Przekazywanie pakietu od nadawcy do odbiorcy

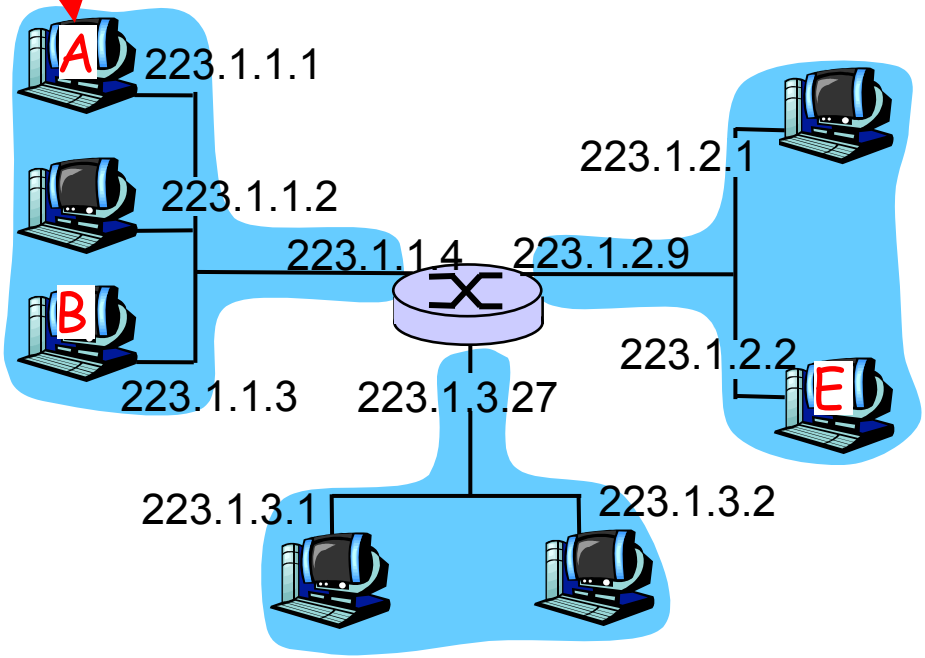
Pakiet IP:

różne pola	adres IP źródła	adres IP celu	dane

- pakiet się nie zmienia podczas przekazywania od źródła do celu
- ważne jest pole adresu celu
- dopiero ostatni ruter wie, czy pakiet dotarł na miejsce. Dlatego potrzeba sposobu na powiadomienie nadawcy o błędzie

tablica rutingu w *A*

Sieć celu	nast. ruter	Odległość
223.1.1		1
223.1.2	223.1.1.4	2
223.1.3	223.1.1.4	2



Przekazywanie pakietu od nadawcy do odbiorcy

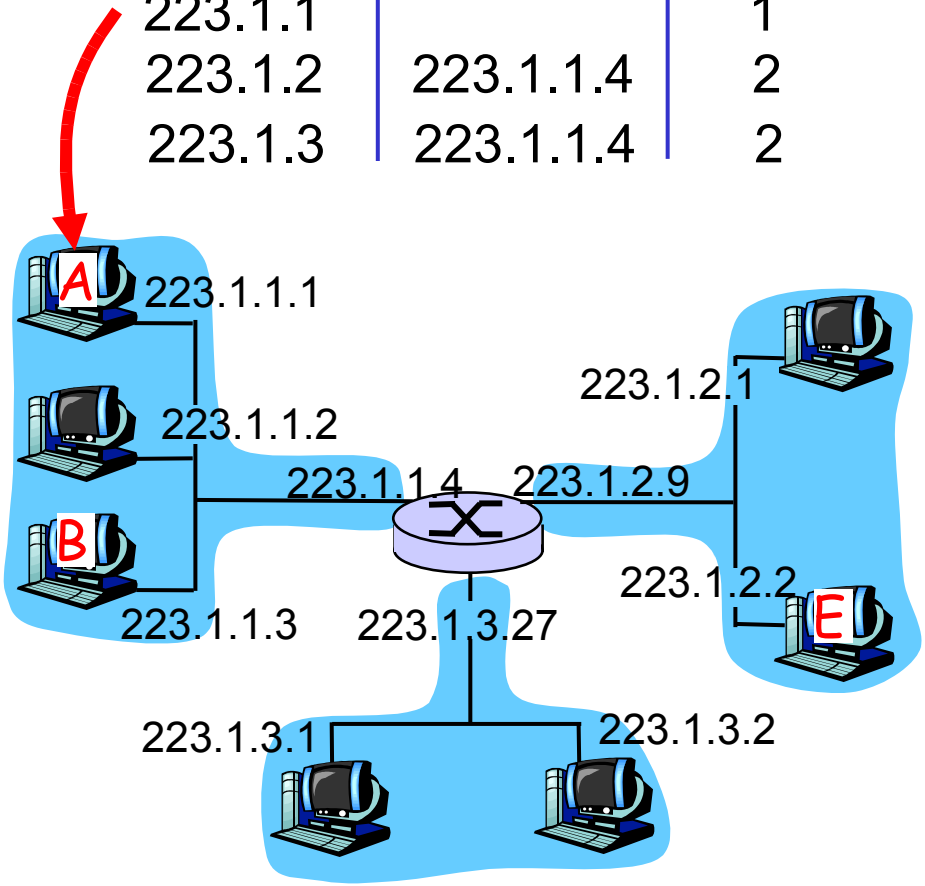
misc fields	223.1.1.1	223.1.1.3	data

Wyślij pakiet IP od A do B:

- poszukaj adresu podobnego do B w tablicy rutingu
- w tablicy zapisano, że B jest w tej samej sieci co A
- warstwa łącza wyśle pakiet bezpośrednio do B w ramce protokołu warstwy łącza
  - B i A są połączone bezpośrednio

tablica rutingu w A

Sieć celu	nast. ruter	Odległość
223.1.1		1
223.1.2	223.1.1.4	2
223.1.3	223.1.1.4	2



Przekazywanie pakietu od nadawcy do odbiorcy

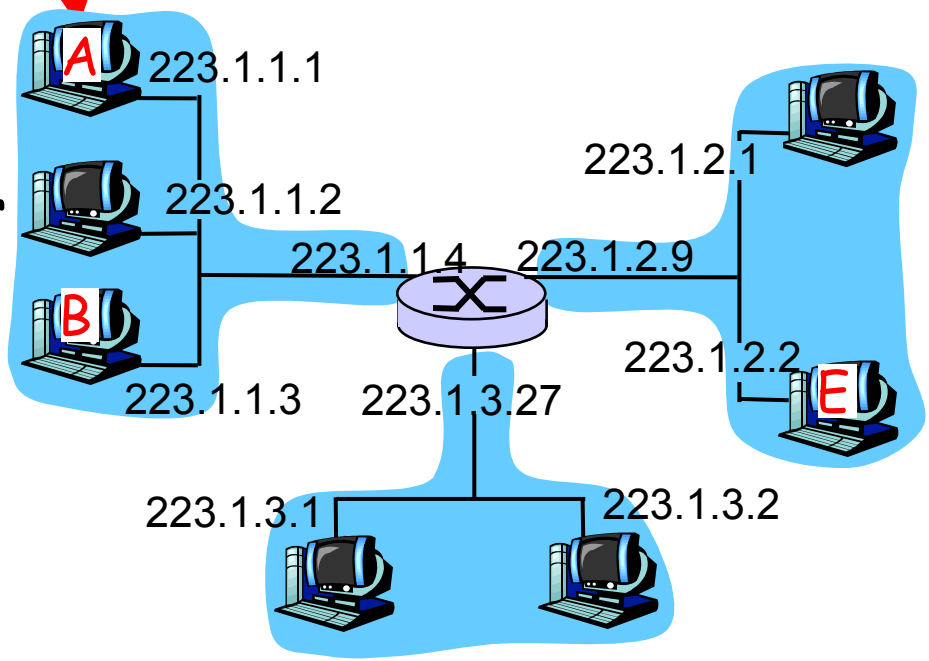
misc fields	223.1.1.1	223.1.2.3	data

Od A do E:

- poszukaj adresu podobnego do E w tablicy rutingu
- E jest w *innej* sieci
  - A, E nie są połączone bezpośrednio
- z tablicy rutingu: następny ruter w kierunku E to 223.1.1.4
- warstwa łącza wysyła pakiet do rutera 223.1.1.4 w ramce protokołu warstwy łącza
- pakiet jest odbierany przez 223.1.1.4
- c.d.n.....

tablica rutingu w A

Sieć celu	nast. ruter	Odległość
223.1.1		1
223.1.2	223.1.1.4	2
223.1.3	223.1.1.4	2



Przekazywanie pakietu od nadawcy do odbiorcy

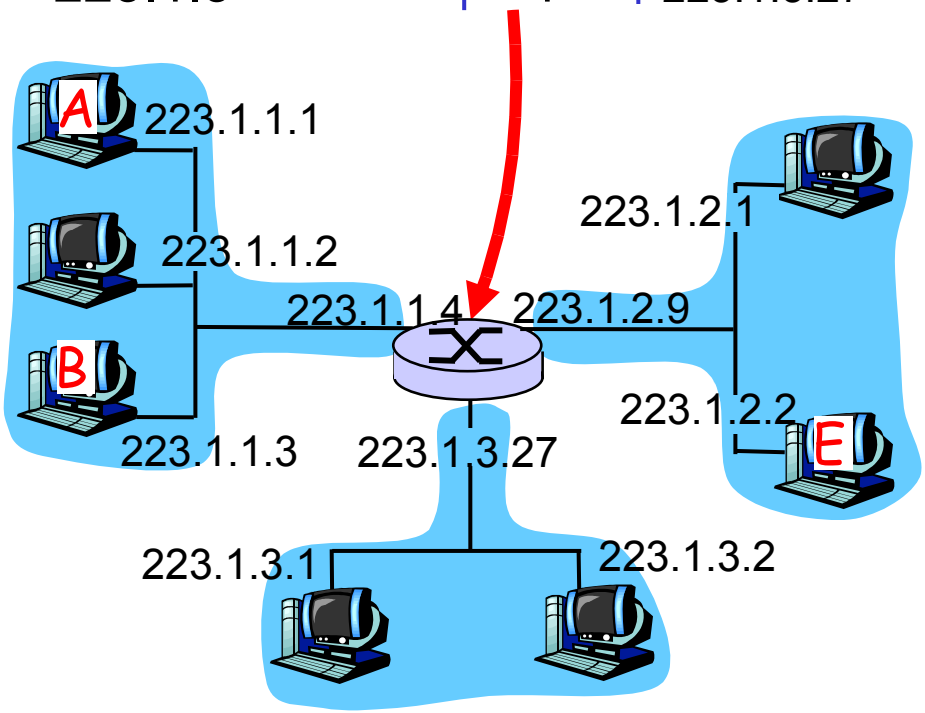
misc fields	223.1.1.1	223.1.2.3	data

Pakiet doszedł do 223.1.4, przeznaczony do 223.1.2.2

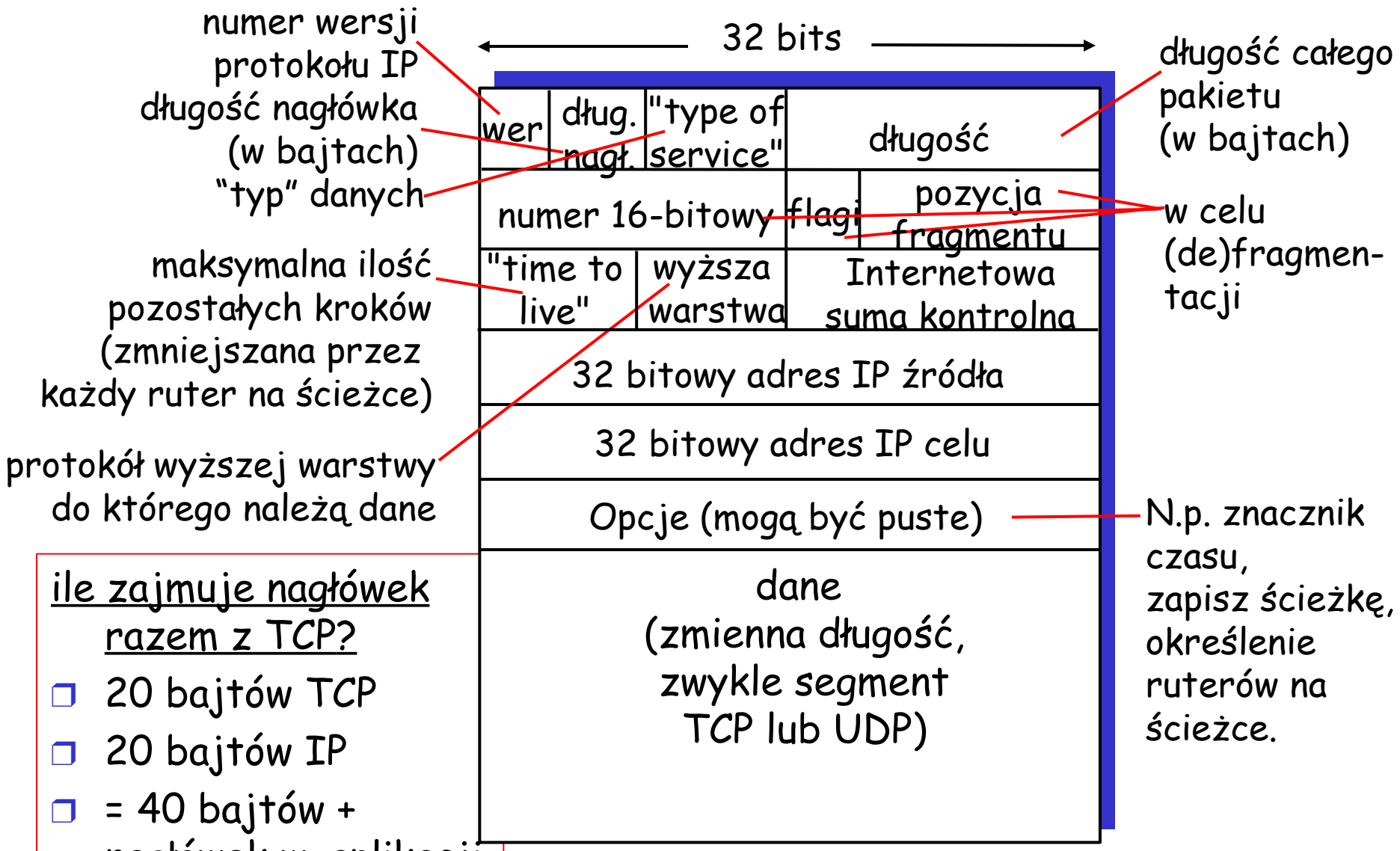
- poszukaj adresu podobnego do E w tablicy rutingu rutera
- E jest w *tej samej* sieci co interfejs 223.1.2.9 rutera
  - ruter i E są połączone bezpośrednio
- warstwa łącza wyśle pakiet do 223.1.2.2 w ramce protokołu warstwy łącza przez interfejs 223.1.2.9
- pakiet dociera do 223.1.2.2 (czyli tam, gdzie trzeba).

tablica rutingu w ruterze

Sieć celu	ruter	odległ.	interfejs
223.1.1	-	1	223.1.1.4
223.1.2	-	1	223.1.2.9
223.1.3	-	1	223.1.3.27



Format pakietu IP

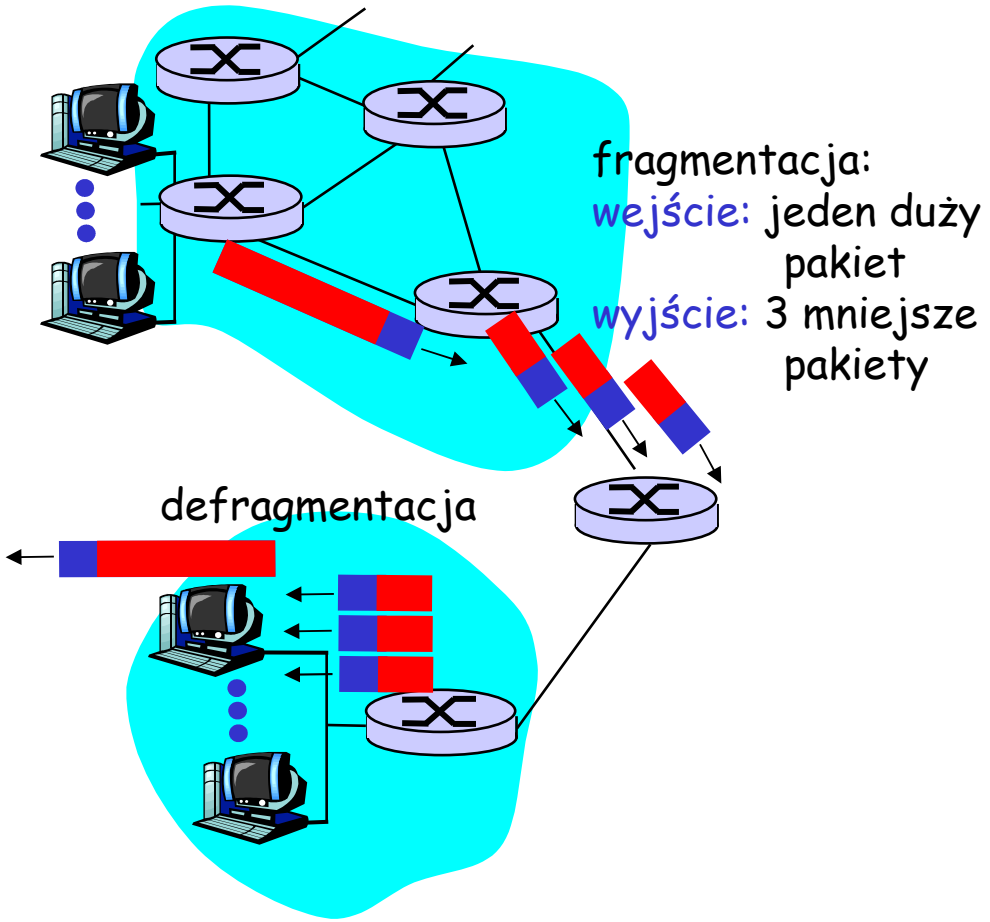


ile zajmuje nagłówek razem z TCP?

- 20 bajtów TCP
- 20 bajtów IP
- = 40 bajtów + nagłówek w. aplikacji

Fragmentacja i defragmentacja *IP*

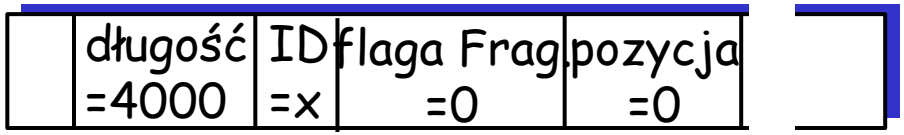
- łącza mają MTU (ang. *maximum transfer size*) – największa możliwa wielkość ramki warstwy łącza.
  - różne typy łącz, różne MTU
- duże pakiety IP są dzielone ("fragmentowane") w sieci
  - jeden pakiet jest dzielony na wiele pakietów
  - "łączone" *dopiero u celu*
  - nagłówek IP używany do rozpoznania, uporządkowania powiązanych fragmentów



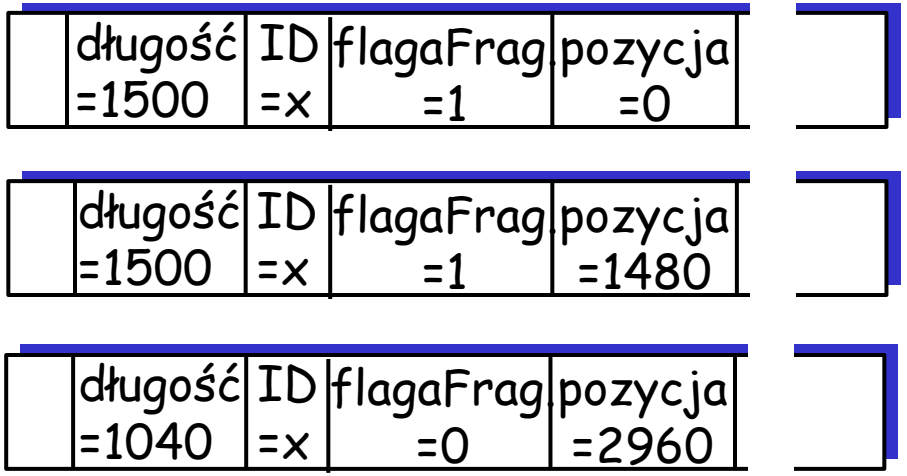
Fragmentacja i defragmentacja *IP*

Przykład

- pakiet 4000 bajtów
- MTU = 1500 bajtów
- MTU: łącza czy ścieżki?



Z jednego dużego pakietu
tworzone są trzy mniejsze pakiety



- fragmentacja może wyglądać inaczej w innych protokołach (lub warstwach)
  - łączenie nie zawsze jest na końcu
  - dzielenie nie zawsze następuje w warstwie wyższej

ICMP: Internet Control Message Protocol

- używany przez hosty, rutery, bramy do komunikacji informacji z warstwy sieci
  - zgłaszanie błędów: niedostępny host, sieć, port, protokół
  - żądanie/odpowiedź echo (używane prze ping)
- podwarstwa sieci "nad" IP:
  - komunikaty ICMP przekazywane w pakietach IP
- komunikat ICMP: typ, kod plus pierwszych 8 bajtów pakietu IP, który spowodował błąd

Typ	Kod	Opis
0	0	odpowiedź echo (ping)
3	0	sieć celu niedostępna
3	1	host celu niedostępny
3	2	protokół celu niedostępny
3	3	port celu niedostępny
3	6	sieć celu nieznana
3	7	host celu nieznany
4	0	spowolnienie źródła (kontrola przeciążenia- nie jest używane)
8	0	żądanie echo (ping)
9	0	ogłoszenie ścieżki
10	0	poszukiwanie rutera
11	0	wygasł TTL
12	0	zły nagłówek IP

Zastosowania ICMP: traceroute i ping

- Jak działa ping?
  - tyle razy, ile chciał użytkownik, wykonaj:
    - włącz zegar
    - wyślij pakiet ICMP, typ 8, kod 0 na adres odbiorcy
    - odbierz pakiet ICMP, typ 0, kod 0, od odbiorcy i zmierz czas RTT
    - Jeśli upłynęło za dużo czasu, zgłoś stratę i nie czekaj na odpowiedź (wykonuj dalej pętlę)
  - podsumuj wyniki: częstość strat

Zastosowania ICMP: traceroute i ping

- Jak działa traceroute?
  - n = 1
  - W pętli, aż nadejdzie pakiet ICMP typ 3, kod 3
    - Włącz zegar
    - Wyślij do odbiorcy 3 pakiety IP z TTL=n, zawierające segment UDP na dziwny port
    - Odbierz 3 pakiety ICMP, typ 11, kod 0
    - Jeśli dla któregoś pakietu zostanie przekroczony timeout, zgłoś stratę i nie czekaj na odpowiedź
    - Pokaż adres IP nadawcy pakietu ICMP (rutera na ścieżce do odbiorcy, o n kroków od nadawcy), czasy RTT lub informację o stratach
    - n = n + 1

DHCP: Dynamic Host Configuration Protocol

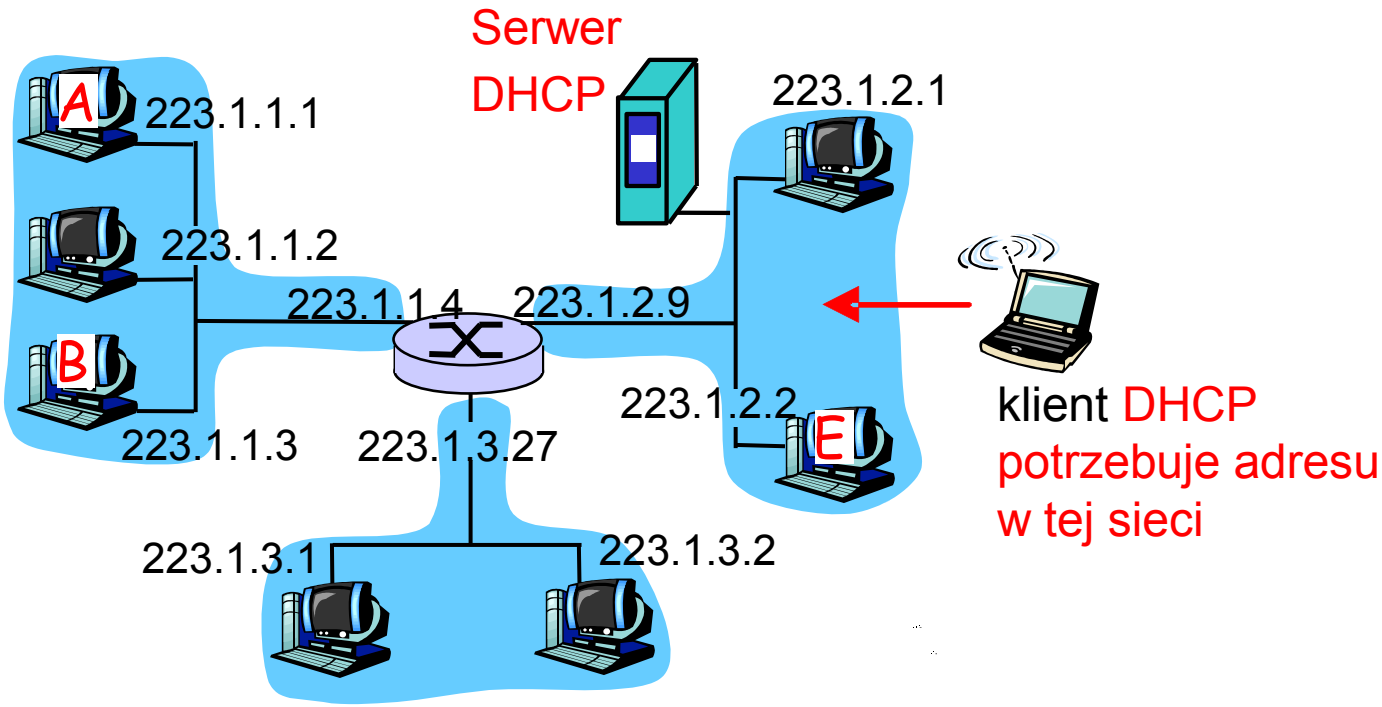
Cel: pozwól hostom *dynamicznie* uzyskiwać adresy IP z serwera w chwili dołączania do sieci

- Można też przedłużyć czas korzystania z adresu
- Pozwala na wielokrotne wykorzystanie adresu (adres jest zajęty tylko, gdy host jest podłączony i włączony
- Obsługa mobilnych użytkowników, chcących dołączyć się do sieci (więcej wkrótce)

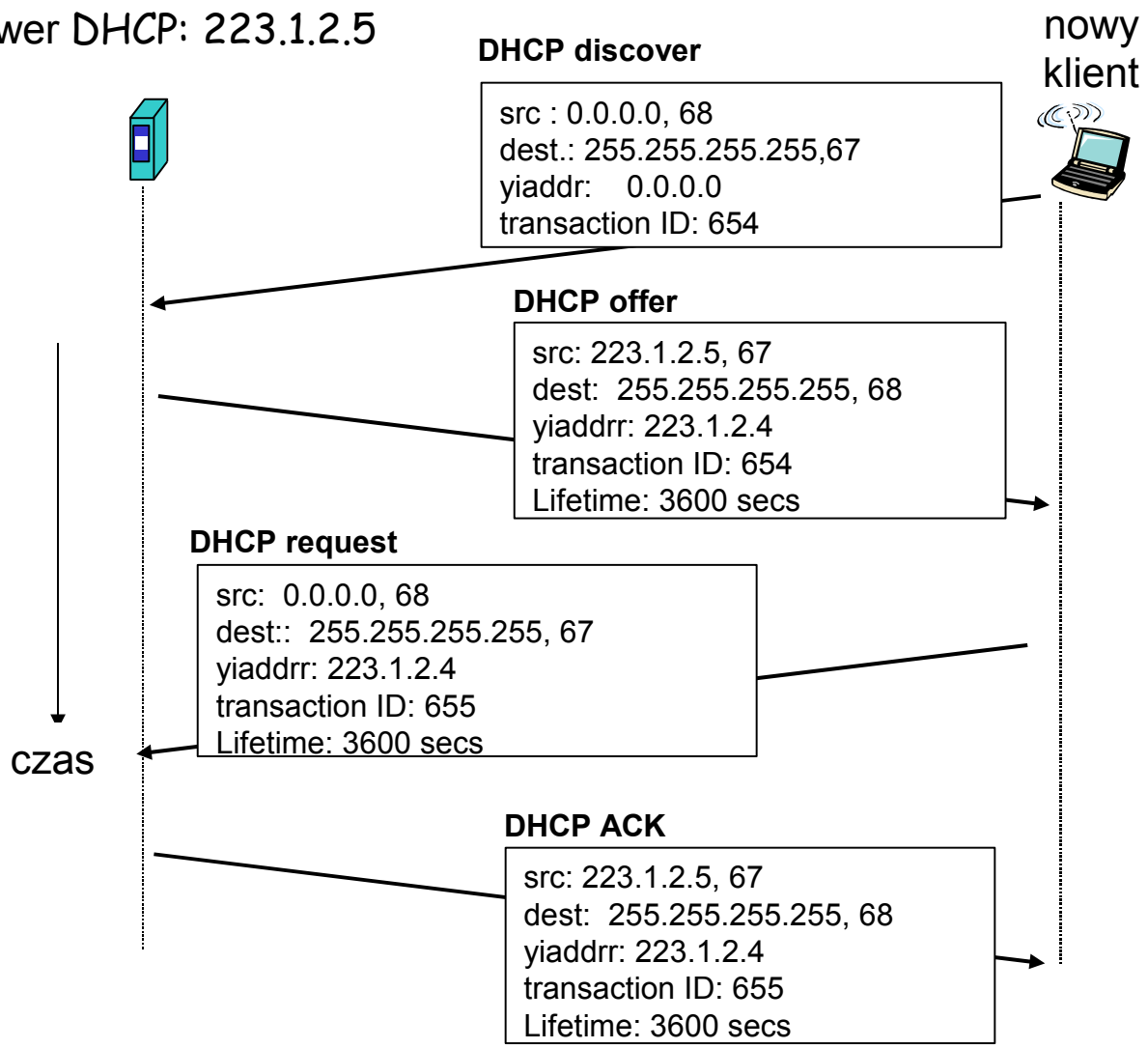
Przegląd DHCP:

- host rozgłasza komunikat "DHCP discover"
- serwer DHCP odpowiada komunikatem "DHCP offer"
- host żąda adresu IP: komunikat "DHCP request"
- serwer DHCP wysyła adres: komunikat "DHCP ack"

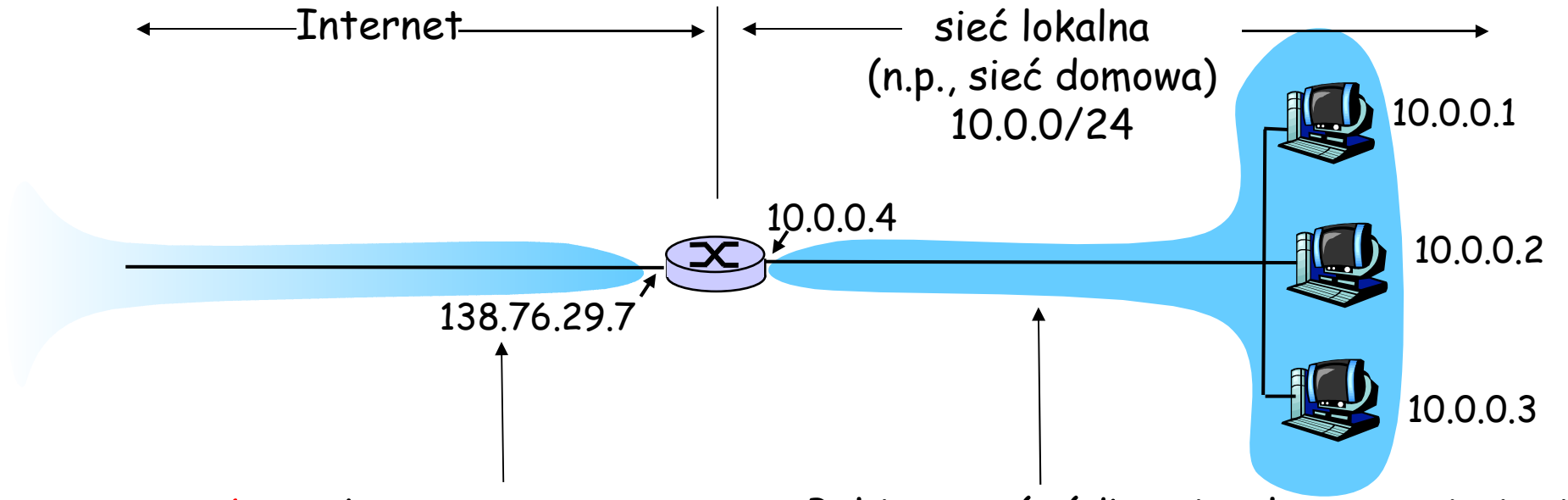
Scenariusz z klientem i serwerem DHCP



Scenariusz z klientem i serwerem DHCP



NAT: Network Address Translation



Wszystkie pakiety *opuszczające* sieć lokalną mają jednakowy adres IP źródła: 138.76.29.7, różne numery portów źródła

Pakiety z źródłem i celem w tej sieci mają adres z puli 10.0.0/24 (jak zwykle)

NAT: Network Address Translation

- Uzasadnienie: lokalna sieć używa tylko jednego adresu IP z punktu widzenia świata zewnętrznego:
  - nie ma potrzeby przydzielać zakresu adresów przez DI: - tylko jeden adres IP jest używany przez wszystkie urządzenia
  - można zmieniać adresy urządzeń w sieci lokalnej bez zawiadamiania świata zewnętrznego
  - można zmienić DI bez zmiany adresów urządzeń w sieci lokalnej
  - urządzenia w sieci lokalnej nie są widoczne ze świata zewnętrznego, dopóki same nie wyślą pakietu (zwiększa bezpieczeństwo).

NAT: Network Address Translation

Implementacja: Ruter NAT musi wykonać następujące czynności dla:

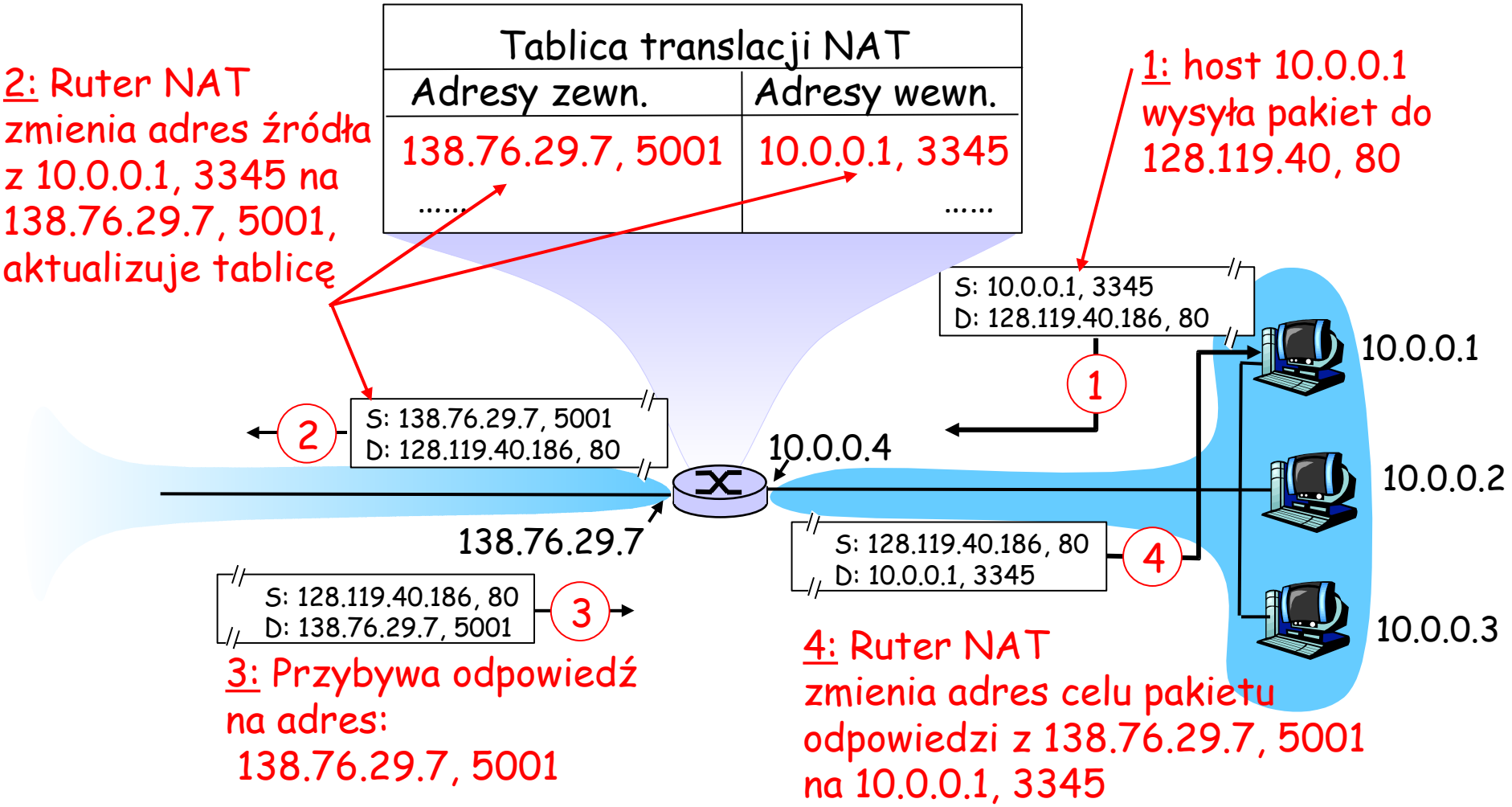
- *wychodzących pakietów: zastąp* (adres IP źródła, numer portu) przez (adres IP NAT, nowy numer portu)

  . . . zdalne hosty będą odpowiadały na adres (adres IP NAT, nowy numer portu) .

- *zapamiętaj (w tablicy translacji NAT)* każdą parę: **część 1**: (adres IP źródła, numer portu) zastąpioną przez **część 2**: (adres IP NAT, nowy numer portu)

- *przychodzących pakietów: zastąp* (adres IP NAT, nowy numer portu) w polach celu przez odpowiednią część 1 pary: (adres IP źródła, numer portu) zapisaną w tablicy translacji NAT

NAT: Network Address Translation



NAT: Network Address Translation

- 16-bitowy numer portu:
  - 60,000 jednoczesnych połączeń z jednego adresu w sieci wewnętrznej!
  - ograniczenie wydajnościowe: rozmiar tablicy translacji
- NAT jest kontrowersyjny:
  - rutery powinny przetwarzać informację warstwy 3
  - zasada koniec-koniec jest naruszona
    - możliwość użycia NAT musi być brana pod uwagę przez projektantów aplikacji, n.p., aplikacji P2P
  - hosty w sieci wewnętrznej nie mogą uruchamiać usług
  - braki adresów powinny być rozwiązane przez IPv6

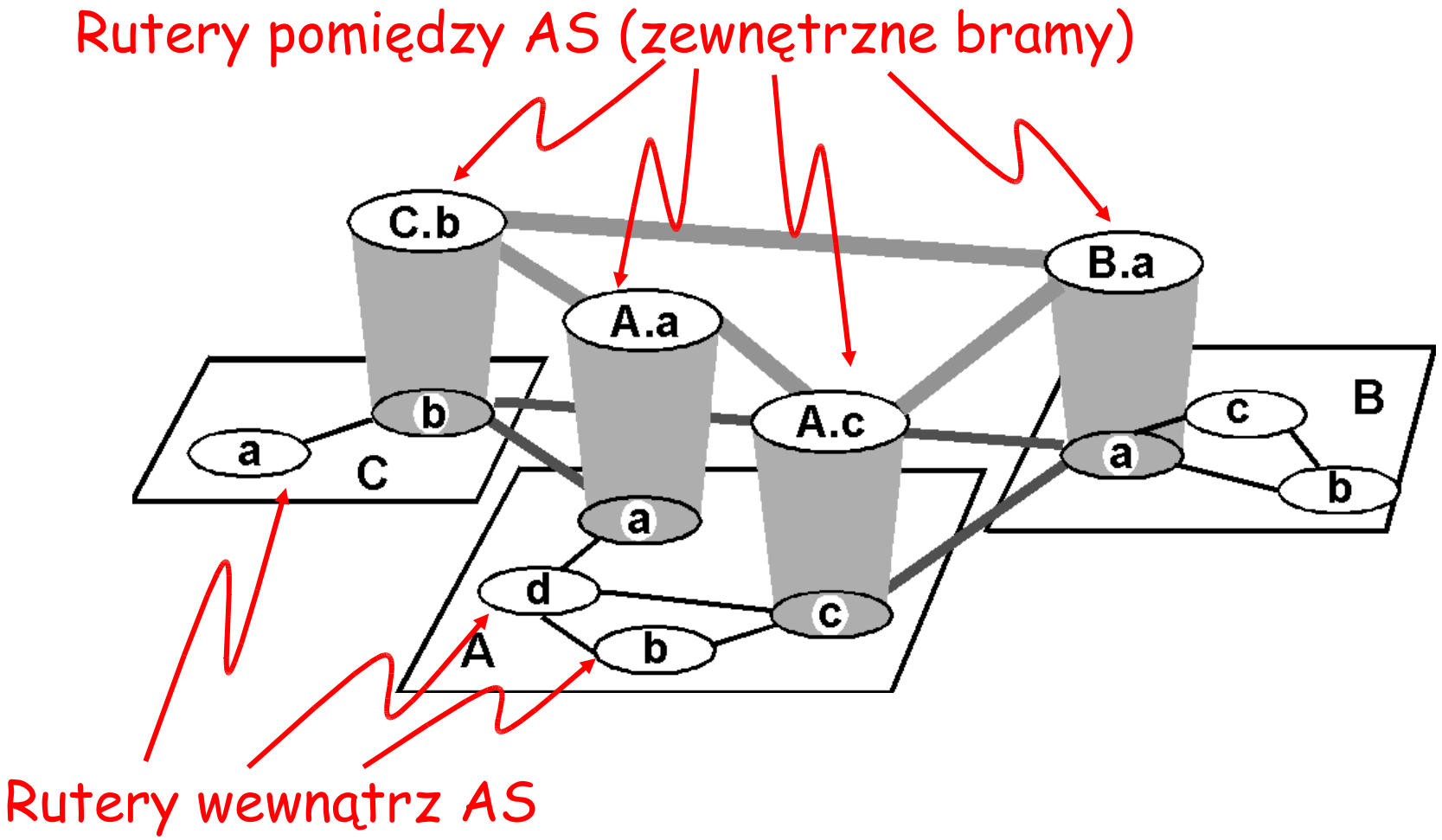
Mapa wykładu

- 4.1 Usługi warstwy sieci z komutacją pakietów
- 4.2 Zasady działania rutingu
- 4.3 Ruting hierarchiczny
- 4.4 Protokół Internetu (IP)
- 4.5 Ruting w Internecie
  - Ruting RIP i OSPF
  - Ruting BGP
- 4.6 Co jest w ruterze
- 4.7 IPv6
- 4.8 Ruting rozsiewczy (multicast)
- 4.9 Mobilność

Ruting w Internecie

- Globalny Internet składa się z Systemów Autonomicznych (AS) połączonych ze sobą:
  - **AS z jednym połączeniem**: mała organizacja: jedno połączenie do innego systemu autonomicznego
  - **AS z wieloma połączeniami**: duża organizacja (bez tranzytu): wiele połączeń z innymi systemami autonomicznymi
  - **AS tranzytowy**: DI poziomu 1 lub 2, łączący wiele systemów autonomicznych
- Dwupoziomowy ruting:
  - **Wewnętrzny:** administrator wybiera algorytm rutingu wewnątrz systemu autonomicznego
  - **Zewnętrzny:** jeden standard rutingu pomiędzy systemami autonomicznymi: BGP

Hierarchia AS w Internecie



Ruting Wewnętrzny

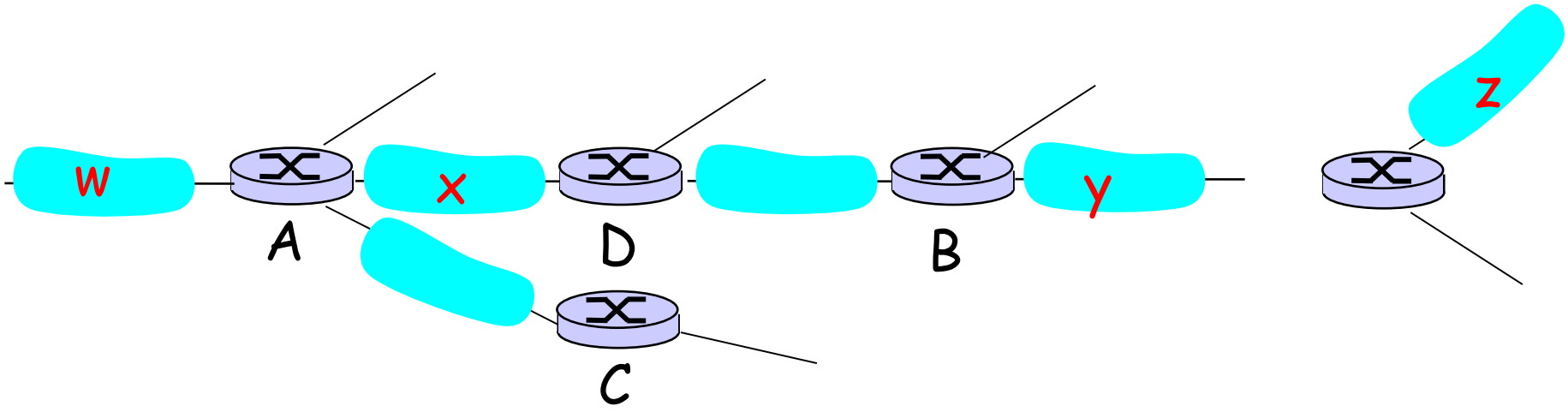
- Interior Gateway Protocols (IGP)
- Najczęściej używane protokołu rutingu wewnętrznego:
  - RIP: Routing Information Protocol
  - OSPF: Open Shortest Path First
  - IGRP: Interior Gateway Routing Protocol (własny protokół firmy Cisco)

RIP (Routing Information Protocol)

- Algorytm wektora odległości
- Był częścią dystrybucji BSD-UNIX w 1982 roku
- Miara odległości: ilość kroków (maksimum = 15 kroków)
  - *Czy potraficie zgadnąć, dlaczego?*

- Wektory odległości: wymieniane przez sąsiadów co 30 sekund przez komunikat odpowiedzi (także nazywany **ogłoszeniem**)
- Każde ogłoszenie: lista najwyżej 25 sieci będących celami w jednym systemie autonomicznym

RIP: Przykład



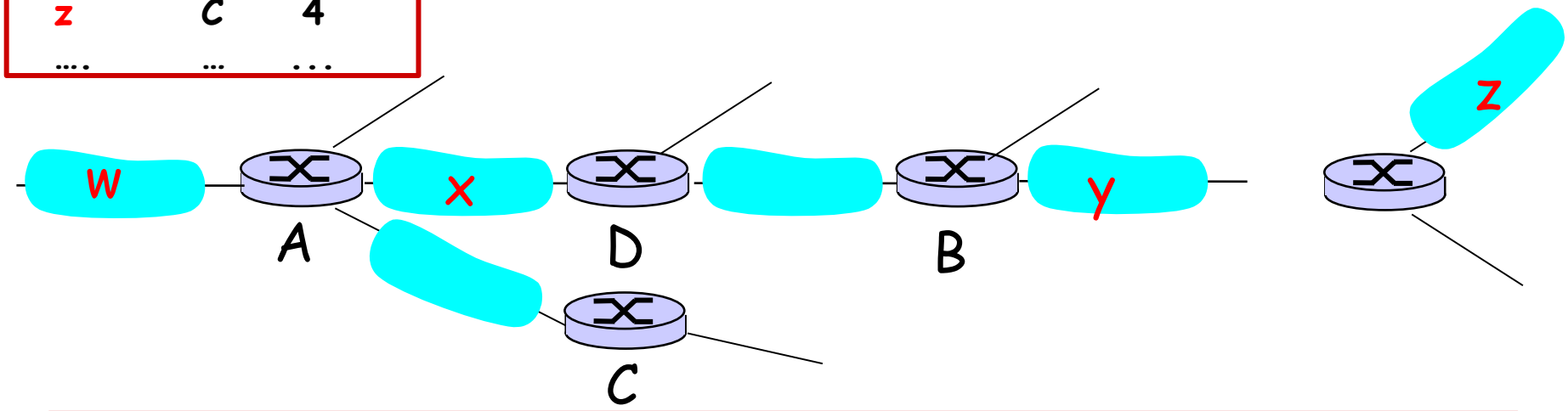
Sieć Celu	Następny ruter	Ilość kroków do celu
w	A	2
y	B	2
z	B	7
x	--	1
….	….	….

Tablica rutingu w D

RIP: Przykład

Cel	Nast.	Kroki
w	-	-
x	-	-
z	C	4
….	…	...

Ogłoszenie od A do D



Sieć Celu	Następny ruter	Ilość kroków do celu
w	A	2
y	B	2
z	~~B~~ A	~~7~~ 5
x	--	1
….	….	....

Tablica rutingu w D

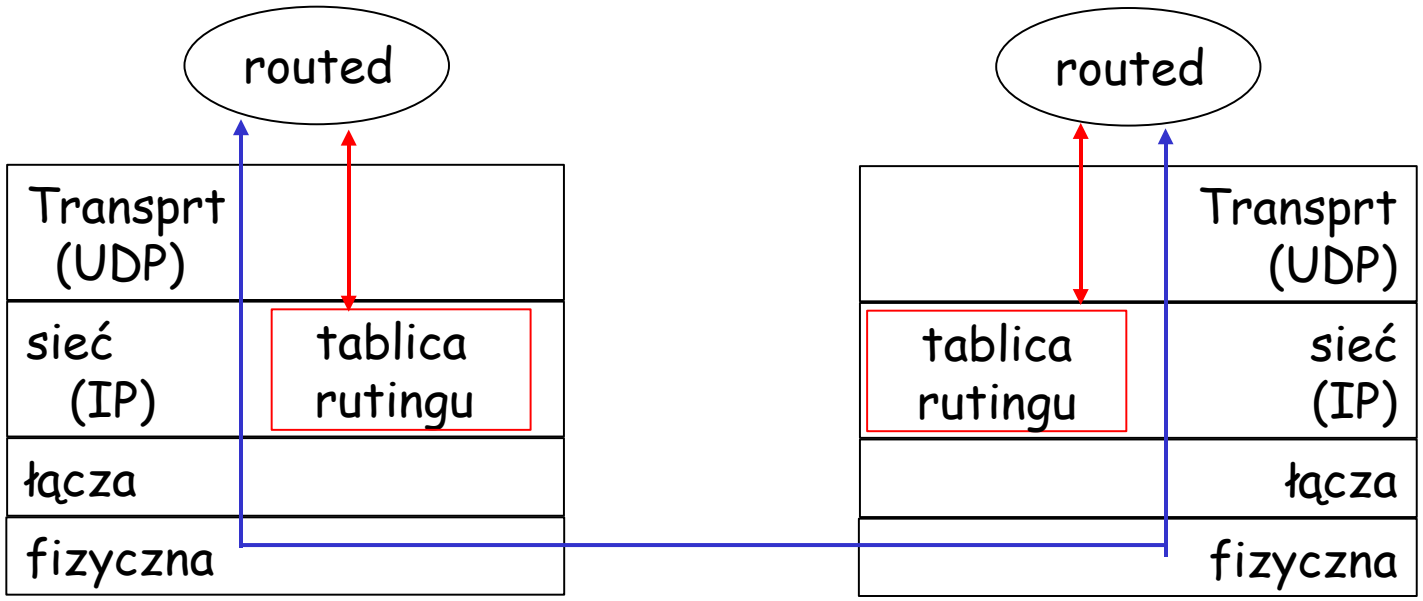
RIP: Awaria łącza i jej naprawa

Jeśli nie ma ogłoszenia przez 180 sekund --> uznaje się, że łącze do sąsiada uległo awarii

- ścieżki przez sąsiada stają się nieważne
- wysyłane jest nowe ogłoszenie do sąsiadów
- następnie, sąsiedzi wysyłają ogłoszenia do swoich sąsiadów (jeśli tablice rutingu uległy zmianie)
- informacja o awarii łącza rozprzestrzenia się szybko w sieci
- zatruty powrót jest używany, żeby uniknąć nieskończonych pętli (odległość nieskończona = 16 kroków)

RIP Przetwarzanie tabel

- Tablice rutingu RIP są zarządzane przez proces **warstwy aplikacji** nazywany route-d (demon)
- ogłoszenia posyłane są w pakietach UDP, okresowo powtarzanych



Przykład tabeli RIP

Ruter: *giroflee.eurocom.fr*

Cel	Brama	Flagi	Ref	Use	Interfejs
127.0.0.1	127.0.0.1	UH	0	26492	lo0
192.168.2.	192.168.2.5	U	2	13	fa0
193.55.114.	193.55.114.6	U	3	58503	le0
192.168.3.	192.168.3.5	U	2	25	qaa0
224.0.0.0	193.55.114.6	U	3	0	le0
default	193.55.114.129	UG	0	143454	

- Trzy podłączone sieci klasy C (sieci LAN)
- Ruter zna drogę tylko do dołączonych sieci
- W celu przesłania "w sieć", używana jest brama domyślna
- Adres multicast ścieżki: 224.0.0.0
- Interfejs loopback (dla testowania)

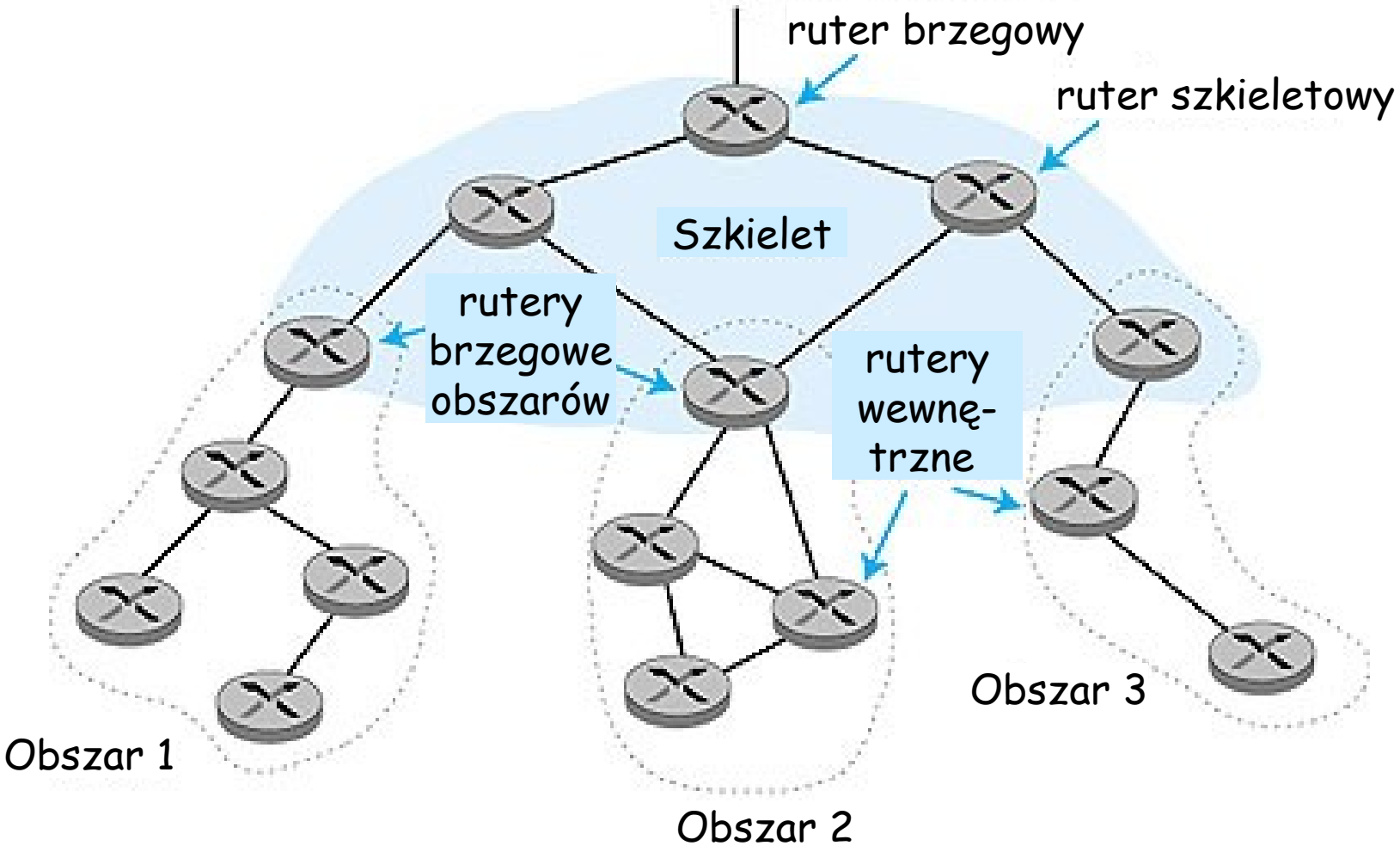
OSPF (Open Shortest Path First)

- "open": otwarty, czyli dostępny dla wszystkich (nieodpłatny, o ogólnie znanej specyfikacji)
- Używa algorytmu stanu łącza
  - rozsyła pakiety (ogłoszenia) SŁ
  - Mapa topologii w każdym węźle
  - Obliczanie ścieżek przy użyciu algorytmu Dijkstry
- Ogłoszenie OSPF ma jeden wpis dla każdego sąsiadującego rutera
- Ogłoszenia są rozsyłane do całego AS (przez zalew)
  - Wysyłane w komunikacie OSPF bezpośrednio przez IP (zamiast TCP lub UDP)

"Zaawansowane" cechy OSPF (niedostępne w RIP)

- Ochrona informacji: każdy komunikat OSPF jest uwierzytelniany (żeby zapobiec złośliwym zmianom)
- Może istnieć wiele ścieżek o tym samym koszcie (w RIP mogła być tylko jedna) – ang. multipath
- Dla każdego łącza, wiele miar kosztu dla różnych rodzajów usług TOS (n.p., koszt łącza satelitarnego dla usług "best effort" jest “niski”; “wysoki” dla usługi czasu rzeczywistego)
- Zintegrowany ruting unicast i multicast:
  - Multicast OSPF (MOSPF) używa tej samej bazy danych o topologii sieci co OSPF
- Hierarchiczny OSPF w dużych sieciach.

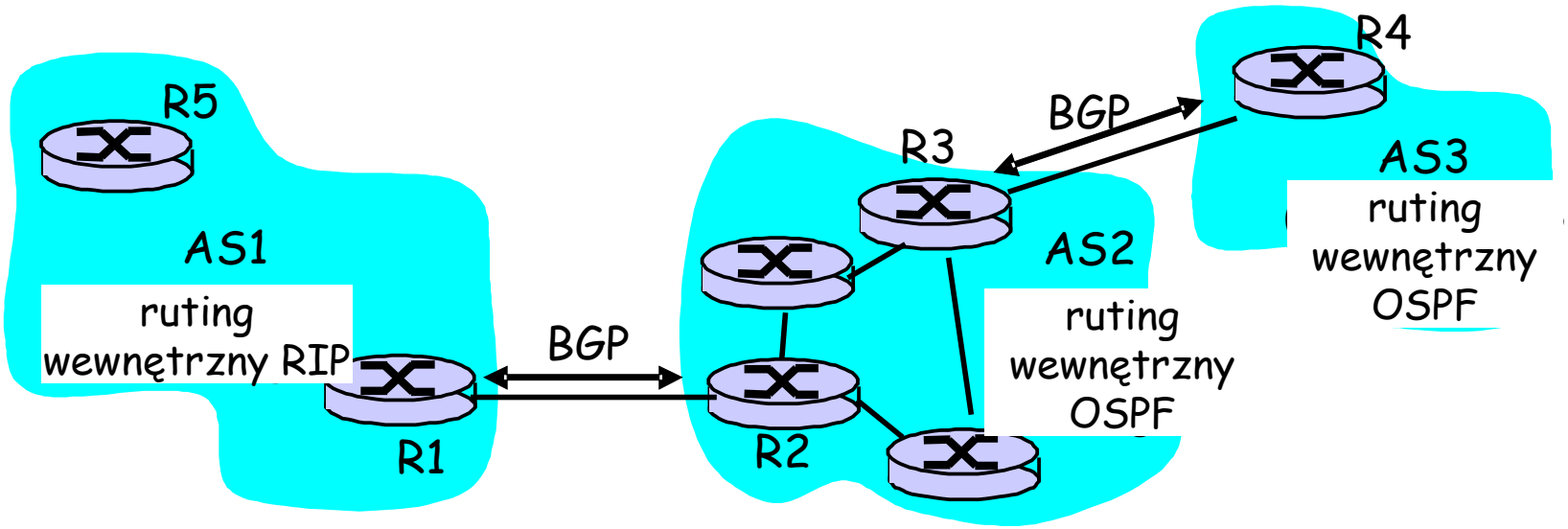
Hierarchiczny OSPF



Hierarchiczny OSPF

- Dwupoziomowa hierarchia: obszar lokalny, szkielet.
  - Ogłoszenia stanu łącza tylko w obszarze lokalnym
  - każdy węzeł ma szczegółową topologię obszaru; zna tylko kierunek (najkrótszą ścieżkę) do sieci w innych obszarach.
- **Rutery brzegowe obszarów:** "podsumowują" odległości do sieci w swoim obszarze, ogłaszają tę informację innym ruterom brzegowym obszarów.
- **Rutery szkieletowe:** realizują ruting OSPF w sieci szkieletowej.
- **Rutery brzegowe:** łączą się z innymi AS.

Ruting pomiędzy AS w Internecie: BGP



Ruting pomiędzy AS w Internecie: BGP

- BGP (Border Gateway Protocol): *standard de facto*
- Protokół **Wektora Ścieżek** :
  - podobny do protokołu Wektora Odległości
  - każda Brama Brzegowa (Border Gateway) rozsyła sąsiadom (partnerom) *całą ścieżkę* (czyli ciąg systemów autonomicznych) do celu
  - BGP rutuje do systemów autonomicznych (AS), a nie poszczególnych hostów
  - N.p., Brama X może wysłać ścieżkę do celu Z:

Ścieżka (X, Z) = X,Y1,Y2,Y3,…,Z

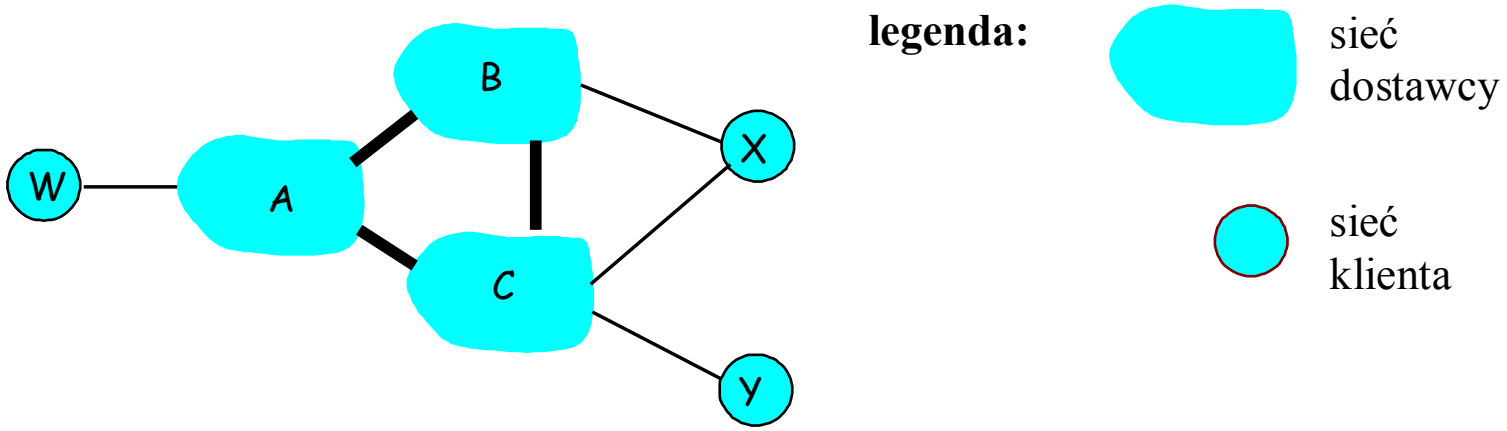
Ruting pomiędzy AS w Internecie: BGP

Przypuśćmy: brama X wysyła ścieżkę do sąsiedniej bramy W

- W może, ale nie musi wybrać ścieżki oferowanej przez X
  - z powodów kosztu, polityki (nie kierować ruchu przez AS konkurencji), w celu unikania pętli.
- Jeśli W wybierze ścieżkę ogłaszaną prze X, to:

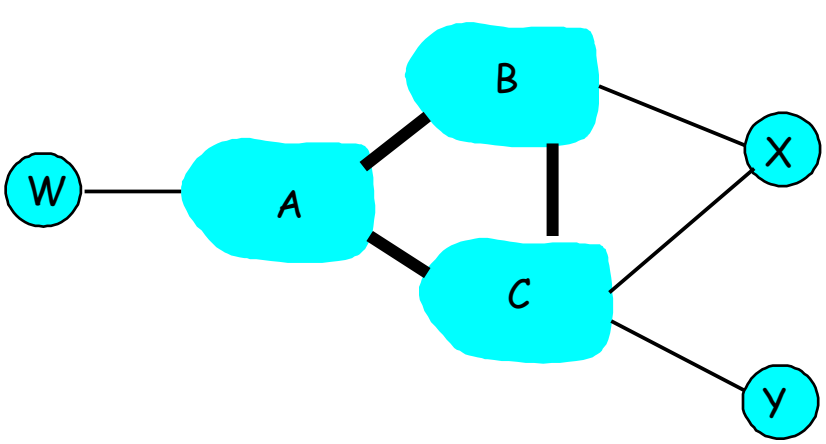
  Ścieżka (W,Z) = w, Ścieżka(X,Z)
- Uwaga: X może sterować przychodzącym ruchem za pomocą ogłoszeń ścieżek wysyłanych do sąsiadów:
  - n.p., nie chcę kierować ruchu do Z -> nie ogłaszam żadnych ścieżek do Z

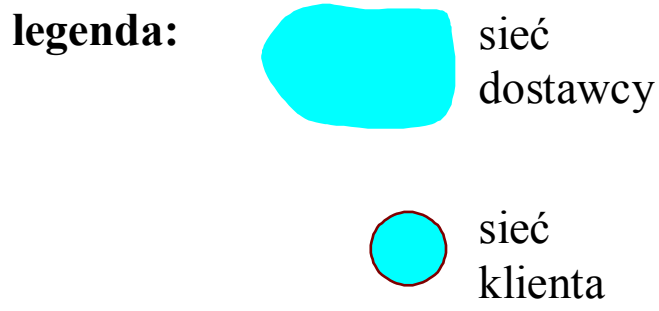
BGP: jak kontrolować, kto do nas rutuje



- A,B,C są sieciami dostawców
- X,W,Y are sieciami klientów
- X jest podwójnie połączona: dołączona do 2 sieci
  - X nie chce przekazywać ruchu z B do C
  - .. zatem X nie ogłosi B ścieżki do C

BGP: jak kontrolować, kto do nas rutuje





- A ogłasza B ścieżkę AW
- B ogłasza X ścieżkę BAW
- Czy B powinien ogłosić C ścieżkę BAW?
  - Na pewno nie! B nie uzyska "zapłaty" za ruting CBAW ponieważ ani W, ani C nie są klientami B
  - B chce zmusić C do rutowania do w przez A
  - B chce rutować *tylko* do/od swoich klientów!

Działanie BGP

Pytanie: co robi ruter BGP?

- Otrzymuje i filtruje ogłoszenia ścieżek od bezpośrednio podłączonych sąsiadów.
- Wybór ścieżek.
  - Żeby kierować ruch do celu X, jaka ścieżka (z wielu ogłoszonych) zostanie wybrana?
- Wysyłanie ogłoszeń ścieżek do sąsiadów.

Komunikaty BGP

- BGP wysyła komunikaty przez TCP.
- Komunikaty BGP:
  - OPEN: otwiera połączenie TCP do sąsiada i uwierzytelnia nadawcę
  - UPDATE: ogłasza nową ścieżkę (lub usuwa starą)
  - KEEPALIVE utrzymuje otwarte połączenie w braku komunikatów UPDATE; także potwierdza komunikat OPEN
  - NOTIFICATION: zgłasza błędy w poprzednim komunikacie; także używane do zamknięcia połączenia

Czemu ruting wewnętrzny i zewnętrzny się różnią?

Polityka:

- Ruting zewnętrzny: administrator chce mieć kontrolę nad tym, kto kieruje ruch przez jego sieć.
- Ruting wewnętrzny: jeden administrator kontroluje całą sieć, więc zagadnienia polityczne są nieistotne

Skalowalność:

- ruting hierarchiczny zmniejsza rozmiar tablic oraz ruch w sieci komunikujący aktualizacje tablic

Wydajność:

- Ruting wewnętrzny: może się skupiać na wydajności
- Ruting zewnętrzny: polityka może być ważniejsza od wydajności

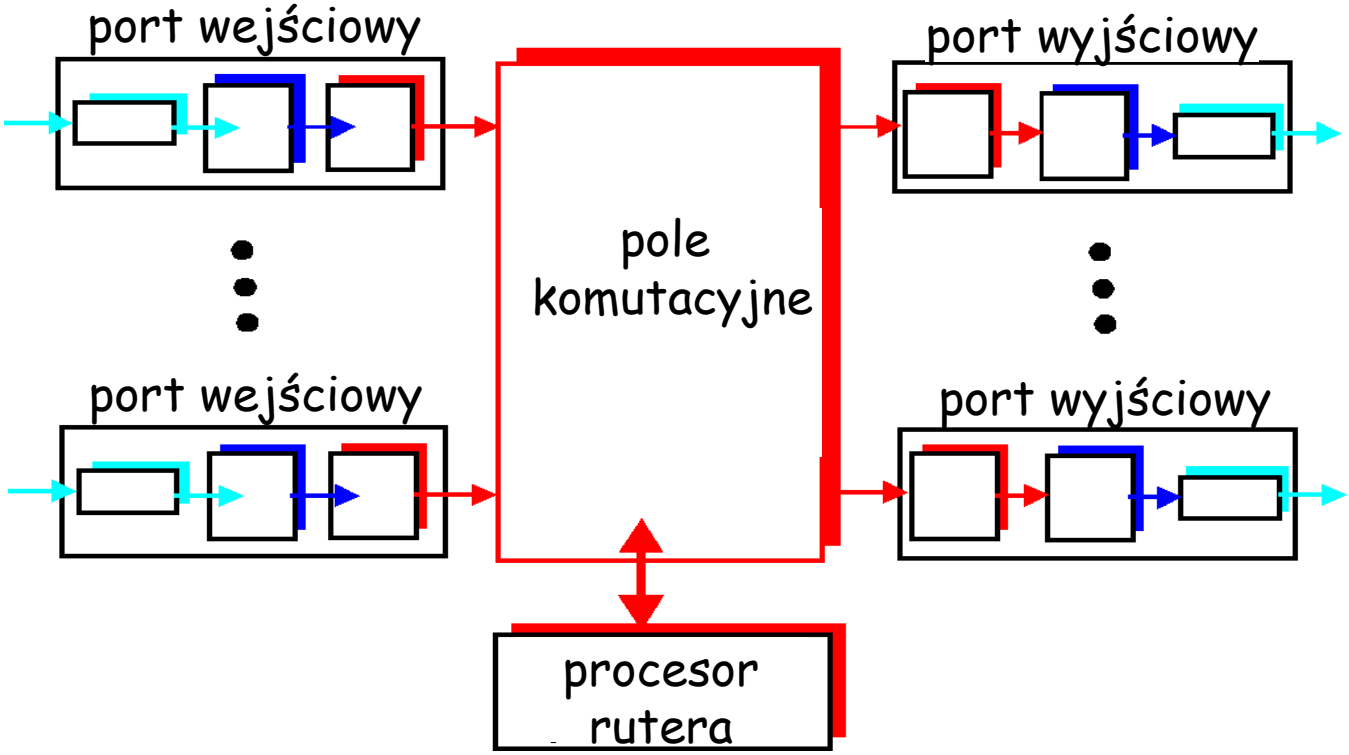
Mapa wykładu

- 4.1 Usługi warstwy sieci z komutacją pakietów
- 4.2 Zasady działania rutingu
- 4.3 Ruting hierarchiczny
- 4.4 Protokół Internetu (IP)
- 4.5 Ruting w Internecie
- 4.6 Co jest w ruterze
- 4.7 IPv6
- 4.8 Ruting rozsiewczy (multicast)
- 4.9 Mobilność

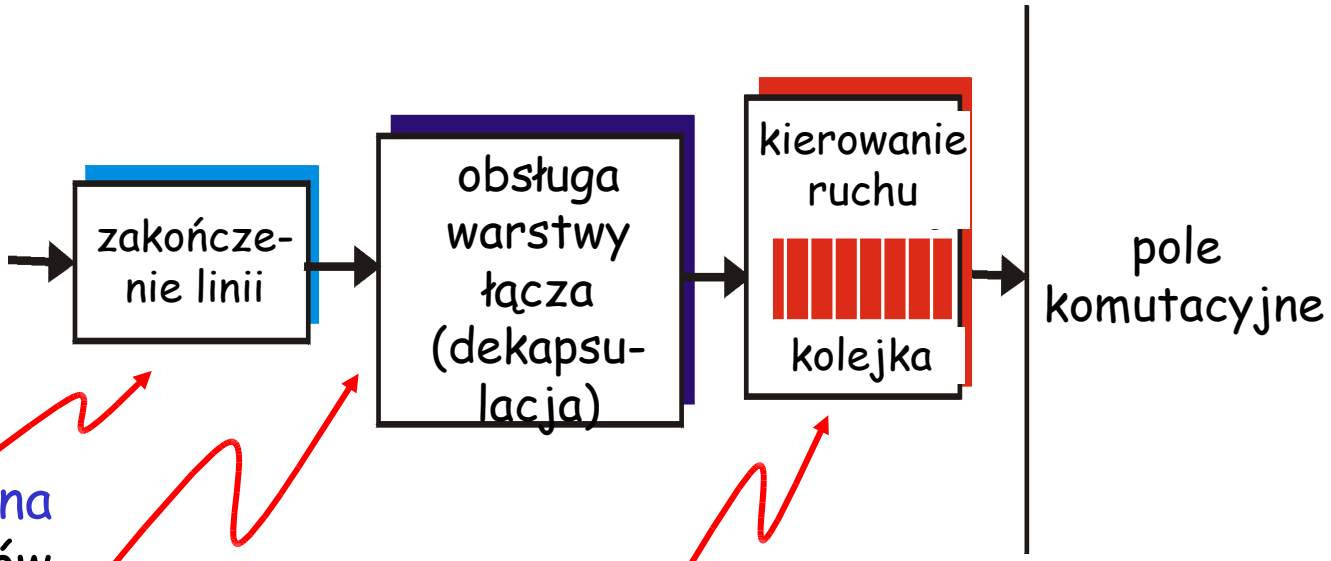
Przegląd architektury rutera

Dwie główne funkcje rutera:

- algorytm rutingu (RIP, OSPF, BGP)
- *przekazywanie* pakietów z łącz wejściowych na wyjściowe



Funkcje portu wejściowego



Warstwa fizyczna
odbiór sygnałów

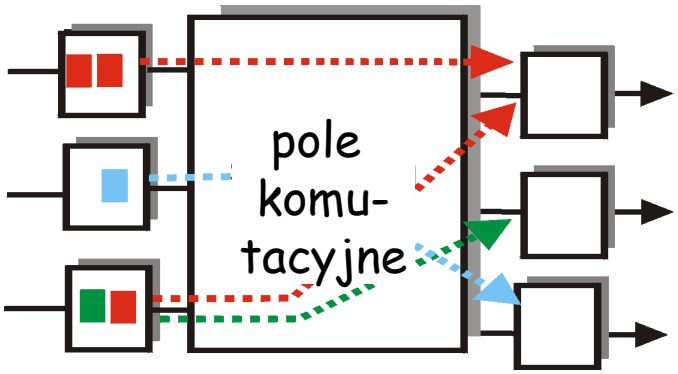
Warstwa łącza:
n.p., Ethernet
(patrz nast. część
wykładu)

Zdecentralizowane przełączanie:

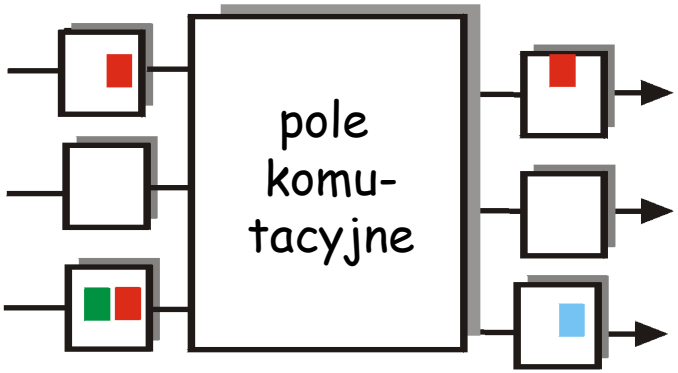
- znając odbiorcę pakietu, znajdź port wyjściowy używając tablicy rutingu w pamięci portu wejściowego
- cel: zakończyć obsługę w porcie wejściowym 'z szybkością łącza'
- kolejkowanie: jeśli pakiety przybywają szybciej niż szybkość przekazywania do pola komutacyjnego

Kolejkowanie w portach wejściowych

- Gdy pole komutacyjne wolniejsze niż połączony ruch z portów wejściowych -> mogą się pojawić kolejki w portach wejściowych
- blokowanie w kolejce: pakiet z przodu kolejki może uniemożliwić przekazanie dalej pakietów za nim
- *opóźnienie i straty spowodowane przez przepełnienie buforów portów wejściowych!*

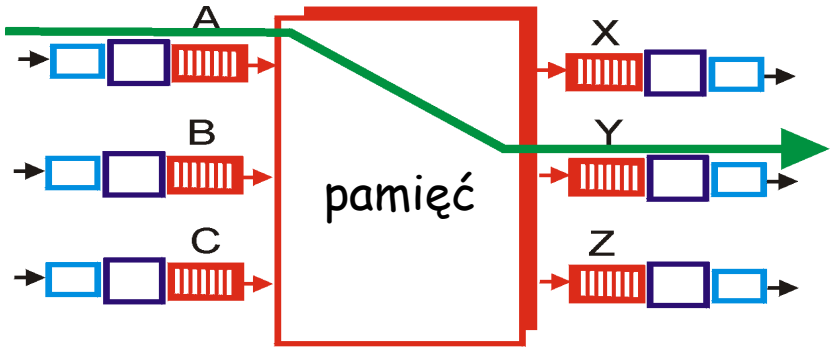


konkurencja o porty wyjściowe: tylko jeden czerwony pakiet może zostać wysłany na raz

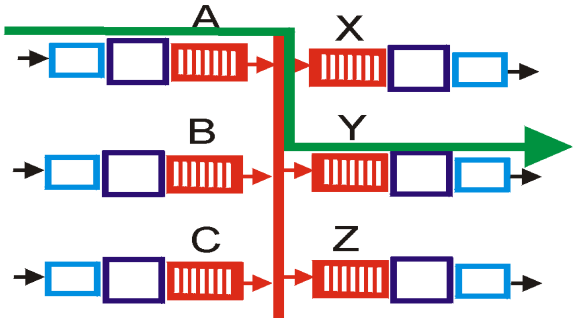


zielony pakiet jest zablokowany w kolejce

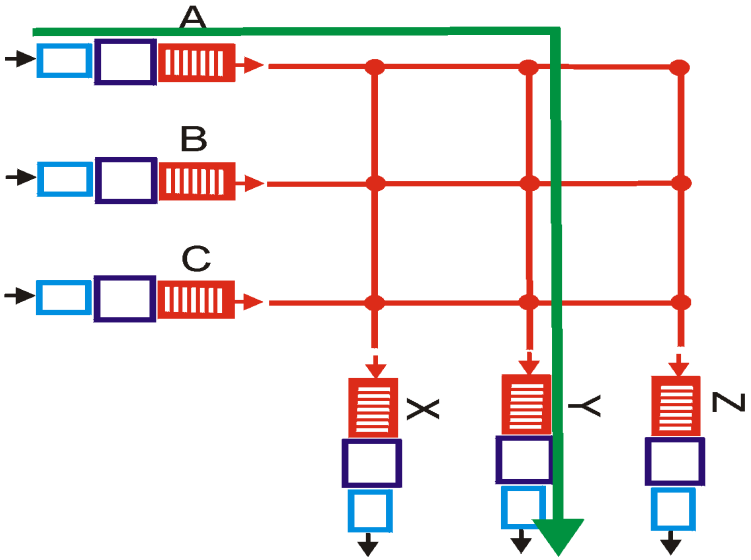
Trzy rodzaje pól komutacyjnych



pamięciowe



szyna

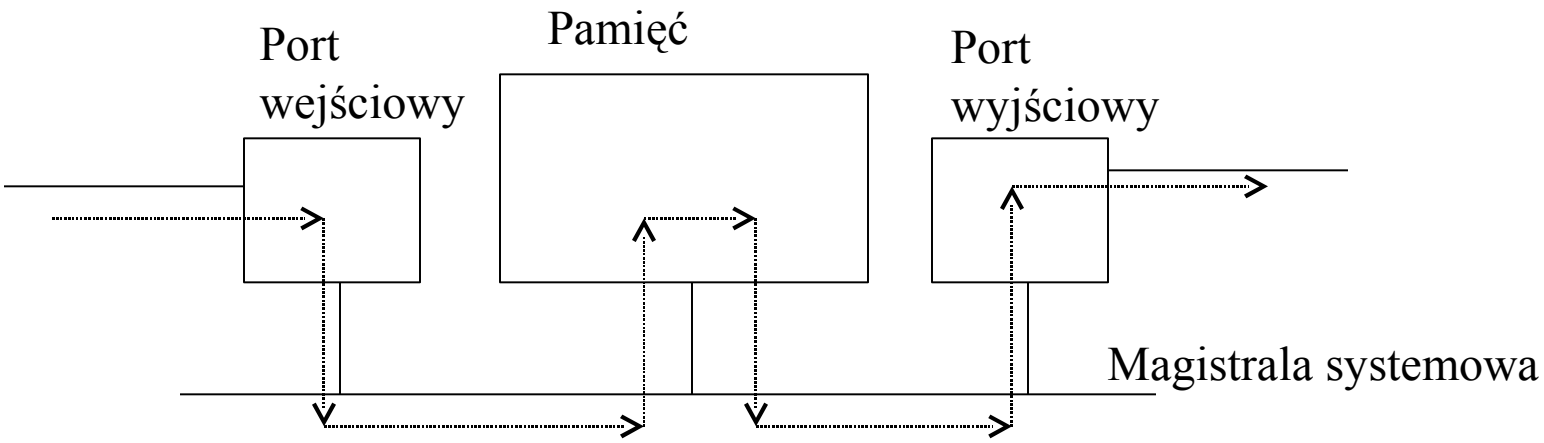


krata

Przełączanie w pamięci

Pierwsza generacja ruterów:

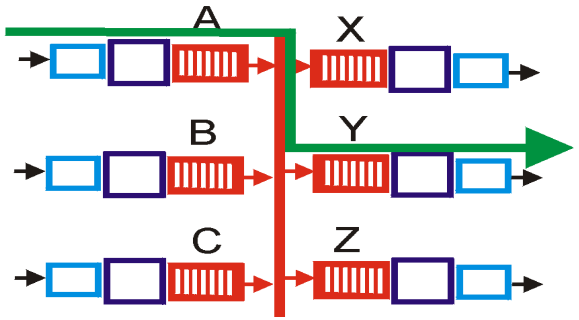
- pakiet kopiowany przez (pojedynczy) procesor rutera
- prędkość ograniczona przez przepustowość pamięci (2 przejścia przez magistralę dla każdego pakietu)



Nowoczesne rutery:

- procesor portu wejściowego zagląda do tablic rutingu, kopiuje pakiet do pamięci
- Cisco Catalyst 8500

Przełączanie za pomocą szyny

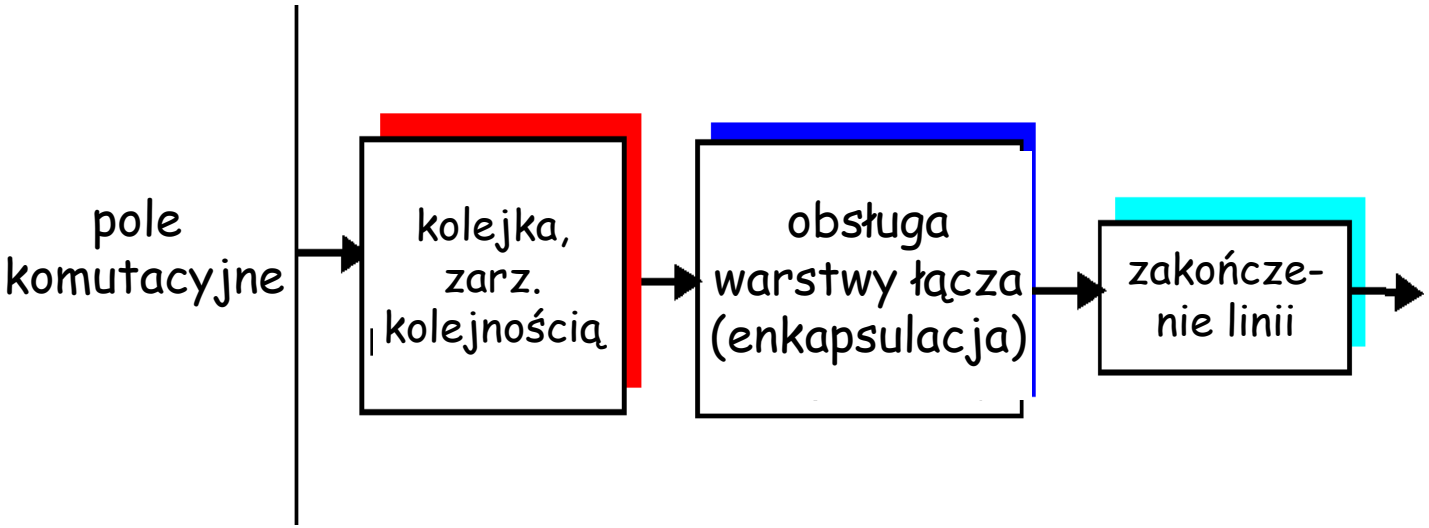


- pakiet przesyłany z pamięci portu wejściowego do pamięci portu wyjściowego przez wspólną szynę
- konkurencja o szynę: szybkość ograniczona przez przepustowość szyny
- szyna 1 Gb/s, Cisco 1900: dostatecznie szybka dla ruterów dostępowych i ruterów małych organizacji (nie dla ruterów regionalnych i szkieletowych)

Przełączanie za pomocą kraty

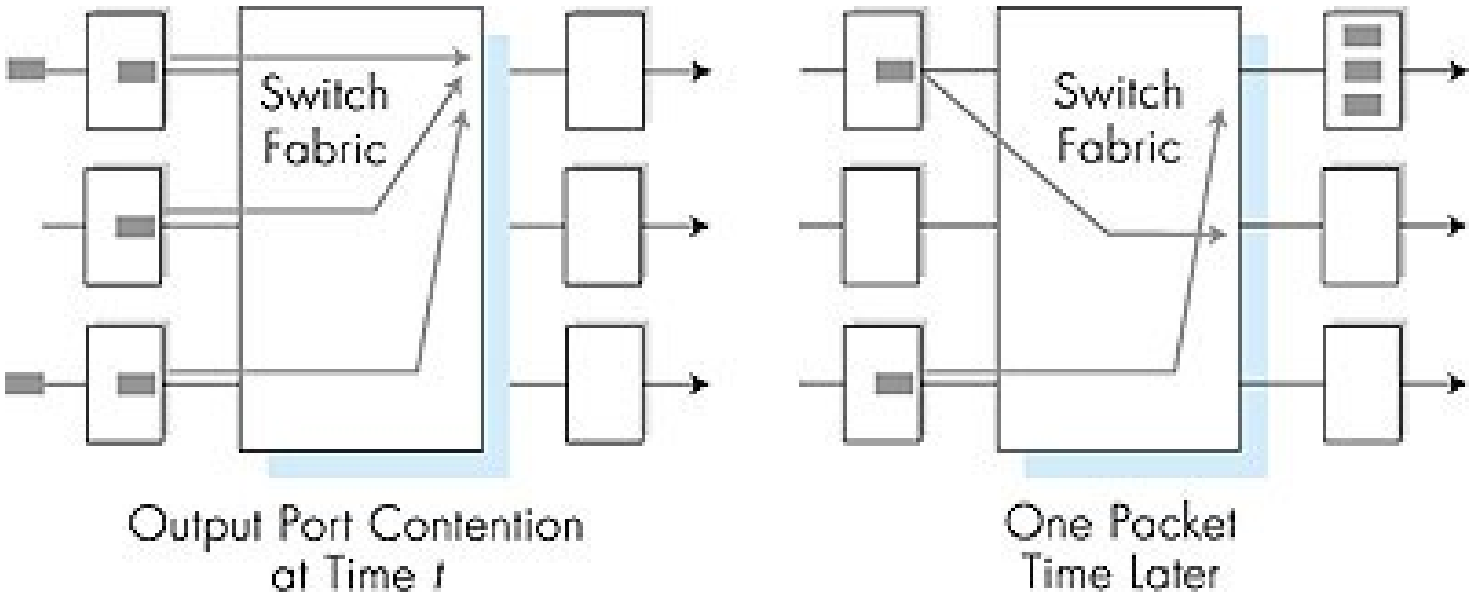
- przezwycięża ograniczenie przepustowości szyny
- sieci Banyan, inne sieci połączeń zaprojektowane początkowo do łączenia procesorów w superkomputerach
- Zaawansowana technologia: podział pakietu na komórki ustalonej wielkości, przełączanie komórek przez kratę.
- Cisco 12000: przełącza z szybkością Gb/s przez kratę

Porty wyjściowe



- *Kolejkowanie* jest potrzebne, gdy pakiety przybywają z pola komutacyjnego szybciej, niż prędkość transmisji łącza
- *Zarządzanie kolejnością* wybiera pakiety z kolejki do transmisji

Output port queueing



- buffering when arrival rate via switch exceeds output line speed
- *queueing (delay) and loss due to output port buffer overflow!*

Mapa wykładu

- 4.1 Usługi warstwy sieci z komutacją pakietów
- 4.2 Zasady działania rutingu
- 4.3 Ruting hierarchiczny
- 4.4 Protokół Internetu (IP)
- 4.5 Ruting w Internecie
- 4.6 Co jest w ruterze
- 4.7 IPv6
- 4.8 Ruting rozsiewczy (multicast)
- 4.9 Mobilność