funkcjonalności można skonfigurować w sieć kratową, w której takie urządzenia przekazują ramki między sobą. Zigbee obejmuje wiele mechanizmów opisanych już w kontekście innych protokołów warstwy łącza danych: ramki identyfikacyjne i potwierdzenia z warstwy łącza danych (podobnie jak protokół 802.11), losowy dostęp do nośnika z wykrywaniem sygnału i odczekiwaniem wykładniczym (podobnie jak protokoły 802.11 i Ethernet) i stały, gwarantowany przydział szczelin czasowych (podobnie jak protokół DOCSIS).

Sieci Zigbee można konfigurować na wiele sposobów. Rozważmy prostą sytuację, w której jedno urządzenie o pełnej funkcjonalności kontroluje wiele urządzeń o ograniczonej funkcjonalności w modelu ze szczelinami czasu i używając ramek identyfikacyjnych. Na rysunku 7.17 pokazany jest scenariusz, w którym sieć Zigbee dzieli czas na powtarzające się ramki nadrzędne. Każda z nich rozpoczyna się od ramki identyfikacyjnej. Ramki identyfikacyjne dzielą ramki nadrzędne na okresy aktywności (kiedy to urządzenia mogą przesyłać dane) i nieaktywności (wtedy wszystkie urządzenia, w tym kontroler, mogą wejść w tryb uśpienia, co zmniejsza zużycie energii). Okresy aktywności obejmują 16 szczelin czasowych. Część z nich jest wykorzystywana przez urządzenia do losowego dostępu do nośnika zgodnie z protokołem CSMA/CA, a część jest przydzielana przez kontroler konkretnym urządzeniom, co zapewnia im gwarantowany dostęp do kanału. Więcej szczegółowych informacji na temat sieci Zigbee znajdziesz w [Baronti 2007, IEEE 802.15.4 2012].



Rysunek 7.17. Struktura ramki nadrzędnej w sieciach Zigbee

7.4. Dostęp do internetu za pomocą sieci komórkowych

W poprzednim podrozdziale wyjaśniliśmy, jak host uzyskuje dostęp do internetu, kiedy znajduje się w hotspocie Wi-Fi w pobliżu punktu dostępowego 802.11. Jednakże większość hotspotów Wi-Fi pokrywa niewielki obszar o średnicy od 10 do 100 metrów. Co zrobić, jeśli desperacko potrzebujemy bezprzewodowego dostępu do internetu, a nie możemy znaleźć żadnego hotspotu Wi-Fi?

Obecnie w wielu regionach świata powszechnie dostępna jest telefonia komórkowa, więc naturalnym rozwiązaniem jest rozbudowanie sieci komórkowych w taki sposób, aby obsługiwały nie tylko transmisję głosu, ale również bezprzewodowy dostęp do internetu. Byłoby najlepiej, gdyby dostęp ten był realizowany z dużą szybkością i zapewniał pełną mobilność, pozwalając użytkownikom na podtrzymanie sesji TCP podczas podróży (na przykład w autobusie lub pociągu). Przy odpowiednio dużej szybkości transmisji użytkownik mógłby nawet brać udział w wideokonferencjach, pozostając w ruchu. Scenariusz ten wcale nie jest taki nieprawdopodobny. Wraz z upowszechnianiem się szerokopasmowego dostępu do internetu (co opisano dalej) coraz częściej oferowana jest szybkość transmisji danych na poziomie kilku megabitów na sekundę.

W tym podrozdziale przedstawimy krótki przegląd bieżących i pojawiających się technologii komórkowego dostępu do internetu. Ponownie skupimy się na pierwszym bezprzewodowym „przeskoku" między telefonem komórkowym a infrastrukturą przewodowej sieci telefonicznej; w podrozdziale 7.7 zastanowimy się, jak rozmowy są przekierowywane do użytkownika poruszającego się między stacjami bazowymi. Z konieczności podamy tylko uproszczony i wysokopoziomowy opis technologii komórkowej. Współczesna komunikacja komórkowa jest bardzo obszernym zagadnieniem, a wiele uniwersytetów oferuje kursy poświęcone różnym jej aspektom. Zainteresowani Czytelnicy mogą sięgnąć po [Goodman 1997; Kaaranen 2001; Lin 2001; Korhonen 2003; Schiller 2003; Palat 2009; Scourias 2012; Turner 2012; Akyildiz 2010], a zwłaszcza po znakomite i kompletne pozycje [Mouly 1992; Sauter 2014].

7.4.1. Omówienie architektury komórkowej

W opisie architektury sieci komórkowych w tym punkcie stosujemy terminologię używaną dla standardów **GSM** (ang. *Global System for Mobile Communications*; ciekawostka dla osób zainteresowanych historią — początkowo akronim GSM utworzono od słów *Groupe Spécial Mobile*, a dopiero potem przyjęto bardziej zanglicyzowaną nazwę, zachowując przy tym litery z pierwotnego akronimu). W latach 80. Europejczycy dostrzegli potrzebę budowy ogólnoeuropejskiego cyfrowego systemu telefonii komórkowej, który zastąpiłby liczne niezgodne ze sobą systemy analogowe. Doprowadziło to do utworzenia standardu GSM [Mouly 1992]. Europejczycy z powodzeniem wdrożyli technologię GSM na początku lat 90., a od tego czasu standard ten stał się królem w świecie telefonii komórkowej. Ponad 80% wszystkich abonentów sieci komórkowych na całym świecie korzysta z technologii GMS.

KĄCIK HISTORYCZNY

SIECI KOMÓRKOWE 4G A BEZPRZEWODOWE SIECI LOKALNE

Wielu operatorów telefonii komórkowej wdraża komórkowe systemy 4G. W niektórych państwach (na przykład w Korei i Japonii) pokrycie sieciami 4G LTE przekracza 90% i jest prawie pełne. W 2015 roku średnia szybkość pobierania w systemach LTE wahała się od 10 Mb/s w Stanach Zjednoczonych i Indiach do blisko 40 Mb/s w Nowej Zelandii. Systemy 4G działają w licencjonowanych pasmach radiowych, a niektórzy operatorzy płacą rządowi za licencje wysokie sumy. Systemy 4G pozwalają użytkownikom uzyskać dostęp do internetu w ruchu, poza terenem zabudowanym, podobnie jak współczesne telefony komórkowe. W wielu sytuacjach użytkownik ma jednocześnie dostęp zarówno do bezprzewodowych sieci lokalnych, jak i do sieci 4G. Ponieważ systemy 4G mają większe ograniczenia i są droższe, wiele urządzeń mobilnych domyślnie używa sieci Wi-Fi zamiast 4G (jeśli obie są dostępne). To, czy jako bezprzewodowe dostępowe sieci brzegowe będą stosowane przede wszystkim bezprzewodowe sieci lokalne, czy systemy komórkowe, pozostaje kwestią otwartą:

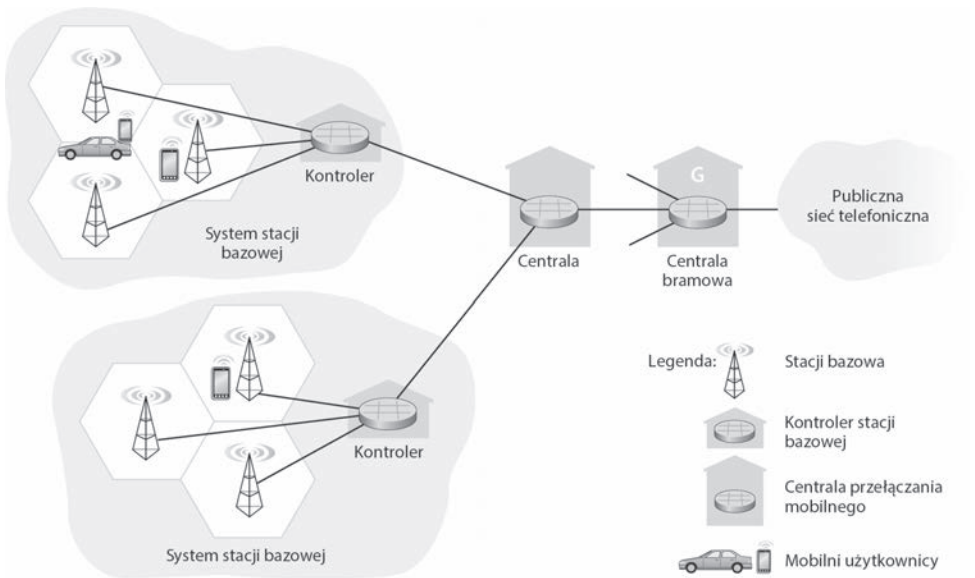
- Rozwijająca się infrastruktura bezprzewodowych sieci lokalnych będzie niemal wszechobecna. Sieci 802.11 operujące z szybkością 54 Mb/s lub wyższą już teraz cieszą się dużą popularnością. Prawie wszystkie laptopy, tablety i smartfony są fabrycznie wyposażone w bezprzewodowe karty 802.11. Niewielkie, energooszczędne karty bezprzewodowe będą również stosowane w nowatorskich urządzeniach internetowych, takich jak kamery bezprzewodowe i cyfrowe ramki na zdjęcia.
- Bezprzewodowe stacje bazowe w sieci lokalnej mogłyby obsługiwać również telefony komórkowe. Już teraz wiele telefonów potrafi łączyć się zarówno z sieciami telefonii komórkowej, jak i z sieciami IP za pomocą usług VoIP podobnych do tych z systemu Skype. W ten sposób można pominąć usługi głosowe i transmisji danych w technologii 4G świadczone przez operatora sieci komórkowej.

Oczywiście wielu innych ekspertów uważa, że systemy 4G nie tylko odniosą sukces, ale również zrewolucjonizują naszą pracę i życie. Zarówno Wi-Fi, jak i 4G mogą stać się dominującymi technologiami bezprzewodowymi, a poruszające się urządzenia będą automatycznie wybierać optymalną metodę dostępu w bieżącej lokacji.

Technologie komórkowe często klasyfikuje się jako należące do różnych „generacji". Pierwsze generacje były projektowane przede wszystkim pod kątem transmisji głosu. Systemy pierwszej generacji (1G) były analogowymi sieciami FDMA przeznaczonymi tylko do komunikacji głosowej. Obecnie zostały niemal całkowicie wyparte przez cyfrowe systemy drugiej generacji (2G). Systemy 2G pierwotnie także zaprojektowano do komunikacji głosowej, jednak później je wzbogacono (2.5G), tak aby obsługiwały transfer zarówno danych (na przykład internetowych), jak i głosu. Obecnie wdrażane są systemy trzeciej generacji (3G), które także obsługują przesyłanie głosu i danych, jednak w systemach tych położono nacisk na funkcje związane z transferem danych i szybsze radiowe łącza dostępowe. Wprowadzane obecnie systemy 4G są oparte na technologii LTE, udostępniają sieci bazowe typu all-IP oraz zapewniają zintegrowaną transmisję głosową i danych z szybkością wielu Mb/s.

Architektura sieci komórkowych. 2G — połączenia głosowe z sieciami telefonicznymi

Przymiotnik *komórkowy* odnosi się do podziału obszaru geograficznego na mniejsze obszary znajdujące się w zasięgu pojedynczej stacji nadawczej i nazywane **komórkami** (widoczne jako sześciokąty po lewej stronie rysunku 7.18). Ze standardem GSM, podobnie jak z omówionym w punkcie 7.3.1 standardem 802.11 Wi-Fi, związana jest określona terminologia. Każda komórka obejmuje **bazową stację przekaźnikową** (ang. *Base Transceiver Station* — **BTS**), która wysyła i odbiera sygnały od mobilnych stacji z danej komórki. Zasięg komórki zależy od wielu czynników, w tym od mocy nadawczej stacji bazowej, mocy nadawczej stacji mobilnych, przeszkód terenowych oraz wysokości anteny stacji bazowej. Choć na rysunku 7.18 każda stacja bazowa znajduje się w środku komórki, w wielu współczesnych systemach stacje są umieszczane na styku trzech komórek, dzięki czemu jedna stacja bazowa z antenami komórkowymi może obsługiwać trzy komórki.



Rysunek 7.18. Komponenty architektury sieci komórkowych GSM 2G

Ze względu na interferencje występujące w powietrzu w standardzie GSM dla systemów komórkowych 2G zastosowano mieszaną technikę FDM/TDM (dla nośnika radiowego). W rozdziale 1. wyjaśniliśmy, że w systemie FDM kanał jest dzielony na pasma częstotliwości, z których każde jest przydzielane określonej komórce. W tym samym rozdziale stwierdziliśmy, że w systemie TDM czas jest dzielony na ramki, a każda z nich — na szczeliny. Poszczególne komórki mogą następnie korzystać z określonej szczeliny w cyklicznej ramce. W mieszanych systemach FDM/TDM kanał jest dzielony na podpasma częstotliwości. W każdym z nich czas jest dzielony na ramki i szczeliny.

Dlatego w mieszanych systemach FDM/TDM, jeśli kanał podzielono na F podpasm, a czas — na T szczelin, kanał może obsługiwać $F \cdot T$ jednoczesnych połączeń. W punkcie 6.3.4 wyjaśniono, że w kablowych sieciach dostępowych też używane są mieszane systemy FDM/TDM. Systemy GSM korzystają z pasm o szerokości 200 kHz, a każde pasmo obsługuje osiem rozmów TDM. Mowa jest kodowana z częstotliwościami 13 kb/s i 12,2 kb/s.

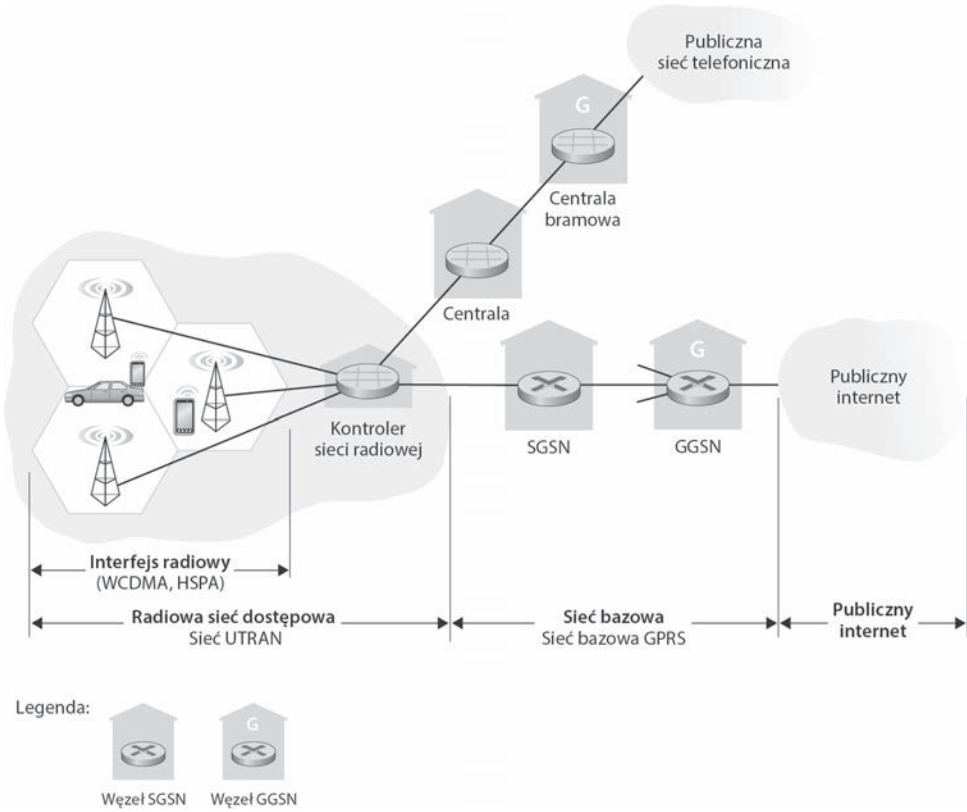
Kontrolery stacji bazowej (ang. *Base Station Controller* — **BSC**) w sieciach GSM mogą fizycznie znajdować się w stacji bazowej, jednak zwykle pojedynczy kontroler obsługuje dziesiątki takich stacji. Funkcja kontrolera polega na przydziale kanałów radiowych stacji bazowych mobilnym użytkownikom, przeprowadzaniu **przywoływania** (wyszukiwaniu komórki, w której znajduje się mobilny użytkownik; ang. *paging*) i transferze mobilnych użytkowników (to zagadnienie pokrótce omawiamy w punkcie 7.7.2). Kontroler i obsługiwane przez niego stacje bazowe składają się na **system stacji bazowej** (ang. *Base Station System* — **BSS**) sieci GSM.

W podrozdziale 7.7 wyjaśniamy, że **centrale przełączania mobilnego** (ang. *Mobile Switching Center* — **MSC**) odgrywają kluczową rolę przy uwierzytelnianiu użytkowników i ewidencjonowaniu ich (czyli określaniu, czy urządzenie przenośne może połączyć się z daną siecią komórkową), rozpoczynaniu i kończeniu rozmów oraz transferze użytkowników. Pojedyncza centrala zwykle obejmuje do pięciu kontrolerów, co oznacza, że obsługuje około 200 000 abonentów. Sieci operatorów komórkowych składają się z wielu central, w tym ze specjalnych central tranzytowych, łączących sieć komórkową danego dostawcy z większą publiczną siecią telefoniczną.

7.4.2. Sieci komórkowe 3G — udostępnianie internetu abonentom sieci komórkowych

W punkcie 7.4.1 koncentrowaliśmy się na łączeniu użytkowników sieci komórkowych z publiczną siecią telefoniczną w celu nawiązania transmisji głosowej. Jednak naturalne jest, że gdy się poruszamy, chcemy też mieć możliwość czytania e-maili, dostępu do internetu, korzystania z usług opartych na lokalizacji (na przykład map lub rekomendacji restauracji), a może nawet strumieniowej transmisji materiałów wideo. To wymaga, aby używany smartfon obsługiwał kompletny stos protokołów TCP/IP (w tym warstw fizycznej, sieciowej, transportowej i aplikacji) oraz łączył się z internetem za pomocą sieci komórkowej. Sieci komórkowe z dostępem do internetu korzystają z trudnego do zrozumienia zestawu konkurencyjnych i wciąż zmieniających się standardów, ponieważ po jednej generacji następuje kolejna (a także generacje „połówkowe"), co powoduje wprowadzenie nowych technologii i usług o nowych akronimach. Sytuację dodatkowo pogarsza to, że nie istnieje żadna oficjalna jednostka, która określa wymogi stawiane technologiom 2.5G, 3G, 3.5G lub 4G. Z tego powodu trudno jest zrozumieć różnice między konkurencyjnymi standardami. W dalszym omówieniu koncentrujemy się na standardach **UMTS** (ang. *Universal Mobile Telecommunications Service*) 3G i 4G, opracowanych w ramach projektu **3GPP** (ang. *3rd Generation Partnership Project*) [3GPP 2016].

Przyjrzyjmy się najpierw na ogólnym poziomie architekturze sieci komórkowych 3G (rysunek 7.19).



Rysunek 7.19. Architektura systemu 3G

Sieć bazowa 3G

Komórkowa bazowa sieć danych 3G łączy sieci z dostępem radiowym z publicznym internetem. Ta sieć bazowa współdziała z komponentami istniejącej głosowej sieci komórkowej (przede wszystkim z centralą), z którą zetknąłeś się już na rysunku 7.18. Ponieważ głosowa sieć komórkowa ma rozbudowaną infrastrukturę (i działają w niej zyskowne usługi), projektanci usług wykorzystujących sieć danych 3G przyjęli proste podejście: *pozostawiają istniejącą głosową sieć bazową GSM bez zmian i dodają nowe mechanizmy związane z danymi, działające równolegle do istniejącej głosowej sieci komórkowej*. Inna możliwość (integracja nowych usług transmisji danych z rdzeniem istniejącej głosowej sieci komórkowej) spowodowałaby trudności opisane w podrozdziale 4.3, gdzie omówiono integrowanie nowych (IPv6) i starszych (IPv4) technologii w internecie.

W sieci bazowej 3G występują dwa rodzaje węzłów: **SGSN** (ang. *Serving GPRS Support Nodes*) i **GGSN** (ang. *Gateway GPRS Support Nodes*). **GPRS** (ang. *Generalized Packet Radio Service*) to wczesna komórkowa usługa transmisji danych w sieciach 2G. Tu omawiamy zmodyfikowaną wersję GPRS z sieci 3G. Węzeł SGSN odpowiada za wymianę datagramów z węzłami mobilnymi w radiowych sieciach dostępowych, z którymi taki węzeł jest powiązany. Węzeł SGNS komunikuje się z centralą głosowej sieci komórkowej z danego obszaru, obsługuje uwierzytelnianie i przekazywanie użytkowników, przechowuje informacje o lokalizacji (w komórce) dotyczące aktywnych węzłów mobilnych i przekazuje datagramy między węzłami mobilnymi w radiowych sieciach dostępowych a węzłami GGSN. Węzeł GGSN pełni funkcję bramy i łączy wiele węzłów SGSN z internetem. Węzeł GGSN jest więc ostatnim elementem infrastruktury sieci 3G, z jakim datagram z węzła mobilnego styka się przed przesłaniem do internetu. Dla świata zewnętrznego węzeł GGSN wygląda jak zwykły router bramowy. Mobilność węzłów 3G w sieci GGSN jest ukrywana przed światem zewnętrznym.

Radiowe sieci dostępowe 3G: obrzeża sieci bezprzewodowych

Radiowa sieć dostępowa 3G to bezprzewodowa sieć pierwszego przeskoku widoczna dla użytkowników technologii 3G. Kontroler sieci radiowej (ang. *Radio Network Controller* — **RNC**) zwykle kontroluje bazowe stacje odbiorczo-nadawcze z kilku komórek, podobne do stacji bazowych z systemów 2G (w słownictwie związanym z sieciami 3G UMTS oficjalnie są one nazywane „Node B", co jest mało opisową nazwą). Łącze bezprzewodowe w każdej komórce łączy węzły mobilne i bazową stację odbiorczo-nadawczą — podobnie jak w sieciach 2G. Kontroler RNC jest połączony z komórkową siecią głosową z przełączaniem obwodów (za pośrednictwem centrali) i z internetem z przełączaniem pakietów (za pośrednictwem węzła SGSN). Dlatego, choć w technologii 3G komórkowa sieć głosowa i komórkowe usługi transmisji danych korzystają z innych sieci bazowych, współużytkują radiowe sieci dostępowe pierwszego i ostatniego przeskoku.

Znaczącą zmianą w sieciach 3G UMTS w porównaniu z sieciami 2G jest to, że zamiast systemu FDMA/TDMA ze standardu GMS wykorzystano w nich technikę CDMA o nazwie Direct Sequence Wideband CDMA (DS-WCDMA) w szczelinach TDMA (ramki ze szczelinami TDMA są dostępne na różnych częstotliwościach — jest to ciekawe zastosowanie wszystkich trzech opisanych wcześniej metod podziału kanału!). Ta zmiana wymaga, aby nowa komórkowa bezprzewodowa sieć dostępowa 3G działała równolegle z siecią systemów stacji bazowych 2G przedstawioną na rysunku 7.19. Usługa transmisji danych związana ze specyfikacją WCDMA to **HSPA** (ang. *High Speed Packet Access*). Ma ona zapewniać szybkość przesyłu do 14 Mb/s. Szczegółowe informacje na temat sieci 3G można znaleźć w witrynie projektu **3GPP** (ang. *3rd Generation Partnership Project*) [3GPP 2016].

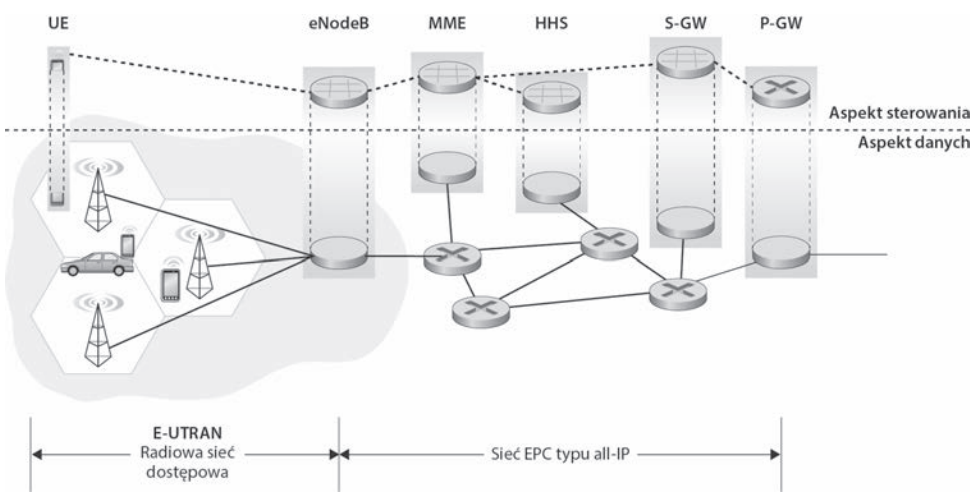
7.4.3. W kierunku sieci 4G — LTE

Systemy komórkowe czwartej generacji (4G) stają się coraz popularniejsze. W 2015 roku w ponad 50 państwach pokrycie sieciami 4G przekraczało 50%. Standard 4G **LTE** (ang. *Long-Term Evolution*) [Sauter 2014] przedstawiony w ramach projektu 3GPP obejmuje dwie ważne innowacje w porównaniu z systemami 3G: sieć bazową typu all-IP i wzbogacone radiowe sieci dostępowe.

Architektura systemu 4G: sieć bazowa typu all-IP

Na rysunku 7.20 pokazano ogólną architekturę sieci 4G, która (niestety) wymaga kolejnego (mało zrozumiałego) zestawu nazw i akronimów reprezentujących komponenty sieci. Nie zatraćmy się jednak w tych akronimach! Na temat architektury 4G można poczynić trzy ważne ogólne obserwacje:

- *Jest to ujednolicona architektura sieciowa typu all-IP*. W odróżnieniu od widocznej na rysunku 7.19 sieci 3G, która obejmuje odrębne komponenty i ścieżki dla głosu oraz danych, przedstawiona na rysunku 7.20 architektura 4G jest architekturą typu all-IP. Oznacza to, że głos i dane są przesyłane w datagramach IP między urządzeniem bezprzewodowym (w żargonie z obszaru sieci 4G jest to **UE** — od ang. *user equipment*) do bramy P-GW łączącej sieć brzegową 4G z resztą sieci. Wraz z pojawieniem się sieci 4G ostatnie pozostałości telekomunikacyjnych korzeni sieci komórkowych zniknęły, co pozwoliło zbudować uniwersalną usługę IP!
- *W sieci 4G obowiązuje wyraźny podział na aspekty danych i sterowania*. W architekturze sieci 4G występuje jednoznaczny podział na aspekty danych i sterowania; przypomina to podział na takie aspekty w warstwie sieciowej z protokołem IP, co opisano w rozdziałach 4. i 5. Funkcje tych aspektów są opisane dalej.



Rysunek 7.20. Architektura sieci 4G

- *Wyraźny podział na radiową sieć dostępową i sieć bazową typu all-IP*. Datagramy IP z danymi użytkownika są za pomocą wewnętrznej (w ramach sieci 4G) sieci IP przekazywane między użytkownikiem (UE) a bramą (P-GW na rysunku 7.20) i zewnętrznym internetem. W tej samej wewnętrznej sieci wymieniane są pakiety kontrolne między komponentami usług sterowania siecią 4G, których zadania są opisane dalej.

Oto podstawowe komponenty architektury 4G.

- Węzeł **eNodeB** to logiczny potomek stacji bazowych 2G i kontrolerów sieci radiowych 3G (Node B), który odgrywa kluczową rolę w sieciach 4G. W aspekcie danych zadanie takich węzłów polega na przekazywaniu datagramów między UE (za pomocą dostępowej sieci radiowej LTE) a P-GW.

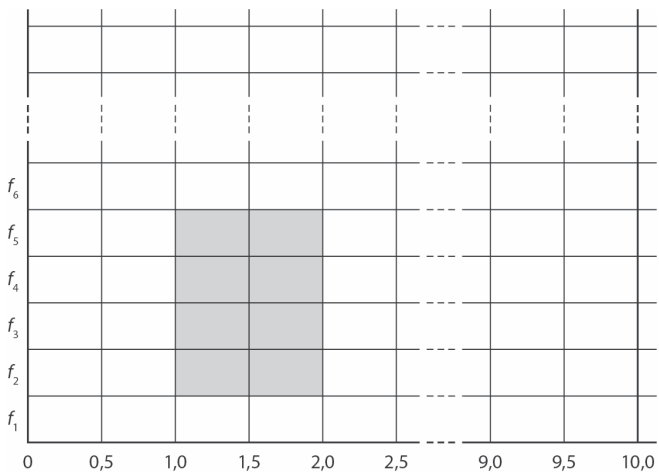
  Datagramy UE są kapsułkowane w węźle eNodeB i tunelowane do bramy P-GW za pomocą sieci **EPC** (ang. *enhanced packet core*) typu all-IP. Tunelowanie danych między eNodeB a P-GW odbywa się podobnie jak opisane w podrozdziale 4.3 tunelowanie datagramów IPv6 między dwoma punktami końcowymi IPv6 przez sieć routerów IPv4. Z tunelami powiązane mogą być gwarancje jakości usług. Na przykład sieć 4G może gwarantować, że w transmisji głosowej między UE a P-GW opóźnienie będzie nie wyższe niż 100 milisekund, a utrata pakietów wyniesie mniej niż 1%. Dla ruchu TCP gwarantowane może być opóźnienie mniejsze niż 300 milisekund i utrata pakietów niższa niż 0,0001% [Palat 2009]. Jakość usług jest opisana w rozdziale 9.

  W aspekcie sterowania węzły eNodeB obsługują na rzecz UE ruch związany z rejestracją i sygnalizowaniem mobilności.
- Brama **P-GW** (ang. *Packet Data Network Gateway*) przydziela adresy IP użytkownikom i wymusza jakość usług. Ponieważ jest punktem końcowym tunelu, odpowiada też za kapsułkowanie i odkapsułkowywanie datagramów w trakcie przekazywania datagramu do użytkownika i w drugą stronę.
- Brama **S-GW** (ang. *Serving Gateway*) to podstawowy punkt obsługi mobilności w aspekcie danych. Cały ruch od użytkowników przechodzi przez bramy S-GW. Taka brama obsługuje też funkcje związane z naliczaniem rachunków i przechwytywaniem ruchu w zgodzie z prawem.
- Jednostka **MME** (ang. *Mobility Management Entity*) zarządza połączeniami i mobilnością na rzecz użytkowników znajdujących się w komórce kontrolowanej przez tę jednostkę. Jednostka MME otrzymuje od serwera HSS informacje o subskrypcjach użytkowników. Mobilność w sieciach komórkowych jest szczegółowo opisana w podrozdziale 7.7.
- Serwer **HSS** (ang. *Home Subscriber Server*) obejmuje informacje o użytkownikach, w tym dane o usługach roamingu i profilach jakości usług oraz dane uwierzytelniające. W podrozdziale 7.7 dowiesz się, że serwer HSS otrzymuje te informacje od dostawcy usług komórkowych użytkownika.

Bardzo zrozumiałe wprowadzenie do architektury sieci 4G i EPC znajdziesz w [Motorola 2007; Palat 2009; Sauter 2014].

Radiowa sieć dostępowa LTE

W sieci LTE w kanale dosyłowym używane są multipleksowanie z podziałem częstotliwości i multipleksowanie z podziałem czasu. Ta technologia to **OFDM** (ang. *orthogonal frequency division multiplexing*) [Rohde 2008; Ericsson 2011]. Określenie „ortogonalne" wynika z tego, że generowane są sygnały przesyłane kanałami o różnej częstotliwości, dlatego interferencje między nimi są bardzo niewielkie — nawet gdy częstotliwości różnią się w bardzo niewielkim stopniu. W LTE każdy aktywny węzeł mobilny otrzymuje szczeliny czasu o długości 0,5 milisekundy w przynajmniej jednej częstotliwości. Na rysunku 7.21 pokazano przydział ośmiu szczelin czasowych w czterech częstotliwościach. Ponieważ węzeł mobilny otrzymuje coraz więcej szczelin (w tej samej częstotliwości lub w różnych), może osiągać coraz wyższą szybkość transmisji. Przydziały (nowe lub ponowne) szczelin dla węzłów mobilnych mogą się odbywać nawet co milisekundę. Można też stosować różne schematy modulacji, aby zmieniać częstotliwość transmisji. Zobacz wcześniejsze omówienie rysunku 7.3 i dynamiczny wybór schematów modulacji w sieciach Wi-Fi.



Rysunek 7.21. Dwadzieścia szczelin o długości 0,5 milisekundy uporządkowanych w ramki o długości 10 milisekund w poszczególnych częstotliwościach. Przydzielanych osiem szczelin jest wyróżnionych ciemniejszym kolorem

Standard LTE nie określa konkretnego sposobu przydziału szczelin czasowych węzłom mobilnym. Decyzja o tym, które węzły mobilne będą mogły przesyłać dane w określonej szczelinie czasowej w poszczególnych częstotliwościach, jest podejmowana przez algorytmy szeregowania definiowane przez dostawców sprzętu LTE i (lub) operatorów sieci. W szeregowaniu oportunistycznym [Bender 2000; Kolding 2003; Kulkarni 2005] kontroler sieci radiowej może w optymalny sposób wykorzystać nośnik bezprzewodowy,

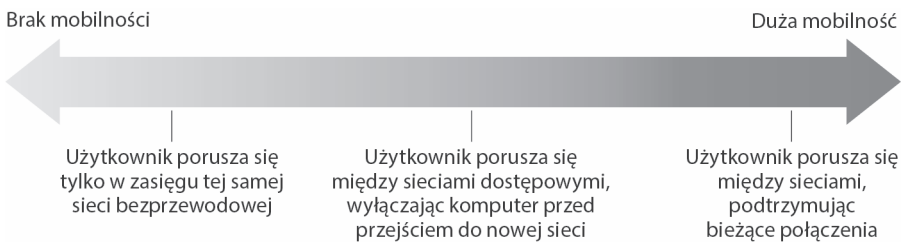
dopasowując protokół warstwy fizycznej do warunków w kanale łączącym nadawcę i odbiorcę oraz wybierając na podstawie warunków w kanale odbiorców, do których pakiety zostaną dostarczone. Do szeregowania transmisji pakietów w kanale dosyłowym można też wykorzystać uprzywilejowanie użytkowników i wykupiony poziom usług (na przykład srebrny, złoty lub platynowy). Obok opisanych wcześniej możliwości sieci LTE dostępny jest też standard LTE-Advanced, który pozwala uzyskać przepustowość kanału dosyłowego na poziomie setek Mb/s dzięki przydziałowi węzłom mobilnym zagregowanych kanałów [Akyildiz 2010].

Inną technologią bezprzewodową 4G z rodziny standardów IEEE 802.16 jest **WiMAX** (ang. *World Interoperability for Microwave Access*). Różni się ona znacznie od LTE. Technologia WiMAX nie została do tej pory spopularyzowana w takim stopniu jak LTE.

7.5. Zasady zarządzania mobilnością

Rozważywszy *bezprzewodową* naturę łączy komunikacyjnych, możemy skupić się na *mobilności* oferowanej przez te łącza. W najszerszym sensie węzeł mobilny to taki, który z czasem zmienia swój punkt przyłączenia do sieci. Ponieważ termin *mobilność* ma wiele znaczeń zarówno w świecie komputerów, jak i telefonii, warto najpierw zastanowić się nad kilkoma wymiarami mobilności.

- *Jaki jest stopień mobilności użytkownika z perspektywy warstwy sieci?* Fizycznie mobilny użytkownik może stawiać różne wyzwania warstwie sieci w zależności od tego, w jaki sposób porusza się między punktami przyłączenia do sieci. Na jednym końcu spektrum z rysunku 7.22 znajduje się użytkownik, który nosi laptop z interfejsem bezprzewodowym w obrębie budynku. Jak wyjaśniono w punkcie 7.3.4, z perspektywy warstwy sieci ten użytkownik *nie jest* mobilny. Co więcej, jeśli użytkownik zawsze jest powiązany z tym samym punktem dostępowym, to nie jest mobilny nawet z perspektywy warstwy łącza.



Rysunek 7.22. Różne stopnie mobilności (z perspektywy warstwy sieci)

Na drugim końcu spektrum znajduje się użytkownik, który mknie autostradą w swoim bmw lub teslą z szybkością 150 kilometrów na godzinę. Przejeżdża on przez różne bezprzewodowe sieci dostępowe i chce zachować nieprzerwane połączenie TCP ze zdalną aplikacją. Ten użytkownik jest *zdecydowanie* mobilny! Między tymi

skrajnościami mamy użytkownika, który przenosi laptop z jednego miejsca (na przykład biura lub akademika) w drugie (na przykład do kawiarni lub sali wykładowej) i chce połączyć się z siecią w nowej lokacji. Użytkownik ten również jest mobilny (choć nie tak jak kierowca bmw!), ale nie musi podtrzymywać połączeń podczas przechodzenia między punktami przyłączeniowymi. Na rysunku 7.22 przedstawiono to spektrum mobilności użytkownika z perspektywy warstwy sieci.

- *Czy adres węzła mobilnego musi pozostawać taki sam?* W telefonii mobilnej numer telefonu — zasadniczo adres telefonu w warstwie sieci — pozostaje taki sam podczas poruszania się między sieciami komórkowymi różnych operatorów. Czy laptop również musi zachowywać swój adres podczas poruszania się między sieciami IP?

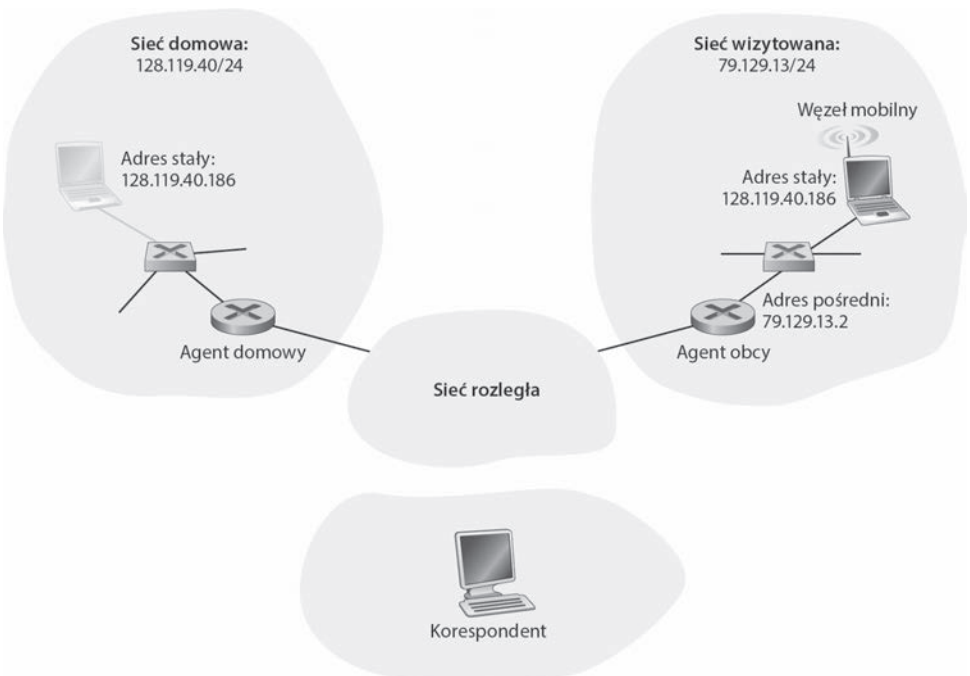
  Odpowiedź na to pytanie w dużej mierze zależy od używanych aplikacji. Kierowca bmw lub tesli, który chce podtrzymywać połączenia TCP podczas jazdy autostradą, powinien zachowywać swój adres. W rozdziale 3. wyjaśniono, że aplikacja internetowa musi znać adres IP oraz numer portu zdalnego węzła, z którym się komunikuje. Jeśli zdalny węzeł będzie mógł zachować swój adres podczas poruszania się, mobilność stanie się niewidoczna z perspektywy aplikacji. Taka przezroczystość jest bardzo korzystna — aplikacja nie musi troszczyć się o potencjalną zmianę adresu IP, a ten sam kod może obsługiwać zarówno połączenia mobilne, jak i stacjonarne. W następnym podrozdziale Czytelnicy dowiedzą się, że przezroczystość tę oferuje protokół Mobile IP, który pozwala węzłowi zachować swój adres IP podczas poruszania się między sieciami.

  Z drugiej strony, skromniejszy użytkownik mobilny może po prostu wyłączać biurowy laptop, zabierać go ze sobą do domu i tam pracować. Jeśli laptop funkcjonuje głównie jak klient w aplikacjach typu klient-serwer (jest używany na przykład do czytania i wysyłania poczty, przeglądania sieci WWW, łączenia się przez Telnet ze zdalnym hostem), jego adres nie jest szczególnie istotny. W zupełności wystarczy tymczasowy adres przydzielany użytkownikowi przez domowego dostawcę usług internetowych. W podrozdziale 4.3 napisaliśmy, że funkcję tę zapewnia protokół DHCP.

- *Czy dostępna jest pomocnicza infrastruktura sieciowa?* We wszystkich powyższych scenariuszach zakładaliśmy, że użytkownik mobilny dysponuje jakąś stałą infrastrukturą, na przykład siecią domowego dostawcy ISP, bezprzewodową siecią dostępową w biurze albo sieciami dostępowymi rozmieszczonymi wzdłuż autostrady. A jeśli taka infrastruktura nie istnieje? Jeśli dwaj użytkownicy znajdują się blisko siebie, czy mogą nawiązać połączenie mimo braku infrastruktury sieciowej? Właśnie takie możliwości oferuje doraźna łączność sieciowa. Ta szybko rozwijająca się dziedzina łączności bezprzewodowej jest obecnie przedmiotem badań, ale jej opis wykracza poza ramy niniejszej książki. Szczegółowe omówienie sieci doraźnych można znaleźć w [Perkins 2000] oraz na stronach WWW grupy roboczej IETF Mobile Ad Hoc Network (manet) [manet 2016].

Aby zilustrować problemy związane z podtrzymywaniem połączeń podczas poruszania się między sieciami, posłużmy się analogią: dwudziestokilkuletnia osoba wyprowadza się z domu, zaczyna mieszkać w akademikach oraz (lub) wynajętych mieszkaniach i często zmienia adres. Jak może się z nią skontaktować stary przyjaciel? Jednym z rozwiązań jest skontaktowanie się z rodziną, ponieważ mobilna osoba zwykle informuje ją o zmianie adresu (choćby po to, aby rodzice wysyłali jej pieniądze na czynsz!). Dom rodzinny ze stałym adresem staje się zatem pierwszym etapem komunikacji z mobilną osobą. Późniejsza komunikacja może odbywać się pośrednio (na przykład poprzez wysyłanie listów do rodziców, którzy następnie przekazują je zaprzyjaźnionej osobie) albo bezpośrednio (poprzez wysyłanie listów na adres uzyskany od rodziców).

W środowisku sieciowym stałym „miejscem zamieszkania" węzła mobilnego (na przykład laptopa lub urządzenia PDA) jest **sieć domowa**, a jednostką, która pełni funkcje zarządzania mobilnością na rzecz tego węzła, jest **agent domowy**. Sieć, w której obecnie rezyduje węzeł mobilny, określa się mianem **sieci obcej** (albo **wizytowanej**), a jednostką, która pełni funkcje zarządzania mobilnością na rzecz tego węzła, jest **agent obcy**. W przypadku pracowników mobilnych siecią domową jest zwykle sieć firmowa, natomiast siecią obcą może być sieć kolegi, u którego są z wizytą. **Korespondent** to jednostka, która chce nawiązać komunikację z węzłem mobilnym. Na rysunku 7.23 zilustrowano te pojęcia, a także omówione niżej kwestie adresowania. Rysunek przedstawia agenty jako procesy działające w routerze, ale mogą one również funkcjonować w hostach lub serwerach.



Rysunek 7.23. Pierwsze elementy architektury sieci mobilnej

7.5.1. Adresowanie

Jak już wspomnieliśmy, aby mobilność użytkownika była niewidoczna dla aplikacji sieciowych, węzeł mobilny powinien zachowywać swój adres podczas poruszania się między sieciami. Kiedy węzeł mobilny rezyduje w sieci obcej, cały ruch zmierzający na stały adres węzła powinien być kierowany do sieci obcej. Jak to osiągnąć? Sieć obca mogłaby informować wszystkie inne sieci, że znajduje się w niej węzeł mobilny. W tym celu można by wykorzystać zwykłą wymianę informacji o routingu wewnątrz- i międzydomenowym, co nie wymagałoby większych zmian w istniejącej infrastrukturze routingu. Sieć obca po prostu informowałaby swoich sąsiadów, że dysponuje bardzo specyficzną trasą do stałego adresu węzła mobilnego (tzn. że może prawidłowo kierować datagramy na stały adres węzła; więcej informacji na ten temat można znaleźć w podrozdziale 4.3). Sąsiedzi rozpowszechnialiby tę informację w całej sieci w ramach zwykłego procesu aktualizowania informacji o routingu i tabel przekazywania. Kiedy węzeł mobilny opuszczałby jedną sieć i dołączał do drugiej, nowa sieć obca ogłaszałaby specyficzną trasę do węzła mobilnego, a stara sieć obca wycofywałaby informacje o routingu dotyczące tego węzła.

W ten sposób można rozwiązać dwa problemy jednocześnie, nie dokonując znaczących zmian w infrastrukturze warstwy sieci. Inne sieci znają lokalizację węzła mobilnego, a routing jest łatwy, ponieważ tablice przekazywania nakazują kierować datagramy do sieci obcej. Podstawową wadą takiego rozwiązania jest jednak mała skalowalność. Gdyby zarządzanie mobilnością było zadaniem routerów sieciowych, musiałyby one przechowywać w tablicach przekazywania wpisy dotyczące milionów węzłów mobilnych i aktualizować je w miarę przemieszczania się węzłów. Niektóre inne wady tego rozwiązania są opisane w punkcie „Problemy" na końcu rozdziału.

Alternatywą (zastosowaną w praktyce) jest przeniesienie funkcji zarządzania mobilnością z rdzenia sieci na jej obrzeże — jest to powtarzający się motyw w naszych studiach nad architekturą internetu. Naturalnym pośrednikiem jest tutaj sieć domowa węzła mobilnego. Podobnie jak rodzice dwudziestokilkuletniej osoby znają miejsce zamieszkania swojego dziecka, tak agent w sieci domowej węzła mobilnego może śledzić sieć obcą, w której aktualnie znajduje się węzeł. Do aktualizowania lokacji węzła mobilnego oczywiście niezbędny jest protokół między węzłem (albo reprezentującym go agentem obcym) a agentem domowym.

Rozważmy teraz bardziej szczegółowo rolę agenta obcego. Najprostszym rozwiązaniem, zilustrowanym na rysunku 7.23, jest umieszczenie agentów obcych w routerach brzegowych sieci obcej. Jednym z zadań agenta obcego jest tworzenie tak zwanego **adresu pośredniego** (ang. *care-of address* — **COA**) węzła mobilnego, w którym część sieciowa wskazuje sieć obcą. Węzeł mobilny ma zatem dwa adresy: **adres stały** (odpowiadający adresowi domu rodzinnego dwudziestokilkulatka z naszej analogii) oraz adres pośredni, czasem nazywany **adresem obcym** (odpowiadający bieżącemu miejscu zamieszkania). W przykładzie z rysunku 7.23 stały adres węzła mobilnego to 128.119.40.186. Podczas wizyty w sieci 79.129.13/24 węzeł mobilny ma adres COA 79.129.13.2. Drugim zadaniem agenta obcego jest informowanie agenta domowego, że węzeł mobilny znajduje się w określonej sieci obcej pod odpowiednim adresem COA. Niebawem Czytelnicy przekonają się, że adres COA pozwala przekierowywać datagramy do węzła mobilnego za pośrednictwem jego agenta obcego.

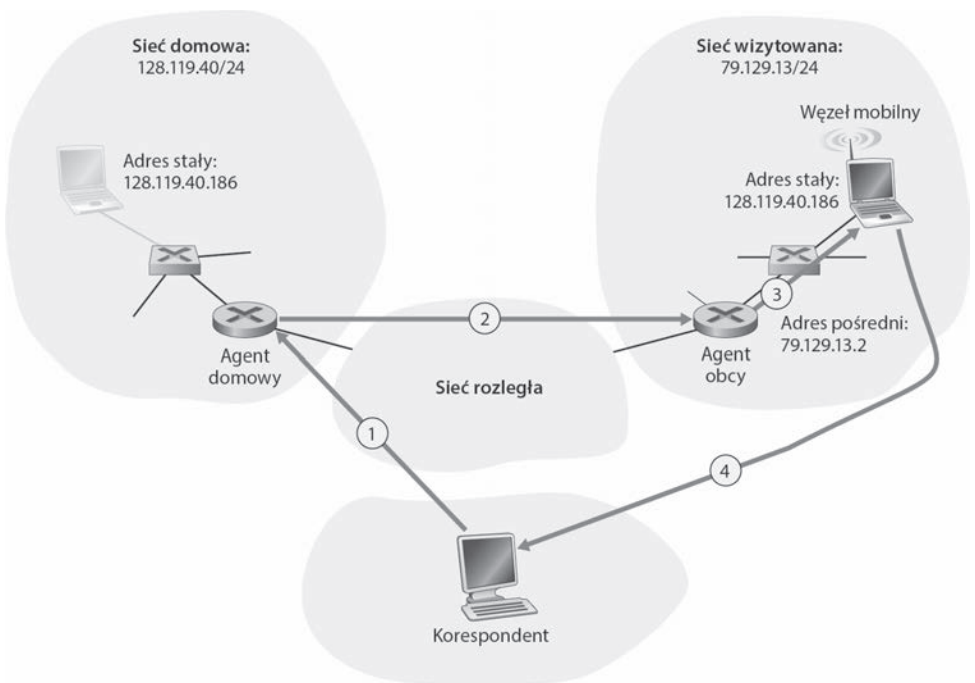
Choć oddzieliliśmy funkcje węzła mobilnego i agenta obcego, warto nadmienić, że agentem może być sam węzeł. Węzeł może na przykład pobrać adres COA w sieci obcej (za pomocą protokołu takiego jak DHCP) i poinformować o nim agenta domowego.

7.5.2. Routing do węzła mobilnego

Wiemy już, jak węzeł mobilny uzyskuje adres COA i jak agent domowy jest informowany o tym adresie. Jednak informowanie agenta domowego o adresie COA rozwiązuje tylko część problemu. Jak adresować i przekazywać datagramy do węzła mobilnego? Ponieważ położenie węzła zna tylko agent domowy (a nie routery w całej sieci), nie wystarczy opatrzyć datagramu stałym adresem węzła i powierzyć go infrastrukturze warstwy sieci. Potrzeba czegoś więcej. Możliwe są dwa rozwiązania, które będziemy nazywać routingiem pośrednim i bezpośrednim.

Routing pośredni do węzła mobilnego

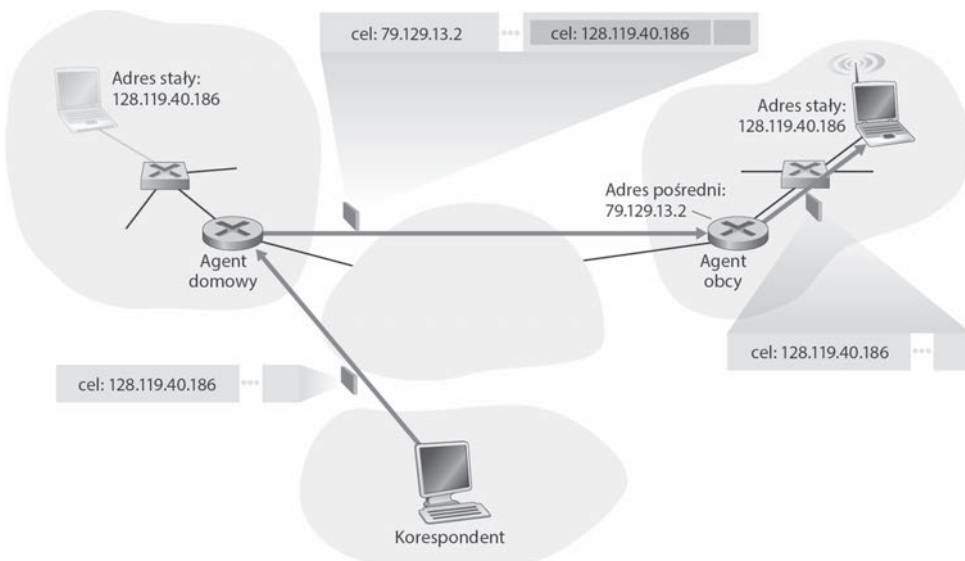
Rozważmy najpierw korespondenta, który chce wysłać datagram do węzła mobilnego. W **routingu pośrednim** korespondent po prostu opatruje datagram stałym adresem węzła mobilnego i wysyła go do sieci, nie dbając o to, czy węzeł znajduje się w sieci domowej, czy też w sieci obcej; mobilność jest zatem całkowicie niewidoczna dla korespondenta. Takie datagramy są najpierw kierowane — w zwykły sposób — do sieci domowej węzła mobilnego. Ilustruje to etap 1. na rysunku 7.24.



Rysunek 7.24. Pośrednie przekazywanie datagramów do węzła mobilnego

Teraz zwróćmy uwagę na agenta domowego. Oprócz współpracy z agentem obcym w celu śledzenia adresu COA węzła mobilnego agent domowy pełni jeszcze jedną, bardzo istotną funkcję. Jego drugim zadaniem jest oczekiwanie na datagramy adresowane do węzłów, których siecią domową jest sieć agenta, a które obecnie znajdują się w sieci obcej. Agent domowy przechwytuje te datagramy i przekazuje je do węzła mobilnego. Proces przekazywania składa się z dwóch etapów. Datagram jest najpierw przekazywany do agenta obcego z wykorzystaniem adresu COA węzła (etap 2. na rysunku 7.24), a następnie od agenta obcego do węzła mobilnego (etap 3. na rysunku 7.24).

Warto dokładniej zbadać to przekierowywanie. Agent domowy będzie musiał opatrzyć datagram adresem COA węzła mobilnego, aby warstwa sieci skierowała datagram do odpowiedniej sieci obcej. Z drugiej strony byłoby dobrze, gdyby datagram korespondenta pozostał nienaruszony, ponieważ aplikacja odbiorcza nie powinna wiedzieć, że datagram został przekierowany przez agenta domowego. Oba te cele zostaną osiągnięte, jeśli agent domowy będzie **kapsułkował** pierwotny datagram korespondenta w nowym (większym) datagramie. Ten większy datagram zostanie opatrzony adresem COA. Agent obcy, który jest „właścicielem" adresu COA, odbierze większy datagram, po czym wyodrębni z niego pierwotny datagram korespondenta i przekaże go (etap 3. na rysunku 7.24) węzłowi mobilnemu. Na rysunku 7.25 przedstawiono pierwotny datagram korespondenta przesyłany do sieci domowej, zakapsułkowany datagram przekazywany do agenta obcego oraz pierwotny datagram dostarczany do węzła mobilnego. Uważny Czytelnik zauważy, że opisane tu kapsułkowanie i odkapsułkowywanie jest identyczne z pojęciem tunelowania omówionym w podrozdziale 4.3 w kontekście transmisji grupowych IP oraz protokołu IPv6.



Rysunek 7.25. Kapsułkowanie i odkapsułkowywanie

Zastanówmy się teraz, jak węzeł mobilny wysyła datagramy do korespondenta. Jest to całkiem proste, ponieważ węzeł może zaadresować datagram *bezpośrednio* do korespondenta (używając swojego stałego adresu jako adresu źródłowego, a adresu korespondenta jako adresu docelowego). Ponieważ węzeł mobilny zna adres korespondenta, nie musi wysyłać datagramu za pośrednictwem agenta domowego. Pokazano to jako etap 4. na rysunku 7.24.

Podsumujmy omówienie routingu pośredniego, wymieniając nowe mechanizmy warstwy sieci niezbędne do obsługi mobilności:

- *Protokół między węzłem mobilnym a agentem obcym.* Kiedy węzeł mobilny łączy się z siecią obcą, rejestruje się w agencie obcym. Podobnie kiedy węzeł opuszcza sieć obcą, wyrejestrowuje się w agencie obcym.
- *Protokół rejestracji między agentem obcym a agentem domowym.* Agent obcy rejestruje adres COA węzła mobilnego w agencie domowym. Agent obcy nie musi jawnie wyrejestrowywać adresu COA, kiedy węzeł mobilny opuszcza jego sieć, ponieważ agent obcy w nowej sieci dokona nowej rejestracji.
- *Protokół kapsułkowania w agencie domowym.* Kapsułkowanie i przekazywanie pierwotnego datagramu korespondenta w datagramie z adresem COA.
- *Protokół odkapsułkowywania w agencie obcym.* Wyodrębnianie pierwotnego datagramu korespondenta z datagramu kapsułkującego i przekazywanie pierwotnego datagramu do węzła mobilnego.

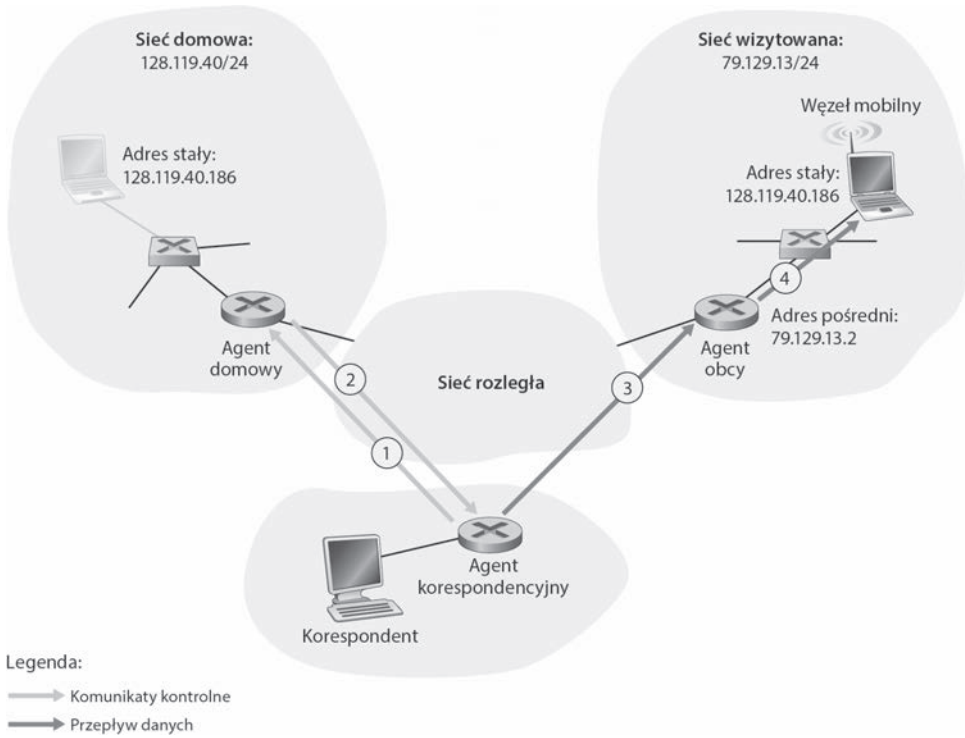
Powyżej opisano wszystkie elementy — agenty obce, agenta domowego i przekazywanie pośrednie — niezbędne do podtrzymywania połączeń podczas poruszania się między sieciami. Aby zrozumieć zależności między tymi elementami, załóżmy, że węzeł mobilny jest podłączony do sieci obcej A, zarejestrował adres COA w sieci A w swoim agencie domowym i odbiera datagramy przekazywane pośrednio przez agenta domowego. Węzeł mobilny przechodzi teraz do sieci obcej B i rejestruje się w jej agencie obcym, który informuje agenta domowego o nowym adresie COA w sieci B. Od tego momentu agent domowy będzie kierował datagramy do sieci obcej B. Z perspektywy korespondenta mobilność jest niewidoczna — datagramy są kierowane do tego samego agenta domowego zarówno przed zmianą sieci, jak i po niej. Z perspektywy agenta domowego przepływ danych jest niezakłócony — przychodzące datagramy są najpierw kierowane do sieci obcej A, a po zmianie adresu — do sieci obcej B. A jak to wygląda z perspektywy węzła mobilnego, który przechodzi do nowej sieci? Jeśli między odłączeniem węzła od sieci A (po którym nie może on już odbierać datagramów za jej pośrednictwem) a przyłączeniem do sieci B (po którym może on zarejestrować nowy adres COA w swoim agencie domowym) upływa krótki czas, zostaną utracone tylko nieliczne datagramy. Jak pamiętamy z rozdziału 3., w połączeniach między węzłami końcowymi może wystąpić utrata pakietów spowodowana przeciążeniem sieci. Zatem utrata kilku datagramów podczas przechodzenia węzła między sieciami nie stanowi poważnego problemu. Jeśli wymagana jest bezstratna komunikacja, mechanizmy wyższych warstw potrafią uporać się z utratą datagramów, bez względu na to, czy wynika ona z przeciążenia sieci, czy też z mobilności użytkownika.

Metoda routingu pośredniego jest używana w standardzie Mobile IP [RFC 5944] omówionym w podrozdziale 7.6.

Routing bezpośredni do węzła mobilnego

Technika routingu pośredniego zilustrowana na rysunku 7.24 bywa niewydajna ze względu na tak zwany **problem routingu trójkątnego** — datagramy adresowane do węzła mobilnego muszą najpierw zostać skierowane do agenta domowego, a następnie do sieci obcej, nawet jeśli istnieje znacznie bardziej efektywna trasa między korespondentem a węzłem. Wyobraźmy sobie najgorszy przypadek: użytkownika mobilnego, który podłączył się do sieci obcej u swojego kolegi. Siedzą oni obok siebie i wymieniają dane przez sieć. Datagramy korespondenta (w tym przypadku odwiedzonego kolegi) są kierowane do agenta domowego użytkownika mobilnego, a następnie z powrotem do sieci obcej!

Routing bezpośredni pozwala zwiększyć wydajność transmisji danych, ale kosztem dodatkowej złożoności. W metodzie routingu bezpośredniego **agent korespondencyjny** w sieci korespondenta najpierw poznaje adres COA węzła mobilnego. W tym celu może odpytać agenta domowego, przy założeniu, że (jak w przypadku routingu pośredniego) węzeł mobilny ma aktualny adres COA zarejestrowany w agencie domowym. Funkcję agenta korespondencyjnego może pełnić sam korespondent (podobnie jak funkcję agenta obcego może pełnić sam węzeł mobilny). Pokazano to jako etapy 1. i 2. na rysunku 7.26. Następnie agent korespondencyjny tuneluje datagramy bezpośrednio na adres COA węzła mobilnego, podobnie jak robi to agent domowy (etapy 3. i 4. na rysunku 7.26).

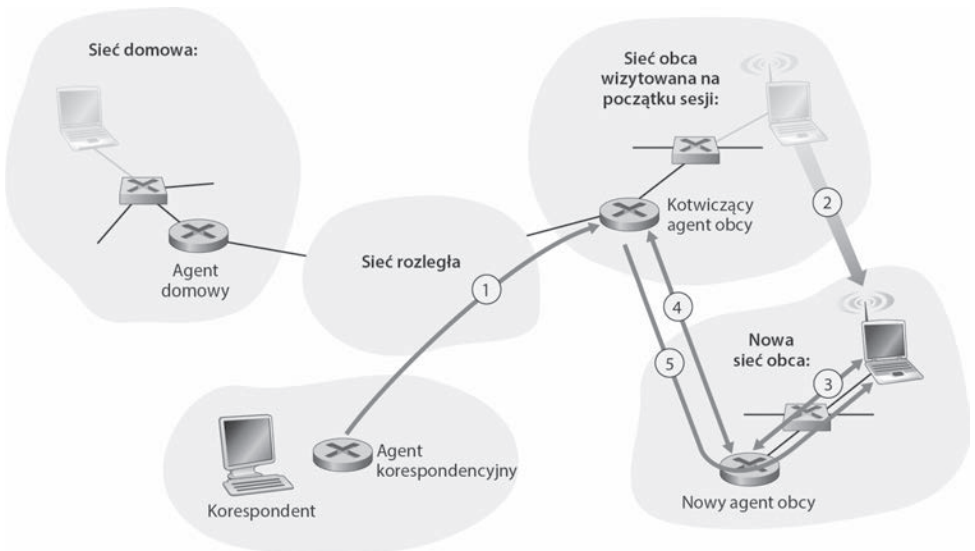


Rysunek 7.26. Routing bezpośredni do węzła mobilnego

Routing bezpośredni rozwiązuje problem routingu trójkątnego, ale wiąże się z dwoma dodatkowymi wyzwaniami:

- Aby agent korespondencyjny mógł odpytać agenta domowego i uzyskać adres COA węzła mobilnego, niezbędny jest **protokół lokalizowania użytkownika mobilnego** (etapy 1. i 2. na rysunku 7.26).
- Jeśli węzeł mobilny przejdzie do nowej sieci obcej, jak będzie można skierować do niej dane? W przypadku routingu pośredniego wystarczyło zaktualizować adres COA przechowywany przez agenta domowego. Jednakże w routingu bezpośrednim agent domowy jest odpytywany przez agenta korespondencyjnego tylko raz, na początku sesji. Zatem aktualizacja adresu COA w agencie domowym, choć potrzebna, nie pozwoli skierować danych do nowej sieci obcej węzła mobilnego.

Jednym z rozwiązań może być dodatkowy protokół, za pomocą którego korespondent będzie informowany o zmianie adresu COA. Alternatywne rozwiązanie, stosowane w praktyce w sieciach GSM, działa w następujący sposób: przypuśćmy, że dane są obecnie przekazywane do węzła mobilnego w sieci obcej, w której węzeł został zlokalizowany na początku sesji (etap 1. na rysunku 7.27). Agenta obcego w tej sieci będziemy określać mianem **kotwiczącego agenta obcego** (ang. *anchor foreign agent*). Kiedy węzeł mobilny przejdzie do nowej sieci (etap 2. na rysunku 7.27), zarejestruje się w nowym agencie obcym (etap 3.), który przekaże agentowi kotwiczącemu nowy adres COA węzła (etap 4.). Kiedy agent kotwiczący otrzyma datagram przeznaczony dla węzła mobilnego, ponownie zakapsułkuje go i wyśle (etap 5.) na nowy adres COA. Jeśli później węzeł mobilny znów zmieni sieć, nowy agent obcy skontaktuje się z agentem kotwiczącym, aby skonfigurować routing do tej sieci.



Rysunek 7.27. Routing bezpośredni i przejście węzła mobilnego do nowej sieci

7.6. Mobile IP

Architektura internetowa i protokoły do obsługi mobilności, zbiorowo określane mianem Mobile IP, są zdefiniowane przede wszystkim w dokumencie RFC 5944 dla protokołu IPv4. Mobile IP to standard elastyczny, uwzględniający wiele różnych trybów komunikacji, na przykład działanie z agentem obcym albo bez niego, wiele sposobów wzajemnego odkrywania się przez agenty i węzły mobilne, używanie jednego lub wielu adresów COA oraz liczne postaci kapsułkowania. Mobile IP jest zatem złożonym standardem, którego szczegółowy opis zająłby całą książkę; jedną z takich pozycji jest [Perkins 1998b]. Naszym skromnym celem będzie tu przegląd najważniejszych aspektów Mobile IP i zilustrowanie jego zastosowań w kilku często spotykanych sytuacjach.

Architektura Mobile IP obejmuje wiele spośród opisanych wyżej elementów, w tym agenty domowe, agenty obce, adresy pośrednie oraz kapsułkowanie i odkapsułkowywanie. Bieżący standard [RFC 5944] zakłada routing pośredni do węzła mobilnego.

Standard Mobile IP składa się z trzech podstawowych części:

- *Odkrywanie agentów.* Mobile IP definiuje protokoły używane przez agenty domowe i obce do oferowania swoich usług węzłom mobilnym, a także protokoły, za pomocą których węzły mobilne zabiegają o usługi agenta domowego lub obcego.
- *Rejestracja w agencie domowym.* Mobile IP definiuje protokoły, których węzeł mobilny i (lub) agent obcy używają do rejestrowania i wyrejestrowywania adresów COA w agencie domowym.
- *Pośredni routing datagramów.* Standard definiuje również sposób przekazywania danych do węzłów mobilnych przez agenta domowego, w tym reguły przekazywania datagramów, reguły obsługi błędów oraz kilka postaci kapsułkowania [RFC 2003, RFC 2004].

W standardzie Mobile IP uwzględniono też kwestie bezpieczeństwa. Niezbędne jest na przykład uwierzytelnianie węzła mobilnego — w przeciwnym razie ktoś o złych intencjach mógłby zarejestrować fałszywy adres COA w agencie domowym, aby przechwytywać wszystkie datagramy kierowane pod adres IP węzła. W Mobile IP bezpieczeństwo osiąga się z wykorzystaniem mechanizmów, które zbadamy w rozdziale 8., więc w poniższym opisie pominiemy te zagadnienia.

Odkrywanie agentów

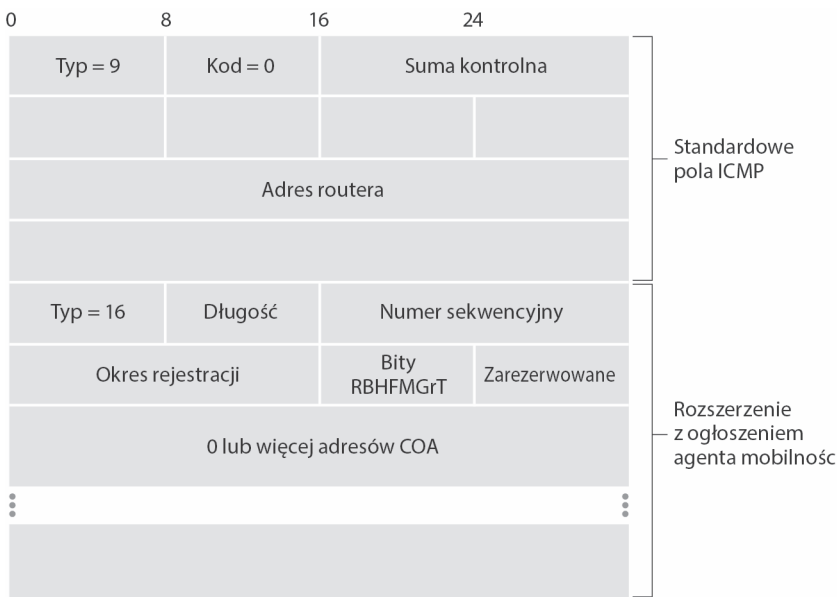
Węzeł Mobile IP, który przybywa do nowej sieci — bez względu na to, czy przyłącza się do sieci obcej, czy też wraca do sieci domowej — musi poznać tożsamość odpowiedniego agenta obcego lub domowego. Właśnie poprzez wykrycie nowego agenta obcego, z nowym adresem sieciowym, oprogramowanie warstwy sieci w węźle mobilnym może ustalić, że węzeł trafił do innej sieci. Proces ten nazywamy **odkrywaniem agentów**. Można zrealizować go dwojako: poprzez ogłaszanie agentów albo poszukiwanie agentów.

W przypadku **ogłaszania agentów** agent obcy lub domowy ogłasza swoje usługi, korzystając z rozszerzenia istniejącego protokołu odkrywania routerów [RFC 1256].

Agent okresowo rozgłasza na wszystkich swoich łączach komunikat ICMP z wartością 9 w polu typu (odkrywanie routera). Komunikat ten zawiera adres IP routera (tzn. agenta), co umożliwia węzłom mobilnym poznanie tego adresu. Rozszerzeniem komunikatu jest ogłoszenie agenta mobilności, które zawiera dodatkowe informacje potrzebne węzłowi mobilnemu. Oto najważniejsze pola tego rozszerzenia:

- *Bit agenta domowego (H).* Wskazuje, że agent pełni w danej sieci funkcję agenta domowego.
- *Bit agenta obcego (F).* Wskazuje, że agent pełni w danej sieci funkcję agenta obcego.
- *Bit wymaganej rejestracji (R).* Wskazuje, że węzeł mobilny w danej sieci *musi* zarejestrować się w agencie obcym. Oznacza to, że węzeł mobilny nie może uzyskać adresu COA w obcej sieci (na przykład za pomocą DHCP) i sam pełnić funkcji agenta obcego, nie rejestrując się w agencie obcym.
- *Bity kapsułkowania (M, G).* Wskazują, czy będzie używane kapsułkowanie inne niż IP w IP.
- *Pola adresu pośredniego (COA).* Lista jednego lub wielu adresów COA oferowanych przez agenta obcego. W omówionym niżej przykładzie adres COA jest związany z agentem obcym, który odbiera datagramy wysyłane na ten adres, a następnie przekazuje je odpowiedniemu węzłowi mobilnemu. Użytkownik mobilny wybiera jeden z dostępnych adresów COA podczas rejestrowania się w agencie obcym.

Na rysunku 7.28 zilustrowano kluczowe pola ogłoszenia agenta.



Rysunek 7.28. Komunikat odkrywania routera ICMP z rozszerzeniem w postaci ogłoszenia agenta

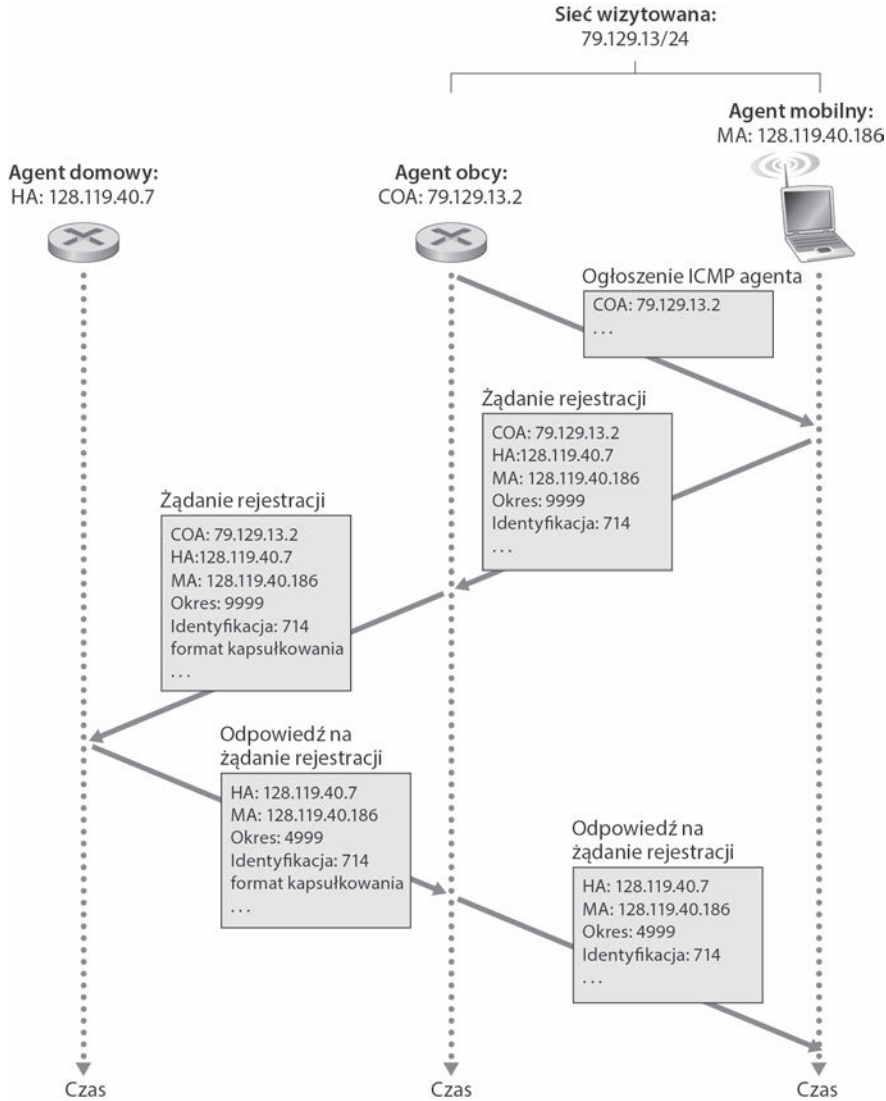
W przypadku **poszukiwania agentów** węzeł mobilny ustala adres agenta bez czekania na ogłoszenie. W tym celu rozgłasza komunikat poszukiwania agenta, który jest po prostu komunikatem ICMP z wartością 10 w polu typu. Agent, który odbierze ten komunikat, prześle swoje ogłoszenie bezpośrednio do węzła mobilnego. Dalej wszystko toczy się tak, jakby agent odebrał spontaniczne ogłoszenie.

Rejestracja w agencie domowym

Kiedy węzeł Mobile IP otrzyma adres COA, konieczne jest zarejestrowanie tego adresu w agencie domowym. Może to zrobić albo agent obcy, albo sam węzeł. Poniżej rozważamy ten pierwszy przypadek. Rejestracja przebiega w czterech etapach:

1. Po odebraniu ogłoszenia od agenta obcego węzeł mobilny wysyła do niego komunikat rejestracyjny Mobile IP. Komunikat rejestracyjny jest wysyłany w datagramie UDP do portu 434. Zawiera on adres COA ogłoszony przez agenta obcego, adres agenta domowego (HA), stały adres węzła mobilnego (MA), żądany okres rejestracji oraz 64-bitowy identyfikator rejestracji. Żądany okres rejestracji to czas (w sekundach), przez jaki rejestracja będzie ważna. Jeśli rejestracja nie zostanie odnowiona po podanym okresie, stanie się nieważna. Identyfikator rejestracji pełni funkcję numeru sekwencyjnego, pozwalając dopasować odpowiedź do żądania rejestracji, jak opisano niżej.
2. Agent obcy odbiera komunikat rejestracyjny i zapisuje stały adres IP węzła mobilnego. Teraz agent obcy wie, że powinien poszukiwać datagramów z osadzonymi datagramami, których adres docelowy odpowiada stałemu adresowi węzła mobilnego. Następnie agent obcy wysyła komunikat rejestracyjny Mobile IP (znów w datagramie UDP) do portu 434 agenta domowego. Komunikat zawiera adresy COA, HA i MA, żądany format kapsułkowania, żądany okres rejestracji oraz identyfikator rejestracji.
3. Agent domowy odbiera żądanie rejestracji i sprawdza jego autentyczność oraz prawidłowość. Agent wiąże stały adres IP węzła mobilnego z jego adresem COA; w przyszłości datagramy przychodzące do agenta domowego i adresowane do węzła mobilnego będą kapsułkowane i tunelowane na adres COA. Agent domowy wysyła odpowiedź rejestracji Mobile IP zawierającą adresy HA i MA, rzeczywisty okres rejestracji oraz identyfikator rejestracji żądania, którego dotyczy ta odpowiedź.
4. Agent obcy otrzymuje odpowiedź rejestracji i przekazuje ją węzłowi mobilnemu.

W tym momencie rejestracja dobiega końca, a węzeł mobilny może odbierać datagramy wysyłane na jego stały adres. Poszczególne etapy tego procesu zilustrowano na rysunku 7.29. Zwróćmy uwagę, że agent domowy podaje okres rejestracji krótszy niż ten, którego zażądał węzeł mobilny.



Rysunek 7.29. Ogłoszenie agenta i rejestracja Mobile IP

Agent obcy nie musi jawnie wyrejestrowywać adresu COA, kiedy węzeł mobilny opuści jego sieć. Odbędzie się to automatycznie, kiedy węzeł przejdzie do nowej sieci (obcej lub domowej) i zarejestruje nowy adres COA.

Standard Mobile IP przewiduje znacznie więcej scenariuszy i funkcji oprócz opisanych powyżej. Zainteresowani Czytelnicy powinni zajrzeć do [Perkins 1998b; RFC 5944].

7.7. Zarządzanie mobilnością w sieciach komórkowych

Zbadaliśmy już zarządzanie mobilnością w sieciach IP, a teraz przyjrzymy się sieciom, które mają jeszcze dłuższą historię obsługi mobilności — sieciom telefonii komórkowej. W podrozdziale 7.4 skupiliśmy się na charakterystyce komórkowych łączy bezprzewodowych, a tu przedstawimy zarządzanie mobilnością w architekturze GSM [Goodman 1997; Mouly 1992; Scourias 2012; Kaaranen 2001; Korhonen 2003; Turner 2012], która jest technologią dojrzałą i szeroko stosowaną. Zasady obsługi mobilności w sieciach 3G i 4G są podobne jak w sieciach GSM. Podobnie jak w przypadku Mobile IP, przekonamy się, że architektura sieci GSM realizuje kilka spośród fundamentalnych zasad wymienionych w podrozdziale 7.5.

W GSM, tak jak w Mobile IP, stosuje się routing pośredni (punkt 7.5.2). Rozmowa korespondenta trafia najpierw do sieci domowej telefonu, a stamtąd do sieci wizytowanej. W terminologii GSM sieć domowa telefonu jest określana mianem **domowej sieci publicznej naziemnej telefonii mobilnej** (ang. *home public land mobile network* — home PMLN). Ponieważ akronim PMLN jest trudny do wymówienia, a ponadto obiecaliśmy sobie unikać „alfabetycznej zupy", domową sieć PMLN będziemy nazywać po prostu **siecią domową**. Jest to sieć operatora telefonii komórkowej, którego abonentem jest użytkownik mobilny (tzn. który wystawia rachunki użytkownikowi). Wizytowana sieć PMLN, którą będziemy nazywać po prostu **siecią wizytowaną**, to sieć, w której obecnie znajduje się telefon.

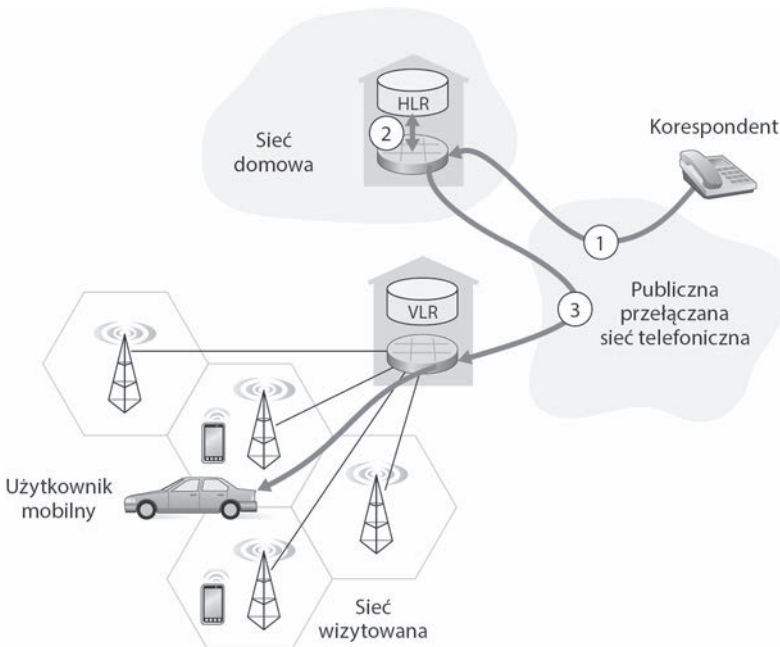
Podobnie jak w przypadku Mobile IP, zadania sieci domowej i wizytowanej są zupełnie inne.

- Sieć domowa utrzymuje bazę danych znaną jako **rejestr stacji własnych** (ang. *home location register* — **HLR**), która zawiera stałe numery telefonów i profile poszczególnych abonentów. Co najważniejsze, HLR zawiera również informacje o bieżącym położeniu tych abonentów. Jeśli użytkownik znajduje się obecnie w sieci komórkowej innego operatora, HLR dostarcza informacji pozwalających uzyskać (za pomocą procesu, który niebawem opiszemy) adres w sieci wizytowanej, na który powinny być kierowane połączenia. Kiedy korespondent chce nawiązać połączenie z użytkownikiem mobilnym, konieczne jest skontaktowanie się ze specjalnym przełącznikiem w sieci domowej, nazywanym **tranzytową centralą przełączania mobilnego** (ang. *Gateway Mobile Switching Center* — **GMSC**). I znów, aby nie mnożyć akronimów, centralę GMSC będziemy tu nazywać **domową centralą MSC**.
- Sieć wizytowana utrzymuje bazę danych znaną jako **rejestr stacji obcych** (ang. *visitor location register* — **VLR**). VLR zawiera wpis dla każdego użytkownika mobilnego, który *obecnie* znajduje się w części sieci obsługiwanej przez dany rejestr. Wpisy VLR pojawiają się więc i znikają, kiedy użytkownicy mobilni przyłączają się do sieci i opuszczają ją. Rejestr VLR jest zwykle zintegrowany z centralą przełączania mobilnego (MSC), która koordynuje zestawianie połączeń z siecią wizytowaną.

W praktyce sieć komórkowa operatora funkcjonuje jako sieć domowa dla jego abonentów oraz jako sieć wizytowana dla użytkowników, którzy mają podpisaną umowę z innym operatorem.

7.7.1. Routing rozmów z użytkownikiem mobilnym

Teraz możemy opisać, jak nawiązywane jest połączenie z mobilnym użytkownikiem GSM w sieci wizytowanej. Poniżej rozważymy prosty przykład; bardziej skomplikowane scenariusze są opisane w [Mouly 1992]. Poszczególne etapy tego procesu są zilustrowane na rysunku 7.30.



Rysunek 7.30. Nawiązywanie połączenia z użytkownikiem mobilnym: routing pośredni

1. Korespondent wybiera numer telefonu użytkownika mobilnego. Numer ten nie odnosi się do konkretnej linii telefonicznej ani lokalizacji (bądź co bądź, numer jest stały, a użytkownik mobilny!). Początkowe cyfry numeru wystarczają do globalnego zidentyfikowania sieci domowej telefonu. Połączenie jest kierowane od korespondenta przez publiczną sieć telefoniczną do domowej centrali MSC w sieci domowej telefonu. Jest to pierwsza część połączenia.
2. Domowa centrala MSC odbiera połączenie i bada rejestr HLR, aby ustalić lokalizację użytkownika mobilnego. W najprostszym przypadku HLR zwraca **numer roamingowy stacji mobilnej** (ang. *mobile station roaming number* — **MSRN**), który będziemy nazywać po prostu **numerem roamingowym**. Zwróćmy uwagę,

że różni się on od stałego numeru telefonu, który związany jest z siecią domową. Numer roamingowy jest efemeryczny: zostaje tymczasowo przydzielony telefonowi podczas rejestracji w sieci wizytowanej. Numer roamingowy pełni funkcję zbliżoną do adresu COA w Mobile IP i — podobnie jak COA — jest niewidoczny dla korespondenta i telefonu. Jeśli rejestr HLR nie zawiera numeru roamingowego, zwraca adres rejestru VLR w sieci obcej. W takim przypadku (niepokazanym na rysunku 7.30) domowa centrala MSC musi odpytać rejestr VLR, aby uzyskać numer roamingowy telefonu. Ale skąd bierze się numer roamingowy albo adres rejestru VLR w rejestrze HLR? Co dzieje się z tymi wartościami, kiedy użytkownik przenosi się do innej sieci wizytowanej? Tymi ważnymi kwestiami zajmiemy się niebawem.

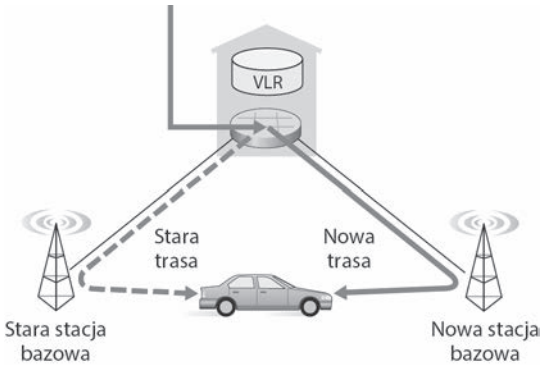
3. Po uzyskaniu numeru roamingowego domowa centrala MSC zestawia drugą część połączenia do centrali MSC w sieci wizytowanej. Proces łączenia dobiega końca; połączenie przebiega od korespondenta do domowej centrali MSC, stamtąd do centrali MSC w sieci wizytowanej, a następnie do stacji bazowej obsługującej użytkownika mobilnego.

Nie wyjaśniliśmy jeszcze, jak w kroku 2. HLR uzyskuje informacje o lokalizacji użytkownika mobilnego. Kiedy telefon komórkowy zostanie włączony albo przeniesiony do części sieci wizytowanej, która jest obsługiwana przez inny rejestr VLR, musi zarejestrować się w sieci wizytowanej. Odbywa się to poprzez wymianę komunikatów sygnałowych między telefonem a rejestrem VLR. Z kolei rejestr VLR wysyła komunikat o aktualizacji położenia telefonu do rejestru HLR. Komunikat ten zawiera albo numer roamingowy, pod którym można skontaktować się z telefonem, albo adres VLR (za pomocą którego później będzie można uzyskać numer telefonu). Podczas tej wymiany rejestr VLR uzyskuje również od rejestru HLR informacje o abonencie i ustala, jakie usługi sieć wizytowana powinna zaoferować użytkownikowi mobilnemu.

7.7.2. Transfery w GSM

Transfer (ang. *handoff*) występuje wtedy, gdy stacja mobilna przełącza się podczas rozmowy z jednej stacji bazowej na drugą. Jak pokazano na rysunku 7.31, początkowo (przed transferem) rozmowa jest kierowana do telefonu poprzez stację bazową (którą będziemy nazywać starą stacją bazową). Po transferze rozmowa zostaje skierowana do innej stacji bazowej (którą będziemy nazywać nową stacją bazową). Należy zwrócić uwagę, że transfer powoduje nie tylko nawiązanie łączności między telefonem a nową stacją bazową, ale również przekierowanie bieżącej rozmowy od punktu przełączania do tej stacji. Załóżmy najpierw, że stara i nowa stacja bazowa są obsługiwane przez tę samą centralę MSC i że przekierowanie zachodzi w tej centrali.

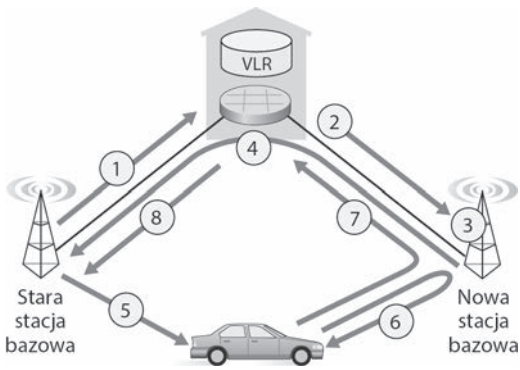
Transfer może być konieczny z kilku różnych powodów: (1) sygnał między bieżącą stacją bazową a telefonem osłabł do tego stopnia, że grozi przerwaniem rozmowy; (2) komórka jest przeciążona, ponieważ obsługuje wiele rozmów. Przeciążeniu można zaradzić przez transfer telefonów do mniej obciążonych, pobliskich komórek.



Rysunek 7.31. Scenariusz transferu między stacjami bazowymi obsługiwanymi przez tę samą centralę MSC

Kiedy telefon jest związany ze stacją bazową, okresowo mierzy siłę sygnału identyfikacyjnego tej stacji, a także innych stacji, które są w jego zasięgu. Pomiary te są przekazywane raz lub dwa razy na sekundę do bieżącej stacji bazowej telefonu. W sieci GSM transfer jest inicjowany przez starą stację bazową na podstawie tych pomiarów, bieżącej liczby telefonów w pobliskich komórkach i innych czynników [Mouly 1992]. Standard GSM nie określa specyficznych algorytmów, które pomagałyby stacji bazowej w podjęciu decyzji o transferze.

Na rysunku 7.32 zilustrowano proces transferu użytkownika mobilnego do nowej stacji bazowej.



Rysunek 7.32. Proces transferu między stacjami bazowymi obsługiwanymi przez tę samą centralę MSC

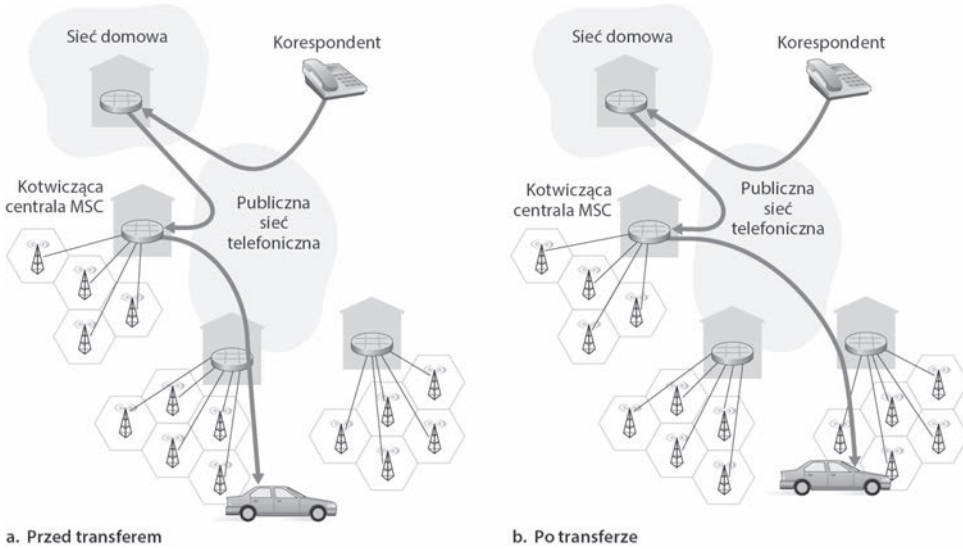
1. Stara stacja bazowa (SB) informuje centralę MSC, że konieczny jest transfer, i określa nową stację (albo zbiór możliwych stacji), do których ma zostać przetransferowany telefon.
2. Centrala MSC zapoczątkowuje zestawianie ścieżki do nowej stacji bazowej, przydzielając zasoby niezbędne do przenoszenia przekierowanej rozmowy, po czym sygnalizuje nowej stacji bazowej, że za chwilę nastąpi transfer.

3. Nowa stacja bazowa przydziela i aktywuje kanał radiowy na użytek telefonu.
4. Nowa stacja bazowa sygnalizuje wizytowanej centrali MSC i starej stacji bazowej, że ścieżka od tej centrali do nowej stacji została zestawiona i że należy poinformować telefon o bliskim transferze. Nowa stacja bazowa dostarcza wszystkich informacji, których telefon będzie potrzebował, aby się z nią związać.
5. Telefon jest informowany, że powinien przeprowadzić transfer. Zauważmy, że do tego momentu telefon był zupełnie „nieświadomy", iż sieć przygotowuje transfer (przydziela kanał w nowej stacji bazowej i zestawia ścieżkę między wizytowaną centralą MSC a nową stacją).
6. Telefon i nowa stacja bazowa wymieniają jeden lub więcej komunikatów, aby w pełni uaktywnić nowy kanał.
7. Telefon wysyła do nowej stacji bazowej komunikat o zakończeniu transferu, który jest przekazywany do centrali MSC. Centrala przekierowuje bieżącą rozmowę do telefonu poprzez nową stację bazową.
8. Zwalniane są zasoby wzdłuż ścieżki wiodącej przez starą stację bazową.

Aby zakończyć opis transferów, rozważmy, co się dzieje, kiedy telefon przechodzi do stacji bazowej związanej z *inną* centralą MSC, a co w przypadku, gdy transfer między centralami MSC zachodzi wielokrotnie. Jak pokazano na rysunku 7.33, w GSM zdefiniowano pojęcie **kotwiczącej centrali MSC**. Jest to centrala, która obsługiwała telefon w momencie rozpoczęcia rozmowy; nie zmienia się ona przez cały czas trwania rozmowy. Bez względu na liczbę transferów między centralami MSC rozmowa jest kierowana od centrali domowej do kotwiczącej, a stamtąd do centrali obsługującej komórkę, w której aktualnie znajduje się telefon. Kiedy telefon przejdzie do obszaru obsługiwanego przez nową centralę MSC, bieżąca rozmowa zostanie przekierowana od centrali kotwiczącej do centrali z nową stacją bazową. Zatem w każdym momencie między korespondentem a telefonem znajdują się najwyżej trzy centrale MSC (domowa, kotwicząca i wizytowana). Na rysunku 7.33 przedstawiono routing rozmowy między centralami MSC.

Rozwiązaniem alternatywnym dla pojedynczego przeskoku między centralą kotwiczącą a bieżącą byłoby tworzenie łańcucha ze wszystkich central MSC odwiedzonych przez telefon — stara centrala MSC po każdym transferze przekazywałaby rozmowę do nowej centrali. Takie łączenie central w łańcuch rzeczywiście może zachodzić w sieciach komórkowych IS-41, z opcjonalnym procesem minimalizacji ścieżki, który pozwala wyeliminować centrale MSC między centralą kotwiczącą a obecnie wizytowaną [Lin 2001].

Zakończmy omówienie mobilności w sieciach GSM porównaniem między systemami GSM a Mobile IP. Z tabeli 7.2 wynika, że choć sieci IP i komórkowe różnią się pod wieloma istotnymi względami, to mają wiele wspólnych elementów funkcjonalnych i cechują się podobnym ogólnym podejściem do obsługi mobilności.



Rysunek 7.33. Routing z wykorzystaniem kotwiczącej centrali MSC

Element GSM	Komentarz dotyczący elementu GSM	Element Mobile IP
System domowy	Sieć, do której należy stały numer telefonu abonenta.	Sieć domowa
Tranzytowa centrala przełączania mobilnego (domowa centrala MSC), rejestr stacji własnych (HLR)	Domowa centrala MSC: punkt kontaktu, który pozwala uzyskać adres użytkownika mobilnego. HLR: baza danych w systemie domowym, która zawiera stały numer telefonu oraz informacje o profilu abonenta i jego bieżącym położeniu.	Agent domowy
System wizytowany	Sieć (inna niż domowa), w której obecnie znajduje się użytkownik mobilny.	Sieć wizytowana
Wizytowana centrala przełączania mobilnego, rejestr stacji obcych (VLR)	Wizytowana centrala MSC: odpowiedzialna za zestawianie połączeń z węzłami mobilnymi znajdującymi się w obsługiwanych przez nią komórkach. VLR: baza danych z tymczasowymi wpisami, które zawierają informacje o każdym obcym abonencie.	Agent obcy
Numer roamingowy stacji mobilnej	Adres umożliwiający przekierowanie rozmowy od centrali domowej do wizytowanej, niewidoczny dla telefonu ani dla korespondenta.	Adres pośredni (COA)

Tabela 7.2. Cechy wspólne obsługi mobilności w systemach IP i GSM

7.8. Wpływ bezprzewodowości i mobilności na protokoły wyższych warstw

W niniejszym rozdziale pokazaliśmy, że sieci bezprzewodowe różnią się znacznie od swoich przewodowych odpowiedników zarówno w warstwie łącza (ze względu na takie cechy kanału bezprzewodowego jak zanikanie, propagacja wielościeżkowa i problem ukrytego terminalu), jak i w warstwie sieci (ze względu na to, że użytkownicy mogą zmieniać punkt przyłączenia do sieci). Czy jakieś istotne różnice występują również w warstwach transportowej i aplikacji? Intuicja podpowiada, że ewentualne różnice powinny być niewielkie, ponieważ zarówno w systemach przewodowych, jak i bezprzewodowych warstwa sieci oferuje wyższym warstwom ten sam model usługowy typu „rób, co w twojej mocy" (best-effort). Podobnie jeśli usługi warstwy transportu są realizowane z wykorzystaniem takich protokołów jak TCP i UDP, to warstwa aplikacji również powinna pozostać niezmieniona. Przekonanie to jest poniekąd uzasadnione — protokoły TCP i UDP mogą funkcjonować (i funkcjonują) w sieciach z łączami bezprzewodowymi. Z drugiej strony, protokoły transportowe — szczególnie TCP — mają bardzo odmienną wydajność w sieciach przewodowych i bezprzewodowych, i właśnie tutaj pojawiają się zasadnicze różnice. Zastanówmy się, dlaczego.

Przypomnijmy, że TCP retransmituje segment, który został utracony lub uszkodzony na ścieżce między nadawcą a odbiorcą. W łączności mobilnej utrata może nastąpić albo wskutek przeciążenia sieci (przepełnienia buforów w routerach), albo transferu (na przykład opóźnienia związanego z przekierowywaniem segmentów do nowego punktu przyłączeniowego stacji mobilnej). We wszystkich przypadkach potwierdzenie TCP wysyłane do nadawcy sygnalizuje tylko, że segment nie został odebrany prawidłowo: nadawca nie wie, czy segment utracono z powodu przeciążenia sieci, podczas transferu czy też wskutek wykrycia błędów bitowych. Niezależnie od przyczyny nadawca zawsze reaguje tak samo: retransmituje segment. Mechanizm kontroli przeciążenia TCP *również* zawsze reaguje tak samo: TCP zmniejsza okno przeciążenia w sposób opisany w podrozdziale 3.7. Bezwarunkowo zmniejszając okno przeciążenia, TCP zakłada, że utrata segmentu była spowodowana przeciążeniem, a nie uszkodzeniem danych lub transferem stacji mobilnej. W podrozdziale 7.2 stwierdziliśmy, że błędy bitowe występują znacznie częściej w sieciach bezprzewodowych niż w przewodowych. Kiedy pojawią się takie błędy albo gdy nastąpi transfer stacji, nie ma powodu, aby nadawca TCP zmniejszał swoje okno przeciążenia (a w konsekwencji szybkość transmisji). Rzeczywiście, może się zdarzyć i tak, że bufory routerów będą puste, a przepływ pakietów między punktami końcowymi nie będzie zakłócany przez przeciążenie.

Na początku lat 90. badacze uświadomili sobie, że (ze względu na wysoki współczynnik błędów na łączach bezprzewodowych oraz możliwość utraty segmentów podczas transferu stacji) mechanizmy kontroli przeciążenia TCP mogą być nieodpowiednie w środowisku bezprzewodowym. Do rozwiązania tego problemu można zastosować trzy ogólne podejścia:

- *Lokalne eliminowanie błędów.* W metodach tego typu błędy bitowe są poprawiane tam, gdzie występują (tzn. na łączu bezprzewodowym). Są to na przykład protokół 802.11 ARQ omówiony w podrozdziale 7.3 albo bardziej zaawansowane techniki wykorzystujące zarówno ARQ, jak i FEC [Ayanoglu 1995]
- *Wykrywanie łączy bezprzewodowych przez nadawcę TCP.* W metodach lokalnego eliminowania błędów nadawca TCP nie wie, że jego segmenty przechodzą przez łącze bezprzewodowe. Gdyby nadawca i odbiorca TCP zdawali sobie sprawę z istnienia łącza bezprzewodowego, mogliby odróżniać straty wynikające z przeciążenia w sieci przewodowej od uszkodzeń oraz strat na łączu bezprzewodowym i korzystać z mechanizmu kontroli przeciążenia tylko w tym pierwszym przypadku. [Balakrishnan 1997] analizuje różne typy TCP, zakładając, że systemy końcowe potrafią dokonać takiego rozróżnienia. [Liu 2003] bada techniki rozróżniania strat w przewodowych i bezprzewodowych segmentach ścieżki między systemami końcowymi.
- *Połączenia dzielone.* W podejściu opartym na podziale połączenia [Bakre 1995] połączenie międzywęzłowe pomiędzy użytkownikiem mobilnym i innym punktem końcowym jest dzielone na dwa połączenia warstwy transportowej. Jedno z nich prowadzi od hosta mobilnego do bezprzewodowego punktu dostępowego, natomiast drugie — od bezprzewodowego punktu dostępowego do innego punktu końcowego (zakładamy tu, że jest to host podłączony przewodowo). Połączenie międzywęzłowe powstaje więc przez scalenie części bezprzewodowej z przewodową. Warstwą transportową w segmencie bezprzewodowym może być standardowe połączenie TCP [Bakre 1995] lub specjalny protokół eliminowania błędów działający nad protokołem UDP. Źródło [Yavatkar 1994] przedstawia badania nad wykorzystaniem protokołów selektywnego powtarzania z warstwy transportowej w połączeniach bezprzewodowych. Pomiary opisane w pracy [Wei 2006] wskazują na to, że dzielone połączenia TCP są powszechnie stosowane w komórkowych sieciach transmisji danych i pozwalają znacznie usprawnić działanie takich sieci.

Nasz opis działania TCP na łączach bezprzewodowych był z konieczności krótki. Szczegółowe omówienie problemów i rozwiązań z tego obszaru znajdziesz w [Hanabali 2005; Leung 2006]. Warto sięgnąć do pozycji wymienionych w bibliografii, aby poznać więcej szczegółów dotyczących tej dziedziny badań.

Omówiwszy protokoły warstwy transportowej, rozważmy teraz wpływ bezprzewodowości i mobilności na protokoły warstwy aplikacji. Tutaj ważną kwestią jest względnie mała przepustowość łączy bezprzewodowych (zilustrowano to na rysunku 7.2). W rezultacie aplikacje, które operują na łączach bezprzewodowych (zwłaszcza komórkowych), muszą traktować pasmo jak towar deficytowy. Na przykład serwer WWW przesyłający dane do przeglądarki WWW w telefonie 4G zapewne nie będzie mógł dostarczyć jej tak wielu obrazów jak przeglądarce dysponującej połączeniem przewodowym. Choć łącza bezprzewodowe stwarzają pewne problemy w warstwie aplikacji, mobilność pozwala na różnorodne zastosowania, w których rozpoznawana jest lokalizacja albo kontekst [Chen 2000; Baldauf 2007]. Mówiąc ogólniej, połączenia bezprzewodowe

i mobilne odegrają kluczową rolę w realizacji wizji powszechnego dostępu do sieci [Weiser 1991]. Nie da się ukryć, że w niniejszym omówieniu wpływu sieci bezprzewodowych i mobilnych na aplikacje sieciowe i ich protokoły dotknęliśmy zaledwie wierzchołka góry lodowej!

7.9. Podsumowanie

Sieci bezprzewodowe i mobilne zrewolucjonizowały telefonię i wywierają coraz silniejszy wpływ na sieci komputerowe. Oferując swobodny, nieskrępowany dostęp do globalnej infrastruktury komunikacyjnej, nie tylko upowszechniają łączność sieciową, ale również pozwalają na oferowanie nowych, ekscytujących usług zależnych od lokalizacji. Ze względu na rosnące znaczenie sieci bezprzewodowych poświęciliśmy niniejszy rozdział zasadom i technologiom łączności oraz architekturom sieciowym obsługującym komunikację bezprzewodową i mobilną.

Zaczęliśmy od wprowadzenia do sieci bezprzewodowych i mobilnych, rozróżniając problemy stwarzane przez *bezprzewodową* naturę łączy komunikacyjnych oraz przez *mobilność*, która z niej wynika. Pozwoliło to lepiej wyizolować, zidentyfikować i opanować kluczowe pojęcia w każdej dziedzinie. Najpierw skupiliśmy się na komunikacji bezprzewodowej, badając właściwości łącza bezprzewodowego w podrozdziale 7.2. W podrozdziałach 7.3 i 7.4 zbadaliśmy aspekty warstwy łącza w bezprzewodowych sieciach lokalnych IEEE 802.11 (Wi-Fi), dwa standardy sieci PAN 802.15 (Bluetooth i Zigbee) oraz komórkowy dostęp do internetu za pomocą technologii 3G i 4G. Następnie zajęliśmy się kwestią mobilności. W podrozdziale 7.5 zidentyfikowaliśmy kilka postaci mobilności, z których każda stwarza inne wyzwania i wymaga innych rozwiązań. Rozważyliśmy problemy lokalizowania użytkownika mobilnego, routingu oraz „transferu" użytkownika, który dynamicznie porusza się między punktami przyłączeniowymi. W podrozdziałach 7.6 i 7.7 wyjaśniliśmy, jak te usługi są zrealizowane w standardzie Mobile IP i w komórkowych sieciach GSM. Wreszcie w podrozdziale 7.8 zbadaliśmy wpływ łączy bezprzewodowych i mobilności na protokoły warstwy transportowej oraz aplikacje sieciowe.

Choć poświęciliśmy cały rozdział na omówienie sieci bezprzewodowych i mobilnych, pełny opis tej ekscytującej, szybko rozwijającej się dziedziny zająłby całą książkę (lub nawet kilka tomów). Zachęcamy więc do bliższego zapoznania się z łącznością bezprzewodową poprzez lekturę książek wymienionych w bibliografii.

Pytania i problemy do rozwiązania

Rozdział 7. Pytania kontrolne

PODROZDZIAŁ 7.1

1. Co oznacza, że sieć bezprzewodowa działa w „trybie infrastrukturalnym”? Jeśli sieć nie funkcjonuje w trybie infrastrukturalnym, w jakim modelu działa? Jakie są różnice między tym podejściem i trybem infrastrukturalnym?
2. Jakie cztery typy sieci bezprzewodowych zidentyfikowaliśmy w klasyfikacji przedstawionej w podrozdziale 7.1? Z których z tych sieci korzystałeś?

PODROZDZIAŁ 7.2

3. Czym różnią się od siebie następujące typy problemów w kanale bezprzewodowym: utrata ścieżki, propagacja wielościeżkowa, interferencja z innymi źródłami?
4. Jakie dwie operacje może wykonać stacja bazowa, aby prawdopodobieństwo utraty transmitowanej ramki nie wzrosło przy oddalaniu się węzła mobilnego od danej stacji?

PODROZDZIAŁY 7.3 I 7.4

5. Opisz rolę ramek identyfikacyjnych w protokole 802.11.
6. Prawda czy fałsz: zanim stacja 802.11 prześle ramkę danych, musi najpierw wysłać ramkę RTS i odebrać odpowiednią ramkę CTS.
7. Dlaczego potwierdzenia stosuje się w sieciach 802.11, ale nie w przewodowym Ethernecie?
8. Prawda czy fałsz: Ethernet i 802.11 używają takiej samej struktury ramki.
9. Opisz działanie progu RTS.
10. Przypuśćmy, że w standardzie IEEE 802.11 ramki RTS i CTS miałyby tę samą długość co standardowe ramki DATA i ACK. Czy użycie ramek RTS i CTS przynosiłoby wówczas jakieś korzyści? Dlaczego (albo dlaczego nie)?
11. W punkcie 7.3.4 omówiono przykład mobilności w standardzie 802.11 w sytuacji, gdy stacja bezprzewodowa przechodzi do innej grupy BSS w obrębie tej samej podsieci. Kiedy punkty dostępowe są połączone przez przełącznik, punkt czasem musi wysłać ramkę ze sfałszowanym adresem MAC, aby przełącznik prawidłowo przekazywał ramki. Dlaczego?
12. Jakie różnice występują między węzłem nadrzędnym w sieciach Bluetooth i stacją bazową w sieciach 802.11?
13. Czym jest ramka nadrzędna w standardzie Zigbee (802.15.4)?
14. Jaką rolę pełni sieć bazowa w architekturze komórkowych sieci transmisji danych 3G?
15. Jaką rolę pełni kontroler RNC w architekturze komórkowych sieci transmisji danych 3G? Jakie jest zadanie kontrolera RNC w komórkowych sieciach głosowych?

16. Do czego służą jednostki eNodeB, MME, P-GW i S-GW w architekturze 4G?
17. Podaj trzy ważne różnice między architekturami sieci komórkowych 3G i 4G.

PODROZDZIAŁY 7.5 – 7.6

18. Czy jeśli węzeł ma bezprzewodowe połączenie z internetem, zawsze jest mobilny? Wyjaśnij odpowiedź. Załóżmy, że użytkownik porusza się wraz z laptopem po domu i zawsze łączy się z internetem za pomocą tego samego punktu dostępowego. Czy z perspektywy sieci jest to użytkownik mobilny? Wyjaśnij odpowiedź.
19. Czym różni się adres stały od adresu pośredniego (COA)? Jaka jednostka przypisuje adresy pośrednie?
20. Rozważmy połączenie TCP obsługiwane przez protokół Mobile IP. Prawda czy fałsz: w fazie nawiązywania połączenia TCP między korespondentem i hostem mobilnym informacje przechodzą przez sieć domową tego hosta, natomiast na etapie transmisji danych komunikacja odbywa się bezpośrednio między korespondentem i hostem mobilnym (z pominięciem sieci domowej).

PODROZDZIAŁ 7.7

21. Do czego służą rejestry HLR i VLR w sieciach GSM? Jakie elementy sieci Mobile IP są podobne do tych rejestrów?
22. Jaką funkcję w sieciach GSM pełni kotwicząca centrala MSC?

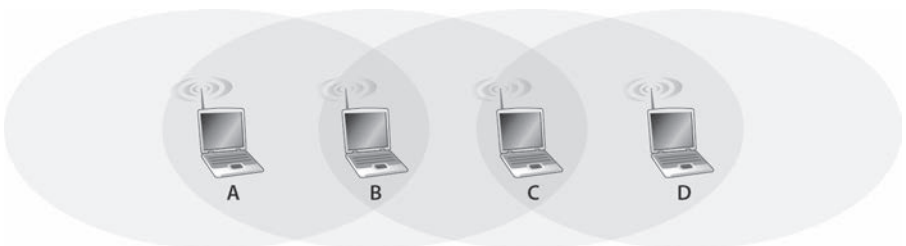
PODROZDZIAŁ 7.8

23. Jakie trzy podejścia można zastosować, aby pojedyncze łącze bezprzewodowe nie powodowało spadku wydajności międzywęzłowego połączenia TCP w warstwie transportowej?

Problemy

1. Rozważ przykład transmisji pojedynczego nadawcy CDMA z rysunku 7.5. Co pojawiłoby się na wyjściu nadawcy (dla dwóch pokazanych bitów danych), gdyby jego kod CDMA miał postać (1, –1, 1, –1, 1, –1, 1, –1)?
2. Rozważ nadawcę 2. z rysunku 7.6. Jaki jest sygnał $Z_{i,m}^{2}$ wysyłany przez niego do kanału (zanim zostanie dodany do sygnału nadawcy 1.)?
3. Przypuśćmy, że odbiorca z rysunku 7.6 chce odebrać dane wysłane przez nadawcę 2. Pokaż (poprzez obliczenie), że odbiorca rzeczywiście będzie mógł wyodrębnić dane nadawcy 2. z sygnału zbiorczego za pomocą kodu nadawcy 2.
4. Załóżmy, że jest dwóch nadawców i dwóch odbiorców. Podaj dwa przykładowe kody CDMA z wartościami 1 i –1 uniemożliwiające odbiorcom wyodrębnienie bitów wysłanych przez dwóch nadawców CDMA.

5. Przypuśćmy, że dwaj dostawcy ISP oferują dostęp Wi-Fi w kawiarence internetowej. Każdy ma własny punkt dostępowy i własny blok adresów IP.
   a. Załóżmy jeszcze, że przypadkowo obaj dostawcy skonfigurowali punkty dostępowe tak, iż operują one w kanale 11. Czy w takiej sytuacji protokół 802.11 całkowicie się załamie? Opisz, co się stanie, gdy dwie stacje — każda związana z innym dostawcą — spróbują jednocześnie wysłać dane.
   b. Załóżmy teraz, że jeden punkt dostępowy operuje w kanale 1., a drugi w kanale 11. Jak zmienią się odpowiedzi?
6. W etapie 4. protokołu CSMA/CA stacja, która pomyślnie prześle ramkę, w celu przesłania następnej ramki wznawia protokół od etapu 2., a nie 1. Dlaczego projektanci CSMA/CA nie chcieli, aby stacja natychmiast wysyłała następną ramkę (nawet jeśli wykryje, że kanał jest wolny)?
7. Przypuśćmy, że stacja 802.11b jest skonfigurowana w taki sposób, że zawsze rezerwuje kanał za pomocą sekwencji RTS-CTS. Załóżmy, że stacja ta chce przesłać 1000 bajtów danych, a inne stacje są w tym momencie bezczynne. Oblicz czas potrzebny na przesłanie ramki i odebranie potwierdzenia jako funkcję okresów SIFS i DIFS, ignorując opóźnienia transmisji i przyjmując brak błędów bitowych.
8. Rozważmy sytuację przedstawioną na rysunku 7.34. Przedstawia on cztery bezprzewodowe węzły — A, B, C i D. Zasięg radiowy tych węzłów ilustrują cieniowane owale. Wszystkie węzły korzystają z tej samej częstotliwości. Nadający węzeł A jest słyszany (odbierany) tylko przez węzeł B. Transmisję węzła B odbierają węzły A i C, dane z węzła C są słyszane przez węzły B i D, a transmisja z węzła D jest odbierana tylko przez węzeł C.



Rysunek 7.34. Scenariusz z problemu 8.

Załóżmy, że każdy węzeł ma nieskończoną liczbę komunikatów, które chce przesłać do każdego z pozostałych węzłów. Jeśli docelową lokalizacją wiadomości nie jest bezpośredni sąsiad, komunikat musi zostać przekazany dalej. Na przykład jeśli A chce przesłać wiadomość do D, najpierw musi przekazać komunikat do B, który następnie prześle go do C, a ten przekaże wiadomość do D. Czas jest podzielony na szczeliny, a transmisja komunikatu zajmuje dokładnie jedną szczelinę (tak jak w szczelinowej wersji protokołu ALOHA). W szczelinie węzeł może wykonać jedną z operacji: *(i)* przesłać komunikat; *(ii)* odebrać komunikat

(jeśli wysłana do niego została dokładnie jedna wiadomość); *(iii)* pozostać bezczynnym. Kiedy węzeł otrzyma jednocześnie więcej niż jedną transmisję, nastąpi kolizja i żaden z przesłanych komunikatów nie zostanie poprawnie odebrany. Można przyjąć, że błędy bitowe nie występują, dlatego jeśli wysłana zostanie dokładnie jedna wiadomość, węzły znajdujące się w zasięgu transmisji nadawcy odbiorą ją prawidłowo.

a. Przyjmijmy, że wszechwiedzący kontroler (znający stan każdego węzła w sieci) może poinstruować węzły, aby wykonały dowolną operację, jakiej zażąda — na przykład wysłały komunikat, odebrały go lub pozostały bezczynne. Jaka — przy założeniu obecności wspomnianego kontrolera — będzie maksymalna szybkość transferu komunikatów z danymi od węzła C do A, jeśli sieć nie ma do przekazania żadnych innych wiadomości między pozostałymi parami węzłów źródłowych i docelowych?

b. Załóżmy, że A wysyła wiadomości do B, a D — do C. Z jaką łączną maksymalną szybkością można przesyłać dane między tymi parami węzłów?

c. Przyjmijmy, że A wysyła komunikaty do B, a C — do D. Z jaką łączną maksymalną szybkością można przesyłać dane między tymi parami węzłów?

d. Przypuśćmy, że łącza bezprzewodowe zastąpiono przewodowymi. Udziel odpowiedzi na pytania z punktów od (a) do (c) dla sieci przewodowej.

e. Ponownie załóżmy, że sieć jest bezprzewodowa. Ponadto dla każdego komunikatu z danymi węzeł docelowy wysyła potwierdzenie ACK do węzła źródłowego (tak jak na przykład w protokole TCP). Przyjmijmy też, że przekazanie każdego takiego potwierdzenia zajmuje jedną szczelinę. Udziel odpowiedzi na pytania z punktów od (a) do (c) przy tych założeniach.

9. Opisz format ramki standardu 802.15.1 Bluetooth. Aby udzielić odpowiedzi, będziesz musiał poszukać informacji poza niniejszą książką. Czy coś w formacie ramki ogranicza liczbę aktywnych węzłów w sieciach 802.15.1 do ośmiu? Wyjaśnij odpowiedź.

10. Rozważmy wyidealizowane warunki działania sieci LTE. W kanale dosyłowym (rysunek 7.21) zastosowano szczeliny czasowe i używanych jest F częstotliwości. W sieci są cztery węzły — A, B, C i D — osiągalne ze stacji bazowej przez kanał dosyłowy działający z szybkością 10 Mb/s, 5 Mb/s, 2,5 Mb/s i 1 Mb/s. Te szybkości są osiągane przy założeniu, że stacja bazowa wykorzystuje wszystkie szczeliny czasu dostępne we wszystkich F częstotliwościach do transmisji danych do tylko jednego węzła. Stacja bazowa ma nieskończoną ilość danych do przesłania do każdego z węzłów i może przekazywać komunikaty do dowolnego węzła w dowolnej szczelinie z użyciem dowolnych z F częstotliwości.

a. Jaka jest maksymalna szybkość, z jaką stacja bazowa może przesyłać dane do węzłów przy założeniu, że w każdej szczelinie może przekazywać je do dowolnie wybranego węzła? Czy zaproponowane rozwiązanie jest uczciwe? Wyjaśnij odpowiedź i zdefiniuj, co rozumiesz przez „uczciwe” rozwiązanie.

b. W przypadku, gdy sieć musi być uczciwa (tzn. każdy węzeł ma otrzymać w każdej sekundzie równą ilość danych), jaka będzie średnia szybkość transmisji stacji bazowej (do wszystkich węzłów) w podramce dosyłowej? Wyjaśnij, w jaki sposób uzyskałeś odpowiedź.

c. Przyjmijmy inne kryterium uczciwości — każdy węzeł może otrzymać w podramce co najwyżej dwa razy tyle danych, co dowolny inny węzeł. Jaka będzie średnia szybkość transmisji stacji bazowej (do wszystkich węzłów) w podramce dosyłowej? Wyjaśnij, w jaki sposób uzyskałeś odpowiedź.

11. W podrozdziale 7.5 przedstawiono kilka rozwiązań, które umożliwiają użytkownikowi mobilnemu zachowanie adresu IP po przejściu do innej sieci obcej. Jedno z nich zakładało, że sieć obca będzie ogłaszać bardzo dokładną trasę do użytkownika mobilnego, a istniejąca infrastruktura routingu będzie rozpowszechniać tę informację w całej sieci. Jedną z wad tego rozwiązania miała być mała skalowalność. Załóżmy, że kiedy użytkownik mobilny przechodzi między sieciami, nowa sieć obca ogłasza dokładną trasę do użytkownika, a stara sieć obca wycofuje swoją trasę. Zastanów się, jak rozpowszechniane są informacje o routingu w algorytmie wektora odległości (zwłaszcza w przypadku routingu międzydomenowego w sieciach o zasięgu globalnym).

a. Czy inne routery będą mogły skierować datagramy do nowej sieci obcej, kiedy tylko rozpocznie ona ogłaszanie trasy?

b. Czy może się zdarzyć, że z perspektywy różnych routerów użytkownik mobilny będzie znajdował się w różnych sieciach?

c. Omów skalę czasu, w której routery wreszcie poznają trasę do użytkownika mobilnego.

12. Przypuśćmy, że korespondent z rysunku 7.23 jest mobilny. Naszkicuj dodatkową infrastrukturę warstwy sieci, która byłaby potrzebna do kierowania datagramów od pierwotnego użytkownika mobilnego do (obecnie mobilnego) korespondenta. Pokaż strukturę datagramów wymienianych między pierwotnym użytkownikiem mobilnym a (obecnie mobilnym) korespondentem, jak na rysunku 7.24.

13. Jaki wpływ ma mobilność na opóźnienia datagramów przesyłanych między źródłowym a docelowym węzłem Mobile IP?

14. Rozważ przykład łańcuchowego łączenia sieci podany na końcu punktu 7.7.2. Przypuśćmy, że użytkownik mobilny odwiedza sieci obce A, B i C, a korespondent inicjuje połączenie z użytkownikiem, kiedy ten znajduje się w sieci A. Podaj sekwencję komunikatów przesyłanych między agentami obcymi oraz między agentami obcymi a agentem domowym, kiedy użytkownik przechodzi z sieci A do B, a następnie do C. Teraz załóż, że sieci nie są łączone łańcuchowo, a korespondent (oraz agent domowy) muszą być jawnie informowani o zmianach adresu COA użytkownika mobilnego. Podaj sekwencję komunikatów wymienianych w drugim scenariuszu.

15. Rozważ dwa węzły mobilne w sieci obcej z agentem obcym. Czy w standardzie Mobile IP dwa węzły mobilne mogą używać tego samego adresu COA? Uzasadnij odpowiedź.
16. Przypomnij sobie, w jaki sposób rejestr VLR aktualizuje rejestr HLR, przekazując mu informacje o bieżącej lokalizacji telefonu. Jakie zalety i wady ma przekazywanie numeru MSRN, a jakie adresu VLR?

Deborah Estrin

Deborah Estrin jest profesorem informatyki na uczelni Cornell Tech w Nowym Jorku i profesorem zdrowia publicznego w Weill Cornell Medical College. Jest założycielką **Health Tech Hub** na Cornell Tech i współzałożycielką startupu non profit **Open mHealth**. W 1985 roku otrzymała dyplom doktora informatyki na M.I.T., a w 1980 roku tytuł BS na UC Berkeley. Początkowo jej badania dotyczyły głównie projektowania protokołów sieciowych, w tym rozsyłania grupowego i routingu między domenami. W 2002 Estrin założyła jednostkę Center for Embedded Networked Sensing (CENS) finansowaną przez NSF i działającą w ramach Science and Technology Center na uczelni UCLA. Organizacja CENS rozpoczęła nowe międzydyscyplinarne badania z zakresu systemów komputerowych, dotyczące dziedzin od sieci czujników monitorowania stanu środowiska po aplikacje typu participatory sensing związane z nauką obywatelską. Obecnie Estrin koncentruje się na rozwiązaniach mobilnych z obszaru zdrowia i na nurcie „small data", wykorzystującym wszechobecność urządzeń mobilnych i interakcji cyfrowych do opieki zdrowotnej i ratowania życia, co opisała w wystąpieniu TEDMED z 2013 roku. Profesor Estrin została wybrana na członka organizacji American Academy of Arts and Sciences (w 2007 roku) i National Academy of Engineering (w 2009 roku). Ponadto należy do organizacji IEEE, ACM i AAAS. Do jej wyróżnień należą: pierwsza nagroda ACM-W Athena Lecturer (2006), nagroda Women of Vision Award for Innovation instytutu Anita Borg Institute (2007), włączenie do galerii sław WITI (2008) oraz doktoraty honoris causa uczelni EPFL (2008) i Uppsala University (2011).

Prosimy Panią o opisanie kilku najbardziej ekscytujących projektów z Pani kariery. Jakie największe wyzwania były z nimi związane?

W połowie lat 90. miałam szczęście pracować w USC i ISI z osobami takimi jak Steve Deering, Mark Handley i Van Jacobson nad projektowaniem protokołów routingu w systemach z rozsyłaniem grupowym (zwłaszcza nad protokołem PIM). Starałam się przenieść wiele wniosków z projektowania architektury rozsyłania grupowego do dziedziny projektowania systemów monitorowania stanu środowiska, gdy po raz pierwszy zaczęłam na poważnie zajmować się praktyką i badaniami multidyscyplinarnymi. Łączenie innowacji z obszarów społecznych i technologicznych to coś, co bardzo interesuje mnie w dziedzinie, której badaniem

zajmuję się ostatnio, czyli w wykorzystywaniu urządzeń mobilnych w opiece zdrowotnej. Wyzwania związane z projektami z tego obszaru są równie zróżnicowane jak same dziedziny, przy czym cechą wspólną napotykanych trudności jest to, że trzeba pamiętać o odpowiedniej definicji problemu w trakcie przechodzenia od projektu do fazy realizacji, prototypów i badań pilotażowych. Żadne z realizowanych zadań nie dało się rozwiązać analitycznie — za pomocą symulacji lub nawet eksperymentów laboratoryjnych. Wymagało to od nas umiejętności zachowania przejrzystej architektury w obliczu skomplikowanych problemów i kontekstów. Wszystkie te prace wymagały też intensywnej współpracy.

Jakie zmiany i innowacje prognozuje Pani w obszarach sieci bezprzewodowych i mobilności?

We wcześniejszym wywiadzie stwierdziłam, że nigdy nie wierzyłam w prognozowanie przyszłości; spekulowałam jednak, że możemy oczekiwać końca zwykłych telefonów (które nie są programowalne oraz służą tylko do rozmów i pisania SMS-ów), ponieważ smartfony zyskują coraz to nowe możliwości i dla wielu osób stają się głównym narzędziem do korzystania z internetu. Po niewielu latach okazało się, że rzeczywiście to się dzieje. Prognozowałam też, że możemy spodziewać się dalszego upowszechniania się wbudowanych kart SIM, za pomocą których różnorodne urządzenia mogą komunikować się z niską szybkością z sieciami komórkowymi. Choć tak się stało, obserwujemy też wiele urządzeń i sprzętu z nurtu „internetu rzeczy", które używają wbudowanych modułów Wi-Fi oraz innych energooszczędnych i mających niewielki zasięg form łączności z lokalnymi koncentratorami. Nie przewidziałam wcześniej powstania dużego rynku na powszechnie dostępne urządzenia ubieralne. Spodziewam się, że do czasu ukazania się następnego wydania tej książki pojawi się wiele rozwiązań wykorzystujących dane z internetu rzeczy i inne ślady cyfrowe.

Jak widzi Pani przyszłość sieci i internetu?

Uważam, że warto spoglądać zarówno wstecz, jak i naprzód. Wcześniej stwierdziłam, że prace nad technologiami NDN i SDN doprowadzą do powstania bogatszej infrastruktury, łatwiejszej w zarządzaniu i modyfikowaniu, a także że są one zgodne z przenoszeniem roli architektury w górę stosu protokołów. U zarania internetu architekturę tworzyło się w warstwie 4. i niższych, a aplikacje były bardziej monolityczne i działały powyżej tych warstw. Obecnie dane i analityka są ważniejsze od transportu. Sieci SDN (bardzo się cieszę, że zostały one opisane w niniejszym siódmym wydaniu tej książki) upowszechniły się w większym stopniu, niż podejrzewałam. Jednak gdy przyjrzeć się wyższym warstwom stosu protokołów, widać, że najpopularniejsze aplikacje (czy to aplikacje mobilne, czy duże systemy takie jak Facebook) coraz częściej działają w izolowanych enklawach. Rozwój nauki o danych i nurtu Big Data mogą być bodźcami do przeniesienia takich aplikacji poza silos z uwagi na wartość, jaką daje łączenie ich z innymi aplikacjami i platformami.

Kto był dla Pani inspiracją w pracy zawodowej?

Na myśl przychodzą mi trzy osoby. Po pierwsze, Dave Clark, „sekretny sos" i cichy bohater społeczności internetowej. W okresie powstawania internetu miałam to szczęście, że mogłam obserwować, jak Clark działa jako „fundamentalna podstawa" w organizacji IAB i jednostkach zarządzających internetem. Był kapłanem trudnych kompromisów i działającego kodu. Po drugie, Scott Shenker — za jego błyskotliwość, integralność i wytrwałość. Dążę (choć rzadko z powodzeniem) do tego, aby równie jasno jak on definiować problemy i rozwiązania. Jest pierwszą osobą, do jakiej

zawsze piszę e-maile z prośbą o wskazówki w drobnych i poważnych kwestiach. Po trzecie, moja siostra, Judy Estrin, która była na tyle kreatywna i odważna, aby wprowadzać pomysły i koncepcje na rynek. Bez podobnych jej osób świat technologii internetowych nigdy nie zmieniłby naszego życia.

Co radzi Pani studentom zainteresowanym karierą w branży informatyki i sieci?

Przede wszystkim warto zbudować solidne podstawy w ramach pracy akademickiej i połączyć je z możliwie wieloma doświadczeniami z pracy praktycznej. W trakcie szukania środowiska pracy dobrze jest szukać możliwości w naprawdę ważnych dla was dziedzinach i gdzie dostępne są inteligentne zespoły, w których można się czegoś nauczyć.

Skorowidz

A

B

C

D

E

F

G

H

I

J

K

L

Ł

M

N

O

P

R

S

Ś

T

U

V

W

Y

Z

Ż

Aplikacje sieciowe, protokoły, internet — wszystko, co musisz wiedzieć.

Zagadnienia związane z sieciami komputerowymi są wyjątkowo złożone. Opanowanie tej tematyki wymaga przyswojenia wielu pojęć oraz zrozumienia licznych protokołów i technologii, które dodatkowo są ze sobą powiązane w zawiły sposób. Konieczne jest również uwzględnienie gwałtownego rozwoju technologii sieciowych i rosnącej złożoności nowych aplikacji. Aby poradzić sobie z tymi zagadnieniami, konieczne jest całościowe ujęcie tematyki sieci komputerowych.

Ta książka jest siódmym, zaktualizowanym i ulepszonym wydaniem znakomitego podręcznika. Zrozumienie zagadnień ułatwia wykorzystana przez autorów metoda omawiania zagadnień „od góry do dołu", od ogółu do szczegółu, a więc prezentują pierwszą warstwę aplikacji, nastepnie kolejne, niższe warstwy — aż do warstwy fizycznej. W książce szczególnie dużo miejsca poświęcono wiedzy o działaniu internetu, jego architekturze i protokołach. Zaprezentowano tu także fundamentalne zasady budowy i działania sieci oraz informacje o podstawowych problemach sieciowych i metodach ich rozwiązywania. W efekcie ten podręcznik pozwala na zdobycie gruntownej wiedzy, umożliwiającej zrozumienie niemal każdej technologii sieciowej.

W tej książce między innymi:

- warstwowość architektury sieciowej
- warstwa aplikacji, w tym strumieniowanie i sieci CDN
- działanie routerów i sterowanie logiką warstwy sieciowej
- bezpieczeństwo sieci
- administrowanie siecią

Dr Jim Kurose pracuje w US National Science Foundation. Wcześniej był redaktorem naczelnym „IEEE Transactions on Communications" i „IEEE/ACM Transactions on Networking". Jest członkiem IEEE i ACM. Zajmuje się protokołami, architekturą sieciową, pomiarami sieciowymi i komunikacją multimedialną.
Prof. Keith Ross jest dziekanem Instytutu Inżynierii i Informatyki uczelni NYU Shanghai. Wcześniej pracował na Uniwersytecie Pensylwanii. Zajmuje się zagadnieniami prywatności, sieci społecznościowych, sieci P2P, pomiarów sieciowych, sieci dystrybucji treści i modelowania stochastycznego.

Helion
helion.pl
HELION S.A.
ul. Kościuszki 1c
44-100 Gliwice
tel.: 32 230 98 63
helion@helion.pl

KOD KORZYŚCI
Sięgnij po więcej! ▶

ISBN 978-83-289-2935-7

9 788328 929357

Cena: 179,00 zł

PEARSON ALWAYS LEARNING