

Rozdział XX

Metody unikania i wykrywania kolizji dla sieci ad-hoc

Bartłomiej ZIELIŃSKI
Politechnika Śląska, Instytut Informatyki
bmw@zeus.polsl.gliwice.pl

Streszczenie

Opisano właściwości łącza bezprzewodowego w sieciach ad-hoc z punktu widzenia rywalizacyjnych protokołów dostępu do łącza. Opisano metody unikania i wykrywania kolizji, które mogą podnieść wydajność sieci ad-hoc.

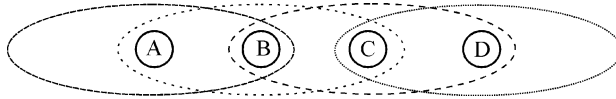
1. Wprowadzenie

Protokoły dostępu do łącza są jednym z najważniejszych aspektów projektowania sieci komputerowych. Jeżeli bowiem protokół taki jest źle dobrany do aplikacji, może nie zapewniać wymaganych parametrów sieci [1].

W sieciach bezprzewodowych występuje szereg zjawisk, nie znanych w sieciach przewodowych, a mających kluczowe znaczenie dla wydajności i stabilności protokołów dostępu do łącza. Ich wpływ jest szczególnie widoczny w sieciach ad-hoc. Sieci te charakteryzują się szybko zmieniającą się i nieregularną strukturą, a także brakiem nadrzędnej stacji sterującej, koordynującej pracę pozostałych stacji [2]. W tej sytuacji stacje sieci ad-hoc mogą posługiwać się niemal jedynie rywalizacyjnymi protokołami dostępu do łącza. Można wprawdzie zastosować protokół z przekazywaniem żetonu, jednak większość czasu pracy sieci będzie wówczas stracona na rekonfigurowanie obiegu żetonu, spowodowane częstymi zmianami struktury sieci.

2. Charakterystyka łącza w sieci ad-hoc

Łącze bezprzewodowe w sieciach ad-hoc stwarza wiele problemów, które nie występują w sieciach przewodowych. Wynikają one głównie z faktu stopniowego zmniejszania się mocy sygnału wraz ze wzrostem odległości. W związku z tym niektóre stacje w sieci ad-hoc mogą znajdować się poza swoim zasięgiem, pomimo że należą do jednej sieci (rys. 1). Dodatkowe utrudnienia wynikają ze zmienności poziomu mocy sygnału, będącej skutkiem m. in. propagacji wielodrogowej.



Rys. 1. Przykładowa bezprzewodowa sieć ad-hoc

Z punktu widzenia protokołów dostępu do łącza najistotniejsze są następujące zjawiska, charakterystyczne dla bezprzewodowych sieci ad-hoc:

- zjawisko ukrytej stacji,
- zjawisko odkrytej stacji,
- efekt przechwytywania,
- interferencje.

2.1. Zjawisko ukrytej stacji

Zjawisko ukrytej stacji (ang. *hidden terminal*) występuje w sieciach, w których pewne stacje znajdują się poza swoim zasięgiem. Stacja jest ukryta, jeżeli znajduje się w zasięgu odbiornika informacji, ale poza zasięgiem nadajnika [3]. W związku z tym nie może ona prawidłowo określić stanu kanału przy użyciu wykrywania nośnej.

Zjawisko ukrytej stacji pokazane jest na rys. 2. Stacja A nadaje do stacji B. Transmisja ta nie zostaje wykryta przez stację C, która zakłada wobec tego, że łącze jest wolne i rozpoczyna transmisję do stacji D. Oczywiście powoduje to w stacji B kolizję między ramkami, pochodzącymi ze stacji A i C. Tak więc zjawisko ukrytej stacji zmniejsza przepustowość łącza przez powodowanie kolizji i konieczność retransmisji.



Rys. 2. Zjawisko ukrytej stacji

W sieciach, w których stosuje się unikanie kolizji przez wymianę informacji sterującej, możliwe jest rozróżnienie ukrytego odbiornika i nadajnika. Ukryty nadajnik jest stacją ukrytą, która ma dane do wysłania (jak stacja C na rys. 2). Ukryty odbiornik jest natomiast adresatem informacji, jednak nie może odpowiedzieć na próbę nawiązania łączności (np. przez wymianę ramek sterujących RTS-CTS), gdyż spowodowałoby to kolizję. Brak odpowiedzi z tej stacji można interpretować na kilka sposobów, np.:

- ramka RTS uległa kolizji lub przekłamaniu,
- ramka CTS uległa kolizji lub przekłamaniu,
- adresat wstrzymuje transmisję,
- adresat jest wyłączony lub znajduje się poza zasięgiem sieci.

W każdym z wymienionych przypadków reakcja stacji usiłującej nawiązać łączność powinna być inna, stąd też konieczność prawidłowego rozpoznania sytuacji, w jakiej znajduje się ukryty odbiornik.

2.2. Zjawisko odkrytej stacji

Zjawisko odkrytej stacji także występuje w sieciach, w których pewne stacje znajdują się poza swoim zasięgiem. Stacja jest odkryta, jeżeli znajduje się w zasięgu nadajnika infor-

macji, ale poza zasięgiem odbiornika [3]. W związku z tym nie może ona prawidłowo określić stanu kanału przy użyciu wykrywania nośnej.

Zjawisko odkrytej stacji pokazane jest na rys. 3. Stacja B nadaje do stacji A. Transmisja ta zostaje wykryta przez stację C, która zakłada wobec tego, że łącze jest zajęte i nie rozpoczyna transmisji do stacji D. Transmisja ta nie zakłócałaby jednak przesyłu ze stacji B do A, ponieważ C i A są poza swoim zasięgiem. Tak więc zjawisko odkrytej stacji zmniejsza przepustowość łącza przez niepotrzebne wstrzymywanie transmisji możliwych do przeprowadzenia.



Rys. 3. Zjawisko odkrytej stacji

Również w ramach zjawiska odkrytej stacji można wyróżnić odkryte nadajniki i odbiorniki.

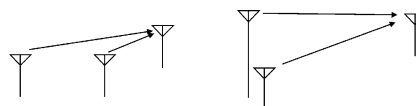
2.3. Efekt przechwytywania

Efekt przechwytywania (ang. *capture effect*) występuje, gdy sygnały pochodzące z różnych nadajników docierają do odbiornika z różną mocą [4]. Może się tak zdarzyć w następujących przypadkach:

- gdy nadajniki o równej mocy znajdują się w różnych odległościach od odbiornika,
- gdy nadajniki o różnych mocach znajdują się w zbliżonej odległości od odbiornika.

W rezultacie, gdy ramki pochodzące z różnych nadajników nałożą się na siebie (ulegną kolizji), jedna z nich – mianowicie ta o największej mocy sygnału – może zostać odebrana prawidłowo, podczas gdy wszystkie pozostałe zostają zagłuszone. Efekt ten prowadzi zatem do podwyższenia stopnia wykorzystania kanału transmisyjnego.

Efekt przechwytywania został przedstawiony na rys. 4.



Rys. 4. Efekt przechwytywania

Skutkiem ubocznym efektu przechwytywania jest niemożność wykrywania kolizji metodą prowadzenia nasłuchu łącza podczas nadawania (jak w sieci Ethernet), ponieważ własny nadajnik zagłusza sygnały pochodzące ze wszystkich pozostałych stacji w sieci.

2.4. Interferencje

Stacja zakłócająca (powodująca interferencje) to taka, która znajduje się poza zasięgiem zarówno nadajnika, jak i odbiornika, jednak wystarczająco blisko, aby zakłócać transmisję między nimi [5]. W przeciwieństwie jednak do przypadku stacji ukrytej i odkrytej nie ma możliwości powiadomienia takiej stacji o możliwości zakłócania innej transmisji.

3. Metody unikania i wykrywania kolizji

W sieciach przewodowych unikanie kolizji metodą wykrywania nośnej jest wystarczające, ponieważ każda stacja może wykryć transmisję z dowolnej innej stacji. Po dodaniu mechanizmu wykrywania kolizji protokół jest wydajny i zachowuje stabilność wystarczającą dla wielu aplikacji nawet przy bardzo wysokich obciążeniach. Niestety, w sieciach bezprzewodowych wykrywanie nośnej nie wystarcza, z powodu występowania zjawisk ukrytej i odkrytej stacji. Co gorsza, wykrywanie kolizji w większości przypadków jest praktycznie niemożliwe. Istnieje zatem potrzeba opracowania nowych metod unikania kolizji, które będą bardziej wydajne niż wykrywanie nośnej w obecności stacji ukrytych i odkrytych.

W sieciach ad-hoc, prócz wykrywania nośnej, można stosować następujące metody unikania kolizji:

- wykrywanie tonu zajętości,
- poprzedzanie przesyłu danych wymianą informacji sterującej,
- wydzielenie osobnego kanału sterującego.

Niektóre z wymienionych metod mogą być połączone z wykrywaniem nośnej, aby podnieść skuteczność unikania kolizji. Możliwe jest także łączenie tych metod.

3.1. Wykrywanie tonu zajętości

Wykrywanie tonu zajętości może zastąpić wykrywanie nośnej. Pasma jest podzielone na dwa kanały [6]:

- kanał komunikatów, używany dla transmisji danych i zajmujący większość pasma,
- kanał zajętości, używany dla sygnalizacji stanu łącza, relatywnie wąski.

Jeżeli stacja ma ramkę gotową do wysłania, przed rozpoczęciem transmisji musi sprawdzić stan kanału. Obecność tonu zajętości świadczy o zajętości łącza, transmisję należy zatem wstrzymać. W przeciwnym przypadku ramkę można natychmiast wysłać.

Sygnał zajętości jest zazwyczaj falą sinusoidalną, którą można wytworzyć na kilka sposobów:

- przez każdą stację, odbierającą dane z kanału komunikatów,
- wyłącznie przez adresata informacji,
- początkowo przez każdą stację wykrywającą transmisję, po rozpoznaniu adresów zaś – wyłącznie przez adresata.

Pierwsza z metod [6] jest najprostsza i bardzo efektywna w zakresie zmniejszenia liczby stacji ukrytych. Niestety, słabością jej jest znaczny i niepotrzebny wzrost liczby stacji odkrytych. Można powiedzieć, że obszar zajęty przez określoną transmisję jest znacznie większy, niż potrzeba.

Aby uniknąć zwiększania liczby stacji odkrytych, wytwarzanie tonu zajętości można ograniczyć do adresata przesyłanej informacji [7]. W ten sposób ramka chroniona jest przed kolizją w pobliżu odbiornika, a więc w jedynym miejscu, w którym naprawdę powinna być chroniona (nie ma potrzeby ochrony w pobliżu nadajnika lub stacji, do których ramka nie jest adresowana). Ta metoda także jest prosta, jednak nie chroni ramki

do chwili rozpoznania jej adresu przeznaczenia. Tak więc na początku transmisji ramka jest narażona na kolizje, wynikające z braku sygnału ochronnego.

Trzecia metoda jest kombinacją opisanych powyżej metod. Charakteryzuje się ona występowaniem dwóch różnych tonów zajętości [8]. Pierwszy z nich jest wytwarzany przez wszystkie stacje znajdujące się w zasięgu nadajnika, gdy tylko zostanie wykryta transmisja w kanale danych. Stan ten trwa do chwili rozpoznania adresu przeznaczenia. Wówczas adresat informacji rozpoczyna wysyłanie drugiego tonu zajętości, którego obecność świadczy o braku kolizji. Pozostałe stacje wyłączają generowanie tonu zajętości, gdy tylko okaże się, że nie są one adresatami przesyłanej informacji.

Warto zauważyć, że ostatnia metoda ma pewne cechy wykrywania kolizji [8]. Jeżeli ramkę podzieli się na preambułę (zawierającą adres docelowy) i rdzeń (zawierający dane), można wykryć kolizję przed rozpoczęciem wysyłania danych. Po wysłaniu preambuły nadajnik czeka na otrzymanie tonu zajętości od adresata. Brak takiego tonu oznacza kolizję lub jakikolwiek inny błąd podczas transmisji preambuły, nie ma zatem sensu wysyłania pozostałej części ramki. Z tego punktu widzenia ten sposób jest podobny do wymiany ramek sterujących.

Pewną wadą każdej odmiany wykrywania tonu zajętości jest możliwość łatwego zablokowania sieci przez zagłuszanie kanału zajętości. Ponadto, metody te nie są łatwe do zaimplementowania w praktyce, ponieważ wymagają dwu- lub trójkanałowych układów nadawczo-odbiorczych.

3.2. Wymiana ramek sterujących

Jako że wykrywanie nośnej nie jest wydajną metodą unikania kolizji w obecności stacji ukrytych lub odkrytych, a wykrywanie tonu zajętości jest kosztowne w implementacji, można poprzedzić transmisję danych wymianą informacji sterującej [9]. Można uznać takie podejście za pewną formę jednokanałowej implementacji wykrywania tonu zajętości, przy czym ton zajętości jest tu zastąpiony przez ramki sterujące, przesyłane w tym samym kanale co dane.

Gdy stacja ma ramkę do wysłania, wysyła najpierw ramkę sterującą, zwaną RTS (ang. *Request To Send*), adresowaną do odbiornika informacji. Ramka powinna zawierać informację dotyczącą długości ramki danych. Jeżeli adresat odbierze tę ramkę poprawnie, odpowiada nadawcy ramką CTS (ang. *Clear To Send*). Ramka ta również powinna zawierać informację o długości ramki danych. Gdy nadajnik poprawnie odbierze ramkę CTS, można przyjąć, że łącze jest zarezerwowane na potrzeby transmisji między tymi dwiema stacjami przez czas wynikający z zawartości ramek RTS i CTS.

Oczywiście, ramki RTS i CTS mogą być odbierane także przez inne stacje znajdujące się w otoczeniu nadajnika i odbiornika. Stacje te można podzielić na następujące grupy:

- stacje odbierające tylko ramkę RTS,
- stacje odbierające tylko ramkę CTS,
- stacje odbierające obie ramki.

Jeżeli stacja odbiera tylko ramkę RTS, znajduje się w zasięgu nadajnika, ale poza zasięgiem odbiornika; można więc przyjąć, iż jest to stacja odkryta. Z drugiej strony, jeśli stacja odbiera tylko ramkę CTS, znajduje się w zasięgu nadajnika, ale poza zasięgiem

nadajnika, jest więc stacją ukrytą. Stacja, odbierająca zarówno RTS, jak i CTS, nie jest ani odkryta, ani ukryta i mogłaby poprawnie rozpoznać stan kanału używając wyłącznie wykrywania nośnej.

Można wykazać [11], że, aby poprawnie chronić ramki danych przed kolizjami spowodowanymi przez stacje ukryte lub odkryte, długość ramki CTS powinna być większa niż RTS.

W niektórych przypadkach celowe jest użycie dodatkowych ramek sterujących. Na przykład, po poprawnej negocjacji RTS-CTS, nadajnik może wysłać ramkę DS (ang. *Data Sending*) [3]. Jest to szczególnie użyteczne, gdy z pewnych powodów odbiorca informacji nie może przyjąć ramki o długości zaproponowanej przez nadawcę. W takim przypadku odsyła w ramce CTS inną długość, jednak informacja ta może nie dotrzeć do wszystkich stacji sąsiadujących z nadawcą (np. stacji odkrytych). Ramka DS pozwala zatem przekazać im nowe ustalenia dotyczące czasu rezerwacji łącza.

Inną techniką godną uwagi jest wytwarzanie potwierdzeń na poziomie podwarstwy dostępu do łącza [3]. Oczywiście wyższe warstwy również mogą wykryć nieprawidłowości transmisji, jednak warstwa dostępu może uczynić to szybciej, zmniejszając czas potrzebny na wykrycie kolizji lub innego błędu oraz zażądanie retransmisji.

Wymiana ramek sterujących jest względnie prostą i wydajną metodą unikania kolizji w obecności stacji ukrytych lub odkrytych. Tym niemniej zastosowana w pojedynkę nie zapewnia wystarczającej skuteczności ze względu na ryzyko kolizji między ramkami sterującymi [10]. Kolizje mogą także wystąpić w środowiskach mobilnych, np. gdy nowa stacja zbliża się do stacji zaangażowanych w transmisję. Stacja taka mogła być poza zasięgiem sieci podczas negocjacji RTS-CTS, ale podczas przesyłu danych znajduje się już w zasięgu. Nie posiadając zatem wiedzy o rezerwacji łącza, może zakłócić przebiegającą transmisję danych. Zjawisko takie można wyeliminować, wspierając wymianę ramek sterujących albo wykrywaniem nośnej [10], albo wykrywaniem tonu zajętości [11].

3.3. Oddzielny kanał sterujący

Zarówno wykrywanie tonu zajętości, jak i wymiana ramek sterujących rozwiązują jedynie problem ukrytego nadajnika i odkrytego nadajnika [5]. Aby rozwiązać także problem ukrytego lub odkrytego odbiornika, należy zapewnić, by:

- ukryty odbiornik mógł przesłać informację o wstrzymaniu transmisji,
- odkryty odbiornik mógł odebrać informację sterującą nawet, gdy odbiera dane.

W obu przypadkach niezbędne jest wprowadzenie dodatkowego kanału sterującego. Pozwala to na uniknięcie kolizji między ramkami danych i sterujących. Ukryty odbiornik może zatem odpowiedzieć na ramkę RTS, nie zakłócając transmisji. Podobnie, odkryty odbiornik może odebrać dowolną informację sterującą i odpowiedzieć na nią.

Warto zauważyć, że zwiększenie zasięgu transmisji polepsza warunki przesyłu informacji. Stacje, które znajdują się poza zasięgiem zarówno odbiornika, jak i nadajnika informacji, mogą znajdować się dostatecznie blisko, aby zakłócać transmisję. Zwiększenie zasięgu kanału sterującego zmienia te stacje w odkryte lub ukryte, tak więc można już poinformować je o stanie łącza.

Podobnie do wykrywania tonu zajętości, osobny kanał sterujący – szczególnie gdy ma większy zasięg – jest względnie trudny w implementacji i wymaga bardziej złożonego (i droższego) układu transmisyjnego. Tak więc, mimo swoich zalet, metoda ta nie jest praktycznie stosowana.

3.4. Wykrywanie kolizji

W bezprzewodowych sieciach ad-hoc najczęściej nie jest możliwe wykrywanie kolizji w sposób znany np. z sieci Ethernet. Dzieje się tak dlatego, że stosowane urządzenia nadawczo-odbiorcze uniemożliwiają podsłuch łącza w czasie nadawania. Gdyby zresztą nawet było to możliwe, ze względu na efekt przechwytywania, sygnał z własnego nadajnika zagłuszałby wszelkie inne sygnały docierające do odbiornika. Tak więc jedynym sposobem na wykrycie kolizji jest potwierdzenie poprawnego odbioru ramki. Brak takiego potwierdzenia może świadczyć jednak nie tylko o kolizji, lecz także o innym błędzie transmisji.

W przypadku, gdy konieczne jest wykrywanie kolizji już w czasie transmisji ramki, można zastosować metodę polegającą na chwilowej przerwie w nadawaniu [13]. Metoda ta może działać w każdym kanale półdupleksowym, wliczając w to kanały radiowe lub optyczne. Ocena stanu kanału odbywa się przy tym na zasadzie wykrywania nośnej.

Po stwierdzeniu, że kanał nie jest zajęty, stacja może rozpocząć transmisję ramki, ale po losowo wybranym czasie przerywa na chwilę i ponownie bada stan kanału. Brak nośnej w tym czasie oznacza brak kolizji i stacja może dokończyć przesyłanie ramki. Natomiast obecność nośnej świadczy o kolizji ramek pochodzących z różnych stacji. Stacja, która wykryła kolizję, kontynuuje transmisję jeszcze przez pewien czas, aby inne stacje także mogły wykryć kolizję. Mechanizm ten nie gwarantuje wykrycia wszystkich kolizji, ponieważ dwie stacje mogą przerwać transmisję i badać stan kanału w tej samej chwili [14].

W niektórych przypadkach, szczególnie gdy czasy przełączania między nadawaniem a odbiorem są większe niż czas propagacji sygnału w kanale, zagłuszanie kanału przez nadajnik może nie wystarczyć, by poinformować wszystkie stacje o wystąpieniu kolizji [15]. Wówczas zagłuszanie powinno być także wykonane przez stacje nie zaangażowane bezpośrednio w transmisję.

Wykrywanie kolizji w sposób znany z sieci Ethernet jest możliwe w niektórych sieciach optycznych, opartych na promieniowaniu rozproszonym [16]. Warunkiem koniecznym jest brak bezpośredniej drogi sygnału z nadajnika do odbiornika w ramach stacji. Do odbiornika docierają wówczas jedynie sygnały odbite, dzięki czemu moc sygnałów pochodzących z różnych stacji jest porównywalna i efekt przechwytywania nie występuje.

4. Zakończenie

Spośród przedstawionych metod unikania i wykrywania kolizji praktyczne zastosowania znalazły tylko niektóre. Wpływ na to ma z pewnością stopień trudności ich implementacji w układzie transmisyjnym, od której zależy koszt takiego układu. Najczęściej stosowane protokoły rywalizacyjne wykorzystują różne odmiany wymiany informacji sterującej połączone z wykrywaniem nośnej sygnału.

Na efektywność protokołu wpływa nie tylko jego architektura, ale także szereg parametrów transmisji, zależnych od środowiska pracy sieci. W pewnych przypadkach wystarczy stosowanie prostych i teoretycznie mało efektywnych protokołów, ponieważ nakład związany z implementacją złożonych mechanizmów unikania lub wykrywania kolizji nie zwróci się w postaci przyrostu wydajności sieci, jeżeli jej obciążenie jest znikome.

LITERATURA

1. Tannenbaum A. S.: *Computer Networks (Fourth Edition)*. Prentice Hall, 2003.
2. Nowicki K., Woźniak J.: *Przewodowe i bezprzewodowe sieci LAN*. Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2002.
3. Bhargavan V., Demers A., Shenker S., Zhang L.: *MACAW: A Media Access Protocol for Wireless LAN's*. Proceedings of the Conference on Communications Architectures, Protocols and Applications SIGCOM '94, London, 1994, s. 212-225.
4. Metzner J.: *On Improving Utilization in Aloha Networks*. IEEE Transactions on Communications, Vol. COM-24, No. 4, Apr. 1976, s. 447-448.
5. Bhargavan V.: *A New Protocol for Medium Access in Wireless Packet Networks*. http://www.janet.ucla.edu/~mbs/courses/1997s/ee298-7/papers/bhargavan_BAPU.ps.
6. Tobagi F. A., Kleinrock L.: *Packet Switching in Radio Channels: Part II – The Hidden Terminal Problem in Carrier Sense Multiple-Access and the Busy-Tone Solution*. IEEE Transactions of Communications, Vol. COM-23, No. 12, Dec. 1975, s. 1417-1433.
7. Wu C., Li V.: *Receiver-Initiated Busy Tone Multiple Access in Packet Radio Networks*. ACM SIGCOMM Computer Communication Review, Vol. 17, No. 5, Oct/Nov 1987, s. 336-342.
8. Gummalla A., Limb J.: *Wireless Collision Detect (WCD): Multiple Access with Receiver Initiated Feedback and Carrier Detect Signal*. Proc. IEEE ICC'00, vol. 1, 2000, s. 397-401.
9. Karn P.: *MACA – A New Channel Access Method for Packet Radio*. Proceedings of ARRL/CRRL Amateur Radio 9th Computer Networking Conference, 1990, s. 134-140.
10. Fullmer C. L., Garcia-Luna-Aceves J. J.: *Floor Acquisition Multiple Access (FAMA) for Packet Radio Networks*. ACM SIGCOMM Computer Communication Review, Vol. 25, No. 4, Oct. 1995, s. 262-273.
11. Fullmer C. L., Garcia-Luna-Aceves J. J.: *Complete single-channel solutions to hidden terminal problems in wireless LANs*. Proceedings of IEEE ICC'97, June 1997, s. 575-579.
12. Deng J., Haas Z. J.: *Dual busy tone multiple access (DBTMA) – A multiple access control scheme for ad hoc networks*, IEEE Transactions on Communications, Vol. 50, No. 6, Jun 2002, s. 975-985.
13. Rom R.: *Collision Detection in Radio Channels*. W: Pickholtz R. L. (ed.): *Local Area & Multiple Access Networks*. Computer Science Press, 1986, s. 235-249.
14. Lo W. F., Mouftah H. T.: *Carrier Sense Multiple Access with Collision Detection for Radio Channels*. Proceedings of 13th International Communications and Energy Conference, 1984, s. 244-247.
15. Fullmer C. L., Garcia-Luna-Aceves J. J.: *FAMA-PJ: A Channel Access Protocol For Wireless LANs*. Proceedings of the 1st Annual International Conference on Mobile Computing and Networking, Berkeley, 1995, s. 76-85.
16. Hołubowicz W., Płóciennik P., Różański A.: *Systemy łączności bezprzewodowej*. Wydawnictwa EFP, Poznań 1996.