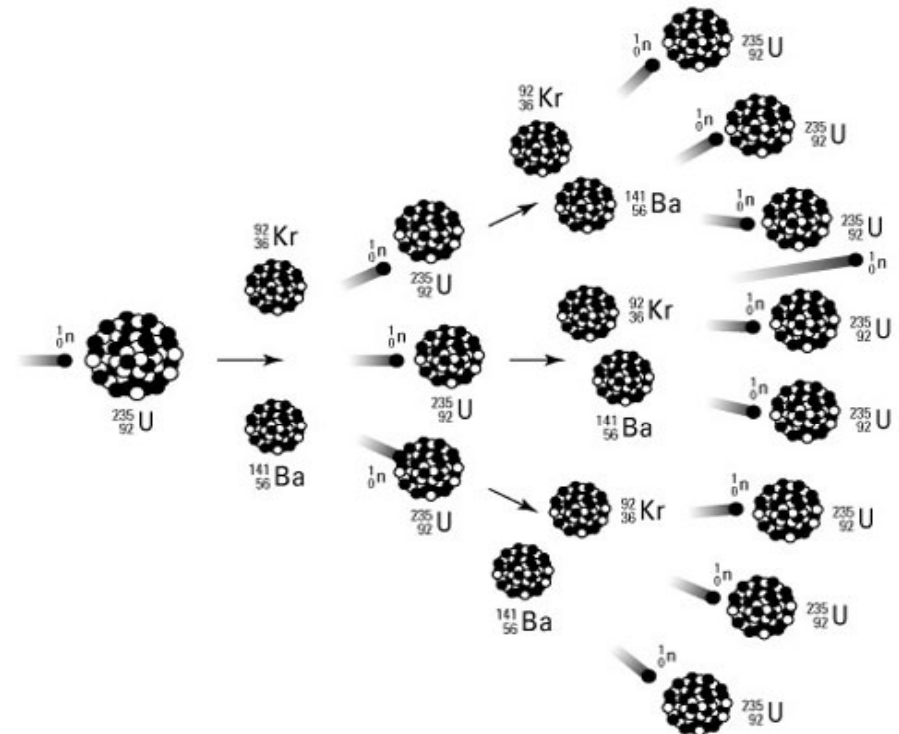


# Multicasty – zastosowanie i działanie



**Bartłomiej Anszperger**

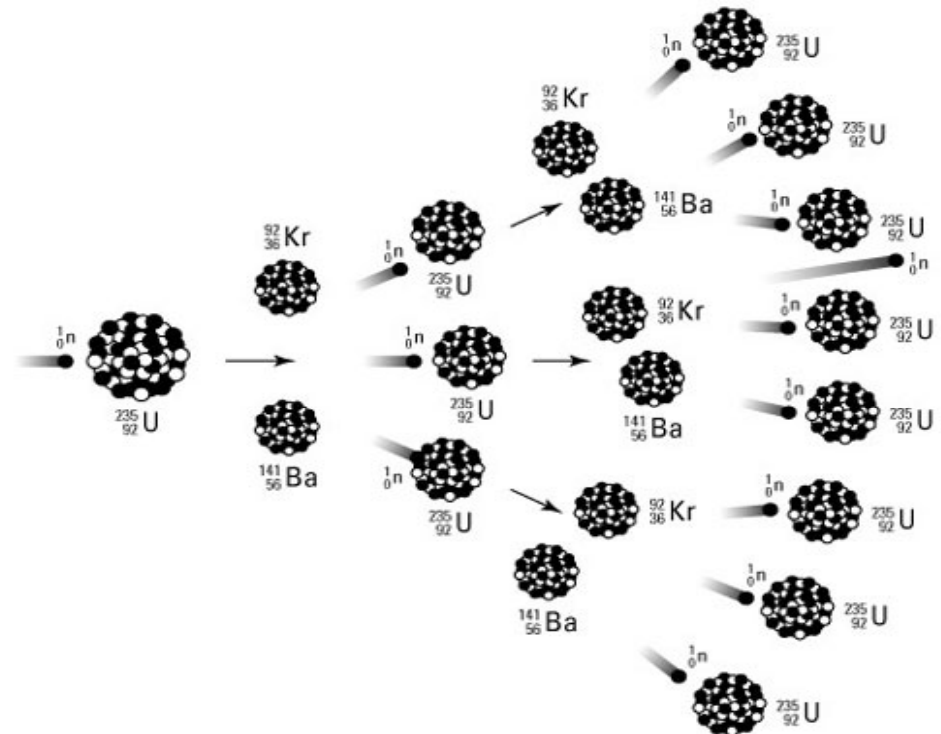
Network Consulting Engineer

[banszper@cisco.com](mailto:banszper@cisco.com)

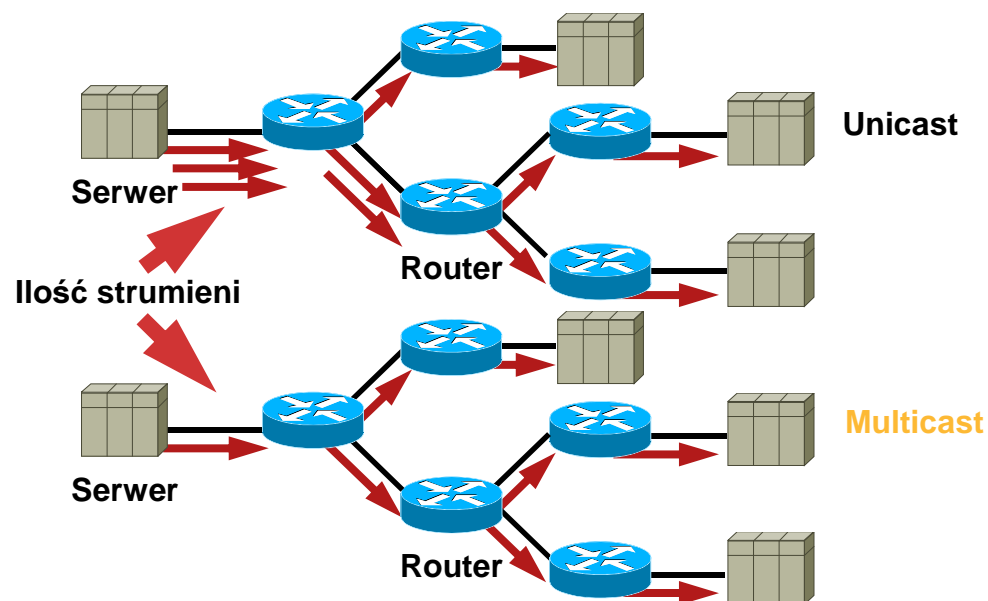
# Agenda

- ❖ Po co nam w ogóle ten cały mutlicast? ;-)
- ❖ Podstawy technologii mutlicast
- ❖ Protokoły PIM
- ❖ Punkty Rendezvous (RP)
- ❖ MPLS a Multicast
- ❖ Multicast w praktyce – czyli co warto robić a o czym zapomnieć

Po co nam w ogóle ten cały mutlicast?



# Ruch typu Unicast a Multicast



## Multicast:

Bezpołączeniowy tryb transmisji danych:

- bazuje na ruchu UDP

- brak gwarancji dostarczenia pakietów do odbiorców (*best effort delivery*)

- brak mechanizmów unikania natłoku

Możliwość pojawienia się zduplikowanych pakietów IP, brak gwarancji kolejności dostarczenia pakietów (*out of order delivery*)

Konieczność sygnalizacji chęci odbioru tego typu ruchu

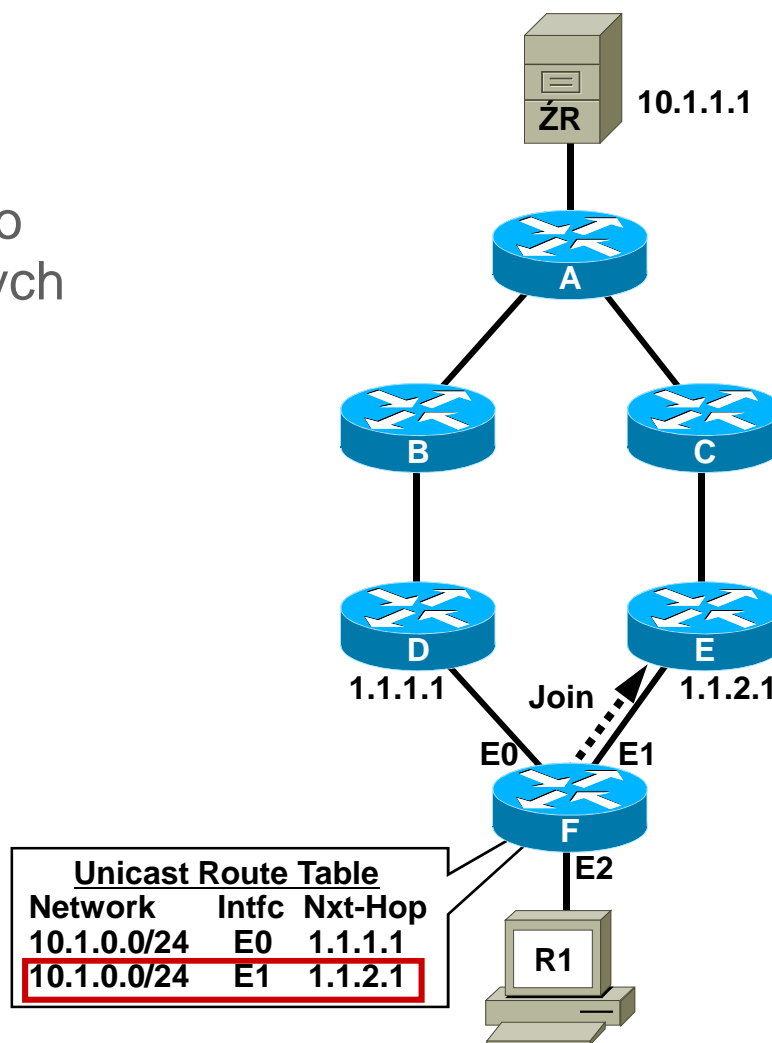
Ograniczenie powielania ruchu poprzez wyznaczenie drzewa dystrybucyjnego którego korzeniem jest źródło



# Mechanizm Reverse Path Forwarding

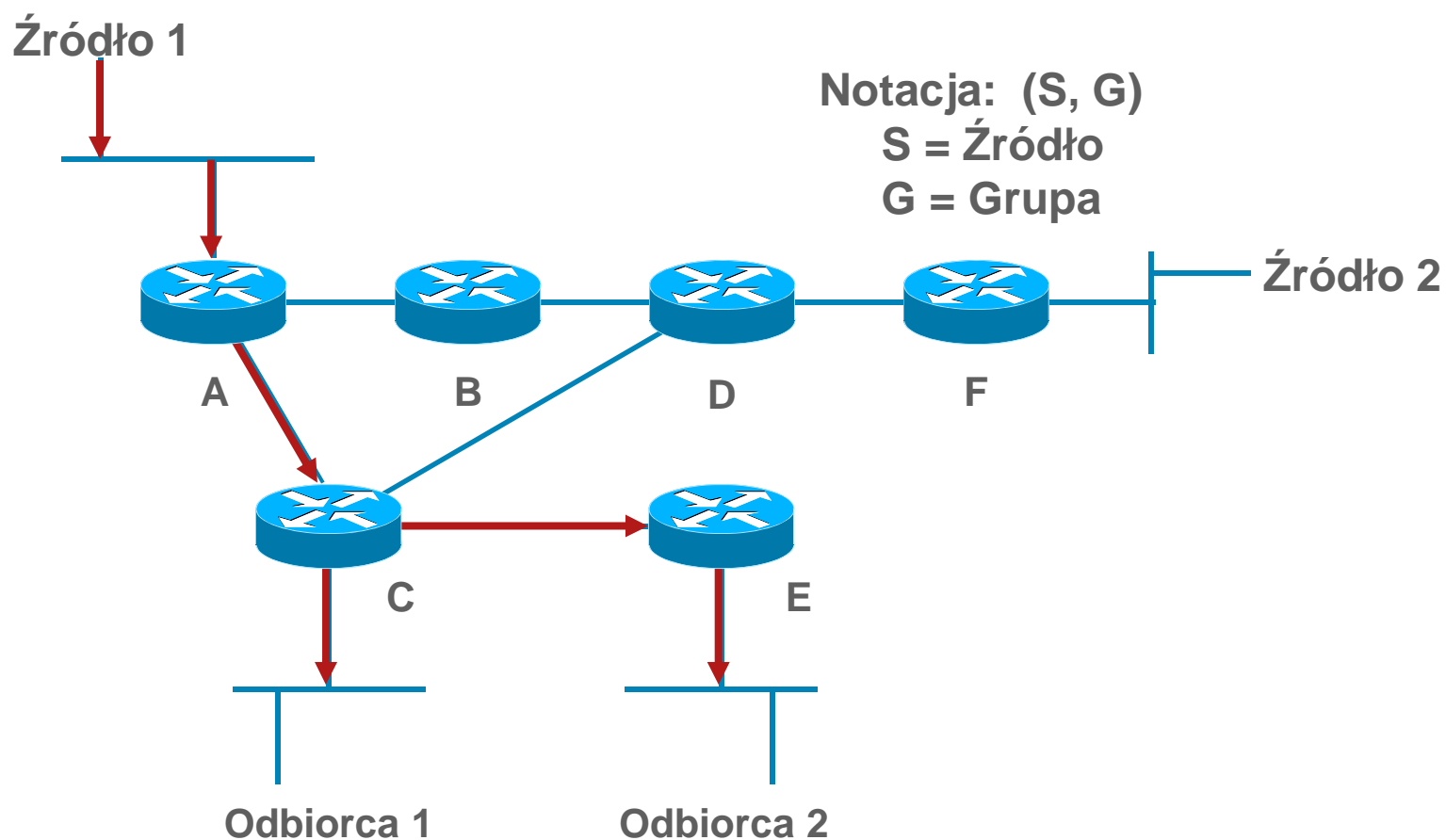
## Sprawdzenie RPF

- ❖ Co jeśli mamy kilka ścieżek o równym koszcie prowadzących do źródła?
  - nie możemy użyć obu!
- ❖ „Tie-breaker”
  - użyj najwyższego adresu IP



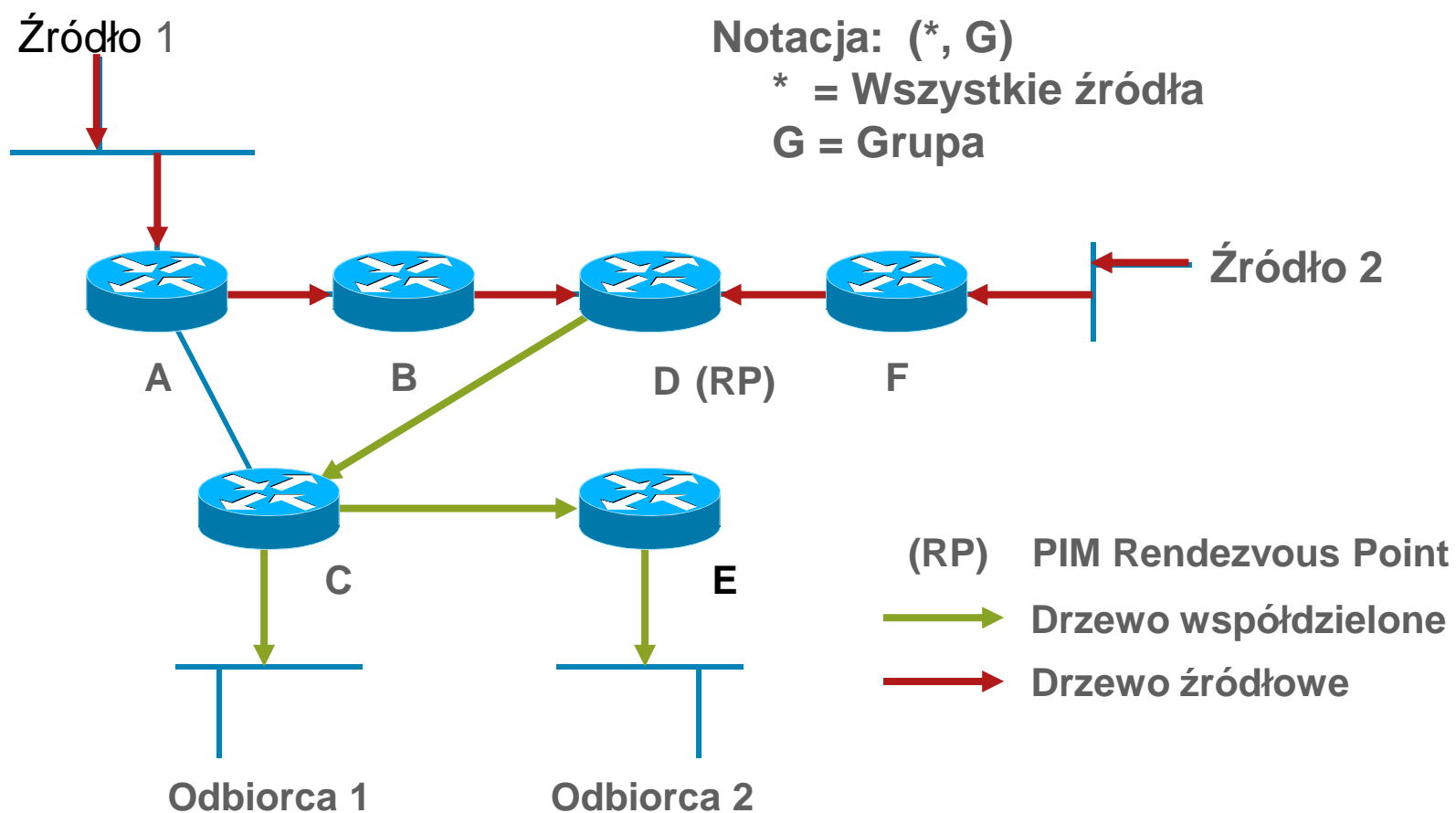
# Drzewa dystrybucyjne

„Drzewo najkrótszej ścieżki/źródłowe”



# Drzewa dystrybucyjne

## „Drzewo współdzielone”



# Drzewa dystrybucyjne

- ❖ Drzewa najkrótszej ścieżki/źródłowe
  - zużywają więcej pamięci  $O(S \times G)$ , ale zapewniają optymalną ścieżkę do źródła sygnału multicastowego, minimalizując opóźnienia
  - wiadomości kontrolne PIM są wysyłane na adres źródła
- ❖ Drzewa współdzielone
  - zużywają mniej pamięci  $O(G)$ , ale ruch jest przesyłany mniej optymalną ścieżką wprowadzając dodatkowe opóźnienia
  - wiadomości kontrolne PIM są wysyłane na adres RP
- ❖ W każdym przypadku drzewa budowane są za pomocą wiadomości PIM Join/Prune

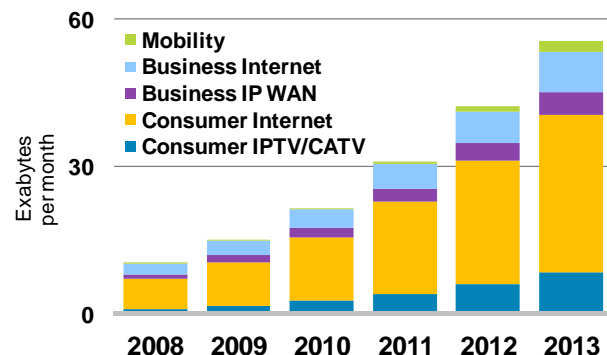
# Przesyłanie ruchu multicast

„Multicast Routing is backwards from Unicast Routing”

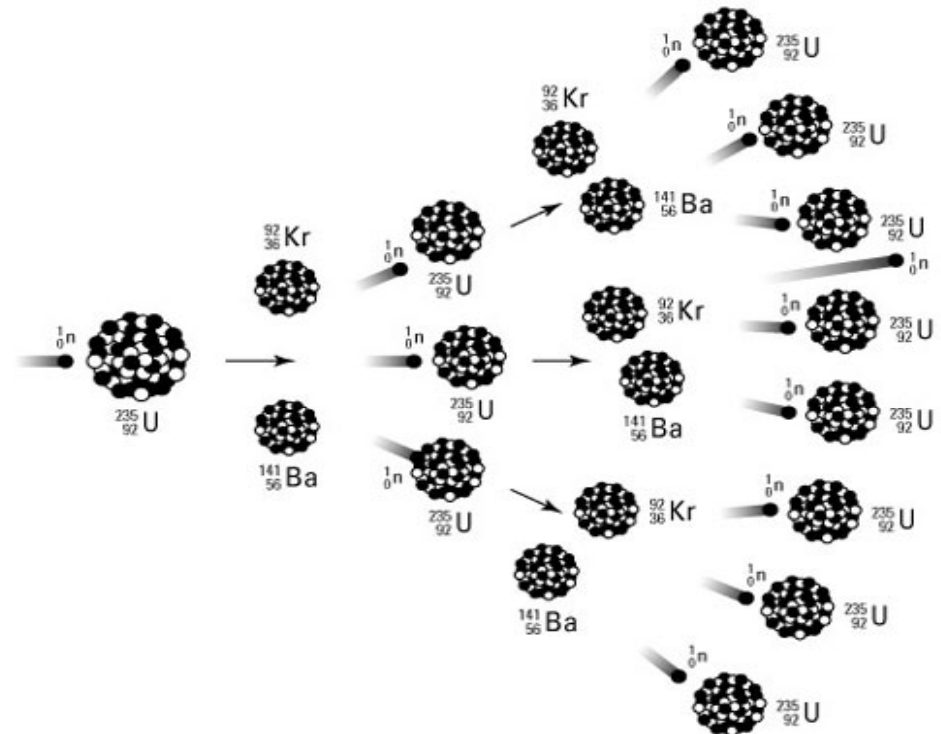
- ❖ Dla ruchu typu unicast transmisja odbywa się na podstawie informacji o docelowym adresie IP
  - adres ten jednoznacznie determinuje miejsce wysłania pakietu (interfejs, adres następnego routera)
  - przesyłanie odbywa się w trybie „hop-by-hop”
- ❖ Dla ruchu multicast istotny jest również adres źródła pakietu:
  - docelowy adres IP nie wskazuje jednoznacznie gdzie wysłać pakiet
  - przesyłanie jest zorientowane połączeniowo – odbiorcy muszą się najpierw „podłączyć” do odpowiedniego drzewa multicast’owego
  - algorytm budowy takich drzew oparty jest o specjalny protokół (np. PIM) i wykorzystuje informacje z tablicy routingu oraz mechanizm RPF (*Reverse Path Forwarding*)
  - drzewa dystrybucyjne umożliwiają stwierdzenie, gdzie należy przesyłać ruch multicastowy dla danej grupy i są budowane w sposób dynamiczny

# Gdzie możemy zastosować multicast?

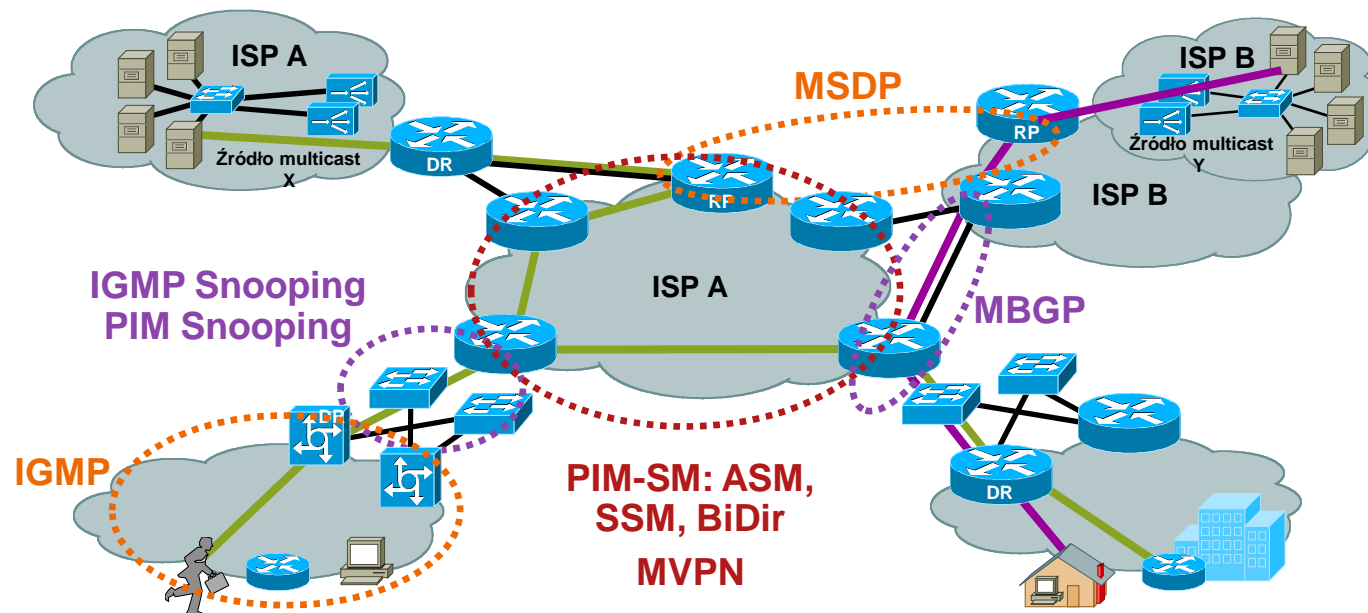
- ❖ Dowolna aplikacja z wieloma odbiorcami
  - jeden do wielu, wielu do wielu
- ❖ Dystrybucja „na żywo” sygnału video
- ❖ Systemy pracy grupowej
  
- ❖ Aplikacje dostarczające dane w sposób periodyczny – technologia “push”
  - notowania giełdowe, wyniki sportowe, gazety, ogłoszenia
- ❖ Replikacja serwerów/serwisów
- ❖ Redukcja zużycia zasobów sieci
  - więcej niż wiele połączeń typu punkt-punkt
- ❖ Wykrywanie zasobów sieciowych np. sąsiednich routerów
- ❖ Rozproszone, interaktywne symulacje (*DIS - distributed interactive simulation*)
  - gry wojenne
  - wirtualna rzeczywistość



# Podstawy technologii multicast



# Komponenty sieci multicast



## ***„Campus Multicast”***

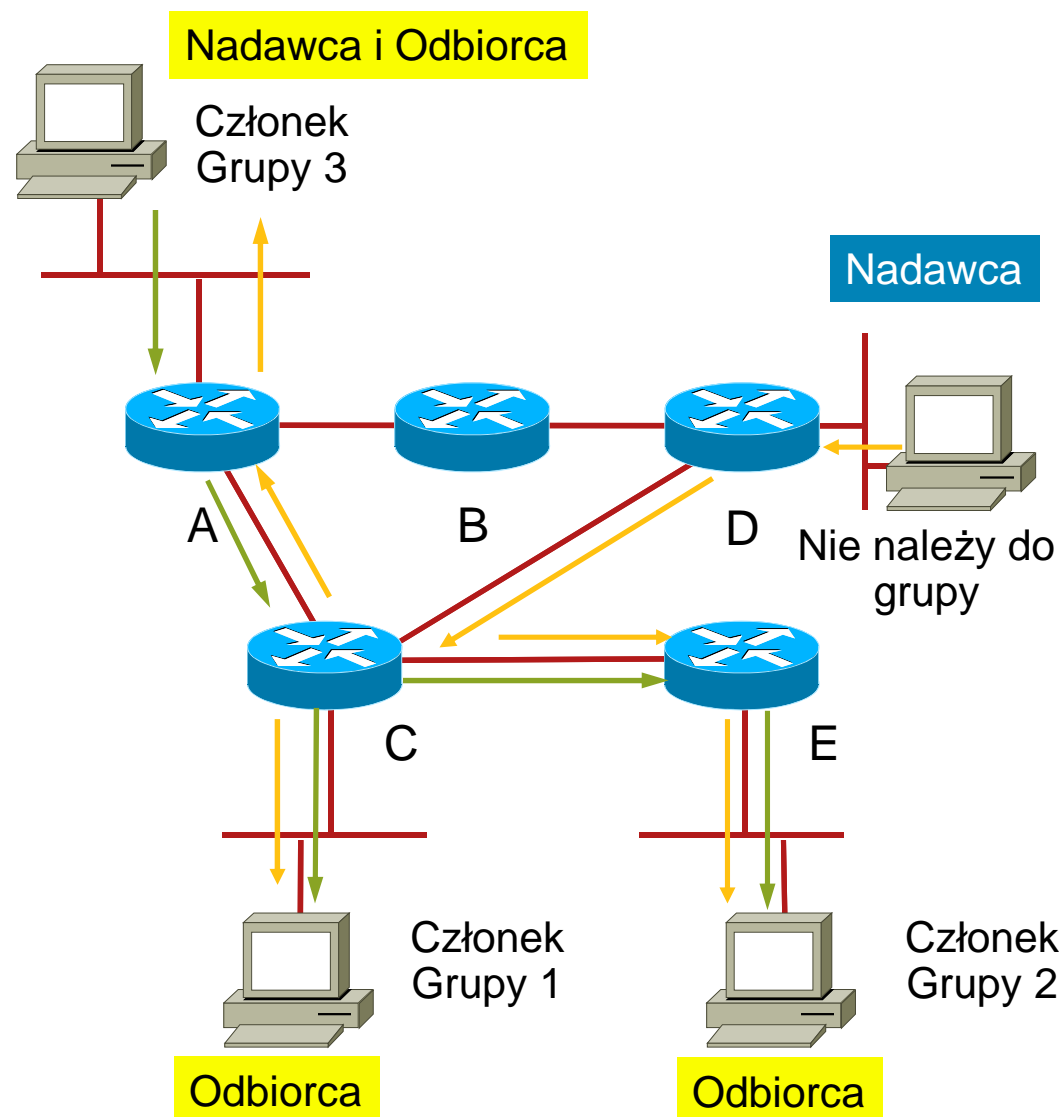
- Stacje końcowe (do routerów)
  - IGMP
- Przełączniki L2 (optymalizacja L2)
  - IGMP snooping i PIM snooping
- Routery (protokół przesyłania multicast’u)
  - PIM sparse mode lub bidirectional PIM

## ***„Interdomain Multicast”***

- Routing ruchu multicast pomiędzy domenami IP
  - MBGP
- Wykrywanie źródeł multicast’owych
  - MSDP z PIM-SM
- Source Specific Multicast
  - SSM

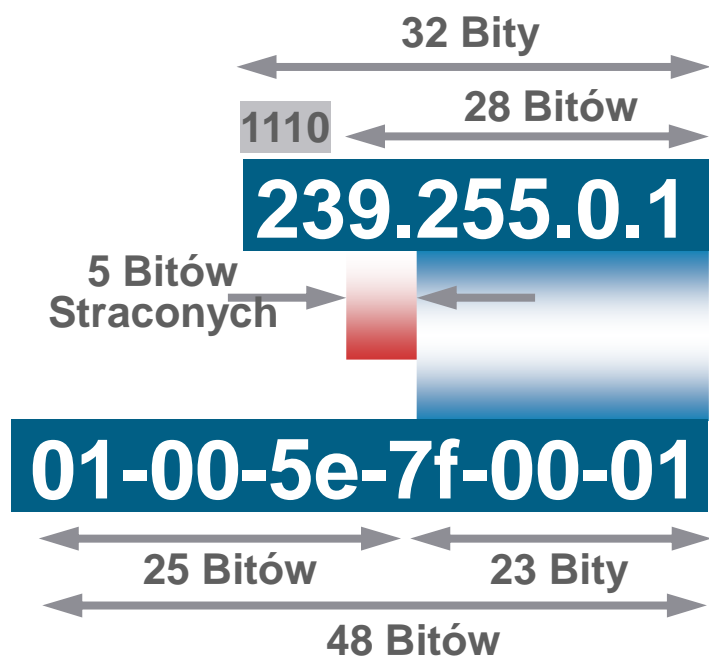
# Konceptcja grupy multicastowej IP

- ❖ Musisz być członkiem grupy by odbierać dane dla tej grupy
- ❖ Jeśli wysyłasz ruch na adres grupy to wszyscy członkowie danej grupy otrzymują te dane
- ❖ Nie musisz należeć do danej grupy by móc nadawać dane do jej członków



# Adresacja ruchu multicast – warstwa 2

**Jakiego adresu MAC użyć dla danej grupy?**



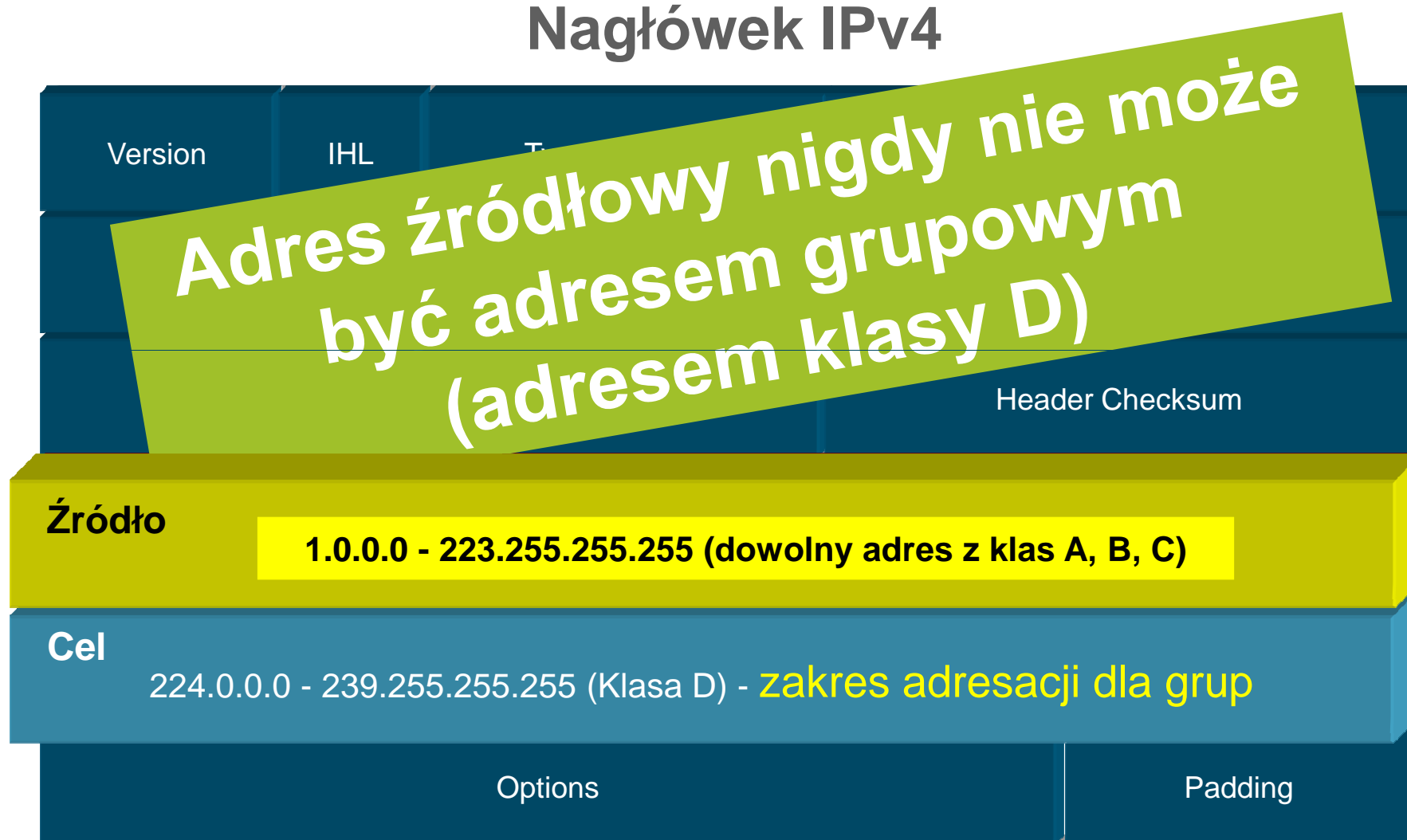
32–adresy IP Multicast



**Uwaga: adresy 32 grup multicastowych są mapowane do tego samego adresu MAC**

# Adresacja ruchu multicast w sieci IPv4

## Nagłówek IPv4



# Adresacja ruchu multicast - Klasa D

- ❖ **Klasa D:** 224/4 = 224.0.0.0 - 239.255.255.255
- ❖ **Zarezerwowane adresy o znaczeniu lokalnym (*link-local*)**
  - 224.0.0.0–224.0.0.255
  - Pakiety zawsze wysyłane z wartością TTL (*time to live*) = 1
  - Przykłady
    - 224.0.0.1      wszystkie systemy w danej podsieci
    - 224.0.0.2      wszystkie routery w danej podsieci
    - 224.0.0.5      wszystkie routery OSPF
    - 224.0.0.13     wszystkie routery PIMv2
    - 224.0.0.22     IGMPv3
- ❖ **Inne zarezerwowane adresy**
  - 224.0.1.0–224.0.1.255
  - Pakiety transmitowane z wartością TTL > 1
  - Przykłady
    - 224.0.1.1      NTP (Network Time Protocol)
    - 224.0.1.32      Mtrace
    - 224.0.1.78      Tibco Multicast1

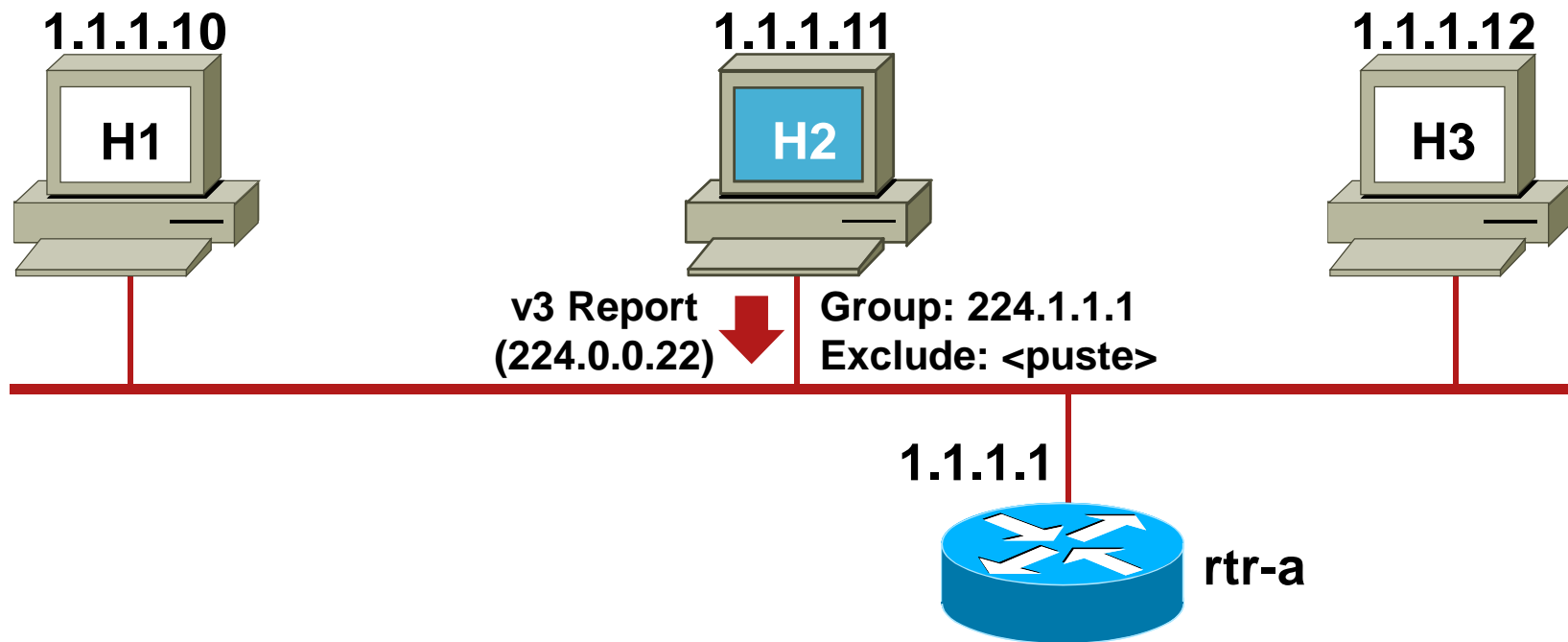
# Adresacja ruchu multicast - Klasa D

- ❖ **Zakres typu „administratively scoped addresses”**
  - 239.0.0.0–239.255.255.255
  - zakres do prywatnego użytku
    - podobny do zakresów opisanych w RFC1918 dla adresów typu unicast
    - nie używany w globalnej sieci Internet
- ❖ **Zakres dla SSM (*Source Specific Multicast*)**
  - 232.0.0.0–232.255.255.255
  - planowany do wykorzystania dla transmisji internetowych
- ❖ **GLOP - RFC 2770**
  - 233.0.0.0–233.255.255.255
  - Zapewnia unikalny zakres (/24) adresów grupowych bazujących na numerze autonomicznym (*AS – autonomus system*) nadawcy

```
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|      233      |          16 bits AS          | local bits |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
```

# Protokół IGMPv3

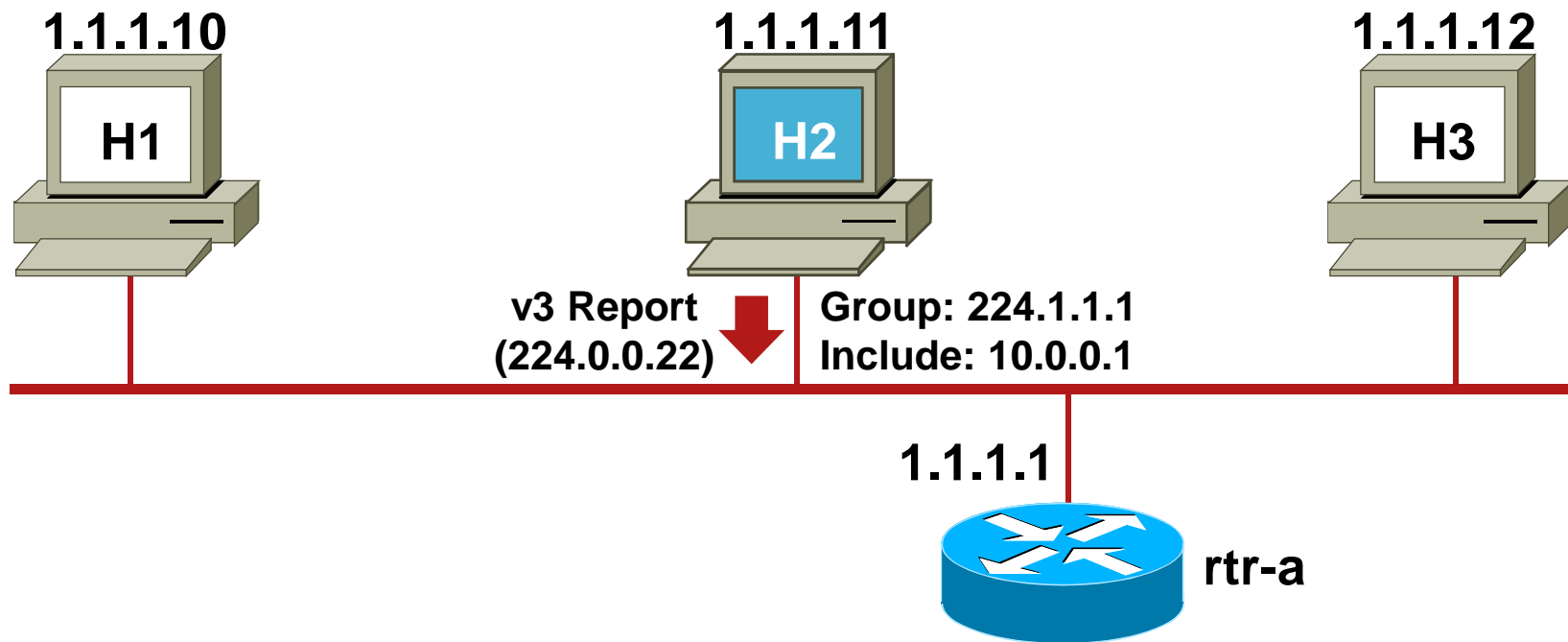
## Subskrypcja do grupy



- ❖ Internet Group Management Protocol (IGMP) służy do komunikacji pomiędzy odbiorcami a routerami celem ustalenia przynależności odbiorców do grup
- ❖ Odbiorca rejestrujący się do grupy wysyła wiadomość *IGMPv3 Report* na adres 224.0.0.22

# Protokół IGMPv3

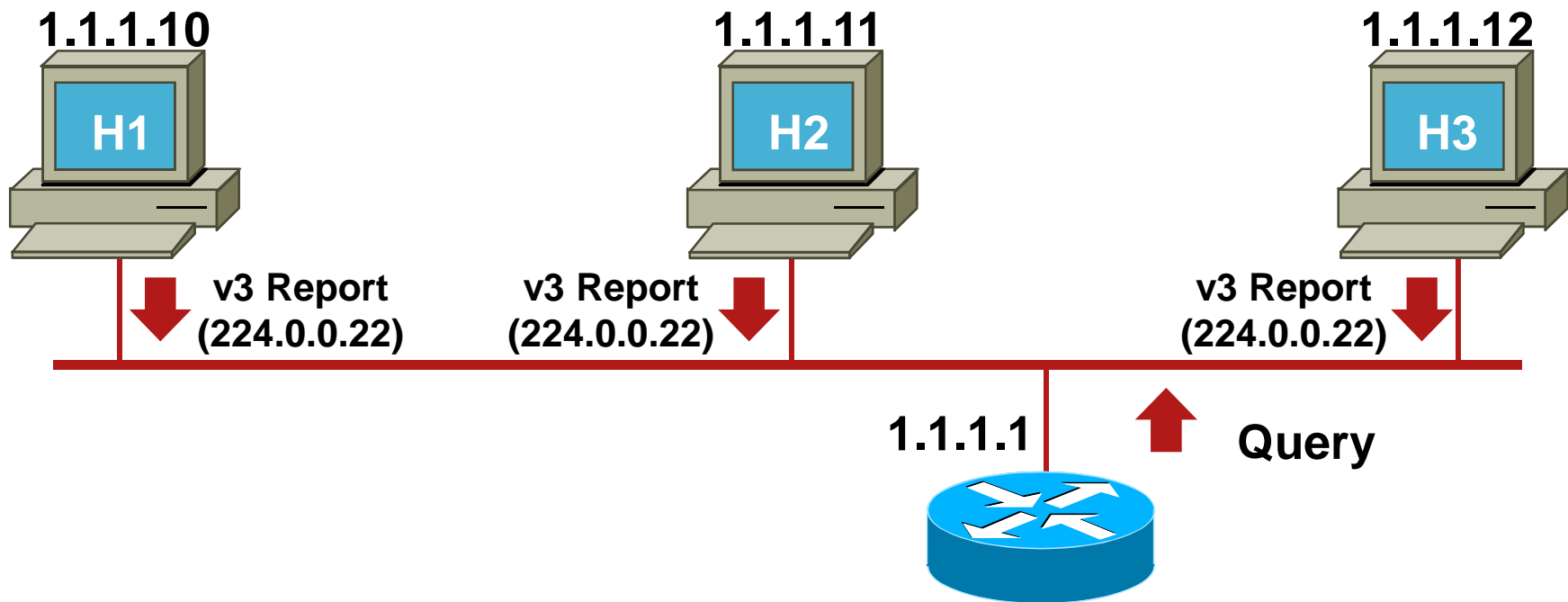
## Subskrypcja do grupy i konkretnego źródła



- ❖ Wiadomość IGMPv3 Report zawiera wymienione konkretne źródła, od których odbiorca chciałby wyłącznie odbierać ruch

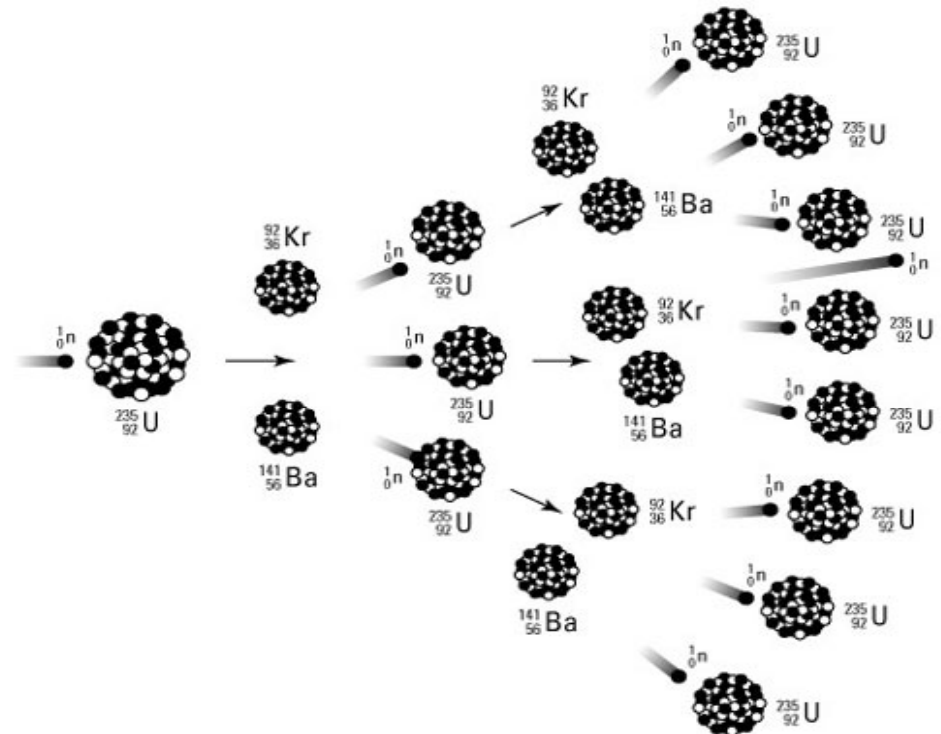
# Protokół IGMPv3

## Sprawdzanie stanu członków grupy



- ❖ Router wysyła cyklicznie wiadomości typu Query na adres 224.0.0.1
- ❖ Wszyscy odbiorcy IGMPv3 odsyłają wiadomość IGMP Membership Report z pełną listą grup, do której są zarejestrowani

# Protokoły PIM



# Stan protokołu PIM

- ❖ Jest to **stan grup multicastowych** w rozumieniu routera, na którym wykonywane jest polecenie
- ❖ Reprezentowany przez wpisy w tablicy routingu multicastowego (mroute)
  - używana przy **podejmowaniu decyzji o przekazywaniu ruchu dalej**
  - składa się z wpisów (\*, G) i (S, G)
  - każdy wpis **zawiera informację RPF**
    - interfejs „wejściowy”
    - sąsiad RPF (router upstream)
  - każdy wpis zawiera **interfejsy wyjściowe**
    - OIL (Outgoing Interface List) – może być pusty (NULL)

# Przykład stanu tabeli mroute

```
R2# show ip mroute
Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,
      L - Local, P - Pruned, R - RP-bit set, F - Register flag,
      T - SPT-bit set, J - Join SPT, M - MSDP created entry,
      X - Proxy Join Timer Running, A - Candidate for MSDP Advertisement,
      U - URD, I - Received Source Specific Host Report
Outgoing interface flags: H - Hardware switched
Timers: Uptime/Expires
Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 224.1.1.1), 2w1d/00:00:00, RP 172.16.25.1, flags: SJC
  Incoming interface: Serial0/1, RPF nbr 172.16.4.1
  Outgoing interface list:
    Ethernet0/1, Forward/Sparse-Dense, 2w1d/00:01:40
    Serial0/0, Forward/Sparse-Dense, 00:4:52/00:02:08

(172.16.8.2, 224.1.1.1), 00:00:10/00:02:59, flags: CJT
  Incoming interface: Serial0/1, RPF nbr 172.16.4.1
  Outgoing interface list:
    Ethernet0/1, Forward/Sparse-Dense, 00:00:10/00:02:49
    Serial0/0, Forward/Sparse-Dense, 00:4:52/00:02:08
```

# Protokoły PIM

## ❖ Tryb dense (gęsty) PIM-DM

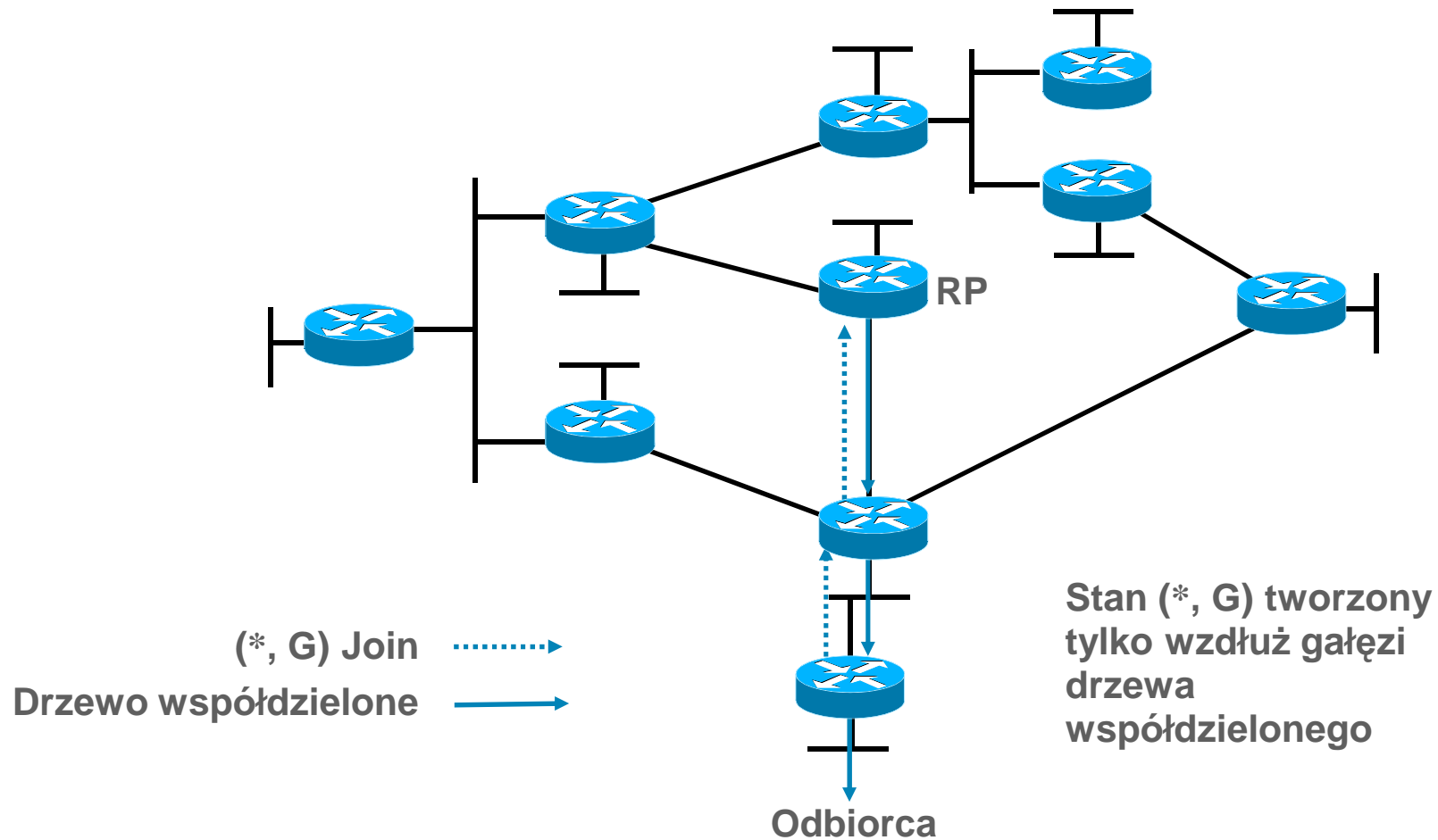
- używa modelu w którym ruch wysyłany jest **domyślnie do wszystkich** i zatrzymywany dopiero gdy wprost sobie to zażyczymy
- **cykl się powtarza** (domyślnie w ciągu 3 minut)

## ❖ Tryb rzadki (sparse) PIM-SM

- używa modelu w którym ruch wysyłany **jest tylko na żądanie**
- wprost **sygnalizowana chęć dołączenia** do drzewa dystrybucji
- używa Rendezvous Point (RP) - nadawcy i odbiorcy „spotykają się” w tym punkcie by powiadomić się o swoim istnieniu
- różne warianty:
  - ASM - Any Source Multicast/RP/SPT/drzewo współdzielone
  - SSM - Source Specific Multicast, brak RP, tylko drzewo źródłowe (SPT)
  - BiDir - Bidirectional PIM, brak SPT, tylko drzewo współdzielone

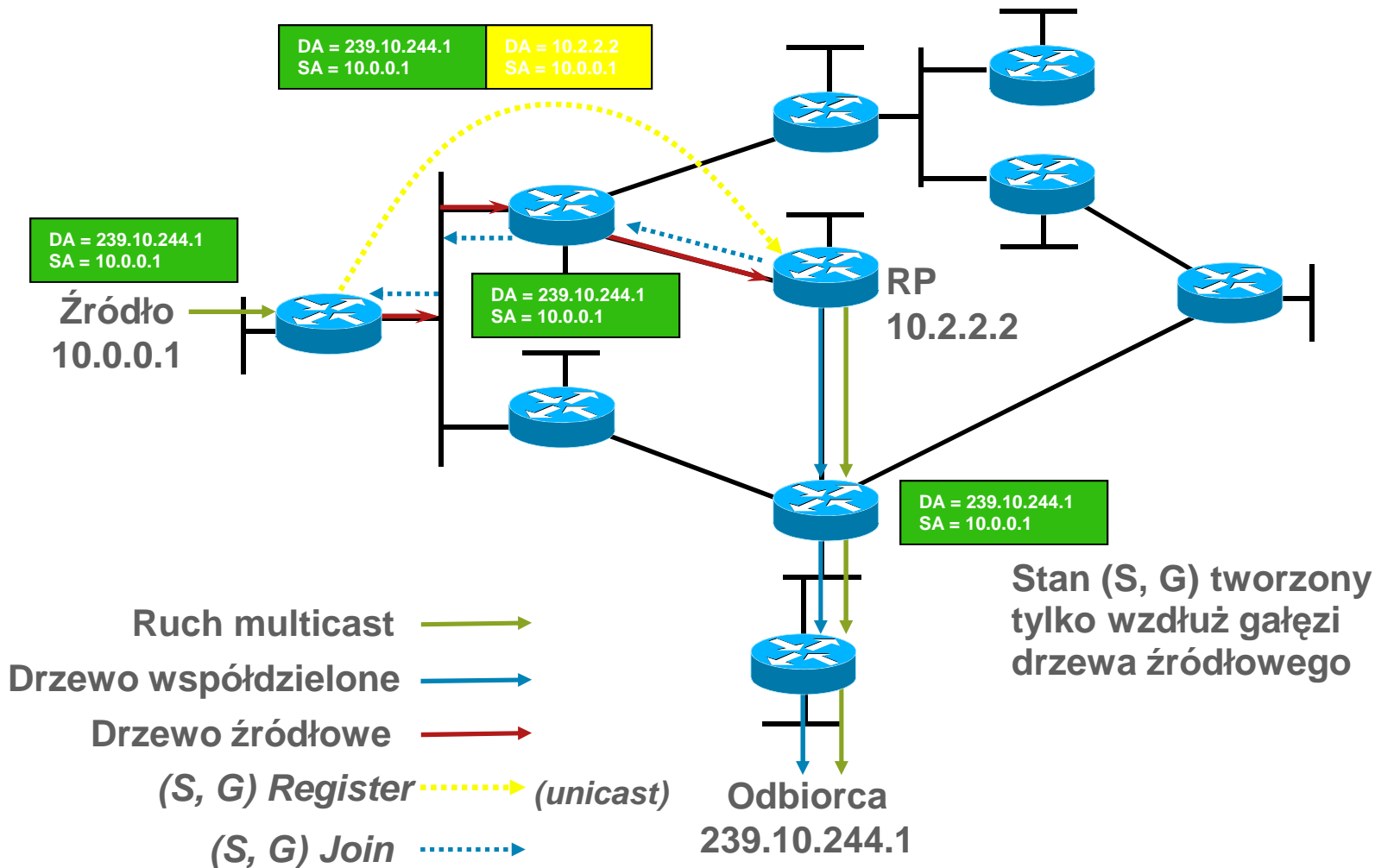
# Protokół PIM-Sparse Mode

## Rejestracja odbiorcy



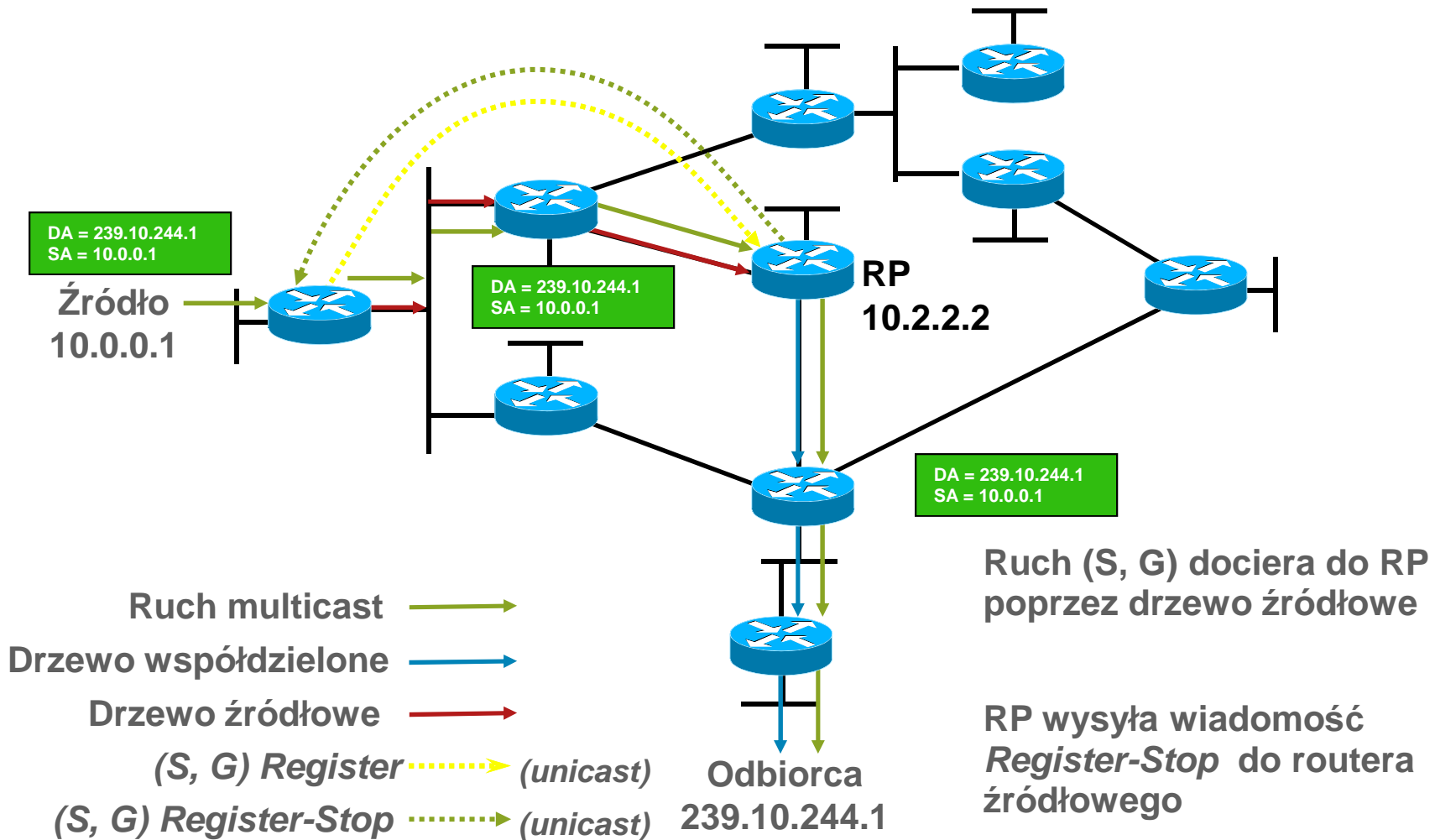
# Protokół PIM-Sparse Mode

## Rejestracja nadawcy (1)



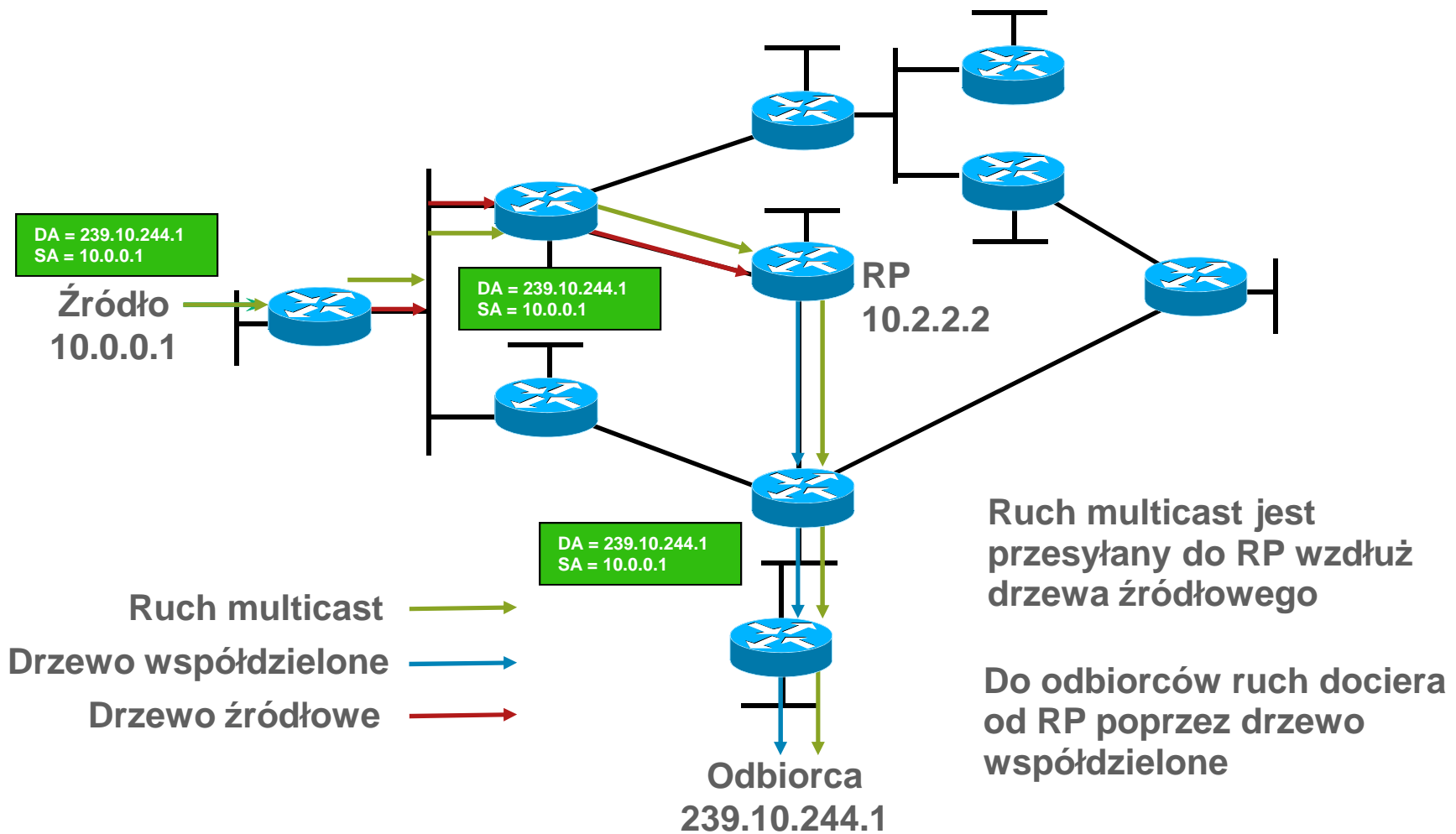
# Protokół PIM-Sparse Mode

## Rejestracja nadawcy (2)



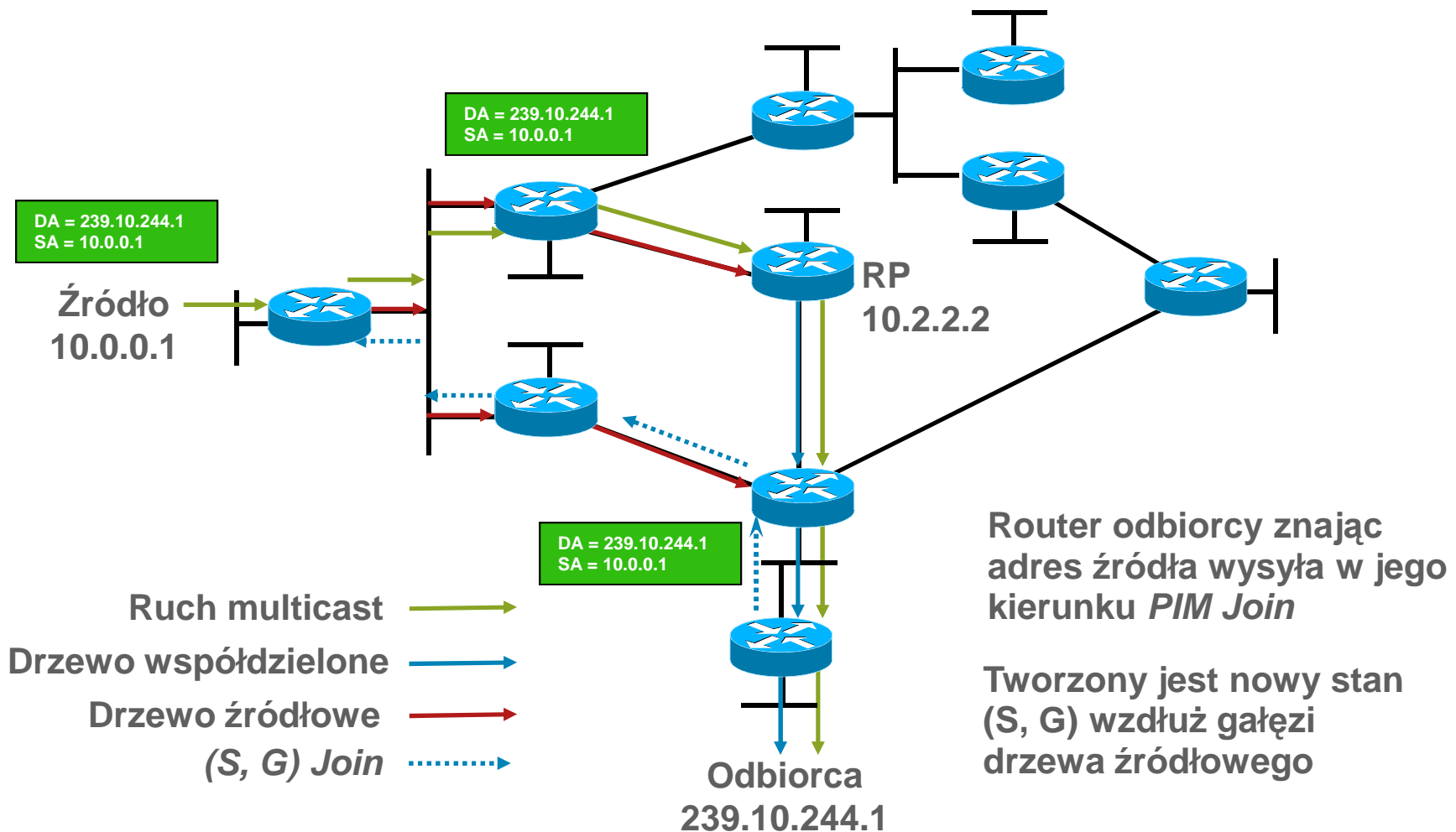
# Protokół PIM-Sparse Mode

## Rejestracja nadawcy (3)



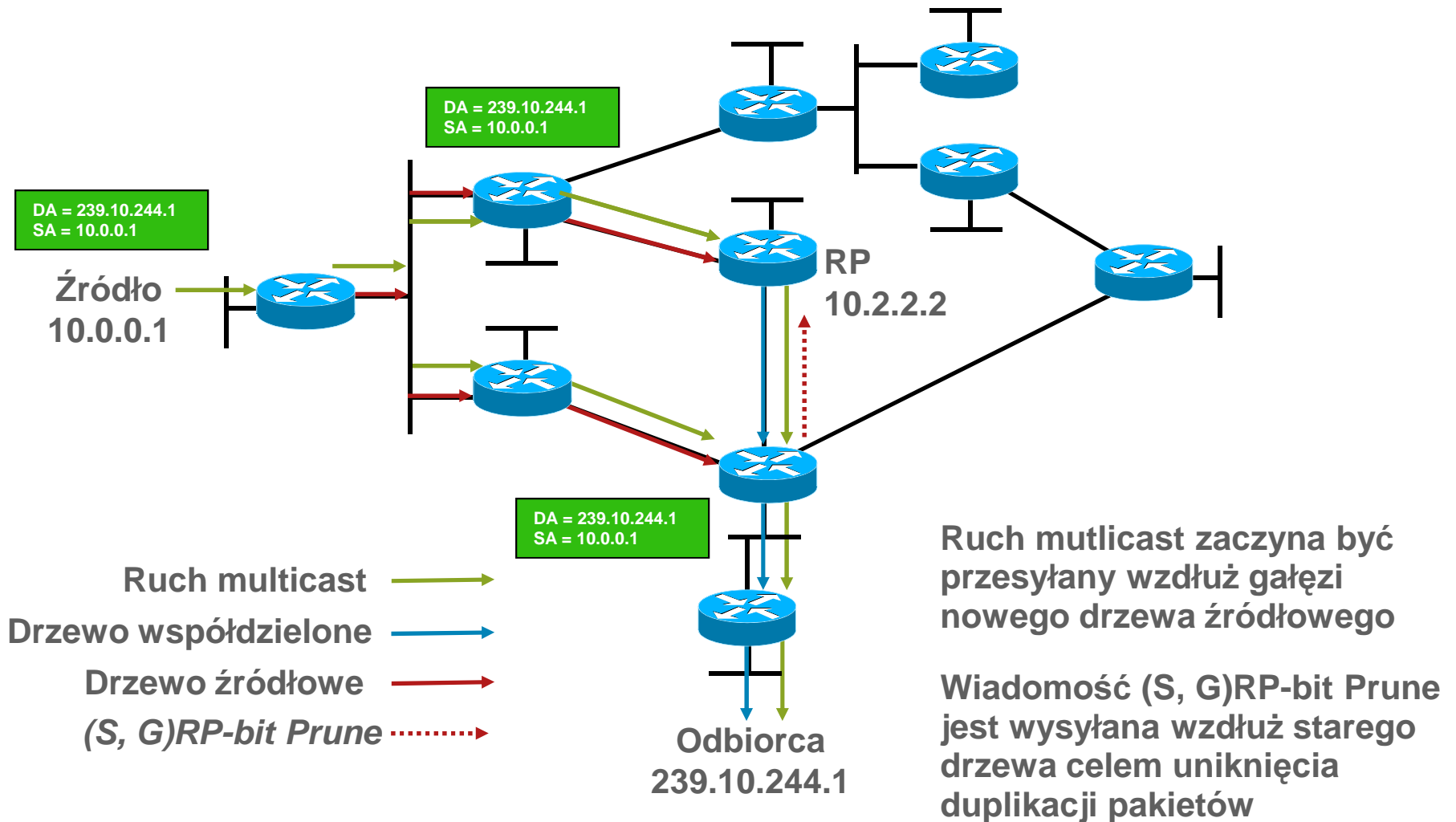
# Protokół PIM-Sparse Mode

## Przełączanie na drzewo źródłowe (1)



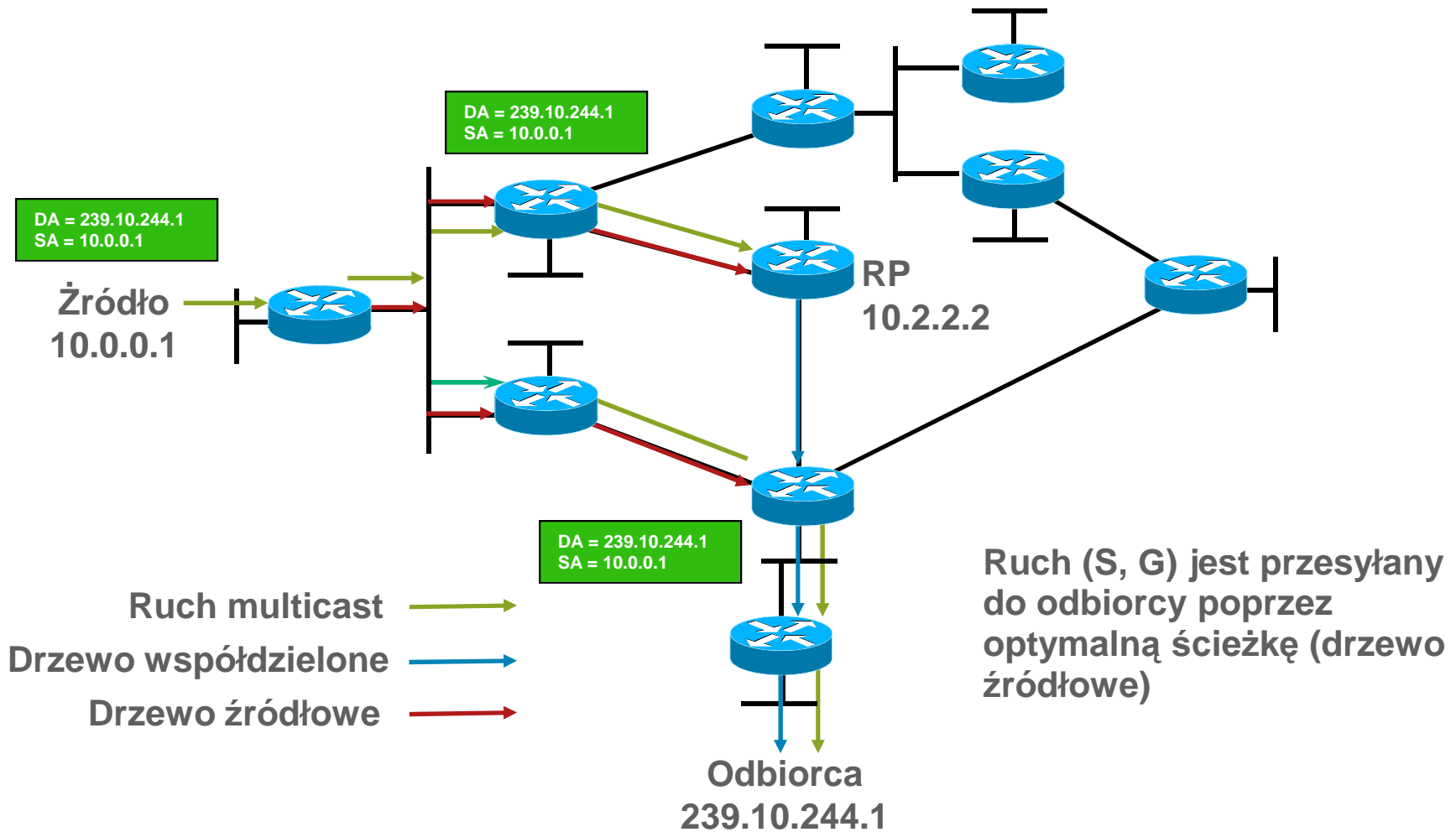
# Protokół PIM-Sparse Mode

## Przełączanie na drzewo źródłowe (2)



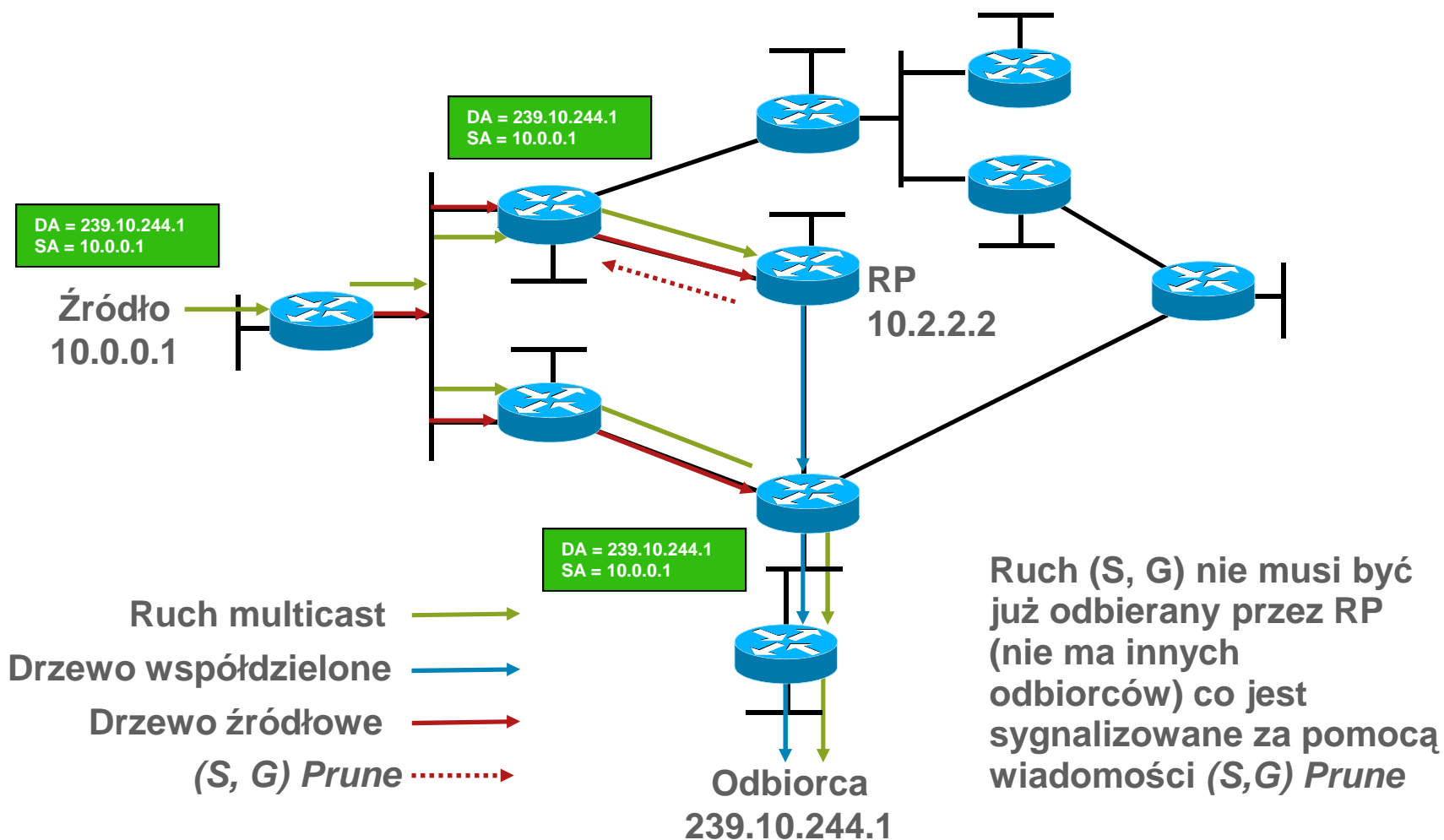
# Protokół PIM-Sparse Mode

## Przełączanie na drzewo źródłowe (3)



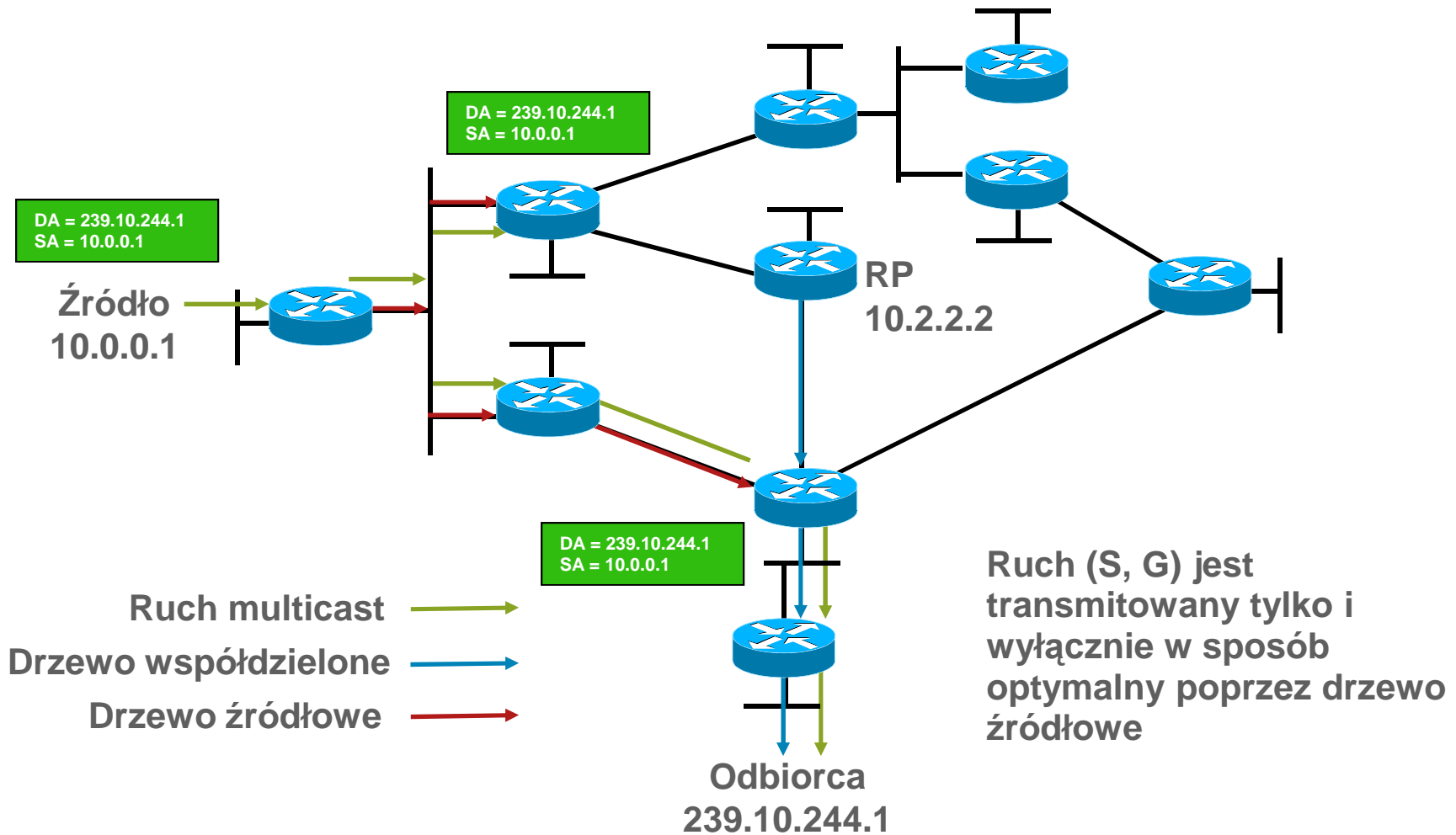
# Protokół PIM-Sparse Mode

## Przełączanie na drzewo źródłowe (4)



# Protokół PIM-Sparse Mode

## Przełączanie na drzewo źródłowe (5)



# Protokół PIM-Sparse Mode

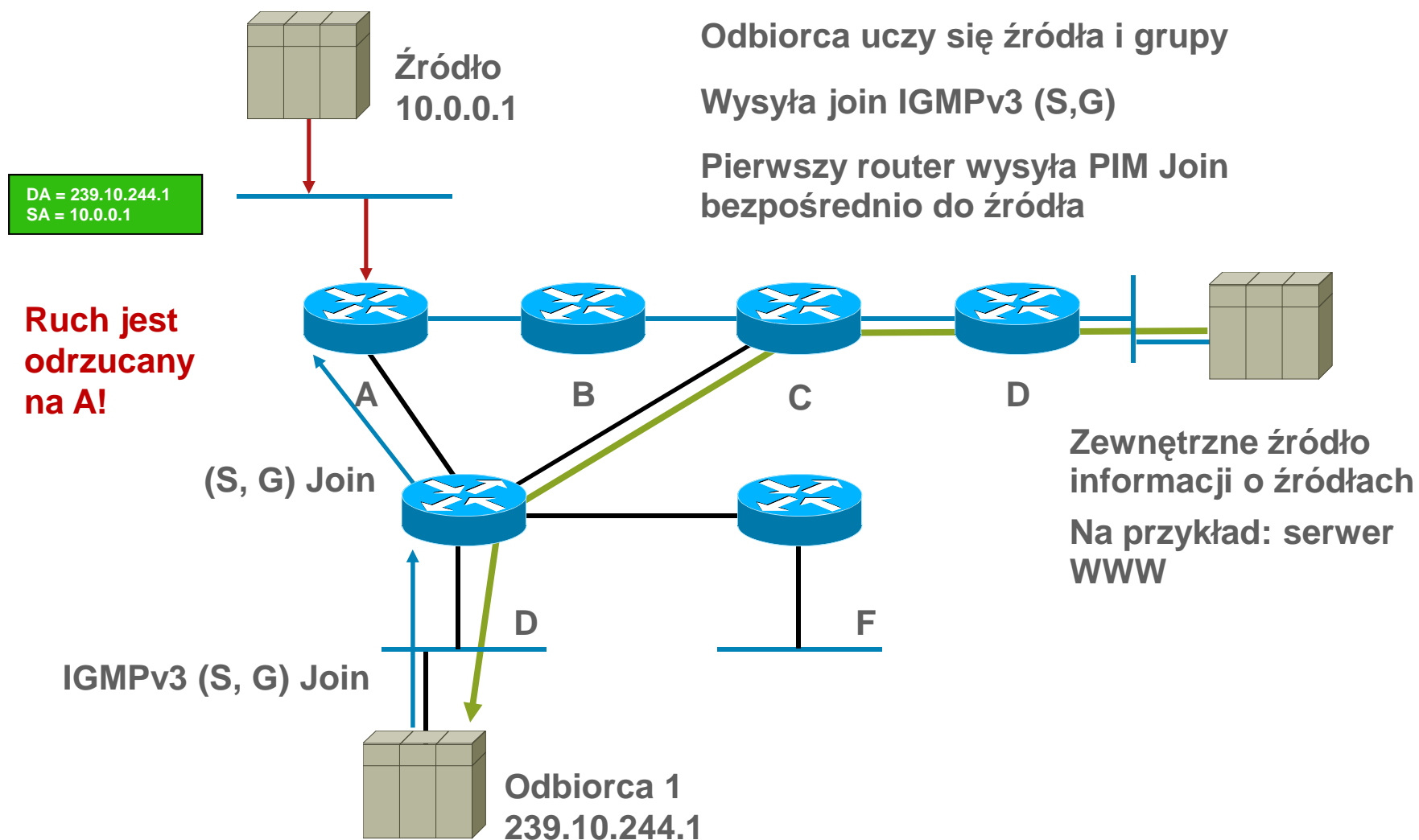
## Podsumowanie

- ❖ Efektywny do dystrybucji multicastów dla **rzadko rozproszonych w sieci** odbiorców
- ❖ Zalety
  - ruch wysyłany jest tylko do odbiorców, którzy wprost wyrazili zainteresowanie treścią – **oszczędzamy pasmo**
  - może **przełączyć sposób konstrukcji drzewa** (a zatem – ścieżkę przez sieć) dla źródeł nadających dużą ilość danych
  - **niezależny od protokołu routingu unicastowego**

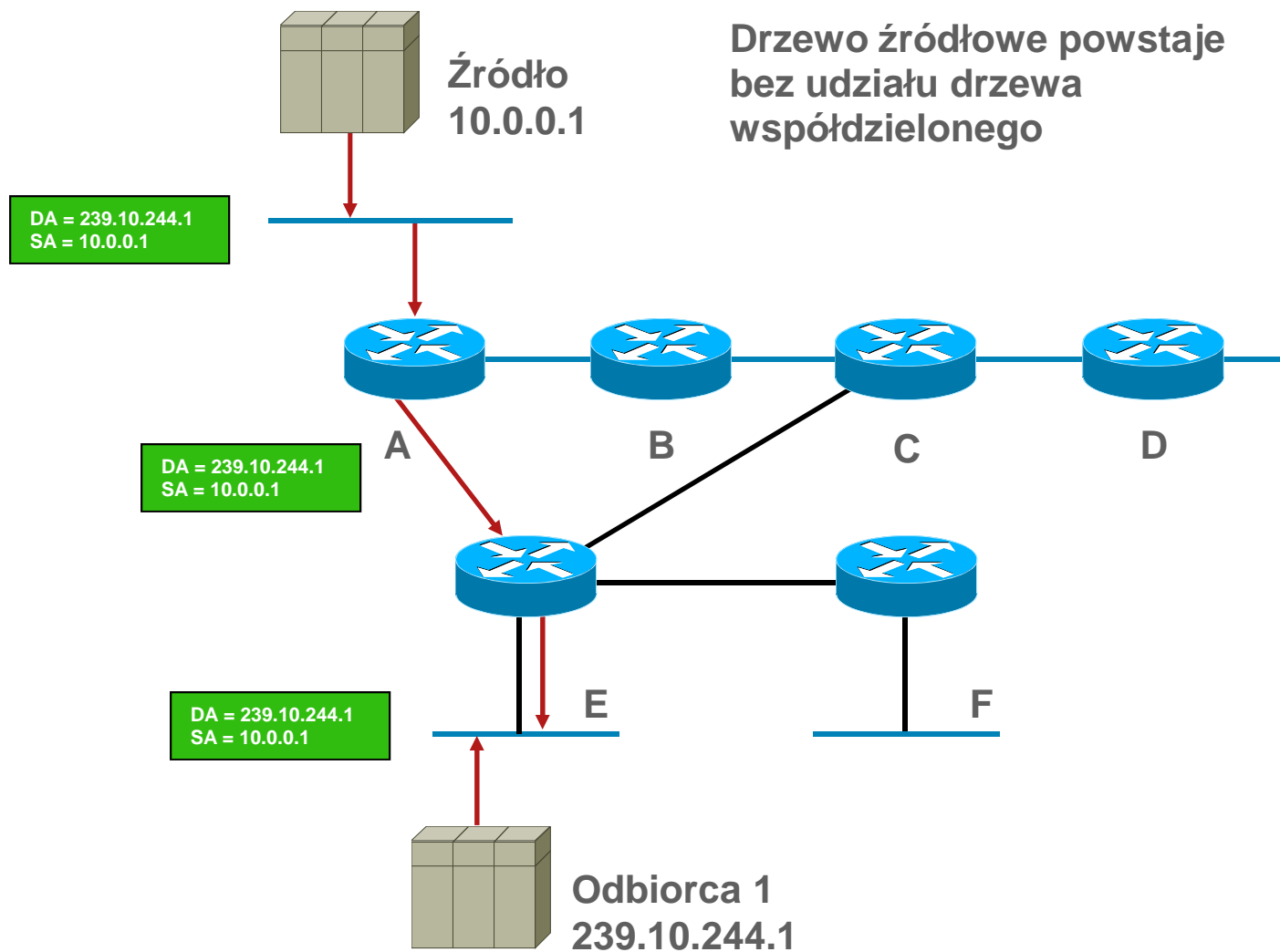
# Protokół PIM SSM - Source Specific Multicast

- ❖ Po co w modelu ASM potrzebujemy współdzielonego drzewa?
  - hosty i router, do którego są **podłączone muszą nauczyć się, gdzie znajduje się aktywne źródło ruchu** dla grupy
- ❖ A co gdybyśmy od razu znali adres źródła które chcemy odbierać?
  - ❖ host może użyć IGMPv3 by zasygnalizować **zapotrzebowanie dla konkretnej pary (S,G)**
  - ❖ współdzielone drzewo i RP nie są w tej sytuacji potrzebne, **różne źródła mogą współdzielić ten sam numer grupy i nie przeszkadzać sobie** na wzajem
- ❖ ...tak powstał: Source Specific Multicast (SSM)
- ❖ RFC 3569: An Overview of Source Specific Multicast (SSM)

# Tryb PIM Source Specific



# Tryb PIM Source Specific



# PIM SSM

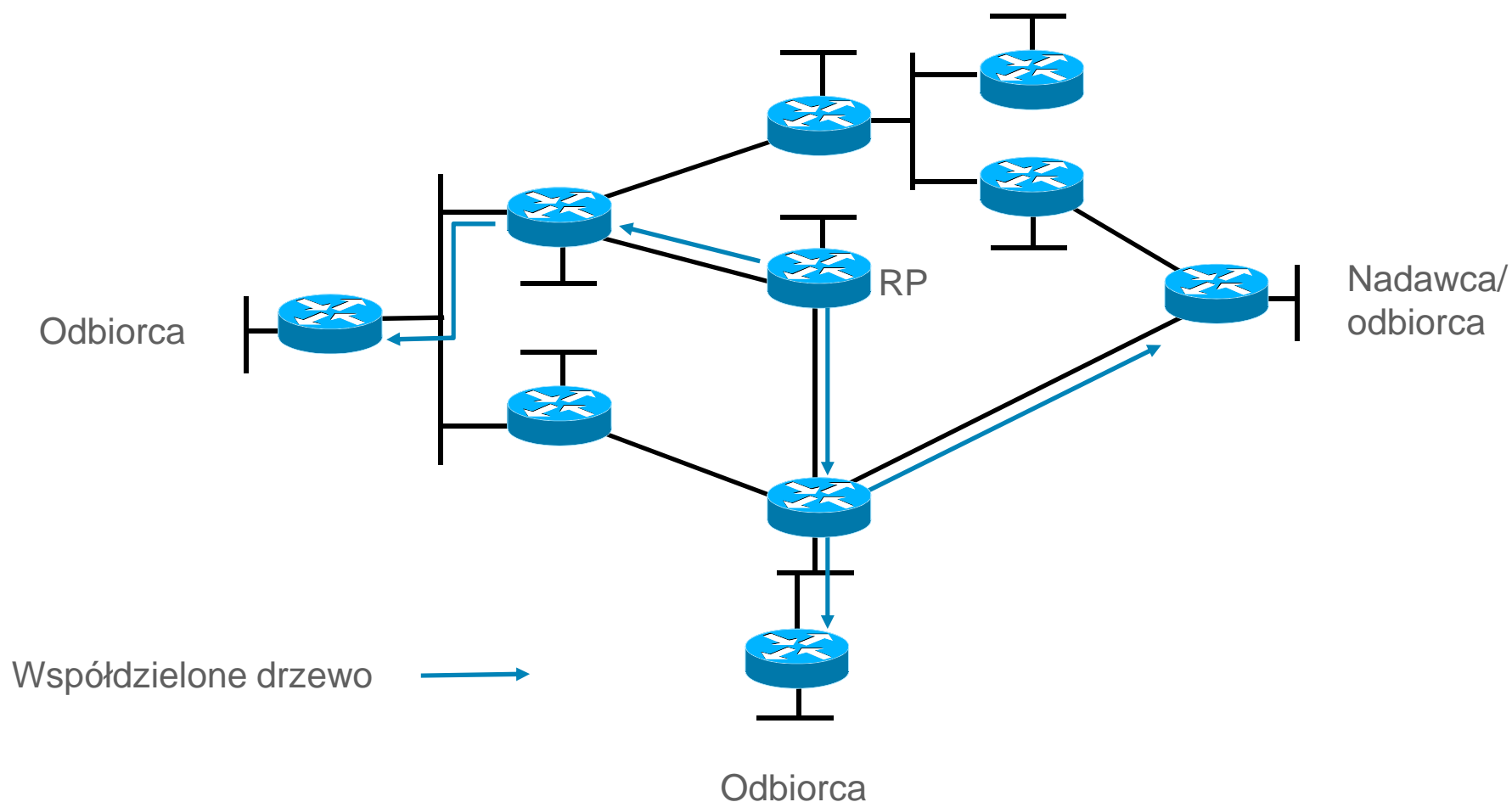
## Podsumowanie

- ❖ Idealny dla aplikacji, w **których jedno źródło wysyła do wielu odbiorców**
- ❖ Używa **uproszczonego protokołu PIM-SM**
- ❖ „Rozwiązuje” problem przydziału adresów multicastowych
  - ruch danych oddzielony **zarówno przez źródło jak i grupę** (nie tylko przez grupę jak w przypadku SM)
  - dostawcy treści mogą używać tych **samych numerów grup** – (S,G) jest unikalne
- ❖ Umożliwia lepsze **zapobieganie atakom DoS**
  - „fałszywe” źródło ruchu nie może nadawać do grupy – nawet jeśli będzie nadawać, być może nie trafi w konkretną parę (S,G)

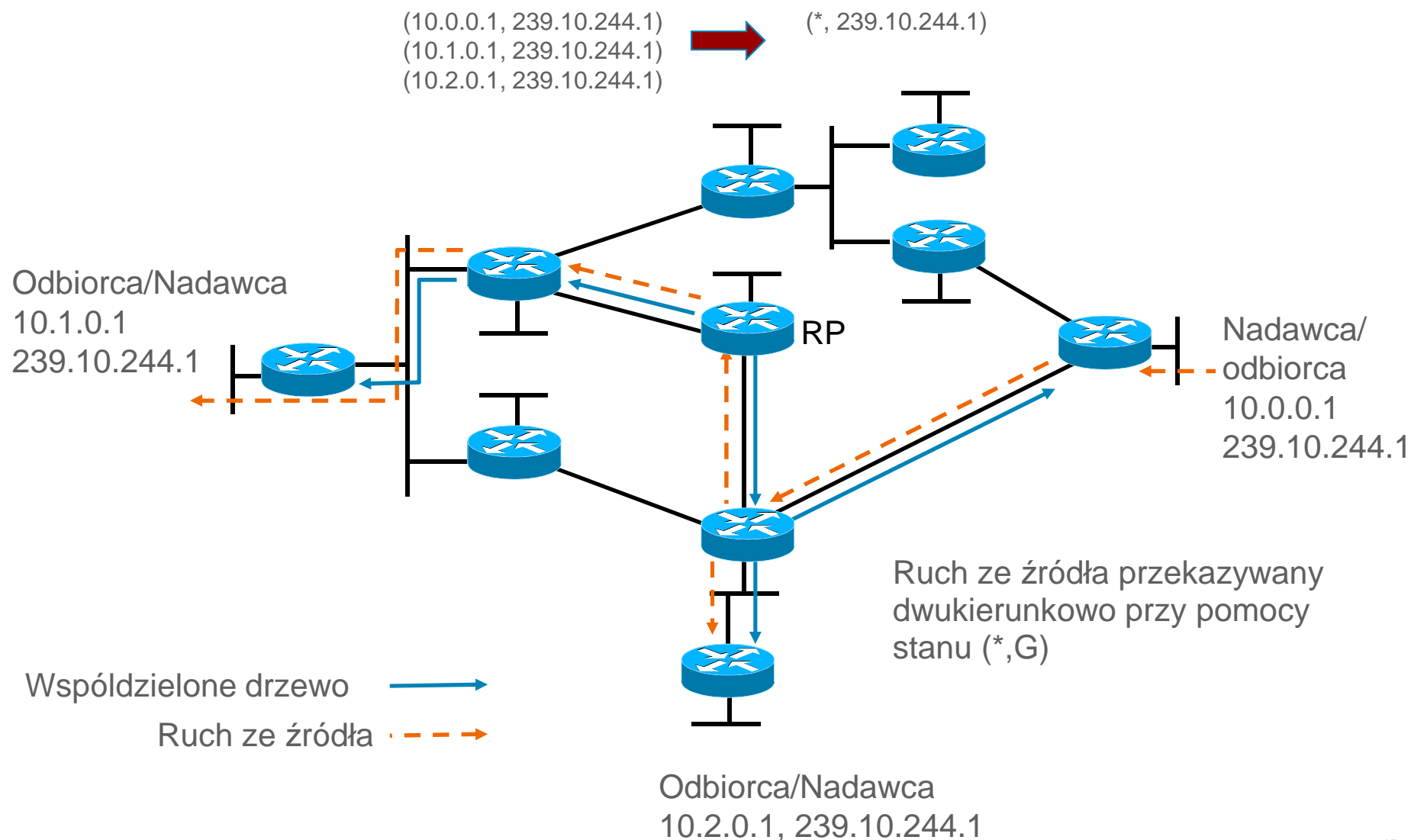
# A co z ruchem multicast typu wiele-do-wielu?

- ❖ Tworzy ogromne tablice (S,G)
  - problem z utrzymaniem **dużych tablic stanów** multicastowych staje się złożony nawet dla platform z szybką pamięcią podręczną
  - **duża ilość interfejsów wyjściowych** (OIL) pogarsza problem w platformach sprzętowych
- ❖ Użycie drzew współdzielonych nieco łagodzi problem
  - **redukcja ilości stanów** (S, G)
  - stan (S, G) tylko wzdłuż drzewa źródłowego do RP
  - niestety nadal zwykle oznacza to **zbyt dużo wpisów** (S, G)

# Bidirectional PIM—Jak to wygląda?



# Bidirectional PIM—Jak to wygląda?

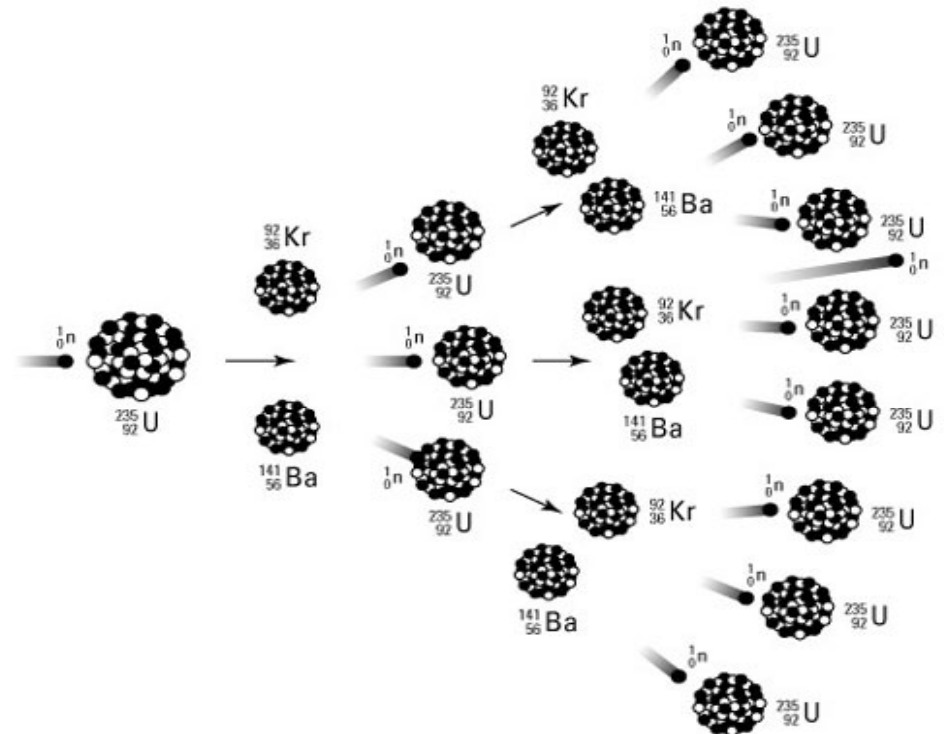


# Bidir PIM

## Podsumowanie

- ❖ W znaczny sposób minimalizuje ilość wpisów w tablicy multicastowej
  - eliminuje **wszystkie stany (S,G)** w sieci
    - **drzewa współdzielone** między źródłami i RP również **są wyeliminowane**
    - ruch ze źródła trafia zarówno „**w górę**” jak i „**w dół**” drzewa **współdzielonego**
  - aplikacje wiele-do-wielu mogą się łatwo skalować
    - dowolna liczba źródeł nie wprowadza **dodatkowego obciążenia dla sieci**

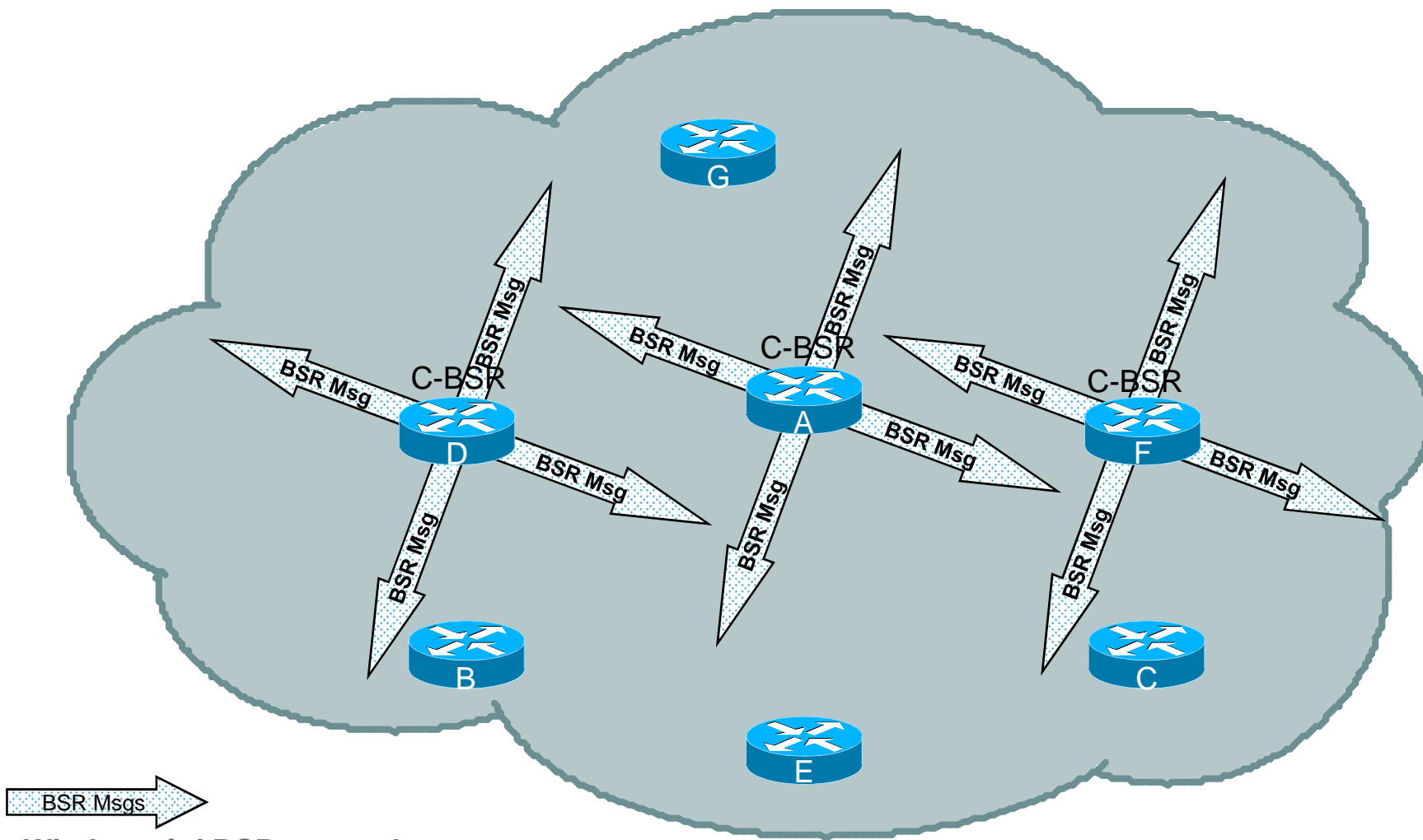
# Punky Rendezvous (RP)



# Skąd sieć wie gdzie jest RP?

- ❖ Konfiguracja statyczna
  - ręcznie na każdym routerze w domenie PIM
- ❖ AutoRP
  - rozwiązanie firmowe Cisco
  - pozwoliło rozpowszechnić zastosowanie PIM-SM
- ❖ PIMv2 BSR
  - draft-ietf-pim-sm-bsr – „standard”

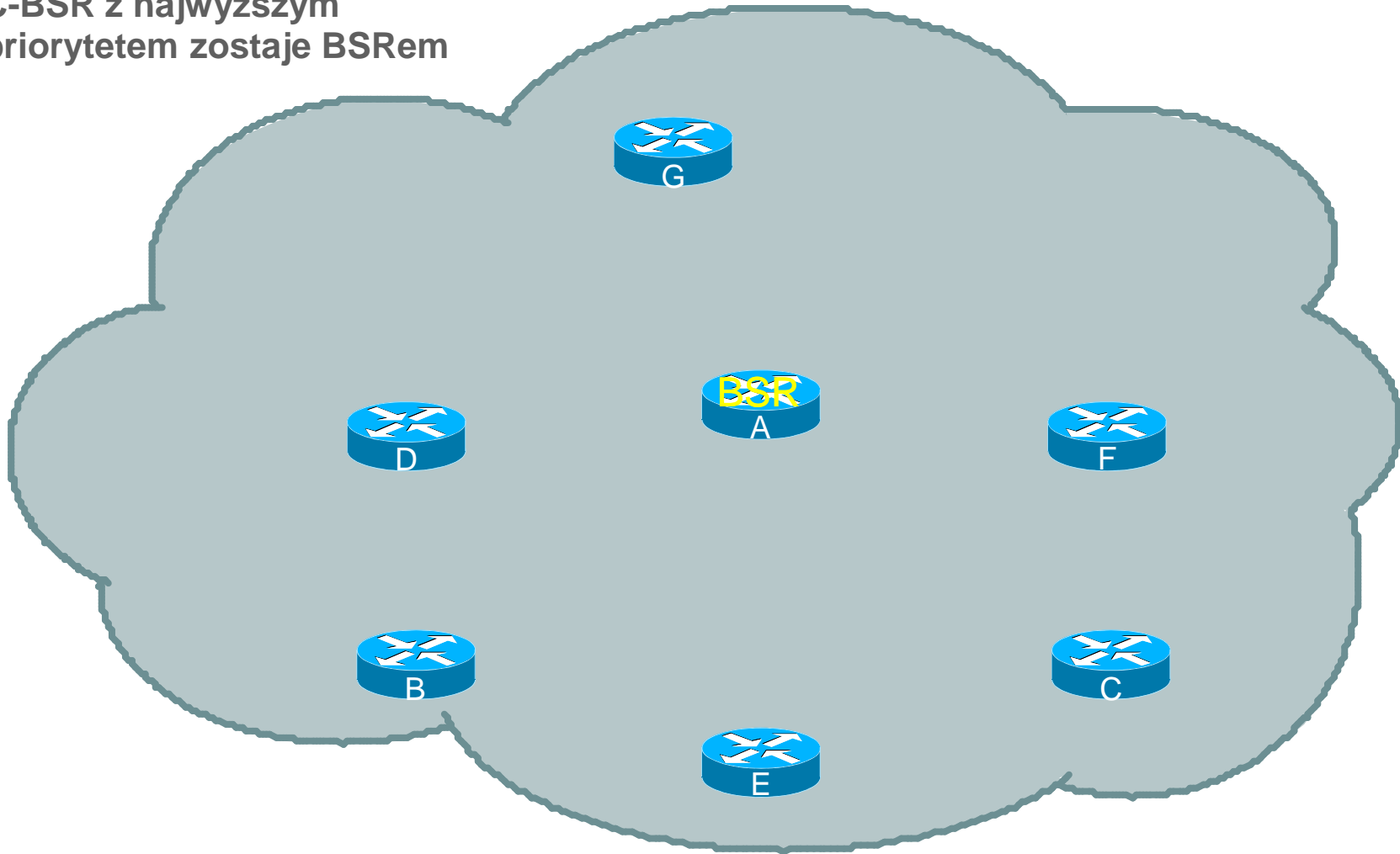
# BSR—z 10,000 metrów



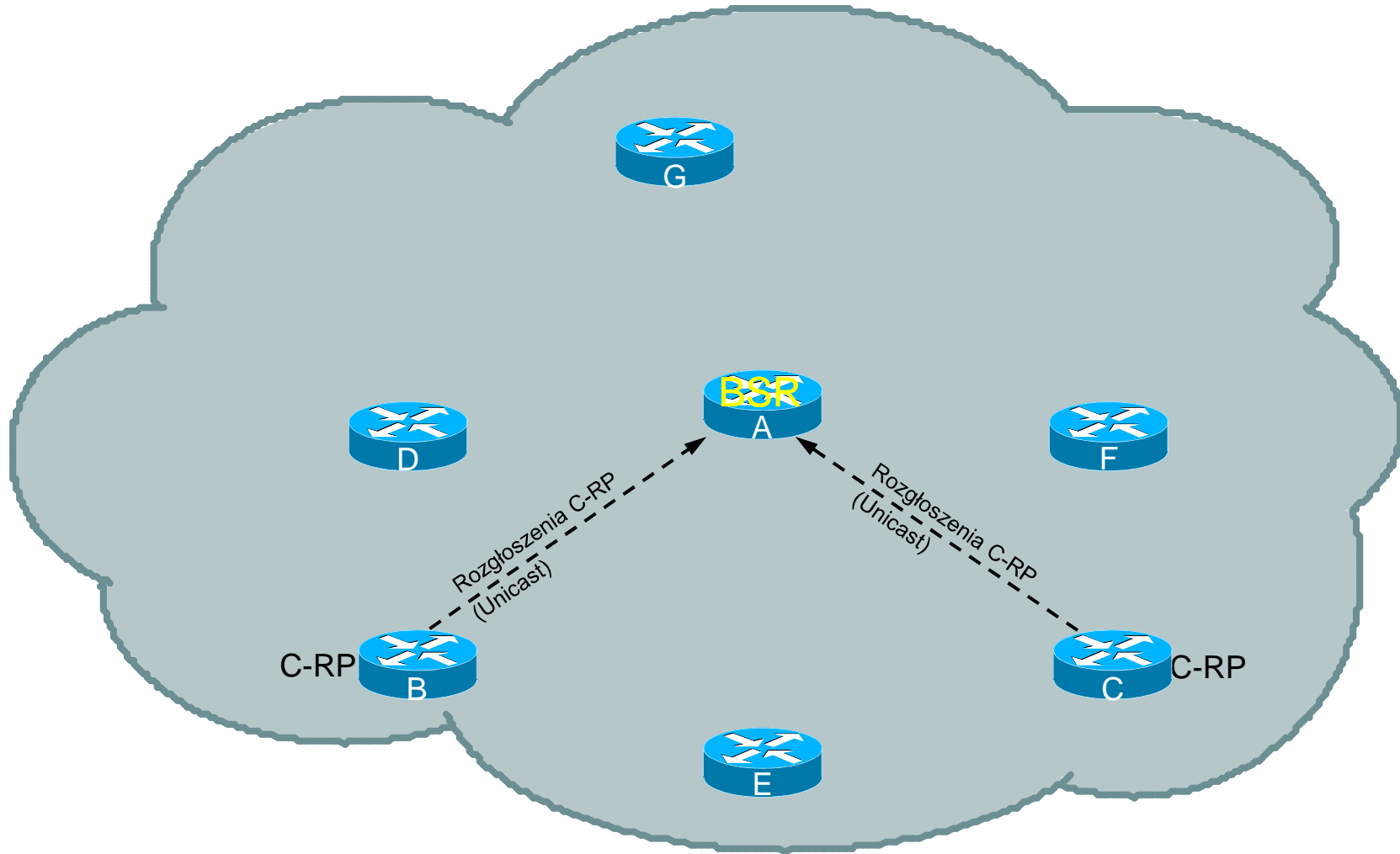
Wiadomości BSR są przekazywane hop-by-hop

# BSR—z 10,000 metrów

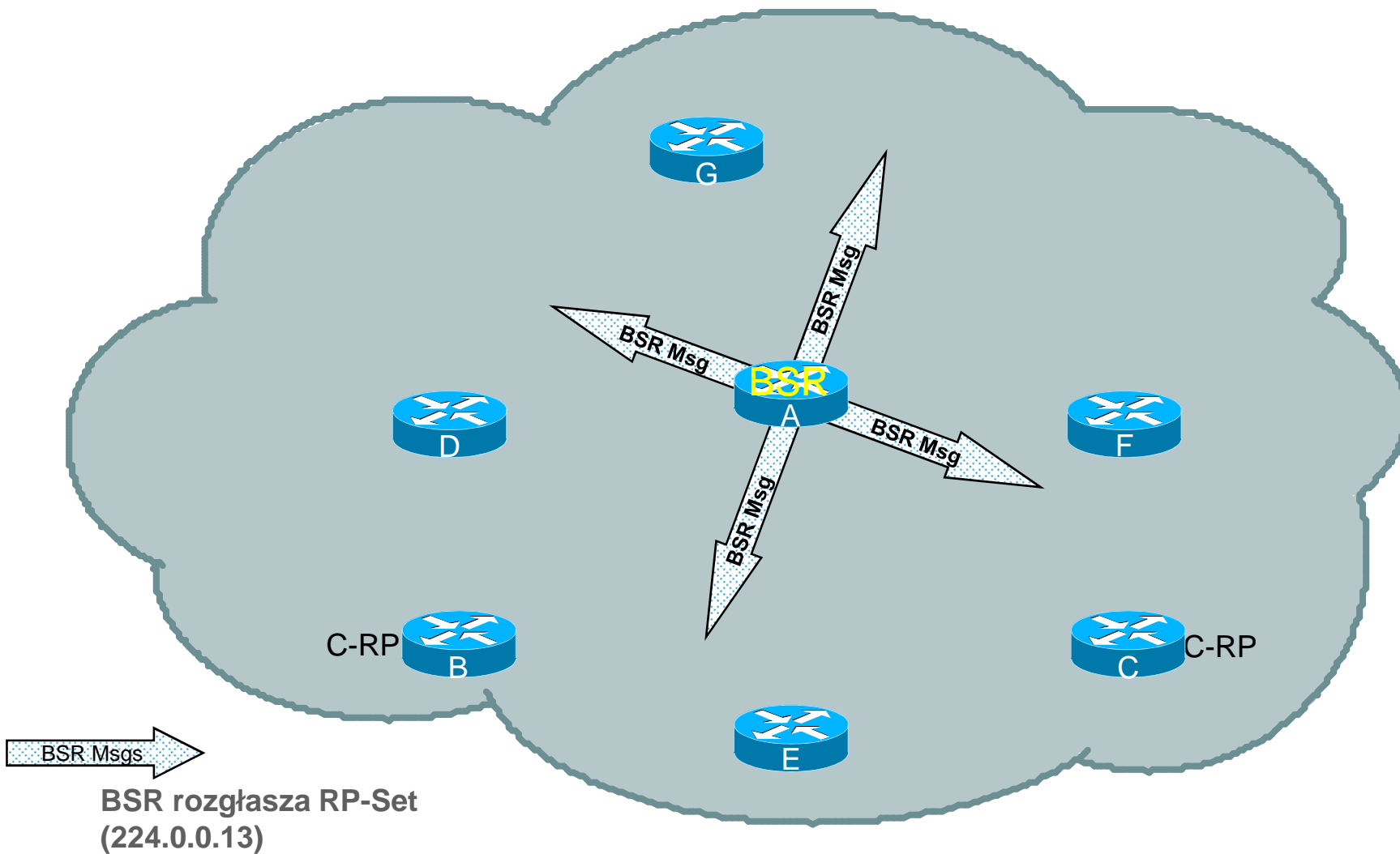
C-BSR z najwyższym priorytetem zostaje BSRem



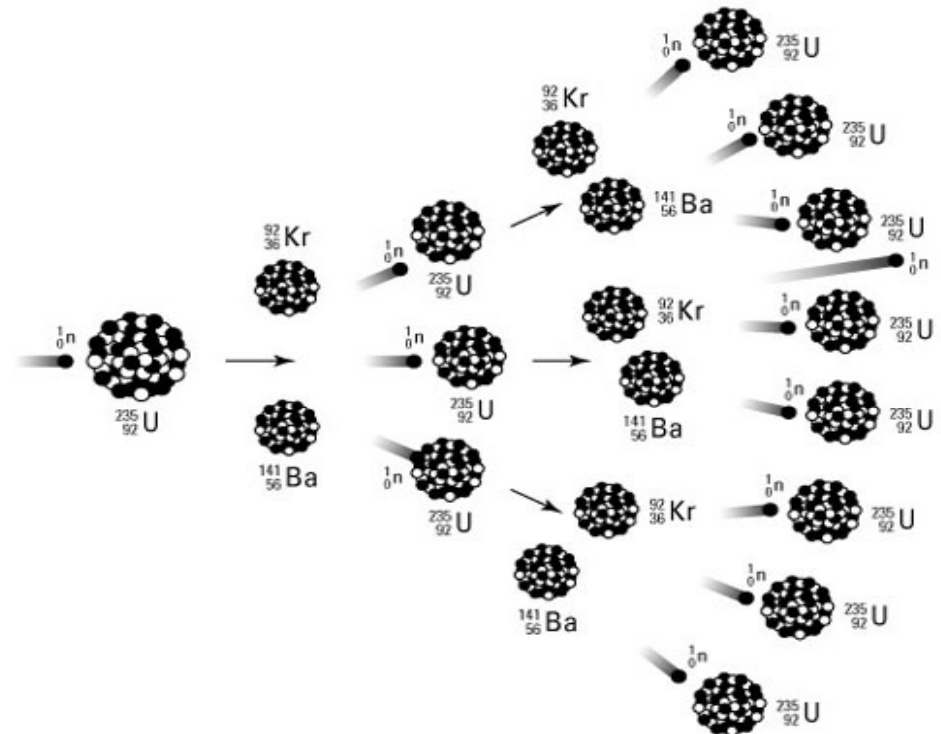
# BSR—z 10,000 metrów



# BSR—z 10,000 metrów



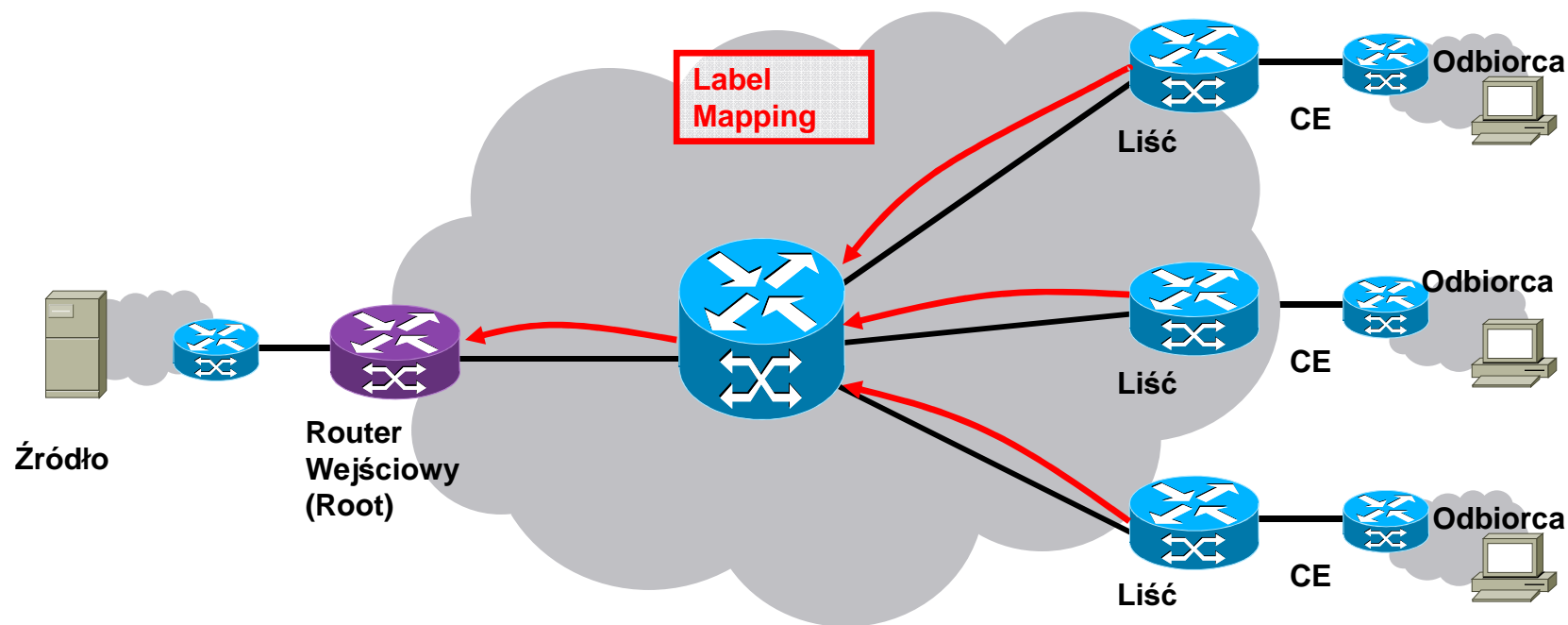
# MPLS a Multicast



# Label Switched Multicast

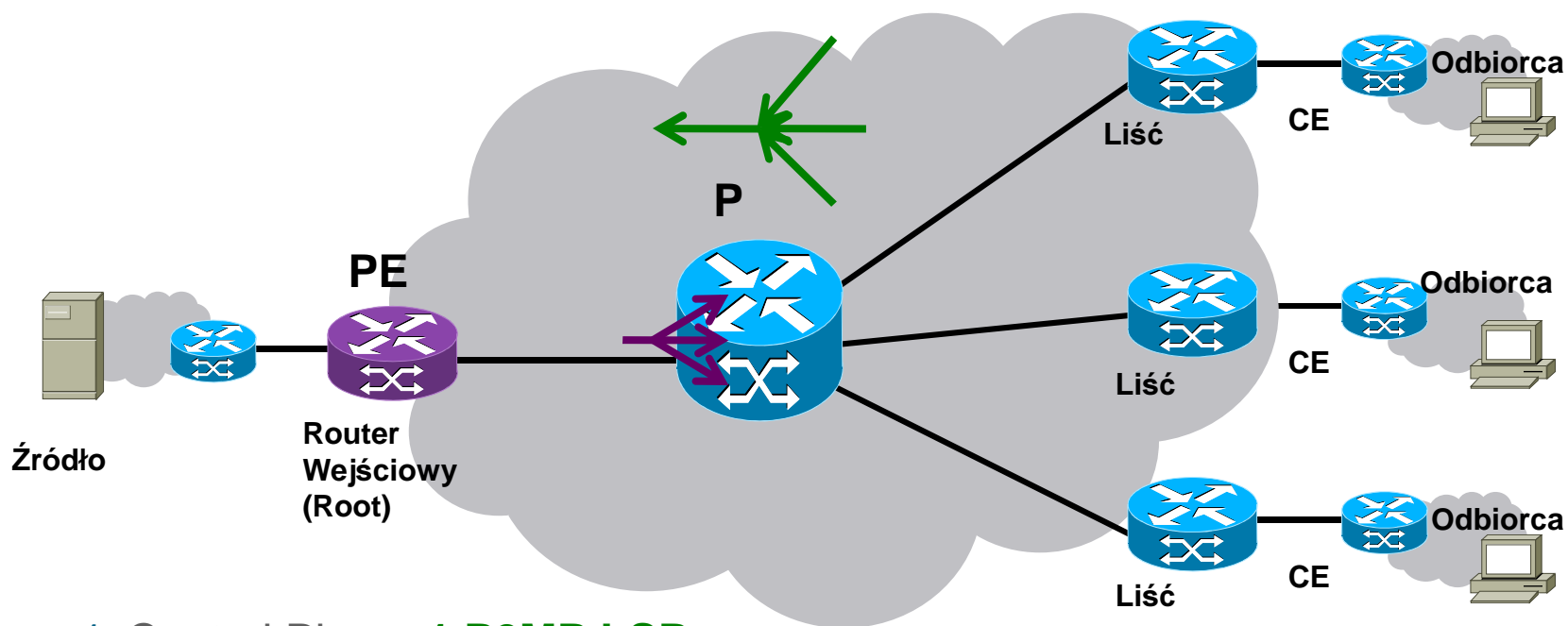
- ❖ pakiety multicastowe IP transportowane z wykorzystaniem **enkapsulacji mpls** (RFC5332)
- ❖ etykiety przydzielane są z tej **samej puli** (etykiety są globalne)
- ❖ alokacja etykiet odbywa się za pomocą **RSVP-TE bądź LDP** które zostały zmodyfikowane by wspierać ścieżki LSP typ P2MP, MP2MP
- ❖ Zalety
  - ❖ współdzielenie *control plane* z ruchem unicast
  - ❖ ten sam paradygmat przesyłu poprzez sieć
  - ❖ możliwość wykorzystania funkcjonalności takich jak
    - ❖ Fast ReRoute (FRR)
    - ❖ rezerwacja pasma

# MLDP P2MP - Sygnalizacja



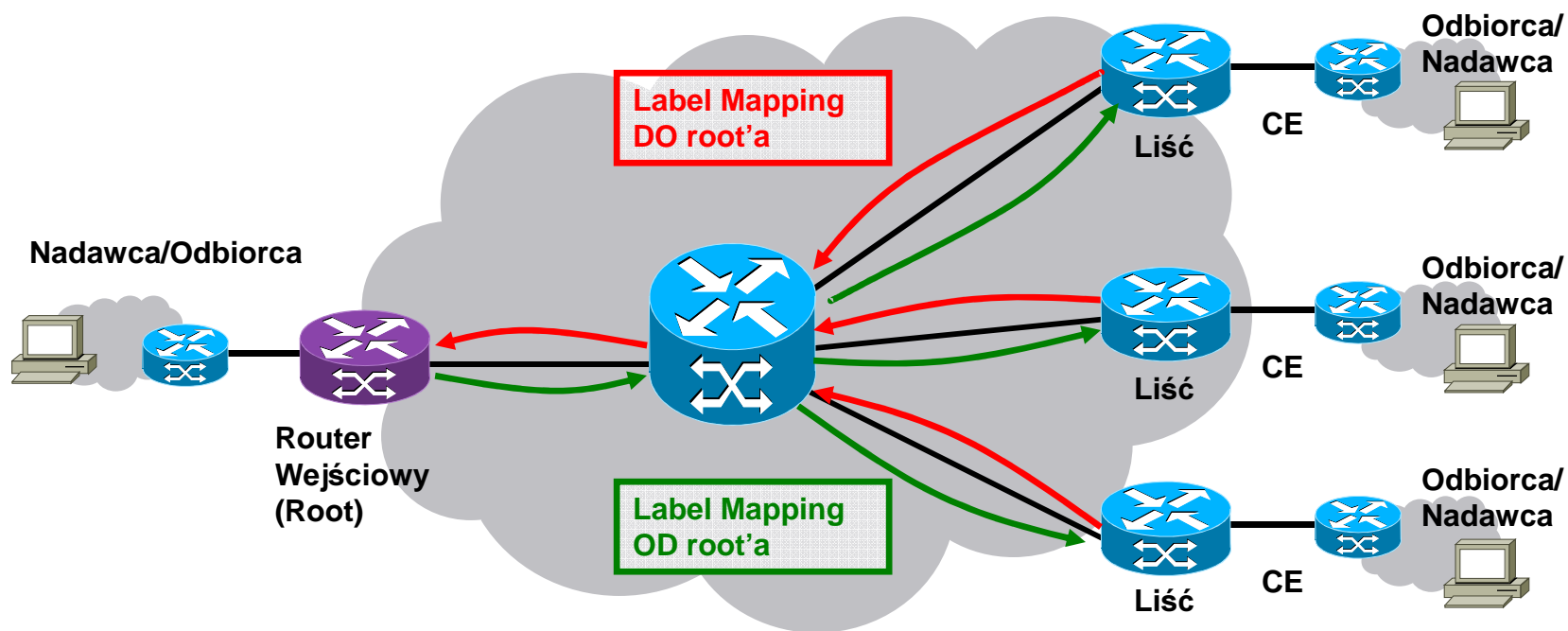
1. Routery wyjściowe (liście) otrzymują komunikat PIM Join
2. Przesyłają komunikat *label mapping* poprzez MLDP i sieć do routera wejściowego (root)
3. Router wejściowy otrzymuje pojedyncze uaktualnienie MLDP

# MLDP P2MP – Stany w sieci



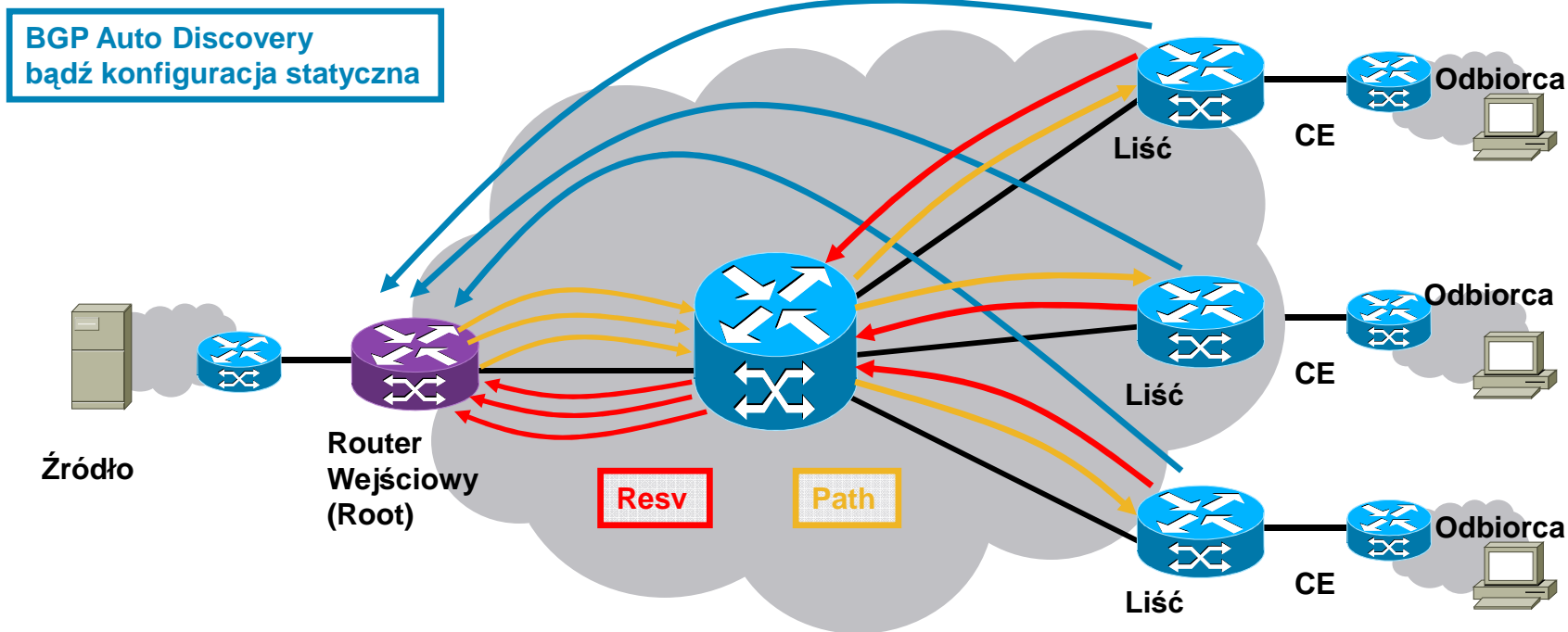
1. Control Plane: **1 P2MP LSP**
2. DataPlane: **1 P2MP LSP (replikacja)**
3. **Kiedy router wyjściowy nie chce już dłużej otrzymywać ruchu multicastowego wysyła odpowiedni komunikat do routera P**

# MLDP MP2MP - Sygnalizacja



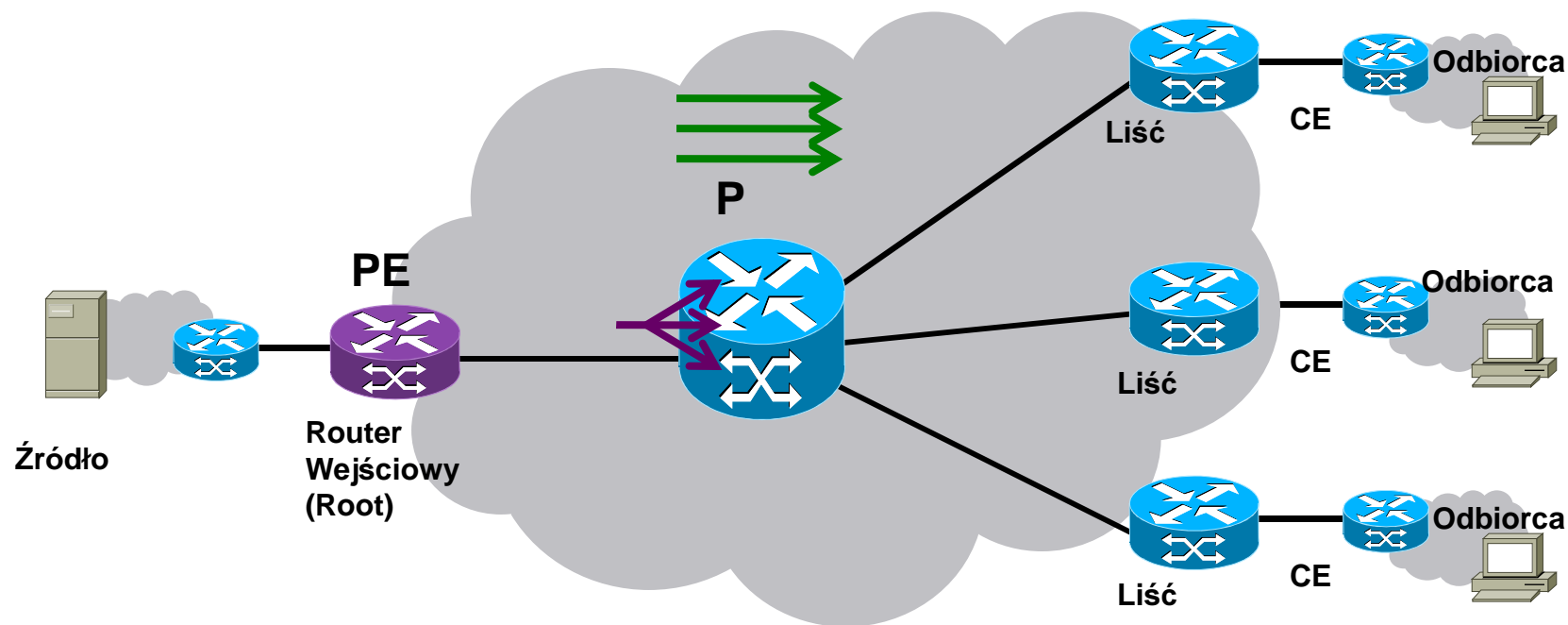
1. Router wyjściowy wysła komunikat MLDP *label mapping* do routera wejściowego (**dokładnie tak jak w P2MP**)
2. Na każdym łączu, wiadomość *label mapping* jest przesyłana w odwrotnym kierunku tworząc dwukierunkową ścieżkę MP2MP LSP

# RSVP-TE - Sygnalizacja



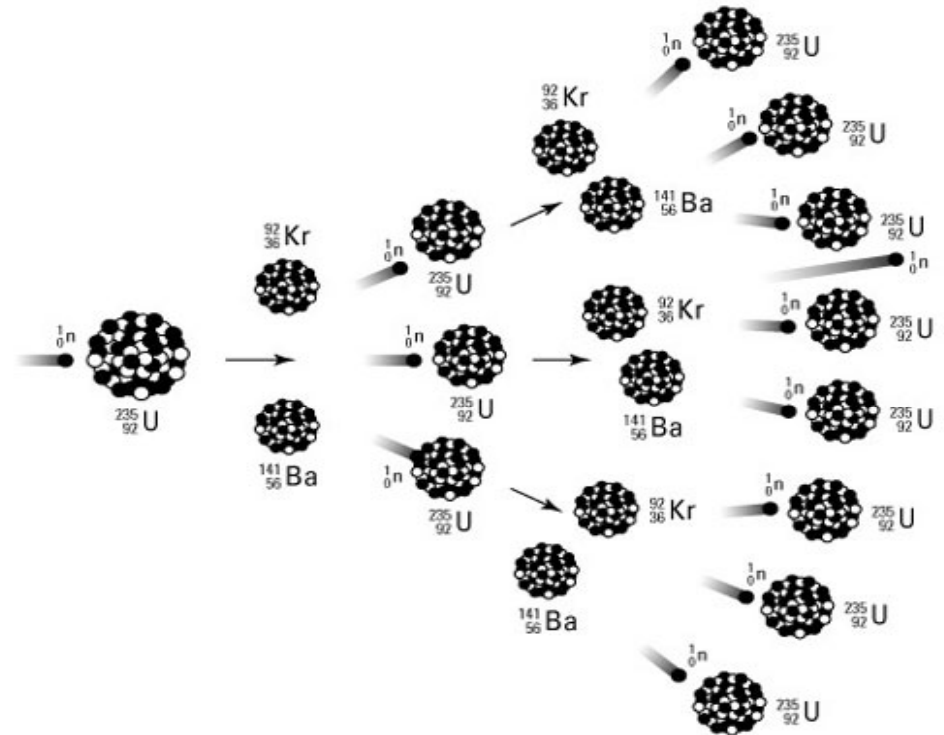
1. Router wyjściowy wysyła wiadomość BGP Auto Discovery do routera wejściowego
2. Router wejściowy (root) wysyła wiadomość typu *RSVP-TE Path* do routerów wyjściowych
3. Routery wyjściowe odpowiadają za pomocą wiadomości *RSVP-TE Resv* (standardowy sposób działania RSVP)

# RSVP-TE - Stany w sieci



1. Control Plane: **3 P2P sub-LSP** od routera wejściowego do wyjściowego
2. Data Plane: **1 P2MP (replikacja)**
3. **Kiedy router wyjściowy nie chce już odbierać ruchu multicastowego, wysyła wiadomość kontrolną do routera wejściowego celem usunięcia LSP do niego**

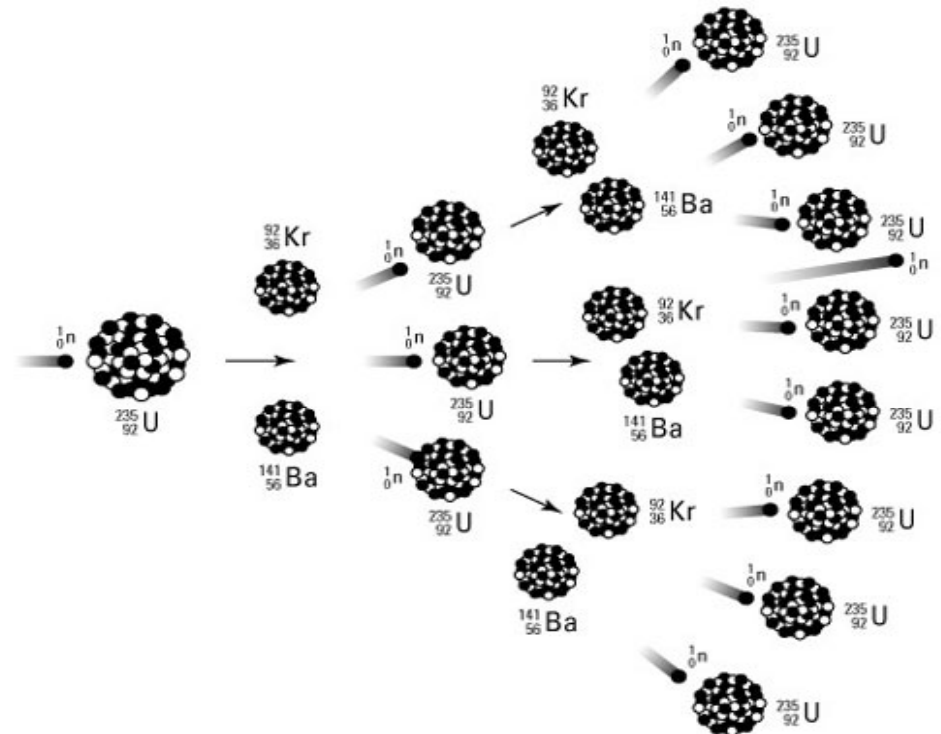
# Multicast w praktyce – czyli co warto robić a o czym zapomnieć



# Praktyczne porady

- ❖ Który PIM?
  - ❖ **SSM** dla aplikacji jeden do wielu - uproszczenie sieci, brak RP
  - ❖ **BiDir** dla aplikacji wiele do wielu (np. mVPN, OTV) – redukcja ilości stanów w sieci
  - ❖ **ASM** do każdego innego zastosowania
- ❖ Multicast w MPLS
  - ❖ **mLDP** dla aplikacji wiele do wielu (np. mVPN)
  - ❖ **P2MP TE** dla aplikacji jeden do wielu
- ❖ Konwergencja sieci multicast **zależy od konwergencji sieci unicast!**
- ❖ Nie zapomnij o **bezpieczeństwie w sieci multicast!**
  - ❖ Zabezpiecz się zarówno od strony odbiorców jak i źródeł
  - ❖ Jeśli to konieczne chroń treści które przesyłasz
  - ❖ Nie zapominaj o atakach DoS (w tym na L2) które ze względu na naturę routingu multicastowego (utrzymywanie stanów) w niezabezpieczonej sieci dosyć łatwo przeprowadzić

# Podsumowanie



# Podsumowanie

- ❖ Multicast stanowi **coraz znaczącą część ruchu przesyłanego w sieciach IP** – przekazy na żywo, TV, aplikacje biznesowe to tylko niektóre przykłady usług, z których korzystamy na co dzień
- ❖ Wszystkie **mechanizmy, protokoły i funkcjonalności** wymagane do skutecznej implementacji sieci multicastowej są w fazie swojej **dojrzałości i stabilności**

**Dziękuję za uwagę!**