SAMODZIELNY ZAKŁAD SIECI KOMPUTEROWYCH
POLITECHNIKA ŁÓDZKA
90-924 Łódź ul. Stefanowskiego 18/22
tel./fax. (42) 636 03 00
e-mail: szsk@zsku.p.lodz.pl

Tomasz Molik

Protokoły rutingu w sieciach rozległych i ich implementacja na ruterach Cisco

praca dyplomowa magisterska

Promotor:
Dr inż. Michał Morawski

Dyplomant:
Tomasz Molik
nr albumu 90162

Łódź, wrzesień 2002

Spis treści

1 Wstęp

Użytkownicy Sieci rzadko myślą o systemach i urządzeniach infrastruktury, które umożliwiają właściwe działanie Internetu. Szybki rozwój Sieci wymusił zmiany w organizacji dostępu do poszczególnych komputerów. Sposoby organizacji tego dostępu to właśnie protokoły routingu. Inaczej mówiąc routing jest procesem znajdowania w sieci ścieżki łączącej nadawcę z żądanym obiektem docelowym. W sieciach IP proces ten sprowadza się do określenia routerów w sieci, przez które muszą przejść pakiety by osiągnąć cel. Dopóki nadawca i odbiorca pozostają w obrębie tej samej sieci lokalnej problem przepływu danych jest rozwiązywany przez technologie, jaka została użyta przy budowie tej sieci. Przykładem może być chociażby Ethernet, który definiuje, w jaki sposób mogą porozumiewać się między sobą stacje znajdujące się wewnątrz tej samej sieci. Kiedy jednak nadawcą jest stacja przyłączona do innej sieci niż stacja odbiorcza w celu przesłania informacji wymagane jest zastosowanie routingu. W takim wypadku pakiety muszą wędrować poprzez routery łączące dwie odległe sieci. Kiedy pakiet dotrze do routera znajdującego się w tej samej sieci co stacja docelowa w celu dostarczenia pakietu do nadawcy stosuje się mechanizmy użytej technologii sieciowej (np. Ethernet).

Pierwsze protokoły obsługiwały tylko pojedyncze podsieci i wymagały ręcznej konfiguracji sprowadzającej się do wpisania statycznej tabeli połączeń, a maksymalna liczba przeskoków wynosiła 15[1]. Kolejne wersje protokołów powstawały, gdy poprzednie nie wystarczały do efektywnego zarządzania ruchem w sieci. Różnorodne urządzenia i zwiększający się transfer zapewniany przez połączone sieci stawiał coraz większe wymagania przed routerami i ich oprogramowaniem, a szybkość przekazywania danych po poszczególnych łączach zaczęła mieć znaczenie dla sposobu nawiązywania połączenia. Wszystkie te czynniki miały wpływ na ewolucję i specjalizację także w warstwie protokołów, które dynamicznie budują trasy połączeń.

Ze względu na złożoność Internetu nie ma możliwości ręcznej konfiguracji tras. Routery robią to samodzielnie, na podstawie wymienianych między sobą informacji o trasowaniu[4]. W dużych sieciach routing dynamiczny jest podstawową metodą zdobywania wiedzy, dzięki której routery poznają topologię sieci oraz budują tabele routingu. Wymiana informacji między routerami odbywa się zgodnie z określonymi algorytmami i przy wykorzystaniu protokołów routingu dynamicznego. Protokoły te muszą uwzględniać w przekazywanych informacjach dane, które pozwolą optymalizować i upraszczać trasy. Muszą także umieć wykryć awarie lub przeciążenia na części łączy. Przed protokołem stawiane jest również wymaganie, by zapobiegał zapętleniu drogi przy automatycznej analizie tras. Jednocześnie, wymieniane informacje o trasach nie mogą powodować przeciążeń łączy. Podstawą trasowania są tablice, które w jednym routerze mogą być tworzone za pomocą kilku protokołów używanych na różnych łączach fizycznych routera.

Obecnie wyróżnia się dwie podstawowe grupy protokołów: wewnątrzdomenowe IGP (Interior Gateway Protocol) oraz pozadomenowe EGP (Exterior Gateway Protocol)[4].

Protokoły grupy IGP służą do wymiany informacji o topologii sieci w obrębie systemu autonomicznego, w jednej domenie administracyjnej. Protokoły EGP natomiast, służą do wymiany informacji pomiędzy dostawcami usług internetowych i transportowych sieciowych. Można obrazowo powiedzieć, że śledzą one główne trasy, a identyfikację ostatecznego odbiorcy powierzają routerom systemów autonomicznych wykorzystujących IGP.

2 Cel i zakres pracy

Protokół trasowania powinien być wybrany w sposób uważny i z uwzględnieniem długoterminowych konsekwencji dokonanego wyboru. Wybranie protokołu bezpośrednio wpływa na rodzaj stosowanego routera oraz na wydajność działania sieci WAN. Znajdujące się w kolejnych rozdziałach opisy protokołów trasowania dynamicznego oraz różnych klas protokołów mają na celu w pełni uświadomić następstwa wybrania każdego z tych rozwiązań. Aby ułatwić wybór, każdy z protokołów zamieszczonych w pracy został dokładnie opisany teoretycznie, a zamieszczone przykłady topologii sieci pomagają lepszemu zrozumieniu stosowanych przez protokoły mechanizmów. Dzięki temu możliwe jest zawężenie wyboru do jednej kategorii lub klasy protokołów.

Następnym krokiem jest określenie, czy w sieci mają być wykorzystane routery jednego czy też kilku producentów. O ile jest to możliwe, zalecane jest korzystanie ze sprzętu jednego producenta, a to za sprawą tego, że otwarte protokoły trasowania zostawiają producentom pewien margines na modyfikacje. Z tego powodu wersja protokołu trasowania danego producenta może nie być w pełni wymienna z protokołem innego producenta. Opisy poszczególnych protokołów trasowania opierają się na ogólnie przyjętych specyfikacjach i nie odnoszą się do konkretnego producenta. Ponieważ jednak, jak już wspomniano, producenci routerów wprowadzają pewne niewielkie zmiany w implementacjach protokołów, dlatego opis wprowadzonych zmian również zawarł się w ramach tejże pracy. Skupiono się na opisie implementacji protokołów na routerach firmy Cisco. Wprowadzone przez producenta zmiany zostały wyraźnie zaznaczone.

Należy pamiętać, że jeśli producent routera zostanie wybrany przed protokołem trasowania, to wynikają z tego ograniczenia wyboru protokołu. Niektóre protokoły trasowania są produktami zastrzeżonymi, co oznacza, że można je uzyskać wyłącznie u jednego producenta. Z drugiej strony, nie wszystkie protokoły trasowania są implementowane przez danego producenta.

Temat protokołów routingu jest tematem bardzo rozległym, poruszającym wiele problemów pojawiających się podczas przesyłania danych przez sieć. Ze względu na złożoność i rozległość zagadnienia, musiały być podjęte pewne ogranicznenia co do zawartości pracy. Główny nacisk położony został na najważniejsze obecnie protokoły trasowania, takie jak RIP, promowany przez firmę Cisco IGRP i EIGRP, protokół stanu łącza OSPF czy protokół zewnętrzny BGP. Część z nich implementowana jest na routerach Cisco. Pewna część opisanych w pracy protokołów nie jest implementowana na routerach firmy Cisco, ale ze względu na ich istotę i znaczenie opisy tych protokołów również zawarły się w pracy. Dodatkowo opisane również zostały pokrótce protokoły multiemisyjne, oraz protokół PNNI stosowany w sieciach ATM. Ze względu na wyznaczone ramy pracy nie uwzględnione zostały takie protokoły jak: stosowany w sieciach ATM Multiprotocol Over ATM (MPOA) czy nowy protokół zewnętrzny Multi Protocol Label Switching (MPLS).

3 Słowniczek pojęć

Drzewo najkrótszych ścieżek - graf skierowany, w skład którego wchodzą sieci i routery. Dostarcza informacji o stanie połączeń w systemie autonomicznym. Krawędź grafu łącząca ze sobą dwa routery odnosi się do sytuacji fizycznego połączenia w sieci typu punkt-punkt. Natomiast krawędź łącząca router z siecią służy opisaniu sytuacji, w której router posiada interfejs do danej sieci.

Exterior Gateway Protocol (EGP) - protokół używany do wymiany informacji o trasowaniu i dostępności sieci pomiędzy systemami autonomicznymi. Protokół ten nazywany jest również protokołem międzydomenowym. Przykładem takiego protokołu jest BGP, IDRP.

Host – maszyna w sieci posiadająca własny adres IP, nazwę i należąca do domeny. Często także: komputer podłączony bezpośrednio do Internetu, pozwalający świadczyć usługi internetowe klientom.

Interfejs – połączenie pomiędzy routerem a jedną z przyłączonych do niego sieci. Z interfejsem związane są pewne informacje dotyczące jego stanu. Informacje te uzyskiwane są od protokołów działających w niższych warstwach jak również od protokołu trasującego. Interfejs posiada przypisany pojedynczy adres i maskę (chyba, że jest to połączenie punkt-punkt nie posiadające numeru).

Interior Gateway Protocol (IGP) - protokół używany do wymiany informacji na temat trasowania i dostępności sieci w obrębie systemu autonomicznego. Przykładem takiego protokołu jest RIP, OSPF, IS-IS.

Maska sieci – 32-bitowa liczba określająca, która część adresu IP jest adresem sieci.

Metryka – parametr przypisany do trasy, określający koszt przesłania pakietu. Metryka opisywać może liczbę routerów, przez które musi przejść pakiet by dotrzeć do odbiorcy. W bardziej protokołach używane są metryki, które uwzględniają takie parametry jak: opóźnienia, koszt, obciążenia, przepustowość i inne. Najważniejszym wymaganiem jest by metryka wyrażała się jako suma metryk poszczególnych przejść przez routery.

Połączenia wirtualne – jest częścią szkieletu systemu autonomicznego. Służy do łączenia dwóch routerów będących routerami szkieletu systemu autonomicznego, które to routery posiadają interfejs do obszaru nie będącego częścią systemu autonomicznego. Protokół traktuje tak połączone routery w taki sam sposób jakby połączone one były siecią punkt-punkt.

Router – przełącznik warstwy trzeciej modelu OSI. Nazywany jest również w literaturze bramą.

Routery sąsiadujące – routery posiadające interfejsy do tej samej sieci. Stosunki pomiędzy sąsiadującymi routerami zależą od protokołu routingu. Przeważnie routery sąsiadujące odnajdują się dynamicznie.

Routery przyległe - są to routery sąsiadujące, które dodatkowo tworzą połączenie logicznie między sobą w celu wymiany informacji o trasowaniu. Nie wszystkie routery sąsiadujące stają się routerami przyległymi.

Router desygnowany – znajduje zastosowanie w sieciach rozgłoszeniowych i NBMA. Spełnia dwie podstawowe funkcje: 1) jest źródłem pakietów opisujących sieć i 2) staje się routerem przyległym dla wszystkich routerów w sieci i tym samym staje się centralnym punktem synchronizacji baz danych.

Sieć – system połączonych elementów, powiązanych ze sobą za pośrednictwem łączy dedykowanych lub komutowanych, mający na celu zapewnienie lokalnej lub zdalnej komunikacji (głosu, obrazu wideo, danych itp.)

Sieć końcowa (stub network) – sieć służąca wymianie danych pomiędzy lokalnymi hostami. Nawet jeśli lokalne hosty posiadają połączenia z innymi sieciami, to te hosty nie służą do przesyłania danych między tymi sieciami.

Sieć punkt-punkt – jest to sieć łącząca pojedynczą parę routerów (i tylko tę parę routerów).

Sieć rozgłoszeniowa (broadcast network) – sieć wspierająca komunikację między wieloma (więcej niż dwoma) routerami, w której istnieje możliwość wysłania pojedynczej wiadomości jednocześnie do wszystkich przyłączonych routerów. Wiadomości, w takim wypadku, adresowane są przy użyciu adresu rozgłoszeniowego. W sieciach rozgłodzeniowych sąsiadujące routery mogą odnajdywać się dynamicznie przy użyciu pakietów Hello. Przykładem takiej sieci jest Ethernet.

Sieć nie-rozgłoszeniowa NBMA (non-broadcast network) – sieć wspierająca komunikację między wieloma (więcej niż dwoma) routerami, ale nie posiadająca możliwości rozgłaszania. Sąsiadujące routery utrzymują kontakt przy użyciu pakietów Hello, jednakże z powodu braku możliwości adresowania rozgłoszeniowego konieczne są pewne czynności konfiguracyjne, dzięki którym możliwe będzie odnajdywanie nowych routerów w sieci. W takich sieciach wiadomości muszą być osobno przesyłane, do każdego z sąsiadujących routerów. Przykładem może być sieć X.25.

System autonomiczny – zbiór routerów zarządzanych przez tą samą jednostkę administracyjną, używających protokołu IGP i tej samej metryki do podejmowania decyzji o trasowaniu pakietów wewnątrz systemu autonomicznego, oraz protokołu EGP służącego do trasowania pakietów, których odbiorca znajduje się w innym systemie autonomicznym. System autonomiczny jest widziany na zewnątrz jako jedna całość, mimo, że w skład systemu autonomicznego może wchodzić wiele sieci. Każdy system autonomiczny posiada swój unikatowy numer identyfikacyjny z zakresu 1 – 65535. Numeru z zakresu 64512 – 65535 są zarezerwowane dla użytku prywatnego i nie powinny być nadawane systemom wchodzącym w skład Internetu.

4 Routing statyczny

Najprostszą formą budowania informacji o topologii sieci są ręcznie podane przez administratora trasy definiujące routing statyczny[4]. Przy tworzeniu takiej trasy wymagane jest jedynie podanie adresu sieci docelowej, interfejsu, przez który pakiet ma zostać wysłany oraz adresu IP następnego routera na trasie.

Routing statyczny ma wiele zalet. Router przesyła pakiety przez z góry ustalone interfejsy bez konieczności każdorazowego obliczania tras, co zmniejsza zajętość cykli procesora i pamięci. Informacja statyczna nie jest narażona na deformację spowodowaną zanikiem działania dynamicznego routingu na routerach sąsiednich. Dodatkowo zmniejsza się zajętość pasma transmisji, gdyż nie są rozsyłane pakiety rozgłoszeniowe protokołów routingu dynamicznego. Dla małych sieci jest to doskonałe rozwiązanie, ponieważ nie musimy posiadać zaawansowanych technologicznie i rozbudowanych sprzętowo routerów[4]. Routing statyczny zapewnia również konfigurację tras domyślnych, nazywanych *bramkami ostatniej szansy* (gateway of the last resort). Jeżeli router uzna, iż żadna pozycja w tablicy routingu nie odpowiada poszukiwanemu adresowi sieci docelowej, korzysta ze statycznego wpisu, który spowoduje odesłanie pakietu w inne miejsce sieci.

Routing statyczny wymaga jednak od administratora sporego nakładu pracy w początkowej fazie konfiguracji sieci, nie jest również w stanie reagować na awarie poszczególnych tras[25].

5 Algorytm wektora długości

Istnieje wiele metod wyszukiwania trasy między nadawcą i odbiorcą[26]. Metody te wykorzystują algorytmy teorii grafów. Dwa podstawowe algorytmy stosowane przy routingu to : algorytm Dijksty i algorytm Forda-Bellmana. Oba algorytmy służą wyznaczeniu najmniejszej odległości od ustalonego wierzchołka *s* do wszystkich pozostałych w skierowanym grafie. Protokoły wektora długości wykorzystują algorytm Forda-Bellmana, natomiast algorytm Dijkstry znalazł zastosowanie w protokołach stanu łącza.

W algorytmie Dijkstry pamiętany jest zbiór Q wierzchołków, dla których nie obliczono jeszcze najkrótszych ścieżek, oraz wektor D[i] odległości od wierzchołka *s* do *i*. Graf wejściowy nie może zawierać krawędzi o ujemnych wagach[15]. Początkowo zbiór Q zawiera wszystkie wierzchołki a wektor D jest pierwszym wierszem macierzy wag krawędzi A. Dopóki zbiór Q nie jest pusty, to pobierany jest z tego zbioru wierzchołek *v* o najmniejszej wartości D[v] i usuwany jednocześnie ze zbioru Q. Dla każdego następnika *i* wierzchołka *v* sprawdzany jest warunek D[i]>D[v]+A[v,i], tzn. czy aktualne oszacowanie odległości do wierzchołka *i* jest większe od oszacowania odległości do wierzchołka *v* plus waga krawędzi (v,i). Jeżeli tak jest, to oszacowanie D[i] jest aktualizowane, czyli przypisywana jest mu prawa strona nierówności (czyli mniejsza wartość). Algorytm kończy działanie w momencie, gdy zbiór Q nie zawiera wierzchołków.

Jak widać z powyższego pseudokodu, algorytm wybiera ze zbioru Q "najlżejszy" wierzchołek, tzn. jest oparty o strategię zachłanną. Wprawdzie metoda zachłanna nie zawsze daje wynik optymalny, jednak algorytm Dijkstry jest algorytmem dokładnym[15].

Istnieje kilka odmian implementacji Dijkstry; najprostsza używa tablicy do przechowywania wierzchołków ze zbiory Q. Inne wersje algorytmu używają kolejki priorytetowej lub kopca Fibonacciego.

Algorytm Forda-Bellmana wymaga, by w skierowanym grafie nie było cykli o ujemnej długości[17]. Warunek nieujemności cyklu jest spowodowany faktem, że w grafie o ujemnych cyklach najmniejsza odległość między niektórymi wierzchołkami jest nieokreślona, ponieważ zależy od liczby przejść w cyklu. Dla zadanego grafu macierz *A* dla każdej pary wierzchołków *u* i *v* zawiera wagę krawędzi (u,v), przy czym jeśli krawędź (u,v) nie istnieje, to przyjmujemy, że jej waga wynosi nieskończoność. Algorytm Forda-Bellmana w każdym kroku oblicza górne oszacowanie $D(v_i)$ odległości od wierzchołka *s* do wszystkich pozostałych wierzchołków v_i. W pierwszym kroku przyjmujemy $D(v_i)=A(s,v_i)$. Gdy stwierdzimy, że D(v)>D(u)+A(u,v), to każdorazowo polepszamy aktualne oszacowanie i podstawiamy D(v):=D(u)+A(u,v). Algorytm kończy się, gdy żadnego oszacowania nie można już poprawić, macierz $D(v_i)$ zawiera najkrótsze odległości od wierzchołka *s* do wszystkich pozostałych. Algorytm można ulepszyć sprawdzając w każdej iteracji, czy coś się zmieniło od poprzedniej i jeśli nie, to można zakończyć obliczenia[17].

Jednym z kryteriów używanych do klasyfikacji metod stosowanych przez protokoły routingu jest rodzaj informacji, jaką potrzebują routery do wyznaczenia trasy. Algorytm wektora odległości opiera swoje obliczenia tras na podstawie niewielkiej ilości informacji o topologii sieci [1]. Każdy uczestnik wymiany (router lub stacja) przechowuje w swojej tablicy routingu dane o wszystkich sieciach w systemie. Dodatkowo oprócz wpisów o sieciach mogą też się pojawić wpisy dotyczące konkretnej stacji w sieci. Algorytm nie robi formalnego rozróżnienia pomiędzy sieciami a pojedynczymi stacjami[26]. Generalnie jednak używa się wpisów o sieciach. Pojedynczy, bowiem wpis w tablicy trasowania informujący nas o docelowej

sieci opisuje jednocześnie wszystkie stacje przyłączone do tej sieci, przez co znacznie zmniejsza rozmiar tablicy. Z powyższych względów w dalszym opisie protokołów wektora odległości pojęcie „sieć" będzie odnosić się zarówno do sieci jak i do pojedynczych stacji.

Każdy wpis w tablic routingu opisujący obiekt docelowy zawiera adres następnego routera na ścieżce, czyli routera, do którego należy przesłać pakiety by tamten skierował je dalej[26]. Dodatkowo wpis zawiera metrykę służącą do określenia długości ścieżki. Pojęcie długości jest umowne i może określać czas przesłania pakietu, koszt pieniężny związany z przesłaniem pakietu daną trasa itp. Nazwa "wektor odległości" pochodzi stąd, iż poszczególne trasy ogłaszane są właśnie jako wektory zawierające te dwie informacje: odległość oraz kierunek i na podstawie tych danych algorytm wyznacza najbardziej optymalną ścieżkę[25].

Aktualne implementacje protokołów wektora odległości przechowują w tablicach routingu następujące informacje[26] :

- adres IP sieci lub stacji docelowej
- adres następnego routera na ścieżce
- interfejs, przez który komunikujemy się z następnym routerem na ścieżce
- metryka określająca odległość do celu
- licznik określający czas, jaki upłynął od ostatniego uaktualnienia o trasie

Dodatkowo w tablicach mogą być przechowywane znaczniki i inne wewnętrzne informacje użyte w danej implementacji.

Początkowo tablica trasowania zawiera informacje o bezpośrednio podłączonych sieciach do routera[1]. Stopniowo jest uzupełniana o wpisy informujące o innych sieciach na podstawie danych otrzymanych od sąsiadujących routerów. Dane te w postaci wiadomości aktualizacyjnych są przesyłane okresowo i są podstawowym źródłem informacji o stanie sieci. Każdy uczestnik wymiany tych uaktualnień wysyła do swoich sąsiadów informacje o sieciach, które posiada w swojej tablicy routingu. Protokoły wektora odległości przechowują w swych tablicach tylko trasy o najmniejszej metryce, wybrane spośród znanych tras[25]. W prostych sieciach metryka opisywać może liczbę routerów, przez które musi przejść pakiet by dotrzeć do odbiorcy, natomiast w bardziej złożonych sieciach można użyć metryki, która uwzględnia takie parametry jak: opóźnienia, koszt, obciążenia, przepustowość i inne. Najważniejszym wymaganiem jest by metryka wyrażała się jako suma „kosztów" poszczególnych przejść przez routery[26].

Algorytm wyznaczania trasy o najmniejszej metryce opiszemy na podstawie przykładu[26]. Przyjmijmy, że router R1 otrzymuje wiadomość o sieci N1 od routera R2 w postaci wektora złożonego z adresu sieci N1 i odległości. Ponieważ jak wspomnieliśmy wcześniej metryka spełnia warunek sumowania kosztów poszczególnych przejść, to dodając do otrzymanej metryki wartość metryki trasy do routera R2 otrzymujemy całkowity koszt K trasy od routera R1 do sieci N1. Router R1 przegląda swoją tablice trasowania w celu znalezienia wpisu odnoszącego się do sieci N1. Jeśli takiego wpisu nie znajdzie oznacza to, że nie posiada informacji o sieci N1 i dodaje nowy wpis, mówiący o tym, że do sieci N1 można dostać się poprzez router R2 i koszt tej trasy jest równy K.

Jeśli istnieje już wpis w tablicy routera R1 trasy do sieci N1 o metryce większej od K to trasa ta jest zastępowana nową trasą prowadzącą przez R2. W przypadku, gdy trasa przez router R2 uznana zostanie za gorszą, trasa ta jest ignorowana i nie ma wpływu na postać tablicy routingu routera R1. Może się również zdarzyć sytuacja, w której dwie trasy będą opisane tą samą wartością metryki. W takim przypadku obie te trasy umieszczane są w tablicy routingu.

Może również zdarzyć się, że R1 posiada wpis o trasie prowadzącej do sieci N1 przez router R2. Otrzymując w tej sytuacji uaktualnienie od R2 odnoszące się do sieci N1, router musi to uaktualnienie zaakceptować i wprowadzić zmiany w swojej tablicy nawet, jeśli metryka się zwiększa, ponieważ R2 jest oryginalnym nauczycielem[26].

5.1 Zmiany topologii sieci

W dotychczasowych rozważaniach zakładaliśmy, że topologia sieci jest niezmienna. W praktyce, topologia ta się zmienia. Routery, a w szczególności łącza często ulegają awariom, a następnie ponownie włączają się do sieci. Aby poradzić sobie z tym zagadnieniem w algorytmie trzeba wprowadzić nieznaczne zmiany[26].

Router, który zauważy zmiany w topologii powiadamia o tym swoich najbliższych sąsiadów[25]. Zmiany te zostaną uwzględnione przy następnym obliczaniu ścieżek. Jednakże, jak wspomniano wyżej, algorytm pamięta tylko najlepsze trasy do danej sieci. Jeśli więc połączenie sieciowe lub router znajdująca się na tej trasie ulegnie awarii, to zmiana ta może nigdy nie zostać odzwierciedlona w tablicy routingu. Dzieje się tak, ponieważ algorytm działa wyłącznie na podstawie danych otrzymanych od innych routerów[26]. Dlatego też, gdy router stanie się niedostępny nie ma możliwości powiadomienia o swojej niedostępności swoich sąsiadów. Trasa pozostaje niezmieniona, mimo, że jest błędna, a przychodzące wiadomości o trasach do danej sieci docelowej są odrzucane z powodu większej metryki[26].

Aby zapobiec możliwości wystąpienia tego typu sytuacji wprowadza się mechanizmy wykrywające przedawnienie tras. Szczegóły tych mechanizmów zależą od konkretnego protokołu. W protokole RIP każdy router uczestniczący w routingu wysyła wiadomość o stanie sieci do każdego swojego sąsiada domyślnie co 30 sekund[2]. Przypuśćmy, że aktualna trasa do sieci N1 używa routera R1. Jeśli nie otrzymamy od R1 żadnej wiadomości w ciągu 180 sekund, możemy przypuszczać, że albo router nie działa poprawnie albo połączenie sieciowe z tym routerem zostało zerwane. Zdarza się, że wiadomość przesyłana siecią nie zawsze dociera do routera-adresata za pierwszym razem, tak więc nie jest dobrym pomysłem uznanie trasy za nieprawidłową w przypadku zagubienia pojedynczej wiadomości[26]. Po upływie 180 sekund oznaczamy trasę jako niedostępną[5]. Jeśli teraz otrzymamy informacje od któregoś z sąsiadów o trasie prowadzącej do sieci N1, to trasa ta zastąpi w tablicy trasę niedostępną. Zwróćmy uwagę, na fakt, że trasa jest oznaczana jako niedostępna dopiero po upływie 180 sekund. Oznacza to, że jeśli trasa stanie się niedostępna (poprzez awarie routera lub łączy), to w ciągu tych 180 sekund prawidłowe trasy prowadzące do N1 (otrzymywane od sąsiadów co 30 sekund) o większych metrykach niż metryka trasy dotychczas używanej są odrzucane[26].

Jak zobaczymy dalej, przydatną staje się możliwość powiadamiania sąsiadów o bieżącej niedostępności trasy prowadzącej do jakiejś sieci. RIP, wraz z innymi protokołami tej klasy, realizuje to poprzez zwykłe wysyłanie aktualnych wiadomości o trasach, zaznaczając daną sieć jako nieosiągalną[3]. Niedostępność sieci docelowej sygnalizowana jest odpowiednią wartością metryki. Wartość ta jest większa niż największa prawidłowa wartość metryki przewidziana w protokole. W implementacji protokołu RIP wartość ta jest równa 16. Liczba ta wydaje się zaskakująco niska. Wartość jej wynika z przesłanek opisach w dalszej części pracy.

5.2 Zapobieganie niestabilnemu trasowaniu

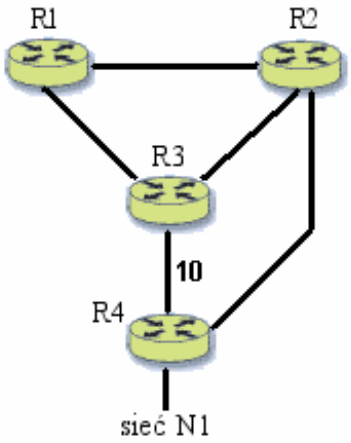
Dotychczas przedstawiony algorytm pozwoli routerom na poprawne wyznaczanie trasy w sieci[26]. Jednakże ciągle jest to algorytm niewystarczający w praktyce. Poniżej przedstawiony przykład pokazuje, że jest on użyteczny do wyznaczenia trasy w skończonym czasie, ale nie gwarantuje, że czas ten będzie wystarczająco mały by trasa była użyteczna. Nie określa również działań na metryce dla sieci nieosiągalnych.

Dla określenia sieci nieosiągalnej używamy wartości, którą w danej metryce będziemy uważać za wartość nieskończoną. Przyjmijmy, że jest to wartość 16.

Przypuśćmy, że sieć staje się niedostępna. Wszystkie routery bezpośrednio podłączone do tej sieci po upływie określonego czasu uznają sieć za niedostępną i przypiszą jej metrykę równą 16. Roześlą one również tą informacje do innych routerów sąsiadujących[26]. Ponieważ te z kolei otrzymały komunikat od routerów bezpośrednio podłączonych do rozważanej sieci to ich metryka będzie zwiększona do wartości równej 17. Routery oddalone o dwa skoki od routera bezpośrednio podłączonego do sieci zwiększą metrykę do 18 itd. Ponieważ jednak ustaliliśmy, że 16 jest wartością maksymalną w naszej metryce, wszystkie te routery oznaczą trasę do omawianej, niedostępnej sieci metryką 16.

Rozważymy teraz problemy związane ze zbieżnością w przypadku uznania trasy za niedostępną. W rozważaniach posłużymy się przykładem. Należy zwrócić uwagę, że opisana w przykładzie sytuacja nie zdarzy się w przypadku poprawnie zaimplementowanego protokołu RIP[26]. Przykład ten służy pokazaniu konieczności stosowania mechanizmów wpływających na poprawę algorytmu trasowania.

Każdy z routerów będzie posługiwał się tablicą routingu, by wyznaczać ścieżki do odpowiednich sieci. Dla zilustrowania przykładu przedstawimy wpisy dla sieci docelowej N1 (Rysunek 1).



Rysunek 1) Przykład połączenia routerów.

R4: bezpośrednie połączenie, metryka = 1
R2: połączenie przez R4, metryka = 2
R3: połączenie przez R2, metryka = 3
R1: połączenie przez R2, metryka = 3

Przyjmijmy, że połączenie z R2 do R4 zostało przerwane. Trasy w tablicach poszczególnych routerów powinny być zmienione w ten sposób, by przechodziły przez router R3. Niestety dostosowanie do nowych warunków zajmie trochę czasu. Zmiany zapoczątkuje router R2, który zauważy niedostępność trasy do R4. Dla uproszczenia przyjmiemy, że wszystkie uaktualnienia są dokonywane w tym samym czasie. Tabela 1 przedstawia wpisy w tablicy poszczególnych routerów po kolejnych uaktualnieniach.

Czas ---------->													
R4 :	bezp	1	bezp	1	bezp	1	bezp	1	...	bezp	1	bezp	1
R2 :	nieosiągalna		R3	4	R3	5	R3	6	...	R3	11	R3	12
R3 :	R2	3	R1	4	R1	5	R1	6	…	R1	11	R4	11
R1 :	R2	3	R3	4	R3	5	R3	6	…	R3	11	R3	12

Tabela 1) Wpisy tablicy routingu po kolejnych uaktualnieniach

R2 zorientuje się o niedostępności trasy poprzez mechanizm liczników, lecz trasa ta pozostanie w systemie przez długi czas. Początkowo R1 i R3, opierając się na swoich tablicach routingu, są przekonane, że do R4 mogą się dostać poprzez router R2, dlatego też trasę tą umieszczają w rozgłaszanych wiadomościach z metryką równą 3. Po pierwszej wymianie informacji zawartych w tablicach routingu, R2 stwierdzi, że do R4 może dostać się używając routera R1, albo routera R3. Nie jest to jednak prawdą, ponieważ trasy ogłaszane przez R1 i R3 prowadzące do R4 nie są już aktualne, ale ani R1, ani R3 nic o tym nie wie[26]. Gdy R1 i R3 odkryją już, że trasa przez R2 nie jest dostępna usuwają ją ze swoich tablic a następnie akceptują przesłane uaktualnienia od siebie wzajemnie i dojść może do sytuacji, w której R1 posiada wpis prowadzący do R4 poprzez router R3, z kolei router R3 posiada wpis mówiący o tym, że do R4 dostać się można przez router R1. W takiej sytuacji pakiety, które mają trafić do sieci docelowej krążą między routerami R1 i R3. Powstała więc pętla między dwoma routerami. Jak widać, po każdym uaktualnieniu wartość metryki trasy prowadzącej do R4 zwiększa się. Dzieje się tak, bo router R1 musi zaakceptować ogłoszenie o trasie do sieci docelowej do routera R3 (mimo że zwiększa się metryka), gdyż router R3 jest oryginalnym nauczycielem, od którego dowiedziały się o sieci docelowej[26]. Tak samo jest w przypadku routera R3.

Router R3 jest bezpośrednio podłączony do routera R4, lecz przez to, że koszt tego połączenia jest równy 10, to wpisuje do swojej tablicy trasę o mniejszej wartość metryki otrzymaną od sąsiadów, czyli trasę przez R1. Tak się dzieje dopóki zwiększająca się metryka trasy prowadzącej do sieci docelowej przez R1 nie osiągnie wartości 12, czyli większej od wartości metryki przypisanej trasie przez router R4, która wynosi 11. Po takim dopiero czasie tablice trasowania poszczególnych routerów są prawidłowo wypełnione[26].

W najgorszym przypadku, gdy część sieci staje się całkowicie niedostępna powolne zwiększanie metryki może trwać, tak długo aż osiągnie wartość równą „nieskończoności", czyli w naszej metryce jest to wartość równa 16[26].

Z tego widać, dlaczego wartość „nieskończoności" jest tak mała. W przypadku utraty łączności z siecią chcielibyśmy, aby problem ewentualnej pętli został wyeliminowany jak najszybciej, dlatego wartość „nieskończoności" powinna być jak najmniejsza[26]. Z drugiej strony musi być ona większa od każdej rzeczywistej trasy mogącej wystąpić w sieci. Wybór wartości „nieskończoności" metryki jest więc kompromisem pomiędzy maksymalnym rozmiarem sieci, a zbieżnością w przypadku powstawania problemu „liczenia do nieskończoności". Twórcy protokołu RIP uznali, że mało prawdopodobne jest praktyczne wykorzystywanie tego protokołu w sieciach o rozpiętości większej niż 15 skoków[26].

Wprowadzono kilka mechanizmów zabezpieczających sieć przed powstawaniem pętli takich jak przedstawiona wcześniej pomiędzy dwoma routerami, jak również przed pętlami rozległymi, obejmującymi swym zasięgiem większe grupy routerów. Przykładowo, protokół RIP używa mechanizmu nazywanego dzieleniem horyzontu (split horizon), oraz mechanizmu aktualizacji wymuszonej (triggered updates)[2].

5.3 Dzielenie horyzontu (split horizon)

Zwróćmy uwagę, że problem powstały w omawianym wcześniej przykładzie powstał, ponieważ routery zaczęły odsyłać do siebie nawzajem pakiety, sądząc, że ten drugi wie, jak dostarczyć pakiet do sieci docelowej. Można by temu zapobiec, gdyby router nie ogłaszał do konkretnego sąsiada sieci, których nauczył się od tegoż sąsiada. W praktyce ogłaszanie takich sieci jest nieużyteczne[26]. Dzielenie horyzontu jest mechanizmem opartym na tym właśnie spostrzeżeniu. Dzielenie horyzontu nazywane jest także blokadą ogłaszania wstecznego albo zakazem „uczenia swojego nauczyciela". Wyróżniamy dwa rodzaje reguł dzielenia horyzontu: proste dzielenie horyzontu (simple split horizon) i dzielenie horyzontu z zatruwaniem wstecznym (split horizon with poisoned reverse). W prostym dzieleniu horyzontu w wysyłanych uaktualnieniach pomija się trasy nauczone od routera, do którego wysyłamy uaktualnienie[1]. W dzieleniu horyzontu z zatruwaniem wstecznym trasy te nie są pomijane, lecz wysyłane z metryką „nieskończoność", czyli oznaczającą nieosiągalność sieci docelowej.

Zastosujmy dzielenie horyzontu z zatruwaniem wstecznym dla wcześniej omawianego przykładu. W momencie, gdy router R1 uzna, że może przesyłać pakiety do sieci docelowej za pośrednictwem R3, to komunikaty o sieciach wysyłane do R3 będą zawierały sieć docelową oznaczoną jako niedostępną. Jeśli trasa przez R3 rzeczywiście prowadzi do docelowej sieci, to albo R3 jest bezpośrednio podłączony do tej sieci, albo przesyła pakiety skierowane do sieci docelowej przez inny router. Routerem tym nie jest jednak na pewno router R1, bo ten przekazał uaktualnienie, w którym oznajmił, że sieć docelowa jest dla niego nieosiągalna. W ten sposób unikamy możliwości zapętlenia w obrębie tych dwóch routerów[26].

Takie rozumowanie jest prawidłowe dla połączenia „point-to-point" routerów[26]. Co jednak stanie się, gdy R1 i R3 połączone są za pomocą sieci rozgłoszeniowej, takiej jak Ethernet i oprócz routerów R1 i R3 występują jeszcze inne routery?

Rozważmy więc taką sytuację. Zgodnie z zasadą dzielenia horyzontu z zatruwaniem ścieżek, gdy router R1 posiada wpis w tablicy o osiągalności sieci docelowej poprzez router R3, powinien wysłać uaktualnienie od R3, w którym oznacza sieć docelową jako niedostępną[26]. Uaktualnienie to wysyła metodą rozgłoszeniową, więc wszystkie routery w obrębie tej sieci otrzymują tą wiadomość i uznają, że router R3 oznaczył sieć docelową jako niedostępną. Nie jest to błędem, ponieważ pozostałe routery w sieci mogą się połączyć z routerem R3 bezpośrednio i wcale nie muszą wiedzieć, że router R1 tak właśnie robi chcąc dostać się do sieci docelowej. Oznacza to, że to samo uaktualnienie może być wysyłane zarówno do R3 jak i do wszystkich innych routerów w tej samej sieci. Oznacza to, że mechanizm dzielenia horyzontu z zatruwaniem wstecznym można stosować używając rozgłoszenia do wymiany informacji o sieciach.

Ogólnie rzecz biorąc, dzielenie horyzontu z zatruwaniem wstecznym jest bezpieczniejsze w użyciu, niż proste dzielenie horyzontu[26]. Jeśli dwa routery posiadają wpisy o trasie wskazującej nawzajem na siebie, to ogłoszenie takiej trasy przez jeden z tych routerów z metryką „nieskończoność" spowoduje natychmiastowo przerwanie powstałej pętli[26]. W przypadku nie ogłaszania tej trasy w ogóle (proste dzielenie horyzontu), trasa ta zostanie usunięta dopiero po upływie pewnego czasu.

W pewnych jednak sytuacjach stosowanie prostego dzielenia horyzontu jest bardziej opłacalne[14]. Wyobraźmy sobie, bowiem sytuacje, w której pojedyncze routery łączą sieci lokalne ze szkieletem sieci. Użyteczną informacją, jaką powinny przesyłać te routery do sieci szkieletowej jest adres sieci lokalnej. I tylko taka informacja znajdzie się w przesyłanych uaktualnieniach, gdy zastosujemy proste dzielenie horyzontu. Z drugiej strony routery otrzymują od szkieletu sieci wiadomości o wszystkich innych podłączonych sieciach. Gdybyśmy zastosowali tu mechanizm dzielenia horyzontu z zatruwaniem ścieżek routery dodatkowo przesyła-

łyby do szkieletu sieci wszystkie trasy, których nauczyły się tą drogą z metryką „nieskończoność”[26]. Z tego względu przesyłane byłoby dużo zbędnych w tym przypadku informacji.

5.4 Aktualizacja wymuszona (triggered updates)

Mechanizm dzielenia horyzontu skutecznie zapobiega powstawaniu pętli w obrębie dwóch routerów[26]. Ciągle możliwe jest jednak powstanie pętli obejmującej trzy lub więcej routerów. Przykładowo router R1 może sądzić, że osiągnie sieć docelową poprzez router R2, router R2 uważa, że może to osiągnąć kierując pakiety do R3, a ten z kolei myśli, że prawidłową trasą jest trasa przez router R1. Taka pętla zostanie wykryta i usunięta dopiero po upływie czasu, jaki potrzebny jest, aby metryka trasy prowadzącej do sieci docelowej osiągnęła wartość „nieskończoność”[26]. Z pomocą przychodzi nam mechanizm aktualizacji wymuszonej, który próbuje przyspieszyć proces wykrywania tego typu pętli.

Aktualizacja wymuszona polega na wysyłaniu komunikatów o zmianie metryki w tablicy routingu za każdym razem, gdy taka zmiana nastąpi[25]. Komunikaty te są wysyłane zaraz po wystąpieniu zmiany i są niezależne od regularnych ogłoszeń o stanie sieci.

Przyjmijmy, że router G wysłał wymuszone uaktualnienie na skutek zmiany metryki trasy w swojej tablicy routingu. Otrzymujący to uaktualnienie router K sprawdza czy uaktualnienie to dotyczy którejś z tras w jego tablicy trasowania. Jeśli tak jest to wprowadza zmiany w metryce trasy na podstawie otrzymanego komunikatu niezależnie od tego czy metryka się zwiększa czy zmniejsza[26]. Ponieważ zmianie uległa metryka trasy to K również wysyła uaktualnienie wymuszone do wszystkich sąsiadujących routerów. W ten sposób dochodzi do kaskadowego wysyłania komunikatów wymuszonych. Łatwo można prześledzić, które z routerów są przesyłają wymuszone uaktualnienia. Aby to pokazać przyjmijmy, że router G uznaje na podstawie liczników trasy sieć docelową N1 za nieosiągalną. Wysyła więc wymuszone uaktualnienie do wszystkich swoich sąsiadów o zmianie metryki tej trasy na wartość „nieskończoność”. Tą wiadomość zaakceptują tylko te routery, których trasy do sieci N1 prowadzą przez router G. Te routery na skutek zmiany metryki w tablicy wyślą wymuszone uaktualnienie dalej[26].
Pozostałe routery posiadające wpis o trasie do sieci docelowej nie przechodzącej przez router G, uznają ogłoszoną przez G trasę za gorszą od posiadanej obecnie i zignorują ją. Ponieważ nie zmieniły one wartości metryki żadnej z tras, toteż nie są zobowiązane do wysyłania wymuszonych uaktualnień.
W ten sposób kaskadowe wymuszone uaktualnienie spowoduje przypisanie wartości „nieskończoność” do wszystkich tras prowadzących do sieci N1 przez router G.

W wyżej przedstawionym scenariuszu wydarzeń można udowodnić, że „liczenie do nieskończoności” nigdy nie wystąpi, ponieważ nieprawidłowe trasy momentalnie są usuwane i nie ma możliwości powstania zapętleń[26]. Niestety rzeczywistość nie zawsze pasuje do tego scenariusza. Podczas przesyłania wymuszonych aktualizacji może zdarzyć się, że będą przesyłane równocześnie regularne komunikaty o stanie sieci. Z tego powodu routery, które nie otrzymały jeszcze komunikatu wymuszonej aktualizacji wysyłać będą do swych sąsiadów informacje o trasach, które już nie istnieją. Zdarzyć się więc może, że taką informacje o nieistniejącej już trasie otrzyma router, który wcześniej dowiedział się o niedostępności tejże trasy z komunikatu wymuszonej aktualizacji i przypisał jej metryce wartość „nieskończoność”. Dochodzi w ten sposób do przekłamań o trasie. W przypadku szybkiego przesyłania komunikatów wymuszonej aktualizacji sytuacja taka jest mało prawdopodobna, jednakże nie mamy stuprocentowej pewności, że „liczenie do nieskończoności” nie wystąpi[3].

6 Protokół RIP

6.1 Wstęp

RIP (Routing Information Protocol) jest protokołem pozwalającym hostom i routerom na wymianę informacji w celu wyznaczania tras w sieciach IP. Jest protokołem wektora odległości, stąd też generalnie odpowiada opisowi z rozdziału 5. RIP może być implementowany zarówno na hostach jak i na routerach[4]. W dalszej części opracowania termin „host" będzie się odnosił do obu.

W założeniu każdy host, na którym zaimplementowano protokół routingu RIP posiada interfejsy do jednej lub więcej sieci. Są to sieci bezpośrednio podłączone do hosta. Protokół opiera swoje działanie na podstawie informacji o tych sieciach. Najważniejszą informacją jest wartość metryki sieci lub inaczej „koszt" trasy prowadzącej przez tą sieć[3]. Wartość metryki zawiera się w przedziale od 1 do 15 włącznie i w większości implementacji przyjmuje wartość 1. Nowsze implementacje, w tym implementacje na routerach Cisco, pozwalają na „ręczne" przypisanie przez administratora wartości metryki do danej sieci[2]. Oprócz metryki, każda z sieci bezpośrednio przyłączonych do routera ma przypisany adres IP i maskę podsieci związaną z tym adresem. Wszystkie te parametry dla danego interfejsu ustala administrator w procesie konfiguracji hosta.

Każdy host używający protokołu RIP posiada tablice routingu, zawierającą wpisy mówiące o wszystkich docelowych sieciach i stacjach, które są osiągalne poprzez system oparty na tym protokole. Każdy taki wpis składa się z pól[26]:

- adres IP docelowej sieci lub stacji
- metryka, opisująca całkowity koszt dotarcia pakietu do odbiorcy
- adres IP następnego routera na ścieżce. Jeśli odbiorca znajduje się w jednej z bezpośrednio podłączonych do routera sieci to adres ten jest ignorowany.
- pole flagi wskazujące, że informacja na temat trasy została ostatnio zmieniona
- liczniki związane z trasą.

6.2 Format przesyłanych wiadomości.

Wiadomości przesyłane są w postaci datagramów UDP. Każdy host, na którym zaimplementowany jest protokół routingu RIP posiada oddzielny proces odpowiedzialny za wysyłanie i przyjmowanie datagramów UDP na i z portu o numerze 520. Specyficzne zapytania lub żądania mogą przychodzić z portów innych niż 520, jednak odpowiedzi na te zapytania kierowane są na port 520[14].

Protokół przewiduje istnienie tzw. „cichych" procesów RIP, które nasłuchują wiadomości, jakie przychodzą od innych hostów, ale same nie wysyłają żadnych komunikatów [26]. Hosty, które używają takich procesów nie pełnią funkcji routera, lecz poprzez nasłuchiwanie komunikatów aktualizacyjnych od routerów monitorują one stan sieci. Nasłuchiwanie to pozwala im budować tablice routingu opisującą aktualną postać systemu. Przejście do stanu, w którym host tylko nasłuchuje i nie wysyła żadnych wiadomości do sieci może być przydatne w sytuacji, gdy host straci połączenie ze wszystkimi oprócz jednej bezpośrednio podłączonymi sieciami. W takiej sytuacji, przestaje on być routerem pośredniczącym w wymianie informacji między sieciami i może zdecydować się na nie wysyłanie wiadomości do sieci, a tym samym stać się tylko biernym obserwatorem[26].

Przejście do stanu „cichego” procesu na routerze nie powinno mieć jednak miejsca, w przypadku, gdy istnieje prawdopodobieństwo, że sąsiednie routery uzależniają wykrycie ponownie dostępnej sieci od komunikatów aktualizacyjnych wysyłanych przez ten router[26].

komenda(1 oktet)	wersja(1 oktet)	musi być zero(2 oktety)
identyfikator rodziny adresów(2 oktety)		musi być zero(2 oktety)
adres IP(4 oktety)		
musi być zero(4 oktety)		
musi być zero(4 oktety)		
metryka(4 oktety)		
. . . . .		

Tabela 2) Format ramki protokołu RIP

Zanim nastąpi interpretacja informacji zawartych w ramce sprawdzane jest pole numeru wersji. Oto wartości, jakie może przyjmować to pole i czynności realizowane w razie ich wystąpienia:

0 ramki, które przyjmują taką wartość pola numeru wersji powinny być ignorowane.

1 gdy numer wersji przyjmuje wartość 1 ramka powinna być przetwarzana. Jeśli jednak którekolwiek z pół oznaczonych jako „musi być zero” przyjmuje wartość niezerową to cała ramka powinna być odrzucona.

>1 ramka powinna być przetwarzana, a wszystkie dane zawarte w polach „musi być zero” powinny zostać zignorowane. W takim przypadku oznacza to, że jest to ramka stworzona przez protokół RIP o numerze wersji większym od 1, w której przesyłane są dodatkowe, specyficzne dla tego protokołu informacje w polach „musi być zero”, które to informacje powinny być zignorowane przez protokół w wersji pierwszej. Wszystkie inne pola powinny być interpretowane tak jak jest to opisane w dalszej części pracy.

Każda ramka zawiera komendę, numer wersji protokołu i rekordy opisujące poszczególne obiekty docelowe. Rozmiar datagramu, czyli ramki, może mieć maksymalnie rozmiar 512 oktetów, nie wliczając w to nagłówków IP czy UDP[26].
W skład pojedynczego rekordu wchodzą pola od „identyfikatora rodziny adresów” do „metryki”. Ze względu na ograniczenie rozmiaru datagramu, w pojedynczej ramce może wystąpić do 25 rekordów. Jeśli takich rekordów do przesłania jest więcej niż 25 tworzona jest kolejna ramka. Ze względu na rodzaj przesyłanych informacji i sposób ich przetwarzania nie jest wymagane specjalne oznaczanie takiej ramki jako zawierającej dalszą część tablicy routingu[26].

W wersji pierwszej protokołu RIP możliwe są następujące wartości pola „komenda”:

1- request – żądanie przesłania całości lub części tablicy routingu
2- response – ramka zawiera całość lub część tablicy routingu nadawcy. Taka ramka może być przesłana w odpowiedzi na komendę „request” lub być ramką aktualizacyjną.
3- traceon – komenda nieaktualna. Wiadomości zawierające taką komendę powinny być zignorowane.
4- traceoff - komenda nieaktualna. Wiadomości zawierające taką komendę powinny być zignorowane.
5- reserved – komenda zarezerwowana przez Sun Microsystems na użytek własny.

Każdy rekord zawiera dwie informacje o docelowej stacji lub sieci, są to: adres i metryka [1]. Format rekordu jest tak przewidziany, aby pozwalał protokołowi RIP przesyłać in-

formacje o trasowaniu dla różnych innych protokołów. Dlatego, każdy rekord zawiera pole „identyfikator rodziny adresów" określający, jaki rodzaj adresu występuje w tym rekordzie. W opracowaniu tym rozważana jest tylko rodzina adresów IP. Dla przyszłych zastosowań, wymagane jest od implementacji protokołu, aby nie uwzględniały informacji zawartych w rekordach, których identyfikator rodziny adresów nie został rozpoznany[26]. Identyfikator rodziny adresów dla sieci IP jest równy 2.

Adres IP jest przeważnie 4-oktetowym adresem zapisanym w porządku sieciowym. Pole metryka musi zawierać wartość z przedziału od 1 do 15 włącznie opisującą bieżącą odległość do obiektu docelowego, lub wartość 16 oznaczającą nieosiągalność tegoż obiektu [14]. Zatem wartość „nieskończoność" używana w rozdziale 5 jest w przypadku protokołu RIP równa 16.

6.3 Maski i adresy sieci

Jak wspomniano już wcześniej, protokoły wektora odległości mogą opisywać trasy zarówno do indywidualnych hostów, jak i do sieci [26]. Protokół RIP umożliwia oba te warianty. Pole „adres IP" w ramce wiadomości może odnosić się do sieci, hostów, lub zawierać specjalne oznaczenie wskazujące na trasę domyślną. W większości przypadków rodzaj informacji znajdującej się w polu adresowym zależeć będzie od przyjętej strategii routingu. Tak więc, wiele sieci jest tak zorganizowana, że przechowywanie tras do poszczególnych hostów jest zbyteczne. Przykładem może być sieć, w której każdy host dostępny jest poprzez ten sam router. Jednakże, sieci zawierające połączenia punkt-punkt wymagają czasami, aby przechowywać trasy do pewnych hostów.
W przypadku, gdy implementacja protokołu RIP nie obsługuje tras do pojedynczych hostów trasy takie, otrzymane od sąsiadujących routerów, są pomijane[26].

Format ramki protokołu RIP nie rozróżnia rodzaju informacji zawartej w polu adresowym. Pole to może zawierać jedną z następujących informacji:

- adres hosta
- numer podsieci
- numer sieci
- 0, oznaczające trasę domyślną

W procesie trasowania RIP jest odpowiedzialny za wybranie, na podstawie wpisów w tablicy, najbardziej odpowiedniej trasy prowadzącej do odbiorcy[1]. Dlatego przeglądając tablice w poszukiwaniu najbardziej odpowiedniego wpisu odpowiadającego odbiorcy przeglądane są najpierw wpisy dotyczące pojedynczych hostów, następnie wpisy dotyczące podsieci i sieci. Jeśli nie znaleziono wpisu pasującego do adresu odbiorcy wybierana jest trasa oznaczona jako domyślna.

Interpretacja danych zawartych w polu adresu zależy od znajomości maski przypisanej do danej sieci. Protokół RIP należy do grupy protokołów klasowych, których podstawową cechą jest to, iż nie ogłaszają one maski podsieci razem z adresem sieci [25]. Stąd mogą się pojawić pewne problemy przy przetwarzaniu adresów dotyczących sieci i hostów. Aby pokazać, jak może zostać zinterpretowana informacja o obiektach docelowych posłużymy się przykładem. Rozważmy sieć 128.6.0.0. Ma ona maskę podsieci postaci: 255.255.255.0. Tak więc 128.6.0.0 jest numerem sieci, 128.6.4.0 jest numerem podsieci, a 128.6.4.1 jest adresem hosta. Jednakże, jeśli host nie zna maski podsieci, interpretacja adresu może być dwuznaczna, ponieważ w przypadku, gdy część hosta ma wartość niezerową, nie ma możliwości stwierdzenia czy adres odnosi się do numeru podsieci czy adresu hosta. Adres ten jako numer podsieci będzie bezużyteczny bez maski, dlatego przyjęte zostanie, że jest to adres reprezentujący

hosta[26]. W celu uniknięcia opisanej dwuznaczności hosty nie mogą rozsyłać informacji o podsieciach do hostów, które przypuszczalnie nie znają odpowiedniej maski tejże podsieci. Ponieważ hosty znają tylko maski podsieci sieci, do których są bezpośrednio podłączone, dlatego wiadomości o podsieci nie mogą być przesyłane poza sieć, której częścią jest rozważana podsieć[26].

Opisaną filtracją przesyłanych informacji zajmują się routery umieszczone na „obrzeżach" sieci podzielonej na podsieci, czyli te routery, które łączą tą sieć z inną siecią[1]. Wewnątrz zdefragmentowanej sieci 128.6.0.0 każda podsieć traktowana jest jako indywidualna sieć, a informacja o tejże podsieci rozsyłana jest przez RIP tylko wewnątrz sieci 128.6.0.0. Routery brzegowe natomiast rozsyłają do innych sieci (innych niż 128.6.0.0) informacje o sieci 128.6.0.0 jako o całości, w postaci pojedynczego wpisu w wiadomości aktualizacyjnej. To oznacza, że routery brzegowe rozsyłają różne informacje do różnych sąsiadów, w zależności czy dany router sąsiadujący jest bezpośrednio podłączony do sieci 128.6.0.0 czy też nie[26].

Do oznaczenia tras domyślnych używa się specjalnego adresu 0.0.0.0.[14] Tras domyślnych używamy, gdy nie jest możliwe umieszczanie wszystkich sieci w tablicy routingu i wiadomościach aktualizacyjnych. Wpis trasy domyślnej w tablicy trasowania jest zwykłym wpisem, w którym adres sieci ma wartość 0.0.0.0. Wpis taki wprowadza administrator systemu, który określa router będący routerem domyślnym oraz metrykę. Wpisów dla tras domyślnych może być kilka. W takim przypadku metryka decyduje o wyborze danego routera domyślnego [26]. Wpisy dla routerów domyślnych są obsługiwane przez RIP dokładnie tak samo jak wpisy dotyczące rzeczywistych numerów sieci.

6.4 Liczniki

Co pewien, określony czas, zwany czasem aktualizacji, proces wyjścia generuje wiadomość aktualizacyjną i wysyła ją do routerów sąsiadujących[1]. W implementacji Cisco protokołu RIP domyślnie jest to 30 sekund[2].. W przypadku wielu routerów wymieniających informacje w obrębie danej sieci wysoce niepożądane jest, aby wiadomości aktualizacyjne przesyłane były w tym samym czasie, zwłaszcza jeśli przepustowość linii łączących jest niska[26]. Może to prowadzić do kolizji w sieciach rozgłoszeniowych. Aby zapobiec takiej synchronizacji stosuje się kilkusekundowe, losowo generowane przesunięcie czasowe dodawane do liczby 30 sekund.

Z każdym wpisem w tablicy routingu związane są dwa liczniki: „timeout" i „garbage-collection time"[26]. Po upływie czasu „timeout" trasa oznaczana jest jako niedostępna, jednakże wpis tej trasy pozostaje w tablicy jeszcze przez pewien czas, aby sąsiednie routery mogły zostać poinformowane o fakcie niedostępności tejże trasy. Po upływie czasu „garbage-collection time" trasa jest usuwana z tablicy routingu.

Licznik „timeout" uruchamiany jest przy operacji wpisu trasy do tablicy i przy każdym otrzymaniu wiadomości aktualizacyjnej dotyczącej trasy, do której odnosi się dany licznik. Po upływie 180 sekund od czasu ostatniej aktualizacji trasy, wpis uznawany jest za niepotrzebny i rozpoczynany jest proces usuwania go z tablicy [14].

Usunięcie wpisu z tablicy routingu może nastąpić z dwóch powodów:

i) upływu czasu
ii) otrzymania wiadomości aktualizacyjnej mówiącej o niedostępności trasy od routera figurującego w tym wpisie jako następny router na trasie

W każdym z tych przypadków podejmowane są następujące kroki:

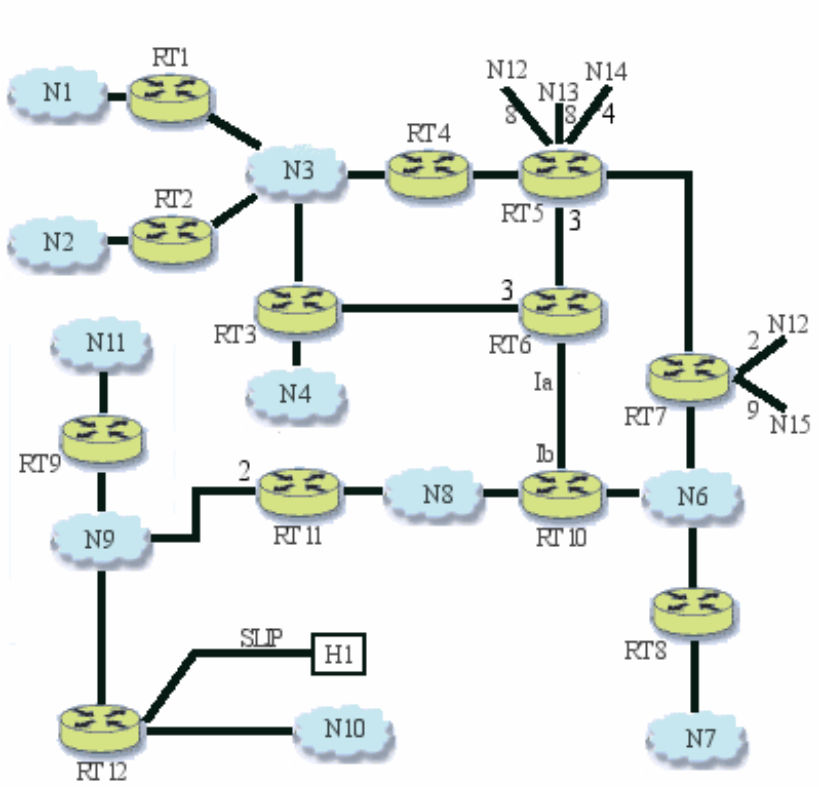
- licznik „garbage collection time" jest ustawiany na 120 sekund
- metryka trasy przyjmuje wartość 16 (nieskończoność)

- ustawiana jest flaga informująca o zmianie wpisu, co sygnalizuje procesowi wyjściowemu nadanie wiadomości wymuszonych (triggered updates)

Do upływu czasu „garbage collection time” trasa jest umieszczana w wiadomościach aktualizacyjnych z metryką równą 16, oznaczając w ten sposób niedostępność trasy. Po upływie tego czasu trasa prowadząca do danego obiektu docelowego jest całkowicie usuwana z tablicy routingu[26]. W przypadku, gdy podczas upływu czasu „garbage collection time” ustalona zostanie nowa trasa do obiektu docelowego, trasa ta jest umieszczana w miejscu trasy czekającej na usunięcie. W tym przypadku licznik „garbage collection time” musi zostać wyzerowany.

6.5 Przykład topologii

Prześledźmy postać tablicy routingu na routerze, na którym zaimplementowany jest protokół RIP. Rysunek 2 przedstawia schemat topologii, w której pracują routery.



Rysunek 2) Przykład topologii sieci dla RIP

Oto postać tablicy routingu RT6:

Sieć docelowa	Metryka	Następny skok
N1	4	RT5
N2	4	RT5
N3	3	RT5
N4	4	RT5
N6	2	RT10
N7	3	RT10
N8	2	RT10
N9	4	RT10
N10	5	RT10
N11	4	RT10
N12	4 4	RT5 RT10
N13	9	RT5
N14	5	RT5
N15	11 11	RT5 RT10

Tabela 3) Tablica routingu routera RT6

Przypuśćmy, że połączenie między routerami RT6 i RT5 zostało przerwane. Po upływie 180 sekund (domyślna wartość implementacji Cisco) trasy, których następnym routerem na trasie jest RT5 zostaną oznaczone jako niedostępne. Te sieci to N1, N2, N3 i N4. Ponieważ w tablicy routingu nie było innych wpisów dotyczących sieci N1, N2, N3 i N4 to router RT6 nie posiada informacji, w jaki sposób kierować pakiety, których odbiorca znajduje się w jednej z tych sieci.

Router RT6 posiada jednak nadal aktywne połączenia z routerami RT3 i RT10 i od nich otrzymuje co 30 sekund wiadomości aktualizacyjne. W wiadomościach tych zawarte są również informacje na temat trasowania do sieci N1, N2, N3 i N4. Router RT3 wysyła wiadomości, w których informuje o koszcie dotarcia do poszczególnych sieci. Oto informacje przesyłane przez RT3 dotyczące omawianych sieci i ich interpretacja na routerze RT6:

Sieć docelowa	Metryka od RT3	Metryka na RT6
N1	2	5 (2+3)
N2	2	5 (2+3)
N3	1	4 (1+3)
N4	1	4 (1+3)

Tabela 4) Metryki przesyłane przez router RT3 i ich interpretacja na routerze RT6

A oto informacje przesyłane przez router RT10:

Sieć docelowa	Metryka od RT10	Metryka na RT6
N1	5	6 (5+1)
N2	5	6 (5+1)
N3	4	5 (4+1)
N4	5	6 (5+1)

Tabela 5) Metryki przesyłane przez router RT10 i ich interpretacja na routerze RT6

Jak widać trasy prowadzące przez router RT3 mają mniejsze wartości metryk od tras prowadzących przez RT10. Z tego też względu trasy nadesłane od RT3 zostaną umieszczone w tablicy routingu. Postać tablicy routingu dotycząca omawianych sieci będzie następująca:

Sieć docelowa	Metryka	Następny skok
N1	5	RT3
N2	5	RT3
N3	4	RT3
N4	4	RT3

Tabela 6) Nowe wpisy w tablicy routingu routera RT6

7 Protokół RIP v2

7.1 Wstęp

Protokół RIPv2 jest rozszerzeniem protokołu RIP w wersji pierwszej [25]. Rozszerzenie to odnosi się głównie do przesyłania dodatkowych, ważnych informacji o trasach w wiadomościach aktualizacyjnych.

7.2 Format przesyłanych wiadomości

Nagłówek ramki stosowanej do ogłaszania tras jest taki sam jak w wersji pierwszej protokołu [33]. Zmieniła się natomiast postać pojedynczego wpisu dla trasy i ma postać:

identyfikator rodziny adresów(2 oktety)	znacznik trasy(2 oktety)
adres IP(4 oktety)	
maska podsieci(4 oktety)	
następny skok(4 oktety)	
metryka(4 oktety)	

Tabela 7) Format ramki protokołu RIP v2

Pola: identyfikator rodziny adresów, adres IP i metryka mają identyczne znaczenie jak w wersji pierwszej protokołu i są opisane w rozdziale 6.2.

7.2.1 Uwierzytelnianie

W RIPv2 wprowadzono możliwość wzajemnego uwierzytelniania routerów wymieniających informacje [25]. Pozwala to wyeliminować z sieci routery nieautoryzowane, od których nie będą akceptowane ogłoszenia. Dla zapewnienia pełnej współpracy ze starszymi urządzeniami, które posługują się tylko wersją pierwszą RIP, dodano komendy pozwalające włączyć pełną zgodność z wersją pierwszą.

Uwierzytelnianie odnosi się do całych ramek przesyłanych w procesie wymiany informacji o sieci. Ponieważ w nagłówku ramki dostępne jest tylko jedno pole dwuoktetowe, a sensowny algorytm uwierzytelniania potrzebować będzie więcej niż dwa oktety, stąd algorytm uwierzytelniania dla RIPv2 używa przestrzeni za nagłówkiem ramki, czyli przestrzeni przeznaczonej na wpisy dla tras[33]. Aby rozróżnić zwykły wpis od informacji potrzebnej do uwierzytelniania w polu „identyfikator rodziny adresów" umieszczana jest wartość 0xFFFF. W takim wypadku postać ramki zawierającej wpisy z danymi do uwierzytelniania rozpoczyna się następująco:

komenda(1 oktet)	wersja(1 oktet)	nieużywane
0xFFFF		rodzaj uwierzytelniania(2 oktety)
informacje uwierzytelniania(16 oktetów)		

Tabela 8) Format ramki z danymi do uwierzytelniania

Ponieważ maksymalna ilość wpisów w pojedynczej ramce wynosi 25, stąd informacje uwierzytelniania mogą zajmować do 24 wpisów[33].

Obecnie uwierzytelnianie opiera się na prostej wymianie hasła. Rodzaj tego uwierzytelniania oznaczony jest liczbą 2. Pozostałe 16 oktetów zawiera zwykły tekst hasła. Jeśli hasło ma mniej niż 16 oktetów, musi zostać dopełnione do 16 oktetów zerami[33].

7.2.2 Znacznik trasy

Zamierzeniem tego pola jest dostarczenie metody pozwalającej odseparować „wewnętrzne" trasy protokołu RIP (trasy do sieci wewnątrz obszaru, na którym stosowany jest RIP) od „zewnętrznych" tras, które mogą być obsługiwane przez protokoły EGP lub inne protokoły IGP[33].

Routery obsługujące protokoły inne niż RIP powinny pozwalać na ustawienie takiej konfiguracji, aby pole znacznika trasy było wypełniane w zależności od pochodzenia informacji o danej trasie. Przykładowo trasy otrzymane od routerów posługujących się protokołem BGP (patrz rozdział 13) powinny posiadać znaczniki tras tegoż protokołu, o wartości ustalonej dowolnie, lub o wartości będącej numerem systemu autonomicznego, z którego nauczył się router danej trasy[33].

Dozwolone jest również inne wykorzystanie pola „znacznik trasy" pod warunkiem jednak, że wszystkie routery systemu autonomicznego używają go w ten sam sposób. Takie rozwiązanie umożliwia wykorzystanie tego pola do określania metod wymiany danych pomiędzy routerami używającymi protokołów BGP i RIP.

7.2.3 Maska podsieci

W polu „maska podsieci" przesyłana jest maska przypisana do adresu IP [25]. Umożliwia to jednoznaczne określenie części hosta w adresie. Przesyłanie maski podsieci wraz z adresem jest znaczącą zmianą w stosunku do RIPv1. W ten sposób w wersji drugiej protokół RIP nadal jest protokołem wektora odległości, ale bezklasowym. Nie występuje już problem sieci nieciągłych, można także wyłączyć automatyczne łączenie tras na granicy sieci głównych[33]. Dzięki rozsyłaniu maski podsieci protokół RIP w wersji drugiej obsługuje sieci VLSM (Variable Length Subnet Masking), czyli te, w których stosuje się różnej długości maskę dla podsieci tej samej sieci głównej.

W przypadku, gdy omawiane pole ma wartość zero, przyjmuje się, że maska dla tego wpisu nie jest określona. W przypadku przesyłania ramek do routerów, na których zaimplementowany jest protokół RIP-1 muszą być zastosowane specjalne zasady, co do ogłaszanych sieci[1]. Dzieje się tak, bo RIP-1 nie uwzględnia pola maski w interpretacji adresu, co może prowadzić do błędów (patrz też opis tego problemu w rozdziale 6.3).
Zasady te są następujące:

i) wewnętrzne informacje o sieci nie mogą być ogłaszane do innych sieci
ii) informacje o specyficznych podsieciach nie mogą być ogłaszane do tych routerów, na których RIP-1 może zinterpretować numery tych podsieci jako adresy hostów

7.2.4 Następny skok

Jest to adres IP, na który powinny być wysyłane pakiety przeznaczone dla opisanego we wpisie odbiorcy[33]. Adres ten musi być siłą rzeczy bezpośrednio osiągalny dla routera. Podstawowym celem stosowania pola „następny skok" jest wyeliminowanie sytuacji, w których pakiety kierowane są niepotrzebnie przez dodatkowe routery. Jest to szczególnie użyteczne, gdy protokół RIP nie jest zaimplementowany na wszystkich routerach w sieci.

Jeśli wartością tego pola jest 0.0.0.0, oznacza to, że routing powinien odbywać się poprzez router, który wysłał to ogłoszenie o trasie.

7.3 Mutliemisja

W wersji drugiej protokołu RIP zoptymalizowano także sposób komunikacji z routerami sąsiednimi[25]. Nadal wykorzystywany jest port 520 protokołu UDP, ale transmisja realizowana jest w drodze multiemisji z wykorzystaniem specjalnej grupy o adresie 224.0.0.9. Dzięki temu ruch związany z protokołem RIP nie obciąża wszystkich urządzeń w danym segmencie, a jedynie routery należące do grupy 224.0.0.9[25].

7.4 Kompatybilność RIP-2 z RIP-1

Jak opisano już w rozdziale 6.2 przetwarzanie ramek uzależnione jest od wartości pola „numer wersji". I tak, ramki wersji 0 muszą być ignorowane, ramki wersji 1 są przetwarzane, pod warunkiem, że wszystkie pola oznaczone jako „musi być zero" przyjmują rzeczywiście wartość zero, a informacje zawarte w ramkach o wersji większej od 1 są przetwarzane w zależności od protokołu, jaki został zaimplementowany na routerze. Oznacza to, że nowe wersje protokołu RIP są całkowicie kompatybilne wstecz[33].

Ponieważ w RIP-2 możliwe jest zastosowanie algorytmów uwierzytelniania, co nie jest przewidziane w wersji pierwszej protokołu, dlatego muszą istnieć zasady określające sposób wymiany ramek w sytuacji stosowania mechanizmów uwierzytelniania przez RIP-2. Jeśli router, na którym zaimplementowany jest RIP-2 skonfigurowany tak, aby przeprowadzane było uwierzytelnianie dla ramek RIP-2, wtedy ramki pochodzące od RIP-1 i ramki od RIP-2, które przeszły pozytywnie proces uwierzytelniania mogą być przetworzone[33]. Ramki RIP-2, których proces uwierzytelniania zakończył się negatywnie są odrzucane. Aby zwiększyć bezpieczeństwo w systemach z uwierzytelnianiem, wiadomości od RIP-1 powinny być odrzucane, gdyż w przeciwnym wypadku uwierzytelnione informacje uzyskane od routerów z RIP-2 będą rozsyłane przez routery z RIP-1 w sposób nie zapewniający uwierzytelniania.

Warto zwrócić uwagę, że zastosowanie uwierzytelniania nie zapobiega przesyłaniu ramek do routerów z RIP-1[33]. A ponieważ dane potrzebne do uwierzytelniania zawarte są we wpisach oznaczonych poprzez wartość 0xFFFF w polu „identyfikator rodziny adresów", dlatego RIP-1 uzna te wpisy, za wpisy należące do innej rodziny adresów i zignoruje je. Dalsze wpisy natomiast, zawierające informacje o sieciach zostaną przetworzone przez RIP-1 w standardowy sposób. Jeśli jest to konieczne, można temu zapobiec wykorzystując multiemisję.

Jeśli router nie używa uwierzytelnianie akceptowane są wiadomości zarówno pochodzące od RIP-1 jak i od RIP-2.

8 Protokół IGRP

8.1 Wstęp

Protokół IGRP (Interior-Gateway Routing Protocol) opracowany został przez firmę Cisco w celu wyeliminowania niektórych ograniczeń protokołu RIP[8]. Głównym celem projektantów było stworzenie protokołu, który spełnia następujące zadania:

- zapewnia stabilny routing, również w dużych i złożonych sieciach,
- posiada skuteczne mechanizmy zapobiegania powstawaniu pętli,
- szybko reaguje na zmiany w topologii sieci,
- nie powoduje nadmiernego, dodatkowego ruchu w sieci,
- rozdziela ruch między kilka tras,
- dynamicznie określa obciążenie i pewność tras.

Takimi właśnie cechami dysponuje protokół IGRP[8].

Aktualna implementacja Cisco protokołu obsługuje jedynie trasy dla sieci opartych na TCP/IP, choć podstawowy projekt przewiduje użycie IGRP dla różnych protokołów[2].

Jako protokół wektora odległości i protokół klasowy IGRP podlega takim samym zasadom pracy, jak protokół RIP i w wielu punktach jest do niego podobny [8]. Ponieważ IGRP jest szeroko stosowany dla realizacji celów podobnych do celów osiąganych poprzez użycie protokołu RIP, dlatego bardzo ważne jest zrozumienie różnic występujących między tymi protokołami. Przy porównywaniu tych protokołów trzeba jednak pamiętać o fakcie, że RIP nie był projektowany w tym samym celu co IGRP[25]. RIP w zamierzeniu, miał za zadanie przeprowadzanie routingu w małych sieciach o jednolitej technologii i w takich sytuacjach sprawdza się całkiem dobrze[25]. Jeszcze kilkanaście lat temu protokół taki jak RIP był wystarczający do obsługi większości rzeczywistych sieci. Jednakże szybki rozwój Internetu i decentralizacja kontroli jego struktur, zaowocowała tym, że zarządzanie systemem jest prawie poza naszymi możliwościami. Takie sytuacje mogą zdarzyć się również w dużych sieciach korporacyjnych. IGRP jest narzędziem przychodzącym nam w takiej sytuacji z pomocą. Oczywiście żadne narzędzie nie rozwiązuje wszystkich problemów. Tak i jest w przypadku IGRP. Omawiany protokół zalicza się do protokołów grupy IGP (internal gateway protocols) i stosowany jest w odniesieniu do pojedynczych, wydzielonych obszarów sieci. Takie obszary kontrolowane są przez protokoły grupy EGP (external gateway protocols), o których mowa będzie w dalszej części pracy (patrz rozdział 13, 14).

8.2 Format przesyłanych wiadomości

Wiadomości aktualizacyjne są rozsyłane przy użyciu datagramów IP z protokołem nr 9 IP (IGP). Pakiety zaczynają się od nagłówka, znajdującym się bezpośrednio za nagłówkiem IP[8].

Oto postać nagłówka:

Wersja(4 bity)	Kod(4 bity)
Numer seryjny (1 oktet)	
Numer systemu autonomicznego(2 oktety)	
Liczba podsieci w sieci lokalnej(2 oktety)	
Liczba sieci w systemie autonomicznym(2 oktety)	
Liczba sieci poza systemem autonomicznym(2 oktety)	
Suma kontrolna nagłówka i danych	

Tabela 9) Format nagłówka wiadomości protokołu IGRP

Numer wersji jest aktualnie równy 1. Pakiety mające inną wartość w tym polu są ignorowane.

Kod określa rodzaj wiadomości i przyjmuje dwie wartości: 1 dla uaktualnień i 2 dla żądań. Szczegóły, co do formatów obu wiadomości będą podane w dalszej części.

Numer seryjny jest numerem zwiększanym za każdym razem, gdy nastąpiła zmiana w tablicy routingu. Numer ten pozwala routerom uniknąć zbędnego przetwarzania wiadomości, które już zostały przetworzone wcześniej. (Uwaga: nie jest to aktualnie implementowane, tzn. numer jest generowany i umieszczany w pakiecie poprawnie, jednak jest on ignorowany na wejściu[8])

W implementacji Cisco router może wchodzić w skład więcej niż jednego systemu autonomicznego[2]. Każdy z tych systemów uruchamia własny protokół IGRP, który to protokół posiada własną tablice routingu. Trasy, których uczy się IGRP z jednego systemu autonomicznego, są ogłaszane tylko w obrębie tego systemu. Pole „numer systemu autonomicznego" pozwala jednoznacznie określić routerowi, która tablica routingu powinna być aktualizowana przy przetwarzaniu tej wiadomości.

Wiadomość aktualizacyjna protokołu IGRP składa się z trzech części, odnoszących się do opisu różnych części topologii sieci:

- wewnętrznej – sekcja ta opisuje trasy do podsieci, ale tylko te podsieci są uwzględniane, które są częścią sieci, do której wędruje wiadomość aktualizacyjna
- systemu autonomicznego – tu znajdują się informacje dotyczące głównych sieci
- zewnętrznej – trasa znajdzie się w tej części wiadomości, jeśli router dowiedział się o tej trasie od innego routera i trasa ta była oznaczona jako zewnętrzna. Implementacja Cisco protokołu przewiduje możliwość oznaczenia sieci jako zewnętrznej. Trasy takie są uznawane jako kandydujące do bycia trasami domyślnymi. Implementacja Cisco wybiera spośród tras zewnętrznych tą z najmniejszą metryką i ta trasę traktuje jako domyślną[8].

Pola „Liczba podsieci w sieci lokalnej", „Liczba sieci w systemie autonomicznym", „Liczba sieci poza systemem autonomicznym" wskazują ile wpisów znajduje się w każdej z sekcji opisanej powyżej.

8.2.1 Rodzaje wiadomości

Wyróżniamy dwa rodzaju wiadomości: żądania i uaktualnienia. Pierwszy rodzaj wiadomości jest żądaniem skierowanym do odbiorcy o przesłanie tablicy routingu. Ten rodzaj wiadomości posiada jedynie nagłówek. Jedynie pola: wersja, kod i numer systemu autonomicznego są wykorzystywane, pozostałe pola zaś przyjmują wartość zero[8]. Odbiorca jest

zobowiązany do przesłania w odpowiedzi normalnej wiadomości aktualizacyjnej do nadawcy żądania.

Drugi rodzaj wiadomości to wiadomości aktualizacyjne. Zawierają one nagłówek, po którym bezpośrednio znajdują się rekordy opisujące trasy. Rekordów może być tyle ile zmieści się w datagramie o rozmiarze 1500 bajtów (włączając nagłówek IP). Przy obecnej postaci rekordów pozwala to na jednorazowe przesłanie 104 rekordów. Jeśli wpisów jest więcej tworzony jest kolejny datagram. Oto struktura pojedynczego rekordu:

Adres IP(3 oktety)
Opóźnienie(3 oktety)
Przepustowość(3 oktety)
MTU(2 oktety)
Pewność(1 oktet)
Obciążenie(1 oktet)
Liczba skoków(1 oktet)

Tabela 10) Struktura rekordu

Adres IP jest adresem obiektu docelowego. W celu oszczędności miejsca, przesyłane są jedynie 3 pierwsze bajty adresu IP, z wyjątkiem sekcji wewnętrznej, gdzie te trzy bajty są interpretowane jako trzy ostatnie bajty adresu obiektu docelowego. Taka oszczędność jest możliwa do zrealizowania, ponieważ w sekcji systemu autonomicznego i sekcji zewnętrznej bajt najmniej znaczący w adresie ma zawsze wartość zero, natomiast w sekcji wewnętrznej zbędny jest bajt najbardziej znaczący adresu, gdyż jest on taki sam jak sieci głównej, której częścią jest dana podsieć[8].

Opóźnienie jest wyrażone w dziesiątkach mikrosekund. Pole to pozwala przesłać wartość z przedziału od 10 mikrosekund do 168 sekund. Jeśli wartość tego pola jest równa 0xFFFFFF, czyli ustawione są jedynki na wszystkich bitach to oznacza to niedostępność trasy.

Przepustowość jest wyrażona jako odwrotność przepustowości liczonej w bitach na sekundę pomnożonej przez stałą 1.0e10. Inaczej mówiąc, jeśli przepustowość ma wartość N Kbps, to w polu tym wpisana zostanie wartość 10000000/N. Taka organizacja zapisu pozwala przesłać przepustowość z przedziału od 1200 BPS do 10 Gbps[8].

MTU jest mierzone w bajtach, natomiast pewność i obciążenie w ułamkach liczby 255.

8.3 Metryka i rozdzielanie ruchu

Jedną z najważniejszych cech odróżniających protokół IGRP od RIP jest zupełnie inny sposób obliczania metryki trasy. W protokole RIP koszt trasy opiera się tylko na liczbie skoków do sieci docelowej. Oznacza to, że nie ma możliwości rozróżnienia jakości tych tras[25]. W takiej sytuacji trasy oparte na wolnym łączu szeregowym i łączu Ethernet są traktowane identycznie.

IGRP również przesyła i monitoruje liczbę skoków, ale tylko w celu sprawdzania, czy trasa nie jest zbyt długa (255 skoków maksymalnie). Liczba skoków nie jest w ogóle brana pod uwagę przy wyliczaniu metryki [25]. W zamian uwzględnia się 4 następujące parametry:

- **Przepustowość** - wartość stała definiowana w konfiguracji interfejsu. Im większa wartość, tym bardziej preferowana trasa. Domyślnie uwzględniana w metryce.
- **Opóźnienie** - wartość stała definiowana w konfiguracji interfejsu. Im mniejsza wartość, tym bardziej preferowana trasa. Domyślnie uwzględniana w metryce.
- **Pewność** - wartość dynamicznie mierzona na poziomie interfejsu i wyrażana jako liczba z przedziału od 1 do 255. Im większa wartość (mniej błędów na interfejsie), tym bardziej preferowana trasa. Domyślnie nieuwzględniana w metryce.
- **Obciążenie** - wartość dynamicznie mierzona na poziomie interfejsu i wyrażana jako liczba z przedziału od 1 do 255. Im mniejsza liczba (mniej obciążony interfejs), tym bardziej preferowana trasa. Domyślnie nieuwzględniana w metryce.[14]

Pewność i Obciążenie są parametrami rzeczywistymi, zmieniającymi się w trakcie pracy interfejsu, oznaczałoby to również "ciągłą" zmianę metryk[8]. Aby temu zapobiec, wprowadzono rozwiązanie, w którym parametry te wyznaczane są z dokładnością do 5 minut na podstawie średniej ważonej z odczytów wykonywanych co 5 sekund.

Implementacja IGRP na routerach Cisco używa następujących wzorów do skalowania wartości przepustowości i opóźnienia[8]:

$BW = (10^7/\text{przepustowość})*256$
DELAY = opóźnienie * 256

Natomiast kompletny wzór, na podstawie, którego wyznacza się metrykę ma postać:

Wzór 1:

$$\text{metryka} = [k1*BW_{IGRP(min)} + (k2*BW_{IGRP(min)})/(256-LOAD) + k3*DELAY_{IGRP(sum)}] * [k5/RLBT + k4]$$

gdzie:
BW – przepustowość
LOAD – obciążenie
DELAY – opóźnienie
RLBT – pewność
k1...k5 – stałe współczynniki; jeśli k5 ma wartość 0, to pewność nie jest brana pod uwagę

Zwróćmy uwagę na stosowane we wzorze stałe od k1 do k5. Jeżeli stała k5 przyjmuje wartość 0 a k4 wartość 1, ostatni składnik wzoru (nawias kwadratowy) nie jest w ogóle uwzględniany. Domyślnie stała k1=k3=1, a pozostałe stałe mają wartość 0, co oznacza, że powyższy wzór przekształca się do następującego:

Wzór 2:

$$\text{metryka} = BW_{IGRP(min)} + DELAY_{IGRP(sum)}$$

Przy obliczaniu wartości przepustowości na całej trasie łączącej nadawcę z odbiorcą bierze się pod uwagę wszystkie występujące na trasie interfejsy wejściowe i wyjściowe i spo-

śród nich wybiera się przepustowość o najmniejszej wartości. Inaczej jest w przypadku obliczania wartości opóźnienie, gdzie wartość wyliczana jest jako suma opóźnień wprowadzanych przez interfejsy wyjściowe[8].

Trasa posiadająca najmniejszą wartość metryki uznawana jest za najlepszą i przez nią kierowany jest ruch. IGRP umożliwia dzielenie ruchu między trasy posiadające wartość metryki z określonego przedziału[14]. Pozwala to równolegle wykorzystywać kilka tras, co w efekcie zwiększa efektywność przesyłu w porównaniu do kierowania ruchu tylko jedną trasą[8]. W celu wyznaczenia przedziału, w którym muszą zawarte być metryki tras wykorzystywanych do równoległego przesyłania pakietów, używa się zmiennej V (variance) określającej odchylenie od minimalnej wartości metryki M (czyli najlepszej trasy). Wszystkie trasy, których metryka jest mniejsza od M*V są wykorzystywane do przesyłania pakietów do odbiorcy[8]. Ruch jest dzielony odwrotnie proporcjonalnie do wartości metryki, czyli inaczej mówiąc większa ilość będzie kierowana trasą o mniejszej metryce. Przykładowo, jeśli wykorzystywane do przesyłu będą trasy o metrykach równych 1 i 3, to na pierwszą z tych tras będzie kierowane 3 razy więcej pakietów niż na trasę o metryce 3. Wartość V ustalana jest przez administratora routera.

Routery firmy Cisco obsługują rozdzielanie ruchu maksymalnie pomiędzy cztery trasy prowadzące do tego samego obiektu docelowego [14]. Domyślnie zmienna V przyjmuje wartość 1, co oznacza, że rozdzielanie ruchu odbywać się będzie pomiędzy trasy o równych wartościach metryki.

Zastosowanie współczynników we wzorze na metrykę pozwala na określenie, które ze składowych pojawiających się we wzorze mają większe znaczenie[25]. Dzięki temu, możemy podejmować decyzje o wyborze trasy na podstawie usług przesyłających dane. I tak na przykład dla ruchu interaktywnego większa waga przypisana będzie opóźnieniu, a przy usłudze FTP przepustowości.

Routery Cisco umożliwiają zmianę standardowych wartości wag przypisanych do różnych usług[14].

8.4 Tablica routingu

Decyzje o trasowaniu IGRP podejmuje na podstawie informacji zebranych w tablicy routingu. Początkową wartość tablicy ustala administrator poprzez dodanie do niej wpisów informujących o bezpośrednio podłączonych sieciach i ich parametrach. Dalsze uzupełnianie tablicy odbywa się poprzez wymianę danych z sąsiadami, podobnie jak to ma miejsce w przypadku protokołu RIP[8]. Częstość, z jaką następuje rozsyłanie danych do sąsiadów uzależniona jest od wartości zmiennej „broadcast time” (patrz też rozdział 6.4) wyrażonej w sekundach. Domyślnie ma ona wartość 90, co oznacza, że co 90 sekund poprzez wszystkie podłączone interfejsy rozgłaszane są wiadomości aktualizacyjne[5].

Bardziej złożona metryka wymaga przesyłania i przechowywania większej ilości danych. Każdy obiekt docelowy znany routerowi posiada miejsce w tablicy składające się z następujących elementów:

- adresu obiektu docelowego,
- listy tras prowadzących do tego obiektu,
- licznika wstrzymania (holddown timer) – gdy obiekt docelowy staje się nieosiągalny (lub metryka wzrasta na tyle, że powoduje to uruchomienie mechanizmu zatruwania ścieżek) obiekt przechodzi w stan wstrzymania (holddown), co powoduje przypisaniu temu licznikowi wartości: czas aktualny + czas wstrzymania. Czas wstrzymania określa jak długo stan wstrzymania powinien trwać. Podczas tego stanu, żadna ścieżka prowadząca do obiektu nie będzie akceptowana,

- czasu ostatniej aktualizacji – jest równy czasowi ostatnio aktualizowanej trasy do obiektu,
- flagi.

Z kolei o każdej trasie wchodzącej w skład listy tras prowadzących do obiektu przechowywane są następujące informacje:

- adres następnego skoku,
- interfejs,
- przepustowość,
- opóźnienie,
- pewność,
- obciążenie,
- metryka - wyliczona na podstawie czterech powyższych parametrów, opisująca całkowity koszt dotarcia pakietu do odbiorcy,
- metryka zdalna – wyliczona tylko i wyłącznie na podstawie wartości odczytanych z ostatniego komunikatu aktualizacyjnego,
- MTU - maksymalny rozmiar pola danych ramki przesyłanej w danym segmencie
- liczba skoków do odbiorcy,
- adres uzyskania informacji o trasie (w praktyce jest to adres następnego skoku)
- czas ostatniej aktualizacji.

Jak większość protokołów routingu IGRP posiada koncepcje tras domyślnych. Jest to bardziej praktyczne, niż umieszczanie w tablicy routingu wszystkich możliwych tras. W przeciwieństwie do RIP, który traktował trasę domyślną jako trasę do rzeczywistej sieci, IGRP oznacza rzeczywiste sieci flagą, jako kandydujące do bycia domyślnymi. Spośród nich wybierana jest ta, której metryka trasy jest najmniejsza i ta trasa staje się trasą domyślną. Takie rozwiązanie jest dużo bardziej elastyczne[8]. Pozwala, bowiem przykładowo na zmianę trasy domyślnej w razie utraty łączności z dotychczasowym routerem, do którego kierowane były pakiety.

Należy w tym miejscu zaznaczyć, że informacje o obiektach docelowych i ścieżkach utrzymywane są w tablicy oddzielnie dla każdego rodzaju usług wspomaganego przez router.

8.5 Liczniki i stałe czasowe

IGRP stosuje 4 stałe czasowe do kontroli rozsyłania i eliminacji tras. Stałe te mogą być ustalane przez administratora. Oto te stałe:

- czas rozgłaszania (broadcast time) – wyrażony w sekundach czas, co jaki rozgłaszane są wiadomości aktualizacyjne. Domyślnie 90 sekund[8].
- czas ważności (invalid time) – jeśli trasa nie została uaktualniona przez ten czas (wyrażony w sekundach) to uznawana jest za niedostępną i usuwana jest z tablicy. Ze względu ma możliwe zagubienie pakietów z uaktualnieniami powinna mieć ta stała wartość kilku cykli rozgłoszeń. Domyślnie jest to wielokrotność 3 cykli, czyli 270 sekund[8].
- czas wstrzymania (hold time)- jest to czas wyrażony w sekundach, określający jak długo żadna ścieżka prowadząca do obiektu nie będzie akceptowana (patrz też rozdział 8.6.4 i 8.4). Stała ta powinna być ustawiona na wielokrotność kilku cykli rozgłoszeń. Domyślnie są to 3 cykle rozgłoszeń plus 10 sekund[8].
- czas usuwania (flush time) – jeśli żadne uaktualnienie nie nadeszło w tym czasie dla danego obiektu docelowego, to wpis w tablicy dotyczący tego obiektu jest usuwany z

tablicy. Należy zwrócić uwagę na różnicę miedzy czasem ważności a czasem usuwania. Czas ważności odnosi się, bowiem do tras, natomiast czas usuwania do obiektów docelowych. Jeśli dla danego obiektu nie ma żadnych tras to obiekt oznaczany jest jako nieosiągalny, ale mimo tego pozostaje w tablicy routingu. Musi on pozostać w celu wymuszenia „wstrzymywania" tej trasy (patrz też rozdział 8.7). Wpis odnoszący się do nieosiągalnego obiektu usuwany jest dopiero po upływie „czasu usuwania" (flush time). Ze względu na zadania, jakie spełnia ten czas powinien on być większy od sumy czasów ważności (invalid time) i wstrzymania (hold time)[8].

8.6 Przetwarzanie informacji trasowania

Protokół IGRP jest protokołem uniwersalnym, który można stosować w odniesieniu do różnych protokołów sieciowych, takich jak TCP/IP czy DECnet. Ponieważ każdy protokół posiada swój system adresacji i format pakietów, dlatego kod implementujący algorytmy protokołu będzie się różnił dla poszczególnych protokołów sieciowych[8]. Podstawowym elementem odróżniającym protokóły od siebie(np. TCP/IP od DECnet) będzie postać pakietu aktualizacyjnego, gdyż uzależniona ona będzie od specyfikacji protokołu sieciowego. W naszych rozważaniach skupimy się na protokole TCP/IP.

Prześledźmy los pakietu, który dotarł do routera poprzez interfejs I. Po odebraniu przesyłki określany jest rodzaj protokołu użytego w tym pakiecie. Jeśli protokół nie zostanie rozpoznany, czyli nie jest znany ten rodzaj protokołu routerowi, to pakiet zostanie odrzucony[8]. Dalsze czynności w dużym stopniu uzależnione są od rodzaju użytego protokołu. Specyfikacja protokołu opisuje szczegóły dotyczące zawartości pakietu, zawiera również procedury służące do określenia przeznaczenia pakietu, porównania adresu przeznaczenia z adresem samego routera, sprawdzenia czy pakiet jest pakietem rozgłoszeniowym oraz czy adres przeznaczenia jest częścią danej sieci.

Wykorzystując procedury specyficzne dla protokołu następuje dalsze przetwarzania pakietu. Sprawdzane jest czy adres przeznaczenia jest którymś z adresów przypisanych do routera lub adresem rozgoszeniowym, jeśli tak, to pakiet przetwarzany jest w sposób określony przez protokół i zawarte w nim informacje[8].

Jeśli adres znajduje się w którejś z sieci bezpośrednio podłączonych do routera to pakiet wysyłany jest prosto do odbiorcy, używając enkapsulacji odpowiedniej do protokołu i użytej technologii sieciowej.

W tym momencie przeglądana jest tablica routingu i określane jest czy istnieje wpis określający trasę do odbiorcy. Jeśli nie, to wysyłany jest komunikat błędu stosowny do protokołu i pakiet jest odrzucany[8]. Jeśli natomiast wiadomo gdzie pakiet powinien być wysłany to tam właśnie jest on kierowany. Używana jest do tego celu odpowiednia enkapsulacja stosowna do protokołu i łącza. Jeśli można wykorzystać do przesyłania kilka tras to pakiety kierowane są na trasy w odpowiednich proporcjach. Należy zwrócić uwagę, że na tym poziomie decyzja o wyborze trasy podejmowana jest na podstawie rodzaju usługi przesyłającej pakiet.

8.7 Stabilność

Ta część pracy opisuje mechanizmy stosowane przez IGRP, które zapobiegają występowaniu błędów w trasowaniu. Błędy mogą powstać na skutek różnych zmian w topologii sieci, takich jak przerwanie łączy, wyłączenie jednego z routerów itp. Temat zabezpieczeń omawiany był już w części poświęconej protokołom wektora odległości (patrz rozdział 5.2). IGRP stosuje omówione już mechanizmy uaktualnień wymuszonych, dzielenia horyzontu i

zatruwania ścieżek. Dodatkowym mechanizmem zastosowanym w IGRP jest „wstrzymywanie uaktualnień" (holddowns), który to mechanizm związany jest bezpośrednio z mechanizmem uaktualnień wymuszonych[8].

Uaktualnienia wymuszone są całkowicie skuteczne, jeśli założymy, że rozgłoszone uaktualnienia dotrą do wszystkich stosownych odbiorców natychmiastowo[8]. Tak się oczywiście nie dzieje. Wystąpić tu mogą dwa problemy. Pierwszy, pakiet zawierający uaktualnienie został zagubiony podczas przesyłu. Drugi, uaktualnienie wymuszone dotrze z pewnym opóźnieniem. W tym przypadku, możliwe jest, że router R1, który nie otrzymał jeszcze wymuszonego uaktualnienia, rozgłosi regularne wiadomości aktualizacyjne. Gdy tak rozgłoszone wiadomości dotrą do routera R2, który otrzymał już uaktualnienie wymuszone, zostaną one przetworzone przez R2 i spowodują umieszczenie w tablicy routingu błędnych wpisów.

W celu wyeliminowania przedstawionego problemu stosuje się „wstrzymywanie uaktualnień"[8]. Polega to na tym, że gdy trasa jest usuwana z tablicy, żadna nowa trasa prowadząca do tego samego obiektu docelowego nie będzie akceptowana przez pewien określony czas. To daje pakietom z uaktualnieniami wymuszonymi czas na dotarcie do wszystkich routerów. Czas wstrzymania musi być tak dobrany, aby umożliwić „fali" wymuszonego uaktualnienia przejście przez całą sieć.

Jeśli pakiet z wiadomością aktualizacyjną nie dotrze do któregoś z routerów, nie spowoduje to uaktualnienia tablicy routingu tego routera i w konsekwencji zatrzymany zostanie łańcuch wymuszonych uaktualnień. Nie jest to jednak wielkim problemem, gdyż przy najbliższym otrzymaniu regularnego komunikatu aktualizacyjnego, zauważone zostaną zmiany w topologii i zmiany te wprowadzone do tablicy routingu spowodują wygenerowanie fali wymuszonych uaktualnień. Z tego względu, czas wstrzymania powinien być dodatkowo zwiększony o czas kilku cykli regularnych rozgłoszeń aktualizacyjnych[8].

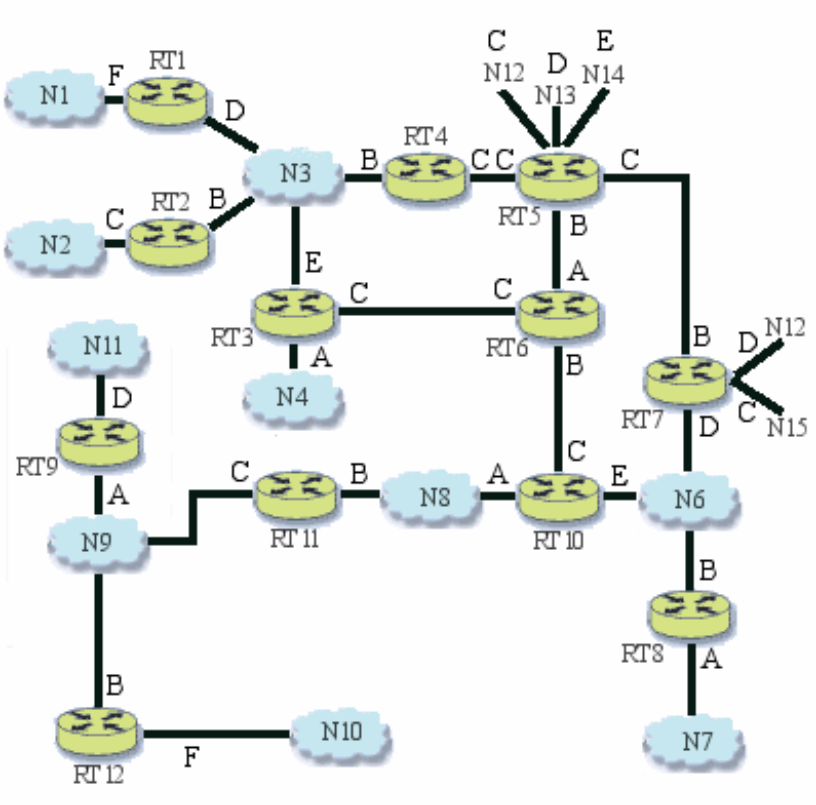
Aby skuteczniej przeciwdziałać pętlom rozległym, trasy, których metryki zwiększają się o określoną wartość są uznawane za zapętlone i oznaczane jako niedostępne. Taka sytuacja występuje, gdy nowa metryka jest większa niż iloczynu V i starej metryki, przy czym jeśli zmienna V jest równa 1 to bierze się do iloczynu wartość 1,1.

Wadą stosowania mechanizmu „wstrzymywania uaktualnień" jest to, że opóźnia on wprowadzanie do tablicy routingu nowych tras w miejscy nieużytecznych tras starych[8]. Przy domyślnych ustawieniach, może potrwać nawet kilka minut zanim router zastosuje nowe trasy po zmianie topologii sieci. Implementacja Cisco 8.2 udostępnia funkcje wyłączenia tego mechanizmu[14]. Mimo tego, z przyczyn wymienionych powyżej, rezygnacja z niego nie jest bezpieczna i może spowodować „liczenie do nieskończoności" (patrz rozdział 5.2)[8]. Wyłączenie „wstrzymywania uaktualnień" powoduje w IGRP zaostrzenie kryteriów uznawania trasy za zapętloną. Za zapętlone uznawane są wtedy trasy, dla których wzrosła liczba skoków. To oczywiście będzie powodować niekiedy usuwanie tras, które są prawidłowe. Jeśli bowiem gdzieś w sieci do trasowania użyty zostanie dodatkowy router, spowoduje to zrost liczby skoków. Niemożliwe jest jednak odróżnienie takiej sytuacji od sytuacji wystąpienia pętli, dlatego najbezpieczniej jest usunąć taką trasę z tablicy routingu, gdy tylko wzrośnie liczba skoków[8]. Jeśli trasa jest prawidłowa zostanie ona ponownie umieszczona w tablicy trasowania przy następnym otrzymaniu komunikatu aktualizacyjnego.

Ze względu na opisane wady, implementacja Cisco protokołu IGRP zapewnia możliwość wyłączenia „wstrzymywania uaktualnień" [14]. Należy jednak zwrócić uwagę na fakt, że „wstrzymywanie uaktualnień" musi być określone dla całego systemu autonomicznego. Inaczej mówiąc wszystkie routery systemu muszą posiadać taką samą politykę stosowania tego mechanizmu[14].

8.8 Przykład topologii

Prześledźmy postać tablicy routingu na routerze, na którym zaimplementowany jest protokół IGRP. Rysunek 3 przedstawia schemat topologii, w której pracują routery.



Rysunek 3) Przykład topologii dla IGRP

Opis oznaczeń wskazujących na rodzaj stosowanych łączy:

Łącze	Przepustowość[kb/s]	Opóźnienie[ms]	Metryka
A	10 000 (1 000)	1 000 (100)	1 100
B	5 000 (2 000)	1 500 (150)	2 150
C	5 000 (2 000)	2 500 (250)	2 250
D	1 000 (10 000)	5 000 (500)	10 500
E	512 (19 531)	10 000 (1 000)	20 531
F	56 (178 571)	20 000 (2 000)	180 571

Tabela 11) Oznaczenia łączy

Poniżej przedstawione zostały trasy zgromadzone w tablicach routingu wybranych routerów. W tablicach tych zawarte zostały istotne parametry tras brane pod uwagę w procesie trasowania. Dodatkowo przy każdej trasie podana została ścieżka łączy, którą będzie podążać pakiet, jeśli zostanie skierowany daną trasą. Duże litery oznaczają interfejsy wyjściowe routerów, a litery małe interfejsy wejściowe. Ścieżka ta nie jest oczywiście częścią składową tablicy routingu, a jej zamieszczenie służy jedynie lepszemu zrozumieniu działania protokołu.

Przepustowość obliczamy jako minimum przepustowości wszystkich interfejsów wejściowych i wyjściowych na trasie, natomiast opóźnienie to suma opóźnień wprowadzanych przez interfejsy wyjściowe.

Oto tablica routingu R6:

Sieć docelowa	Następny skok	Przepustowość	Opóźnienie	Metryka	Ilość skoków		Ścieżka
N1	RT5	178 571	2 500	181 071	4		AbCcBdF
N2	RT5	2 000	750	2 750	4		AbCcBbC
N3	RT5	2 000	500	2 500	3		AbCcB
N4	RT3	2 000	350	2 350	2		CcA
N6	RT5	10 000	850	10 850	3		AbCbD
N7	RT5	10 000	950	10 950	4		AbCbDbA
N8	RT10	2 000	250	2 250	2		BcA
N9	RT10	2 000	500	2 500	3		BcAbC
N10	RT10	178 571	2 500	181 071	4		BcAbCbF
N11	RT10	10 000	1 000	11 000	4		BcAbCaD
N12	RT5	2 000	350	2 350	2		AbC
N13	RT5	10 000	600	10 600	2		AbD
N14	RT5	19 531	1 100	20 631	2		AbE
N15	RT5	2 000	600	2 600	3		AbCbC

Tabela 12) Tablica routingu routera R6

Poniżej przedstawione zostały trasy zgromadzone w tablicy routingu R3, które mają znaczący wpływ na stan tablicy routingu routera R6. Znakiem # oznaczono trasy o najmniejszych wartościach metryki. Znak * przy w kolumnie następnego skoku oznacza, że router jest bezpośrednio przyłączony do danej sieci.

Sieć docelowa	Następny skok	Przepustowość	Opóźnienie	Metryka	Ilość skoków		Ścieżka
N1	RT1	178 571	3 000	181 571	2		EdF
	#RT6	178 571	2 750	181 321	5		CcAbCcBdF
N2	RT2	19 531	1 250	20 781	2		EbC
	#RT6	2 000	1 000	3 000	5		CcAbCcBbC
N3	*	19 531	1 000	20 531	1		E
	#RT6	2 000	750	2 750	4		CcAbCcB

Tabela 13) Wybrane wpisy tablicy routingu routera R3

Poniżej przedstawione zostały trasy zgromadzone w tablicy routingu R10, które mają znaczący wpływ na stan tablicy routingu routera R6

Sieć docelowa	Następny skok	Przepustowość	Opóźnienie	Metryka	Ilość skoków		Ścieżka
N6	*	19 531	1 000	20 531	1		E
	#RT6	10 000	1 100	11 100	4		CbAbCbD
N7	RT8	19 531	1 100	20 631	2		EbA
	#RT6	10 000	1 200	11 200	5		CbAbCbDbA

Tabela 14) Wybrane wpisy tablicy routingu routera R10

Przyjmijmy teraz, że połączenie między routerami RT6 i RT5 zostało przerwane na skutek awarii linii łączącej. Jedynymi trasami routera RT6 prowadzącymi do sieci N1, N2, N3, N6, N7, N12, N13, N14 i N15 były trasy prowadzące przez router RT5. W sytuacji braku

połączenia z routerem RT5 sieci te oznaczane są jako niedostępne. Brak łączności między RT5 i RT6 wpłynie również na stan tablic routingu pozostałych routerów.

W tablicy routingu RT3 znajdują się po dwie trasy prowadzące do sieci N1 i N2. Do obu tych sieci trasą preferowaną (o mniejszej metryce) była trasa prowadząca przez router RT6 i RT5. Ponieważ połączenie między RT6 i RT5 zostało zerwane, to trasa ta jest już nie aktualna. Mimo tego sieci N1 i N2 nie zostają oznaczone jako nieosiągalne, ponieważ inna trasa, prowadząca przez RT1 (do sieci N1) i przez RT2 (do sieci N2), jest znana.

Po zmianie topologii sieci część tablicy routingu RT3 będzie postaci:

Sieć docelowa	Następny skok	Przepustowość	Opóźnienie	Metryka	Ilość skoków		Ścieżka
N1	#RT1	178 571	3 000	181 571	2		EdF
N2	RT2	19 531	1 250	20 781	2		EbC
N3	*	19 531	1 000	20 531	1		E
N6	RT4	19 531	2 000	21 531	4		EbCcCbD
	#RT6	19 531	1 400	20 931	3		CcBcE
N12	#RT4	19 531	1 500	21 031	3		EbCcC
	RT6	19 531	1 800	21 331	5		CcBcEdBcC
N13	#RT4	19 531	1 750	21 281	3		EbCcD
	RT6	19 531	2 050	21 581	5		CcBcEdBcD
N14	#RT4	19 531	2 250	21 781	3		EbCcE
	RT6	19 531	2 250	21 781	5		CcBcEdBcE
N15	RT4	19 531	1 750	21 281	4		EbCcCbC
	#RT6	19 531	1 650	21 181	4		CeBcEdC

Tabela 15) Wybrane wpisy tablicy routingu routera R3 po zmianach w topologii sieci

Oto część tablicy RT10 po zmianach, które zaszły w sieci:

Sieć docelowa	Następny skok	Przepustowość	Opóźnienie	Metryka	Ilość skoków		Ścieżka
N6	*	19 531	1 000	20 531	1		E
N15	RT7	19 531	1 250	20 781	2		EdC

Tabela 16) Wybrane wpisy tablicy routingu routera R10 po zmianach w topologii sieci

Router RT6 po utracie połączenia z RT5 ma oznaczone trasy do niektórych sieci jako niedostępne. Informacje o dostępności tych sieci otrzyma router RT6 od swoich pozostałych sąsiadów, czyli od routerów RT3 i RT10. Wcześniej router RT6 nie otrzymywał takich wiadomości, ponieważ dla routerów RT3 i RT10 router RT6 był oryginalnym nauczycielem rozważanych, niedostępnych już tras. Aby to lepiej zrozumieć, posłużmy się konkretną sytuacją zaistniałą w naszej topologii. Jak pokazuje tabela 14 router RT10 posiadał przed zmianą dwa wpisy odnoszące się do sieci N6. Jedna trasa prowadziła przez RT6, a druga trasa była bezpośrednim podłączeniem do sieci N6, ale o bardzo słabej przepustowości łącza. W takiej sytuacji trasą bardziej preferowaną była trasa przez RT6. Ponieważ ogłaszane są tylko najlepsze trasy, to RT10 ogłaszał sieć N6 z trasą prowadzącą przez RT6. Ogłoszenie to jednak nie było wysyłane do routera RT6 ze względu na mechanizm dzielenia horyzontu. Oznacza to, że do tej pory router RT6 nie wiedział o bezpośrednim połączeniu routera RT10 z siecią N6.

Po zmianach w sieci, router RT10 (tabela 16) uznał, że w takiej topologii najlepszym rozwiązaniem trasowania będzie kierowanie pakietów bezpośrednio do sieci N6 poprzez wolne łącze E. I taką też trasę będzie ogłaszał do swych sąsiadów. Stąd też RT6 dowie się o tym połączeniu i odzyska możliwość trasowania pakietów do sieci N6.

Podobnie jak w omówionym powyżej przykładzie router RT6 odzyska dostęp do pozostałych sieci N2, N7, N12, N13, N14 i N15.

Postać tablicy routera RT6 zmieni się. Oto wpisy tablicy, które ulegną zmianie:

Sieć docelowa	Następny skok	Przepustowość	Opóźnienie	Metryka	Ilość skoków		Ścieżka
N1	RT3	178 571	3 500	182 071	3		CcEdF
N2	RT3	19 531	1 500	21 031	3		CcEbC
N3	RT3	19 531	1 250	20 781	2		CcE
N4	RT3	2 000	350	2 350	2		CcA
N6	RT10	19 531	1 150	20 681	2		BcE
N7	RT10	19 531	1 250	20 781	3		BcEbA
N12	RT3	19 531	1 750	21 281	4		CcEbCcC
	#RT10	19 531	1 550	21 081	4		BcEdBcC
N13	RT3	19 531	2 000	21 531	4		CcEbCcD
	#RT10	19 531	1 800	21 331	4		BcEdBcD
N14	RT3	19 531	2 500	22 031	4		CcEbCcE
	#RT10	19 531	2 300	21 831	4		BcEdBcE
N15	RT10	19 531	1 400	20 931	3		BcEdC

Tabela 17) Wybrane wpisy tablicy routingu routera R6 po zmianach w topologii sieci

9 Protokół EIGRP

9.1 Wstęp

EIGRP (Extended Interior-Gateway Routing Protocol) jest wzbogaconą wersją protokołu IGRP. EIGRP pozostał protokołem wektora odległości i opiera się na tych samych, co IGRP informacjach opisujących odległość[10]. Protokół EIGRP jest protokołem hybrydowym, stanowiącym połączenie protokołów wektora odległości i protokołów stanu łącza. Protokół routingu hybrydowego stosuje wektor odległości dla wyznaczenia najlepszych ścieżek do punktu docelowego. Jednakże różni się od większości protokołów wektora odległości, ponieważ uaktualnienia bazy danych są wyzwalane przez zmiany topologii sieci[4].

W stosunku do IGRP znacząco zmieniły się mechanizmy odpowiedzialne za skuteczne trasowanie pakietów, co jest konsekwencją chociażby szybkiej zbieżności gwarantowanej przez protokół[10]. Do wyznaczania tras stosowany jest algorytm DUAL (Distibuted Update Algorithm) pozwalający wyeliminować ewentualne pętle mogące powstać w czasie trasowania pakietów[12]. W przypadku zmian zachodzących w topologii algorytm odpowiedzialny jest za synchronizacje routerów, na których decyzje ma wpływ zmiana w sieci. Na tych routerach algorytm odpowiednio modyfikuje tablice routingu w celu dostosowania do bieżącego obrazu połączeń.

Protokół EIGRP składa się z czterech podstawowych komponentów[14]:

- procesu wyszukiwania/odszukiwania sąsiednich routerów
- niezawodnego protokołu transportowego
- procesu realizującego algorytm DUAL
- modułów wspierających protokół

Proces wyszukiwania/odszukiwania sąsiednich routerów jest procesem używanym do dynamicznego odszukiwania sąsiadów w bezpośrednio przyłączonych sieciach[10]. Dzięki temu procesowi możliwe jest również wykrycie braku połączenia z sąsiednim routerem. Wykrywanie i sprawdzanie poprawnego działania sąsiednich routerów osiągane jest przez okresowe wysyłanie pakietów Hello o małej wielkości. Wysyłanie pakietów realizowane jest przez każdy aktywny router. Odbiorca takiego pakietu dowiaduje się w ten sposób, że posiada połączenie z routerem-nadawcą i router-nadawca pracuje poprawnie.

Do przekazywania pakietów informacyjnych przekazywanych między routerami EIGRP stosowany jest niezawodny protokół transportowy[10]. Jego zastosowanie gwarantuje poprawne dotarcie pakietów EIGRP do sąsiadujących routerów z zachowaniem odpowiedniej kolejności. Dla niektórych pakietów stosowanych przez EIGRP nie jest konieczne użycie niezawodnego przesyłania danych, dlatego w celu poprawy skuteczności działania protokołu niezawodny protokół transportowy jest stosowany tylko tam gdzie jest to niezbędne. Przykładem może być rozgłoszeniowa sieć Ethernet, gdzie nie jest konieczne wysyłanie pakietów Hello oddzielnie do wszystkich sąsiadujących routerów. Zamiast tego EIGRP rozgłasza pojedynczy pakiet Hello z zaznaczeniem, że pakiet ten nie wymaga potwierdzenia. Inne pakiety, takie jak pakiety uaktualnień, wymagają potwierdzeń, co jest zaznaczone w przesyłanym pakiecie[10].

Proces realizujący algorytm DUAL jest odpowiedzialny za analizę tras otrzymanych od sąsiadujących routerów i podejmowanie odpowiednich czynności w zależności od zaistniałych sytuacji. Najbardziej korzystne i pozbawione pętli trasy wyznaczane są na podstawie wartości metryk[3]. Wybrane przez DUAL trasy umieszczane są w tablicy routingu. Każda

trasa znajdująca się w tablicy routingu posiada wykonalny następny skok. Oznacza to, że następny skok na ścieżce prowadzącej do obiektu docelowego jest skokiem do routera, który ogłosił trasę i skok ten nie spowoduje powstania pętli. W przypadku, gdy nie ma wykonalnego następnego skoku związanego z daną trasą, a któryś z routerów ogłosił trasę jako dostępną, to musi zostać przeprowadzony proces ponownej analizy tras według algorytmu. Czas potrzebny na ponowne wyznaczenie wykonalnego następnego skoku ma wpływ na czas zbieżności protokołu EIGRP[10]. Mimo tego, że przeprowadzane obliczenia nie obciążają bardzo procesora i czas obliczeń nie jest duży, to algorytm unika zbyt często powtarzanych obliczeń. Gdy topologia sieci się zmienia, DUAL szuka wykonalnego następnego skoku i używa najlepszego skoku spośród obecnie dostępnych, bez każdorazowego dokonywania obliczeń od początku.

Moduły wspierające zapewniają prawidłową komunikację w warstwie sieci[10]. Przykładowo moduł IP-EIGRP jest odpowiedzialny za wysyłanie i odbieranie pakietów EIGRP kapsułkowanych wewnątrz pakietów IP. Zadaniem tego modułu jest również analiza składniowa pakietu EIGRP i informowania algorytmu DUAL o nadejściu nowych informacji. Z drugiej strony IP-EIGRP zwraca się do DUAL, by ten dokonał odpowiednich decyzji trasowania i umieścił te decyzje w tablicy routingu IP. Do zadań IP-EIGRP należy również dystrybucja tras nauczonych poprzez inne protokoły routingu.

9.2 Format przesyłanych wiadomości

Protokół EIGRP używa pięciu rodzajów pakietów[10]:

- pakiety Hello – jak już wspomniano wcześniej, służą one do odnajdowania i utrzymywania połączeń między sąsiednimi routerami. Rozsyłane są poprzez multiemisję i nie wymagają potwierdzeń. Pakiety Hello nie zawierające żadnych danych, używane są również jako pakiety potwierdzeń. Pakiety potwierdzeń są kierowane do konkretnych routerów (na konkretny adres) i niosą ze sobą niezerowy numer potwierdzenia.
- pakiety uaktualnień – używane do przesyłania informacji o dostępności obiektów docelowych. Gdy nowy router sąsiadujący zostanie wykryty przesyłane są do niego pakiety uaktualnień, by mógł w ten sposób zbudować tabele topologii. W takiej sytuacji pakiety są kierowane do pojedynczego routera. Pakiety te mogą być również wysyłane przy użyciu multiemisji, na przykład w sytuacji, gdy zmieni się koszt którejś z tras. Pakiety uaktualnień są zawsze przesyłane przy użyciu niezawodnego protokołu transportowego.
- pakiety zapytań i pakiety odpowiedzi – te rodzaje pakietów wykorzystywane są, gdy któryś z obiektów przechodzi w stan aktywny. Zapytania są zawsze rozsyłane poprzez multiemisję. Wyjątkiem jest sytuacja, w której router odpowiada na zapytaniem wysyłając zapytanie – w takim wypadku pakiet jest kierowany na konkretny adres. Pakiety odpowiedzi są generowane po otrzymaniu pakietu zapytania, w celu poinformowania, że router posiada trasę w tablicy routingu prowadzącą do danego obiektu docelowego. Odpowiedzi są kierowane do routera, który nadesłał zapytanie. Zarówno pakiety zapytań jak i odpowiedzi przesyłane są przy użyciu niezawodnego protokołu transportowego.
- pakiety żądań – wykorzystywane są do uzyskiwania specyficznych informacji od jednego lub więcej sąsiednich routerów. Mogą być przesyłane poprzez multiemisję lub do konkretnego routera. Niezawodny protokół transportowy nie jest tu używany.

9.3 Metryka

Metryka używana przez protokół EIGRP jest identyczna z metryką używaną w IGRP [12]. Oznacza to, że metryka obliczana jest według tych samych wzorów i te same parametry brane są pod uwagę. Metoda obliczania metryki opisana jest w rozdziale 8.3.

9.4 Tabele przyległości, tabele topologii, tablice trasowania

Każdy router przechowuje informacje o sąsiednich routerach w tabeli przyległości[10]. Router wysyłając pakiet Hello oznajmia swoją obecność w topologii. Gdy otrzyma pakiet Hello od innego routera to routery uznają się za sąsiadujące. Nowo odnaleziony router wpisywany jest do tabeli przyległości. W przesyłanych między routerami pakietach Hello określony jest czas przetrzymania HoldTime. Czas ten określa ile maksymalnie może upłynąć czasu między otrzymaniem poprawnych pakietów Hello od danego sąsiada. Inaczej mówiąc, przez czas określony parametrem HoldTime sąsiedni router uznawany jest za aktywny i działający. Po upływie tego czasu (jeśli nie nadejdzie pakiet Hello) licznik związany z danym routerem przyległym uruchamia odpowiedni fragment algorytmu DUAL informując w ten sposób o zmianie w topologii. Czas wstrzymania jest domyślnie brany jako trzykrotność czasu wysyłania pakietów Hello[10].

Wpis dotyczący danego sąsiada zawiera dodatkowo informacje służące niezawodnemu przekazywaniu pakietów między routerami[14]. Numer sekwencyjny pakietów jest wykorzystywany do sprawdzania poprawności kolejności odbieranych pakietów oraz do potwierdzeń odbioru pakietów. Lista transmisji służy przechowywaniu i kolejkowaniu wysyłanych pakietów. Pakiet utrzymywany jest na liście do momentu otrzymania potwierdzenia poprawnego odbioru danego pakietu. Z każdym sąsiadującym routerem związany jest licznik służący oszacowaniu najkorzystniejszego odstępu czasu między retransmitowanymi pakietami.

Tabela topologii danego routera zawiera wszystkie obiekty docelowe, które ogłoszone zostały jako obiekty osiągalne przez routery sąsiednie[10]. Każdy wpis opisujący obiekt docelowy zawiera adres obiektu i listę routerów, które ogłosiły tenże obiekt. Dla każdego routera na liście przechowywana jest metryka ogłoszona przez ten router. Router ogłasza taką wartość metryki, jaka związana jest z daną trasą znajdującą się w jego tablicy routingu. Oznacza to, że trasa ogłaszana przez sąsiedni router, jest trasą używaną przez ten router do trasowania pakietów w kierunku obiektu docelowego. Uzupełnianiem tabeli topologii zajmują się moduły wspierające protokoły. Spośród tras znajdujących się w tabeli topologii algorytm DUAL wybiera najlepsze trasy i umieszcza je w tablicy routingu. Trasowanie pakietów opiera się o informacje zawarte właśnie w tablicy routingu. Z trasami zawartymi w tablicy routingu związana jest metryka, będąca sumą metryki otrzymanej od routera sąsiedniego i kosztu dotarcia do tego routera. Dodatkowo przechowywana jest również metryka zgłoszona, czyli metryka, która otrzymana została od routera ogłaszającego trasę. Pod względem przechowywanych w tablicy routingu parametrów opisujących trasy EIGRP nie różni się od swojego poprzednika – protokołu IGRP (patrz rozdział 8.4)[10].

EIGRP rozróżnia dwa rodzaje tras: wewnętrzne i zewnętrzne [12]. Trasy wewnętrzne, to trasy prowadzące do obiektów wewnątrz lokalnego systemu autonomicznego. Zewnętrzne trasy natomiast, to te trasy, które otrzymane zostały od innych protokołów routingu, lub umieszczone zostały w tablicy routingu statycznie. Trasy zewnętrzne opisane są dodatkowymi parametrami:

- identyfikator routera EIGRP ogłaszającego trasę,

- numer systemu autonomicznego, w którym znajduje się obiekt docelowy,
- znacznik administratora,
- identyfikator zewnętrznego protokołu,
- metryka uzyskana od zewnętrznego protokołu,
- flagi.

9.5 Przetwarzanie informacji trasowania

Decyzja o trasowaniu pakietów podejmowana jest w oparciu o informacje zgromadzone w tablicy trasowania[10]. Tablicę trasowania wypełnia algorytm DUAL, wybierając trasy spośród wszystkich tras znajdujących się w tabeli topologii. Spośród tych tras używana jest ta, której wartość metryki jest najmniejsza. Metrykę tą nazywamy metryką wykonawczą.

O umieszczeniu trasy w tablicy trasowania decyduje wynik porównania metryki wykonawczej z metryką zgłoszoną trasy. Każda trasa, której metryka zgłoszona jest mniejsza od metryki wykonawczej umieszczana jest w tablicy routingu. Trasa spełniająca powyższy warunek uznawana jest jako posiadająca tzw. wykonalny następny skok.

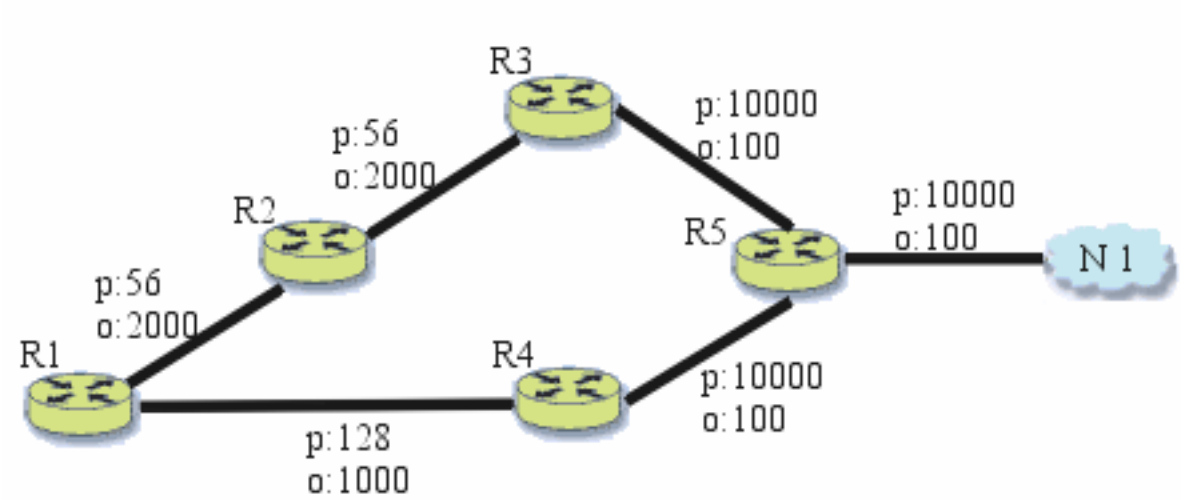
W przypadku, gdy aktualnie używana trasa stanie się niedostępna, wybierana jest inna trasa znajdująca się w tablicy routingu i prowadząca do tego samego obiektu docelowego. W wyborze tym nie uczestniczy algorytm DUAL. Odwołanie do algorytmu następuje dopiero wtedy, gdy brak jest w tablicy routingu trasy prowadzącej do danego obiektu docelowego[12]. W takiej sytuacji obiekt przechodzi w stan aktywny i rozpoczyna się proces obliczeniowy mający na celu znalezienie wykonalnego następnego skoku. Pierwszą czynnością tego procesu jest rozesłanie zapytań do wszystkich sąsiadów o dany obiekt docelowy. Sąsiedni router otrzymując takie zapytanie może odesłać odpowiedź, jeśli posiada trasę prowadzącą do obiektu. Gdy router sąsiedni nie posiada wpisu odnoszącego się do szukanego obiektu, to obiekt ten zostaje wpisany do tablicy routingu i przechodzi w stan aktywny. Tym samym router, który otrzymał zapytanie kieruje pytania do swych sąsiadów. Proces ten jest rekurencyjny. Jeśli okaże się, że żaden z routerów sąsiednich nie posiada trasy prowadzącej do obiektu, to zapytanie jest odsyłane do routera pytającego. Gdy wszystkie odpowiedzi sąsiednich routerów dotrą do routera nadającego zapytanie rozpoczyna się właściwa cześć obliczeń algorytmu DUAL, której celem jest wytypowania wykonalnych następnych skoków. Jeśli w ramach tych obliczeń umieszczone zostaną trasy prowadzące do rozważanego obiektu do tablicy routingu, to obiekt ze stanu aktywnego przechodzi do stanu pasywnego. Obiekt pozostaje w stanie pasywnym tak długo, jak długo znajdują się w tablicy routingu trasy prowadzące do tego obiektu[10].

Jeśli utracone zostanie połączenie z sąsiednim routerem, to wszystkie obiekty, do których jedynymi trasami w tablicy routingu są trasy prowadzą przez ten router, przejdą w stan aktywny.

W przypadku, gdy wszystkie zapytania o dany obiekt docelowy wrócą z odpowiedziami niedostępności tego obiektu, to obiekt przechodzi w stan pasywny.

Protokół EIGRP nie jest protokołem hierarchicznym[12]. Oznacza to, że nie jest przewidziane w protokole dzielenie systemu autonomicznego na obszary. Jednakże, ze względu na to, że EIGRP jest protokołem bezklasowym, a algorytm trasowania pozwala na filtracje i agregacje tras, pozwala to uruchomić kilka procesów routingu na pojedynczym routerze i tak skonfigurować protokoły EIGRP, że osiągnięty zostanie hierarchiczny podział obszaru[12].

Prześledźmy zasady działania mechanizmu trasowania na przykładzie poniższej topologii (rysunek 4).



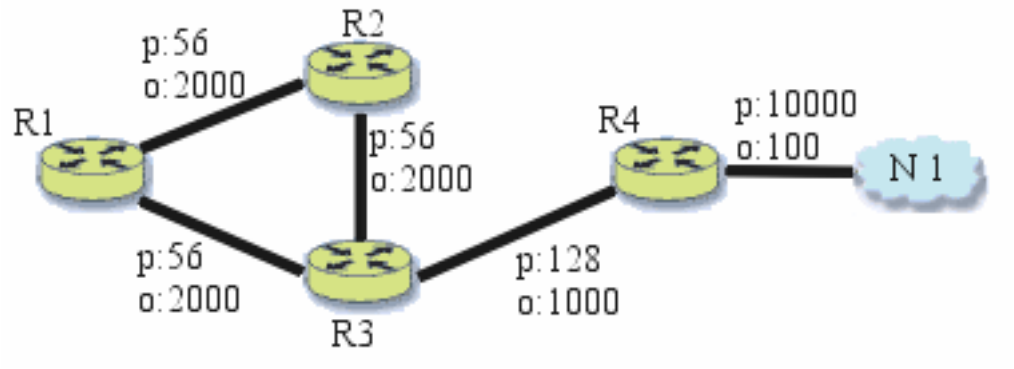
Rysunek 4) Zasady działania mechanizmu trasowania w EIGRP (na rysunku literą „p" oznaczona jest przepustowość łącza, natomiast litera „o" odnosi się do opóźnienia).

Załóżmy, że śledzimy działanie routera R1, który trasuje pakiety mające dotrzeć do sieci N1. Na rysunku pokazane są parametry poszczególnych łączy. Jak widać na rysunku, do sieci N1 prowadzą dwie trasy: przez router R2 z metryką równą 46789376 (metryki obliczone na podstawie wzoru 2 z IGRP, rozdział 8.3) i przez router R4 z metryką 20307200. Router R1 wybiera mniejszą spośród tych dwóch metryk i metryka ta staje się metryką wykonawczą. Oznacza to, że trasa prowadząca przez router R4 będzie trasą wykorzystywaną do przesyłania pakietów, których odbiorca znajduje się w sieci N1. Oczywiście router R4 kwalifikuje się jako wykonalny następny skok. Spójrzmy, czy trasa prowadząca przez R2 również posiada wykonalny następny skok. Metryka zgłoszona przez router R2 jest równa 46277376 i jest większa od wartości metryki wykonawczej. Oznacza to, że router R2 nie spełnia kryteriów bycia wykonalnym następnym skokiem na trasie. Tak więc w tablicy routingu umieszczona zostanie tylko trasa prowadząca przez router R4.

Przyjmijmy teraz, że połączenie routera R1 z routerem R4 zostało zerwane. Trasa do sieci N1 przechodząca przez router R4 była jedyną trasą w tablicy routingu, dlatego musi zostać uruchomiony algorytm DUAL w celu znalezienia nowego sposobu dotarcia do sieci N1. Router R1 rozsyła zapytania do wszystkich sąsiadujących routerów (w naszym przypadku jedynie do routera R2) o sieć N1. Router R2 posiada połączenie z siecią N1, dlatego odpowiada pozytywnie na zapytanie. Ponieważ router nie posiada w tym momencie lepszej trasy, niż ta prowadząca przez router R2, dlatego trasa ta wprowadzona zostaje do tablicy routingu. Metryka związana z tą trasą staje się metryką wykonawczą. W takim układzie router R2 kwalifikuje się jako wykonalny następny skok, bo wartość metryki zgłoszonej przez router R2 (46277376) jest mniejsza od wartości metryki wykonawczej równej 46789376.

9.6 Stabilność

Wartość metryki wykonawczej i metryk zgłoszonych jest wykorzystywana w procesie detekcji pętli w topologii. W celu przedstawienia zasad wykrywania pętli posłużymy się ponownie przykładem. Rysunek 5 przedstawia rozważany układ routerów.



Rysunek 5) Wykrywanie pętli w topologii w EIGRP

Załóżmy dodatkowo, że wyłączone jest dzielenie horyzontu na routerach. Przy tym założeniu router R3 posiada trzy ścieżki prowadzące do sieci N1: przez R4, przez R2 (ścieżka ma postać: R2, R1, R3, R4) i przez R1 (ścieżka ma postać: R1, R2, R3, R4). Jeśli R3 akceptowałby wszystkie te trasy, skutkowałoby to powstaniem pętli, jednak stosując się do zasad wyboru wykonalnego następnego skoku można tego uniknąć. Całkowity koszt dotarcia do sieci N1 przez router R4 wynosi 20281600, przez router R2 koszt ten jest równy 47019776, a przez R1 – 47019776. Najmniejszą spośród tych metryk jest metryka prowadząca przez router R4, dlatego ta trasa instalowana jest w tablicy routingu, a metryka wykonawcza ma od tego momentu wartość 20281600. Trasy przez routery R1 i R2 są rozważane do umieszczenia w tablicy routingu. Jednak ze względu na to, że wartość metryk zgłoszonych związanych z tymi trasami jest większa od metryki wykonawczej routery R1 i R2 nie mogą zostać uznane jako wykonalne następne skoki i rozważane trasy nie mogą zostać umieszczone w tablicy routingu[12].

Przyjmijmy teraz, że połączenie między routerami R3 i R4 zostało przerwane. Router R3 nie posiada w tablicy routingu innych pozycji opisujących trasy do sieci N1, dlatego rozsyła do swoich sąsiadów zapytania. Router R2 otrzymuje zapytanie i widząc, że zapytanie to pochodzi od routera, który jest wykonalnym następnym skokiem na trasie do sieci N1, szuka innych, alternatywnych tras. W tabeli topologii routera R2 znajduje się wpis mówiący, że router R1 posiada możliwość dotarcia do sieci N1, ale zgłoszona z tą trasą metryka nie jest mniejsza od metryki wykonawczej (prowadzącej przez router R3), dlatego router R2 oznacza trasę jako niedostępną, wprowadza obiekt (sieć N1) w stan aktywny i rozsyła do sąsiadów (w tym przypadku, tylko do routera R1) zapytanie o sieć N1.

Router R3 wysyła zapytanie o N1 również do R1. Router R1 przeprowadza czynności analogiczne do czynności przeprowadzanych przez R2, skutkiem czego także R1 oznacza trasę do sieci N1 jako niedostępną. Routery R1 i R2 odsyłają do routera R3 zapytanie, co oznacza, że również R3 jest zmuszony do uznania trasy za niedostępną. W tym momencie routery posiadają już aktualny stan topologii sieci, a trasy prowadzące do N1 przechodzą w stan pasywny.

9.7 Kompatybilność EIGRP z IGRP

Protokół EIGRP jest zgodny ze swoim poprzednikiem – protokołem IGRP[10]. Nie jest konieczne rozgraniczanie działania tych dwóch protokołów. Umożliwia to czerpanie korzyści z jednoczesnego wykorzystania obu protokołów, współpracujących ze sobą. EIGRP może być uruchamiane w strategicznych miejscach sieci, bez przerywania pracy protokołu IGRP. Trasy nauczone od routerów IGRP są wykorzystywane przez IGRP i odwrotnie. Metryki tras są obliczane na podstawie tych samych zależności i wzorów, dlatego mogą również być bezpośrednio porównywane, bez dodatkowych przekształceń[12]. Trasy pochodzące od IGRP są traktowane przez EIGRP jako trasy zewnętrzne, co powoduje, że trasy te mogą być traktowane w szczególny sposób, poprzez przypisanie im odpowiednich znaczników. Domyślnie na routerach Cisco trasy IGRP mają pierwszeństwo przed trasami nauczonymi od EIGRP. Może to jednak zostać zmienione w konfiguracji routera, bez konieczności ponownego uruchamiania protokołu trasowania[1].

Również konfiguracja routera posługującego się EIGRP jest taka sama jak w przypadku IGRP (z dodatkiem oczywiście dodatkowych możliwości). Pliki konfiguracyjne wykorzystywane przy IGRP mogą być użyte w procesie konfiguracji routera z obsługą EIGRP.

Routery Cisco zapewniają automatyczną wymianę informacji między IGRP i EIGRP[14]. Implementacja umożliwia wyłączenie tego automatycznego procesu, bądź to na wszystkich interfejsach, bądź też, tylko na wybranych interfejsach.

9.8 Przykład topologii

Do zilustrowania działania protokołu EIGRP posłużymy się tą samą topologią, co przy omawianiu IGRP (rysunek 3, rozdział 8.8). Oznaczenia wprowadzone do opisu tras są identyczne jak w przykładach topologii z poprzednich rozdziałów. Obliczana metryka tras nie zmienia się i jest identyczna jak w IGRP. Ponieważ EIGRP używa innego mechanizmu klasyfikacji tras, dlatego wpłynie to na postać tablicy routingu każdego z routerów w sieci. Dokładniej mówiąc, niektóre trasy mogą się nie znaleźć w tablicy routingu, jeśli ich metryka (zgłoszona przez router ogłaszający trasę) będzie większa od metryki wykonawczej (patrz dokładny opis rozdział 9.5 i 9.6)[12]. Poniżej przedstawione są parametry tras brane pod uwagę podczas umieszczania ich w tablicach trasowania poszczególnych routerów.

Router R6

Sieć Docel.	Nast. skok	Przepus.	Opóź.	Metryka	Liczba skoków	Przepus. zdalna	Opóź. zdalne.	Metryka zdalna	TR	Ścieżka
N1	RT5	178 571	2 500	181 071	4	178 571	2 400	180 971	T	AbCcBdF
N2	RT5	2 000	750	2 750	4	2 000	650	2 650	T	AbCcBbC
N3	RT5	2 000	500	2 500	3	2 000	400	2 400	T	AbCcB
N4	RT3	2 000	350	2 350	2	1 000	100	1 100	T	CcA
N6	RT5	10 000	850	10 850	3	10 000	750	10 750	T	AbCbD
N7	RT5	10 000	950	10 950	4	10 000	850	10 850	T	AbCbDbA
N8	RT10	2 000	250	2 250	2	1 000	100	1 100	T	BcA
N9	RT10	2 000	500	2 500	3	2 000	350	2 350	T	BcAbC
N10	RT10	178 571	2 500	181 071	4	178 571	2 350	180 921	T	BcAbCbF
N11	RT10	10 000	1 000	11 000	4	10 000	850	10 850	T	BcAbCaD
N12	RT5	2 000	350	2 350	2	2 000	250	2 250	T	AbC
N13	RT5	10 000	600	10 600	2	10 000	500	10 500	T	AbD
N14	RT5	19 531	1 100	20 631	2	19 531	1 000	20 531	T	AbE
N15	RT5	2 000	600	2 600	3	2 000	500	2 500	T	AbCbC

Tabela 18) Tablica routingu routera R6

Router RT3

Sieć Docel.	Nast. skok	Przepus.	Opóź.	Metryka	Ilość skoków	Przepus. zdalna	Opóź. zdalne.	Metryka zdalna	TR	Ścieżka
N1	RT1	178 571	3 000	181 571	2	178 571	2 000	180 571	T	EdF
	#RT6	178 571	2 750	#181321	5	178 571	2 500	181 071	T	CcAbCcBdF
N2	RT2	19 531	1 250	20 781	2	2 000	250	2 250	T	EbC
	#RT6	2 000	1 000	#3 000	5	2 000	750	2 750	T	CcAbCcBbC
N3	*	19 531	1 000	20 531	1	-	-	-	T	E
	#RT6	2 000	750	#2 750	4	2 000	500	2 500	T	CcAbCcB

Tabela 19) Wybrane wpisy tablicy routingu routera R3

Router RT10

Sieć Docel.	Nast. skok	Przepus.	Opóź.	Metryka	Ilość skoków	Przepus. zdalna	Opóź. zdalne.	Metryka zdalna	TR	Ścieżka
N6	*	19 531	1 000	20 531	1	-	-	-	T	E
	#RT6	10 000	1 100	#11 100	4	10 000	850	10 850	T	CbAbCbD
N7	RT8	19 531	1 100	20 631	2	1 000	100	1 100	T	EbA
	#RT6	10 000	1 200	#11 200	5	10 000	950	10 950	T	CbAbCbDbA

Tabela 20) Wybrane wpisy tablicy routingu routera R10

Metryka zdalna jest metryką otrzymaną od routera ogłaszającego trasę. Pole TR to pole mówiące czy dana trasa znalazła się w tablicy routingu routera (T), czy też nie (N). O tym czy trasa znajdzie się w tablicy routingu decyduje wynik porównania metryki wykonawczej z metryką zdalną (patrz też rozdział 9.6). Metryki wykonawcze oznaczone są znakiem #. Jak widać z powyższych tabel, wszystkie trasy ogłaszane zostały umieszczone w tablicach routingu. Po zmianie topologii sieci zobaczymy, że nie zawsze tak musi być.

Podobnie jak poprzednio załóżmy, że nastąpiła awaria łącza między RT5 i RT6. Oto informacje o trasach rozważane przez routery po zmianie.

Router RT6

Sieć Docel.	Nast. skok	Przepus.	Opóź.	Metryka	Ilość skoków	Przepus. zdalna	Opóź. zdalne.	Metryka Zdalna	TR	Ścieżka
N1	RT3	178 571	3 500	182 071	3	178 571	3 000	181 571	T	CcEdF
N2	RT3	19 531	1 500	21 031	3	19 531	1 250	20 781	T	CcEbC
N3	RT3	19 531	1 250	20 781	2	19 531	1 000	20 531	T	CcE
N4	RT3	2 000	350	2 350	2	1 000	100	1 100	T	CcA
N6	RT10	19 531	1 150	20 681	2	19 531	1 000	20 531	T	BcE
N7	RT10	19 531	1 250	20 781	3	19 531	1 100	20 631	T	BcEbA
N12	RT3	19 531	1 750	21 281	4	19 531	1 500	21 031	T	CcEbCcC
	#RT10	19 531	1 550	21 081	4	19 531	1 400	20 931	T	BcEdBcC
N13	RT3	19 531	2 000	21 531	4	19 531	1 750	21 281	T	CcEbCcD
	#RT10	19 531	1 800	#21 331	4	19 531	1 650	21 281	T	BcEdBcD
N14	RT3	19 531	2 500	22 031	4	19 531	2 250	21 781	T	CcEbCcE
	#RT10	19 531	2 300	#21 831	4	19 531	2 150	21 681	T	BcEdBcE
N15	RT10	19 531	1 400	20 931	3	19 531	1 250	20 781	T	BcEdC

Tabela 21) Wybrane wpisy tablicy routingu routera R6 po zmianach w topologii sieci

Router RT3

Sieć Docel.	Nast. skok	Przepus.	Opóź.	Metryka	Ilość skoków	Przepus. zdalna	Opóź. zdalne.	Metryka Zdalna	TR	Ścieżka
N1	#RT1	178 571	3 000	181 571	2	178 571	2 000	180 571	T	EdF
N2	RT2	19 531	1 250	20 781	2	2 000	250	2 250	T	EbC
N3	*	19 531	1 000	20 531	1	-	-	-	T	E
N6	RT4	19 531	2 000	21 531	4	10 000	1 000	11 000	T	EbCcCbD
	#RT6	19 531	1 400	#20 931	3	19 531	1 150	20 681	T	CcBcE
N12	**#RT4**	**19 531**	**1 500**	**#21 031**	**3**	**2 000**	**500**	**2 500**	**T**	**EbCcC**
	RT6	**19 531**	**1 800**	**21 331**	**5**	**19 531**	**1 550**	**21 081**	**N**	**CcBcEdBcC**
N13	#RT4	19 531	1 750	#21 281	3	10 000	750	10 750	T	EbCcD
	RT6	19 531	2 050	21 581	5	19 531	1 800	21 331	T	CcBcEdBcD
N14	#RT4	19 531	2 250	21 781	3	19 531	1 250	20 781	T	EbCcE
	RT6	19 531	2 250	21 781	5	19 531	2 300	21 831	T	CcBcEdBcE
N15	RT4	19 531	1 750	21 281	4	2 000	750	2 750	T	EbCcCbC
	#RT6	19 531	1 650	21 181	4	19 531	1 400	20 931	T	CeBcEdC

Tabela 22) Wybrane wpisy tablicy routingu routera R3 po zmianach w topologii sieci

Trasa ogłoszona przez RT6 prowadząca do sieci N12 nie została umieszczona w tablicy routingu routera RT3, ponieważ metryka zdalna (21 081) ma większą wartość od metryki wykonawczej (21 031).

Router RT10

Sieć Docel.	Nast. skok	Przepus.	Opóź.	Metryka	Ilość skoków	Przepus. zdalna	Opóź. zdalne.	Metryka Zdalna	TR	Ścieżka
N6	*	19 531	1 000	20 531	1	-	-	-	T	E
N15	RT7	19 531	1 250	20 781	2	2 000	250	2 250	T	EdC

Tabela 23) Wybrane wpisy tablicy routingu routera R10 po zmianach w topologii sieci

10 Protokół OSPF wersja 2

10.1 Wstęp

Protokół routingu dynamicznego OSPF (Open Shortest Path First) jest protokołem stanu łącza (link-state protocol) [31]. Routery, na których zaimplementowany jest protokół należący do rodziny protokołów stanu łącza, utrzymują bazę danych, w której przechowują informacje na temat aktualnej topologii systemu autonomicznego, czyli inaczej mówiąc informacje o stanie połączeń. Baza ta ma taką samą postać i zawartość u wszystkich routerów uczestniczących w procesie wymiany informacji. Każda część bazy, odnosząca się do konkretnego routera, przechowuje informacje dotyczące lokalnych połączeń dla tegoż routera. Pojedynczy router powiadamia inne routery o stanie swoich lokalnych połączeń rozsyłając wiadomości w postaci pakietów LSA (link-state advertisement), które to pakiety wędrując niezmienione od routera do routera, przemierzają cały system autonomiczny. Wszystkie routery używają dokładnie tego samego algorytmu, który na podstawie bazy danych buduje drzewo najkrótszych ścieżek, przy czym router, na którym działa algorytm, jest korzeniem drzewa[25]. Na podstawie tego drzewa można wyznaczyć trasy do wszystkich obiektów docelowych w danym systemie autonomicznym.

10.2 Format przesyłanych wiadomości

Protokół routingu OSPF działa bezpośrednio w oparciu o IP, używając protokołu 89 IP[31]. OSPF nie zapewnia mechanizmów fragmentacji. Jeśli takowa jest wymagana używana jest fragmentacja IP. Protokół OSPF został jednak tak zaprojektowany, że duże pakiety protokołu mogą być dzielona na kilka mniejszych pakietów, i taka praktyka jest rekomendowana[31]. Pakiety protokołu routingu powinny być przesyłane z polem TOS protokołu IP równym zero, aby pakiety protokołu routingu były przesyłane w pierwszej kolejności.

OSPF wyróżnia pięć rodzajów pakietów:

- pakiety Hello – używane do odszukiwania sąsiednich routerów i utrzymywania połączeń między nimi,
- pakiety opisu bazy danych (database description) – używane do przesyłania między wymieniającymi, przyległymi routerami skróconych informacji o bazach stanów łączy,
- pakiety żądania stanu łącza (link state request) - używane do przesyłania między przyległymi routerami informacji o stanu łącza,
- pakiety aktualizacji stanu łącza (link state update) – przesyłane są za ich pomocą ogłoszenia o nowym stanie łączy. Przesyłane są one tylko do sąsiadujących routerów, ale mogą zawierać w sobie pakiety LSA od różnych routerów.

Z wyjątkiem pakietów Hello, wszystkie inne pakiety są przesyłane tylko między routerami przyległymi, co oznacza, że pakiety przebywają tylko jeden skok (z wyjątkiem tych przesyłanych poprzez połączenia wirtualne).

10.2.1 Nagłówek pakietów protokołu OSPF

Gdy router zamierza wysłać pakiet, pierwszą czynnością jest wypełnienie standardowego nagłówka pakietu OSPF [31]. W skład nagłówka wchodzą pola:

- wersja (1 oktet) – określa numer wersji protokołu,
- typ (1 oktet) – wskazuje na typ pakietu OSPF, taki jak pakiet aktualizacji, czy pakiet Hello,
- rozmiar (2 oktety) – jest to mierzony w bajtach rozmiar całego pakietu OSPF, łącznie z nagłówkiem,
- identyfikator routera (4 oktety) – określa router, który jest nadawcą pakietu,
- identyfikator obszaru (4 oktety) – opisuje obszar, dla którego przeznaczony jest pakiet,
- suma kontrolna (2 oktety) – suma kontrolna całego pakietu wraz z nagłówkiem,
- rodzaj uwierzytelniania (2 oktety) – określa rodzaj procedury, jaka powinna być zastosowana w celu uwierzytelnienia pakietu,
- dane uwierzytelniające (8 oktetów)

10.2.2 Nagłówek pakietów LSA

Nagłówek każdego pakietu LSA zawiera:

- pole typu(LS type),
- identyfikator części systemu (link state ID), do której odnosi się pakiet oraz
- pole wskazujące na router (advertising router), od którego pochodzi pakiet.

Te trzy pola jednoznacznie określają pakiet LSA[31]. Jednocześnie, w systemie może być obecnych kilka wersji tego samego pakietu LSA. W celu określenia tych, które są najbardziej aktualne sprawdzane są kolejne pola umieszczone w nagłówku pakietu:

- numer sekwencyjny (LS sequence number) – jest to 32-bitowa liczba całkowita używana do detekcji pakietów starszych i zdublowanych. Im większy numer, tym bardziej aktualne są informacje zgromadzone w pakiecie.
- czas życia **(**LS age) – jest to 16-bitowa liczba bez znaku określająca w sekundach czas życia pakietu. W chwili utworzenia pole ma wartość 0. Router jest odpowiedzialny za zwiększanie tego licznika przy każdorazowym przesyłaniu pakietu do innego routera, jak również podczas przechowywania pakietu w bazie. Wartość czasu życia nigdy nie jest zwiększana, gdy osiągnie wartość maksymalnego czasu życia. W przypadku, gdy czas życia pakietu LSA osiągnie po dopuszczalną wartość maksymalną pakiet jest rozsyłany po sieci a następnie usuwany.
- suma kontrolna **(**LS checksum**)**– jest to suma kontrolna całego pakietu LSA. Do sumy kontrolnej nie jest brany czas życia, aby podczas zwiększania czasu życia nie było konieczne wyliczanie od początku sumy kontrolnej. Pole to nie może przyjmować wartości zerowej. Inaczej mówiąc umieszczanie sumy kontrolnej w nagłówku pakietu jest obowiązkowe.

Dodatkowo nagłówek zawiera jeszcze pola:

- opcje (options) – pole to określa zakres usług, jakie jest w stanie obsłużyć obszar, do którego odnosi się pakiet
- długość (length) – określa w bajtach rozmiar całego pakietu LSA

Czas Życia	Opcje	Typ
Identyfikator		
Router ogłaszający		
Numer Sekwencyjny		
Suma Kontrolna	Długość	

Tabela 24) Format nagłówka pakietów LSA

10.2.3 Typy pakietów LSA

Wyróżniamy następujące typy pakietów LSA:

1 - pakiety LSA routerów (router-LSAs)

2 - pakiety LSA sieci (network-LSAs) – zawierają listę routerów przyłączonych do sieci. Rozsyłane są w obrębie całego pojedynczego obszaru.

3 , 4 - pakiety LSA skrócone (summary-LSAs) – rozsyłane są przez brzegowe routery obszarów i rozsyłane w całym obszarze związanym z danym pakietem LSA. Każdy pakiet tego typu opisuje trasy do obiektów na zewnątrz obszaru, ale wewnątrz systemu autonomicznego. Typ 3 opisuje trasy do sieci, a typ 4 trasy do routerów brzegowych systemu autonomicznego.

5 - pakiety LSA zewnętrzne (AS-external-LSAs) – rozsyłane są przez brzegowe routery systemu autonomicznego i rozsyłane w całym systemie autonomicznym. Każdy pakiet opisuje obiekty w innym systemie autonomicznym. W tych pakietach mogą być również rozsyłane trasy domyślne dla systemu autonomicznego.

Pakiet LSA jest dodawany do bazy danych routera, jeśli a) otrzymany został w procesie rozsyłania pakietu w całej sieci lub b) został stworzony i rozesłany przez tenże router. Pakiet LSA jest usuwany z bazy, jeśli a) został nadpisany przez nowszą wersję pakietu otrzymają od sąsiadującego routera lub b) router sam utworzył i rozesłał nowszą wersję pakietu lub c) czas życia pakietu LSA osiągnął maksymalną dopuszczalną wartość.

10.2.4 Pakiety LSA routerów

Każdy router w sieci jest źródłem tego typu pakietów[31]. Opisują one stan interfejsów przyłączeniowych routera do obszaru i są rozsyłane w całym pojedynczym obszarze. Postać pakietów przedstawia tabela 25.

Nagłówek LSA					
0	Bit V	Bit E	Bit B	0	Ilość opisanych przyłączeń
. . .					
ID przyłącza					
Opis przyłącza					
Typ		Ilość TOS			Metryka
. . .					

Tabela 25) Format pakietów LSA

Opis pól:

Bit V - jeśli ustawiony, oznacza to, że router jest końcowym routerem połączenia wirtualnego,
Bit E - jeśli ustawiony, oznacza to, że router jest brzegowym routerem systemu autonomicznego,
Bit B - jeśli ustawiony, router jest brzegowym routerem obszaru,
Typ - opisuje typ przyłącza. Wyróżniamy cztery typy: 1 – połączenie typu punkt-punkt z innym routerem, 2 – przyłącze do sieci tranzytowej, 3 – przyłącze do sieci końcowej, 4 – połączenie wirtualne,
Ilość TOS – określa ilość różnego rodzaju metryk związanych z interfejsem, w zależności od rodzaju usługi, do której odnosi się metryka,
Metryka - koszt użycia tego przyłącza.

10.2.5 Pakiety LSA sieci

Pakiety sieci są pakietami typu drugiego pakietów LSA. Pakiety te są generowane przez router desygnowany w celu opisu sieci rozgłoszeniowej lub sieci NBMA. Pakiety te zawierają listę wszystkich przyłączonych do sieci routerów, wraz z routerem desygnowanym. W polu identyfikatora (w nagłówku pakietu LSA) umieszczany jest adres IP routera desygnowanego[31].

Ponieważ wartość metryki opisującej trasę od sieci do wszystkich routerów ma wartość zerową, dlatego pole metryki w pakiecie LSA sieci jest zbędne.

Nagłówek LSA
Maska sieci
. . .
Przyłączony Router
. . .

Tabela 26) Format pakietów LSA sieci

Opis pól:

Maska sieci - maska przypisana do sieci opisywanej przez ten pakiet
Przyłączony Router - identyfikator routera przyłączonego do sieci. Na liście znajdą się tylko te routery, które są routerami przyległymi (patrz rozdział 10.7) do routera desygnowanego. Identyfikator routera desygnowanego również figuruje na tejże liście. Ilość routerów, jaka znajduje się na tej liście jest jednoznacznie określana poprzez analizę pola długości pakietu LSA.

10.2.6 Skrócone pakiety LSA

Pakiety LSA tego typu są generowane przez brzegowe routery obszarów i opisują obiekty docelowe w obszarach. Skrócone pakiety LSA to pakiety o numerze 3 i 4. Typ o numerze 3 jest używany, gdy obiektem docelowym jest sieć. W tym przypadku pole identyfikatora zawiera numer IP sieci. Gdy obiektem docelowym jest router brzegowy systemu autonomicznego używany jest typ o numerze 4 pakietu LSA. Identyfikator w polu nagłówka pakietu LSA informuje o numerze identyfikacyjnym routera.

Nagłówek LSA	
Maska Sieci	
0	Metryka
. . .	
TOS	Metryka TOS
. . .	

Tabela 27) Format pakietów LSA skróconych

Opis pól:

Maska sieci - dla typu 3 jest maska sieci docelowej. Dla typu 4 wartość tego pola musi przyjmować wartość zerową.
TOS - rodzaj usługi IP, do której odnosi się metryka.

10.2.7 Zewnętrzne pakiety LSA

Źródłem tych pakietów są routery brzegowe systemu autonomicznego, które za ich pomocą dostarczają informacji od obiektach znajdujących się poza systemem [31]. Pole identyfikatora zawiera numer opisywanej sieci. Pakiety te służą również do opisu tras domyślnych. W tym przypadku pole identyfikatora ma wartość 0.0.0.0, a maska sieci przyjmuje postać 0.0.0.0.

Maska Sieci		
Bit E	0	Metryka
Następny Adres		
Znacznik Zewnętrznej Trasy		
Bit E	TOS	Metryka TOS
Następny Adres		
Znacznik Zewnętrznej Trasy		
. . .		

Tabela 28) Format pakietów LSA zewnętrznych

Opis pól:

Maska Sieci - maska przypisana do ogłaszanego obiektu.
Bit E - określa rodzaj metryki użytej do opisu trasy. Jeśli ustawiony, oznacza to, że używamy metryki drugiego rodzaju. Jeśli bit ma wartość zero to metryka jest rodzaju pierwszego.
Metryka - koszt związany z ogłaszaną trasą (interpretacja zależy od ustawienia bitu E).
Następny Adres - na ten adres kierowane będą pakiety, których odbiorca jest opisany tym pakietem LSA. Jeśli pakiet opisuje trasę domyślną, to pakiety będą kierowane do routera, który ogłosił taką trasę domyślną.
Znacznik Zewnętrznej Trasy - pole nie używane przez OSPF. Może być wykorzystane do komunikowania się routerów położonych w różnych systemach autonomicznych.

TOS - rodzaj usługi IP, do której odnosi się metryka, następny adres i znacznik zewnętrznej trasy.

Jak wspomniano, pakiety zewnętrzne służą przesyłaniu informacji na temat tras domyślnych. Generowaniem informacji o trasach domyślnych zajmują się brzegowe routery systemu autonomicznego[14]. W implementacji Cisco protokołu, router, który generuje trasy domyślne staje się automatycznie routerem brzegowym systemu autonomicznego. Jednakże, domyślnie, router brzegowy systemu autonomicznego nie generuje tras domyślnych.

10.3 Metryka

OSPF rozróżnia dwa rodzaje metryk opisujących trasy zewnętrzne[31]. Pierwszy rodzaj metryki wyraża się w tych samych jednostkach, co koszt tras przypisanych do interfejsów routera. Rodzaj drugi metryki jest rozważany jako większy od metryki każdej ścieżki w obrębie systemu autonomicznego. Ten rodzaj metryki stosuje założenie, że głównym kosztem trasowania jest trasowanie między systemami autonomicznymi i eliminuje w ten sposób potrzebę każdorazowego obliczania kosztów tras na zewnątrz systemu uwzględniając wewnętrzne koszty tras w systemie autonomicznym. Jako przykład stosowania metryk załóżmy, że routery RT7 i RT5 ogłaszają informacje o swoich zewnętrznych trasach przy użyciu metryki pierwszego rodzaju. Każdy router w sieci, chcąc obliczyć koszt trasy na zewnątrz systemu dokona tego dodając koszt dotarcia do routera brzegowego (w tym przypadku RT5 lub RT7) i koszt zewnętrznej trasy ogłaszanej przez tenże router. Jeśli dwa routery brzegowe ogłaszają tą samą trasę zewnętrzną, to wybierana jest trasa o mniejszym koszcie[31]. Na przykładowym schemacie routery RT5 i RT7 ogłaszają zewnętrzne trasy do sieci docelowej N12. Przyjmując, że pakiet wychodzi od routera RT6 preferowaną trasą jest trasa przez router RT7, gdyż jej koszt wynosi 10 (8+2), natomiast trasa przez router RT5 ma wartość metryki równą 14 (6+8). Poniższa tabela pokazuje, jakie wpisy zostaną uwzględnione w tablicy trasowania routera RT6 w trakcie badania tras zewnętrznych:

Obiekt docelowy	Następny skok	Odległość
N12	RT10	10
N13	RT5	14
N14	RT5	14
N15	RT10	17

Tabela 29) Wpisy uwzględnione przez router RT6

W przypadku drugiego rodzaju metryki, brany jest pod uwagę tylko koszt trasy ogłaszany przez routery brzegowe. Oznacza to, że w wyliczaniu kosztu trasy zewnętrznej nie jest brany pod uwagę koszt dotarcia do routera brzegowego. Przyjmijmy, że routery RT5 i RT7 ogłaszają trasy zewnętrzne z metryką drugiego rodzaju. W tym przypadku cały ruch pakietów, które mają dotrzeć do sieci N12, będzie kierowany poprzez router RT7, bo $2 < 8$. W przypadku, gdy do sieci docelowej prowadzi więcej niż jedna trasa, to o wyborze trasy decyduje koszt dotarcia do routera brzegowego[31].

W systemie autonomicznym równocześnie mogą być używane oba rodzaje metryk. W przypadku znajomości wartości obu metryk dla danej trasy, decyzja o trasowaniu podejmowana jest na podstawie metryki rodzaju pierwszego.

10.4 Tablica routingu

Tablica routingu zawiera informacje potrzebne do przesłania pakietu IP najlepszą dostępną drogą. Każdy wpis w tablicy opisuje zbiór ścieżek do pojedynczego obiektu docelowego. W procesie trasowania pakietu IP wpis tablicy dostarcza najlepszą trasę do obiektu docelowego, jaka aktualnie istnieje, co sprowadza się do wyznaczenia następnego skoku na trasie. OSPF udostępnia także mechanizm ścieżek domyślnych [31].

W każdym routerze istnieje pojedyncza tablica routingu. Tablica składa się z wpisów opisujących obiekty docelowe. Pierwsza część pół wpisu opisuje obiekt docelowy i składa się z pól:

- typ obiektu docelowego (destination type)– jest to albo „sieć" albo „router". W procesie trasowania pakietów IP używane są tylko wpisy odnoszące się do sieci. Sieć jest w tym przypadku adresem IP, na który mogą być kierowane pakiety. Może to być adres sieci, podsieci lub pojedynczego hosta. Wpisy dotyczące routerów wykorzystywane są wyłącznie jako pośrednie kroki w procesie budowy tablicy trasowania. Wpisy te dotyczą routerów brzegowych obszarów i routerów brzegowych systemu autonomicznego. Wpisy dla routerów brzegowych obszarów są wykorzystywane w trakcie wyznaczania tras między obszarami, a routery brzegowe systemu autonomicznego w przypadku wyznaczania tras zewnętrznych, prowadzących poza system autonomiczny.
- identyfikator obiektu docelowego (destination ID)– zależy od typu obiektu. Dla sieci jest to związany z siecią adres IP, a dla routera jest to identyfikator routera w systemie OSPF.
- maska adresu (address mask) – jest to maska związana z adresem sieci. Dla routerów nie jest wykorzystywane to pole.
- możliwości (optional capabilities) – pole wykorzystywane dla opisu routerów. Określa możliwości i usługi oferowane przez docelowy router.
- obszar (area) – wskazuje na obszar, od którego pochodzi informacja o łączu. Pole to nie jest używane dla tras zewnętrznych.

Pozostała część wpisu tablicy routingu opisuje zbiór ścieżek prowadzących do obiektu docelowego. Oznacza to, że każda ścieżka zawarta we wpisie posiada te same parametry. Parametry te są następujące:

- typ ścieżki (path-type) – określa typ ścieżki prowadzącej do obiektu docelowego. Wyróżniamy cztery takie typy: wewnątrzobszarowy, międzyobszarowy, zewnętrzny rodzaju pierwszego i zewnętrzny rodzaju drugiego. Ścieżka wewnątrzobszarowa oznacza, że obiekt docelowy znajduje się w jednym z obszarów bezpośrednio połączonych z routerem. Ścieżki międzyobszarowe wskazują, że trasa prowadzi do innego obszaru OSPF. Są wyznaczane na podstawie analizy skróconych pakietów LSA. Typ zewnętrzny określa, że ścieżka prowadzi poza system autonomiczny. Ścieżka wyznaczana jest na podstawie otrzymanych zewnętrznych pakietach LSA.
- koszt (cost) – koszt stanu łącza ścieżki do obiektu, liczony jako suma kosztów wszystkich odcinków na trasie. Dla wszystkich ścieżek, oprócz ścieżek zewnętrznych drugiego rodzaju, jest to całkowity koszt trasy. Dla ścieżek zewnętrznych drugiego rodzaju jest to koszt trasy wewnątrz obszaru autonomicznego.
- koszt typu drugiego (type 2 cost) – używany tylko w przypadku tras zewnętrznych drugiego rodzaju i określa koszt części ścieżki leżącej poza systemem autonomicznym. Koszt ten jest ogłaszany przez brzegowe routery systemu autonomicznego i jest bardziej znaczącą częścią całkowitego kosztu trasy.

- źródło informacji (link state origin) – określa pakiet LSA (routera lub sieci), od którego pochodzi informacja o obiekcie docelowym. Pole używane tylko dla ścieżek wewnątrzobszarowych.

Jeśli istnieje kilka ścieżek o tych samych parametrach typu ścieżki i kosztu prowadzących do tego samego obiektu, to umieszczane są one w obrębie tego samego wpisu, a rozróżniane są na podstawie:

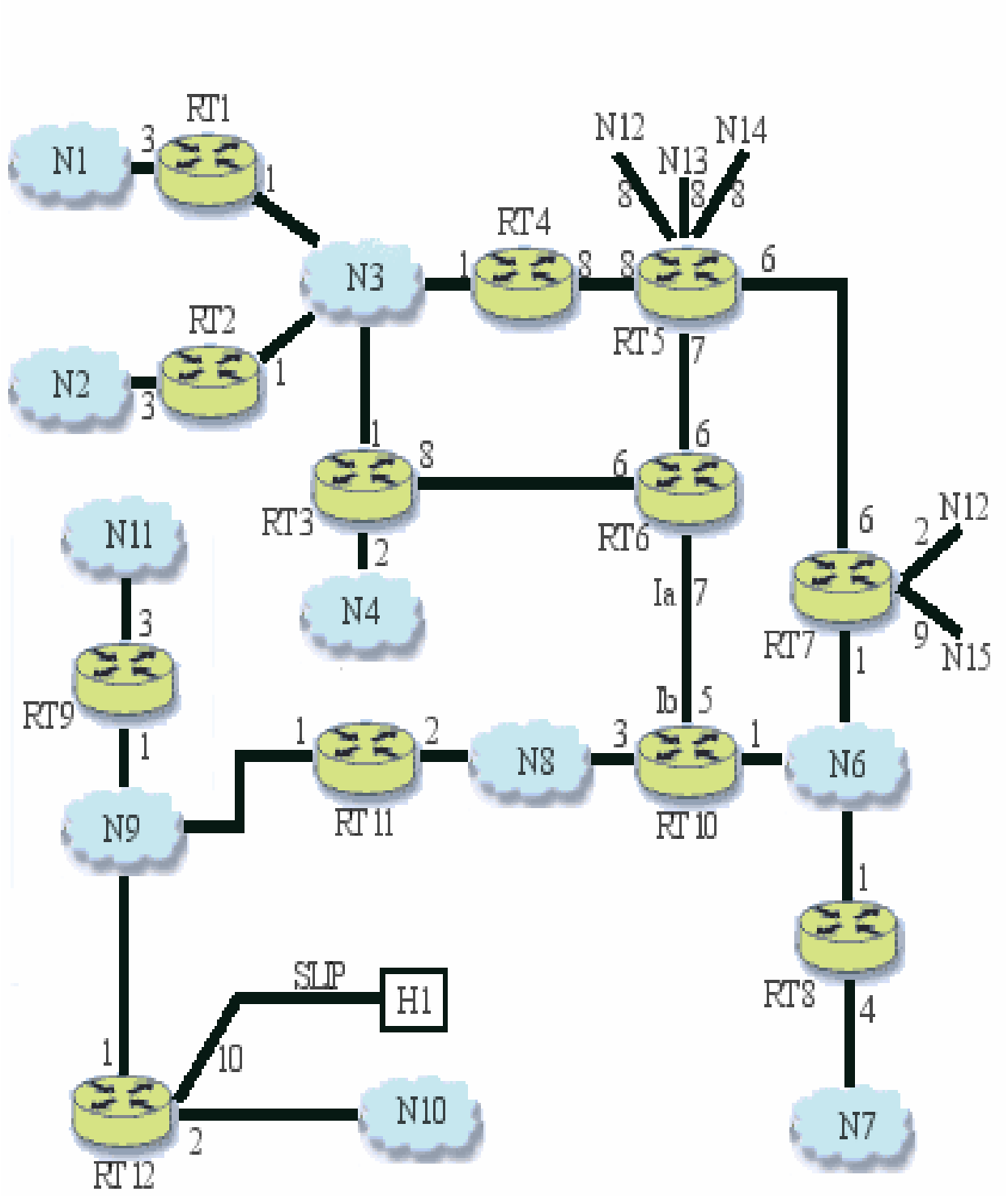
- następnego skoku (next hop) – jest to interfejs, przez który powinien zostać wysłany pakiet. W sieciach rozgłoszeniowych i NBMA następny skok określa również adres IP następnego routera (jeśli taki jest) na ścieżce do obiektu docelowego.
- routera ogłaszającego (advertising router) – używane tylko dla ścieżek międzyobszarowych i zewnętrznych. Określa identyfikator routera, od którego otrzymano skrócony pakiet LSA lub zewnętrzny pakiet LSA opisujący trasę.

10.5 Baza danych połączeń i przetwarzanie informacji trasowania

Baza danych zawierająca informacje o stanie połączeń w systemie autonomicznym służy do budowy grafu skierowanego w skład, którego wchodzą sieci i routery[1]. Krawędź grafu łącząca ze sobą dwa routery odnosi się do sytuacji fizycznego połączenia w sieci typu punkt-punkt. Natomiast krawędź łącząca router z siecią służy opisaniu sytuacji, w której router posiada interfejs do danej sieci.

W celu lepszego zobrazowania budowy, przetwarzania i postaci bazy danych posłużymy się przykładową topologią sieci. Rysunek 6 przedstawia schemat rozważanej sieci[31].

Prostokąt oznaczony jako H1 oznacza pojedynczy host podłączony bezpośrednio poprzez SLIP do routera RT12. Z tego względu router RT12 jest odpowiedzialny za rozgłaszanie informacji o trasie prowadzącej do H1. Linie łączące routery odpowiadają fizycznym połączeniom punkt-punkt. W naszej przykładowej sieci, jedynym połączeniem, w którym przypisane są adresy do poszczególnych interfejsów, jest połączenie między RT6 i RT10. Z kolei routery RT5 i RT7 posiadają połączenie z innym systemem autonomicznym, wykorzystując do tego celu protokół routingu BGP (patrz też rozdział 13). Na schemacie widoczne są sieci, których nauczyły się te routery od BGP. Z każdą trasą związany jest pewien koszt, który przypisany jest wyjściowemu interfejsowi każdego routera. Im mniejszy koszt danej trasy, tym większe prawdopodobieństwo, że tam właśnie będzie kierowany ruch pakietów[31]. Koszt przypisany jest także do tras pochodzących z zewnątrz, czyli przykładowo do tras nauczonych od BGP.



Rysunek 6) Przykładowa topologia sieci dla OSPF

Tabela 30 opisuje przedstawianą topologie systemu autonomicznego. Routery i sieci umieszczone w nagłówkach kolumn interpretujemy jako źródła pakietów, a routery i sieci umieszczone w nagłówkach wierszy jako adresatów, do których kierowane są pakiety.

	RT1	RT2	RT3	RT4	RT5	RT6	RT7	RT8	RT9	RT10	RT11	RT12	N3	N6	N8	N9
RT1													0			
RT2													0			
RT3						6							0			
RT4					8								0			
RT5				8		6	6									
RT6			8		7					5						
RT7					6									0		
RT8														0		
RT9																0
RT10						7								0	0	
RT11															0	0
RT12																0
N1	3															
N2		3														
N3	1	1	1	1												
N4			2													
N5																
N6							1	1		1						
N7								4								
N8										3	2					
N9									1		1	1				
N10												2				
N11									3							
N12					8		2									
N13					8											
N14					8											
N15							9									
H1												10				

Tabela 30) Opis topologii systemu autonomicznego przez protokół OSPF

Każda sieć i każdy router posiada swój wiersz i kolumnę. Jeśli przecięcie kolumny A i wiersza B oznaczone jest liczbą X to oznacza to, że w grafie wierzchołek A łączy się z wierzchołkiem B i koszt tego połączenia wynosi X.

Baza stanu połączeń jest budowana na podstawie pakietów LSA generowanych przez routery[31]. Otoczenie pojedynczego routera lub sieci jest reprezentowane przez oddzielny pakiet LSA. Oto przykładowa, graficzna postać takich pakietów generowanych przez router RT12 i sieć N9:

	RT12	N9	N10	H1
RT12				
N9	1			
N10	2			
H1	10			

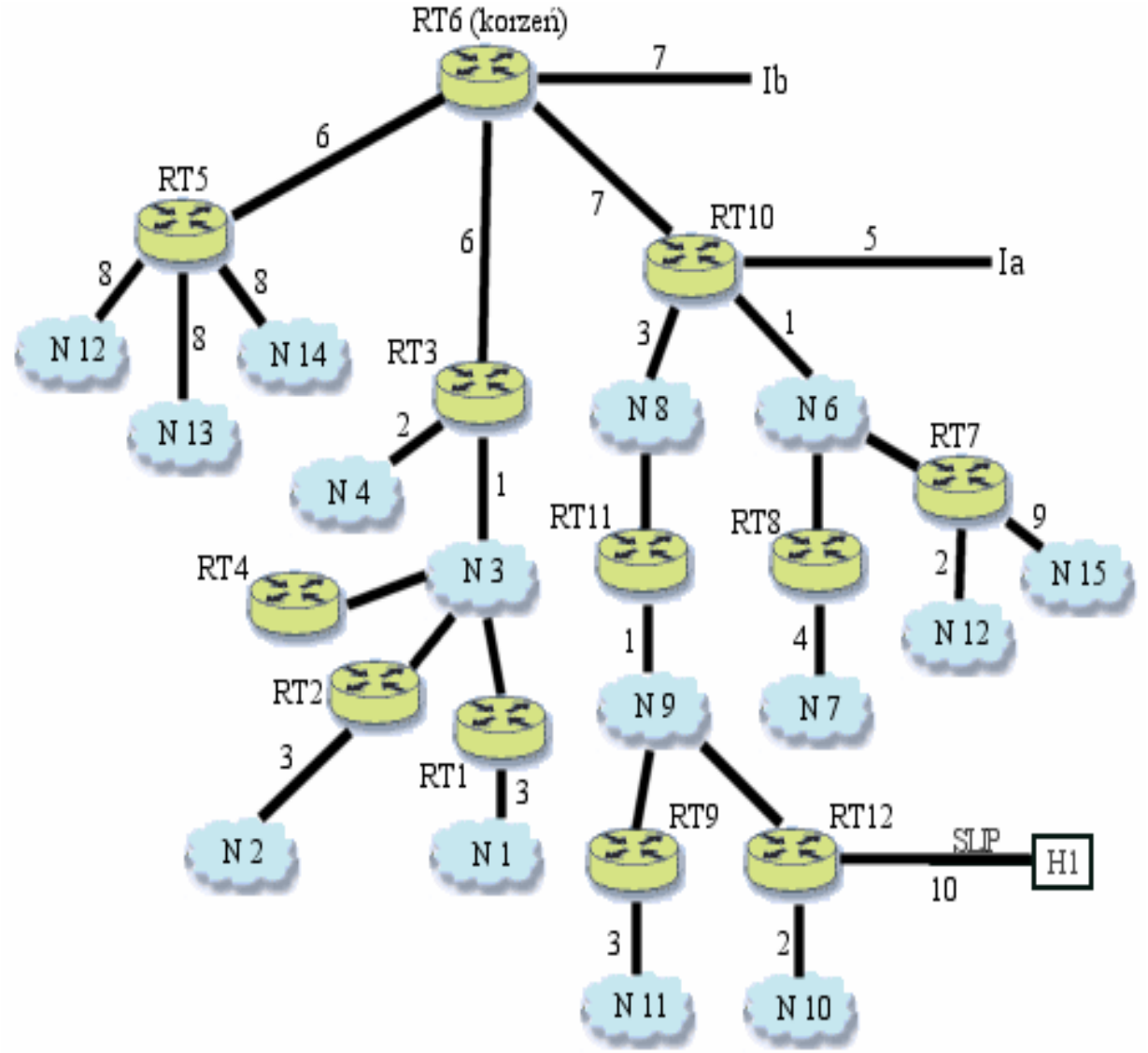
Tabela 31) LSA routera pochodzące od RT12

	RT9	RT11	RT12	N9
RT9				0
RT11				0
RT12				0
N9				

Tabela 32) LSA sieci pochodzące od N9

Interpretacja zapisu w powyższych tabelach jest identyczna jak interpretacja tabeli 30. Warto w tym miejscu zwrócić uwagę na fakt, że w pakiecie opisującym sieć (tabela 32) na przecięciu sieci N9 i routerów występują zera. Oznacza to, że koszt trasy prowadzącej od sieci do routera jest równy zero. Jest to logiczne rozwiązanie, ponieważ koszt trasy jest związany tylko z interfejsem wyjściowym routera[31].

Na podstawie informacji zebranych w bazie stanu połączeń budowane algorytmem Dijkstr'y drzewo najkrótszych ścieżek, które w naszym przypadku, dla routera RT6 będzie miało postać:



Rysunek 7) Drzewo najkrótszych ścieżek dla OSPF

natomiast zbudowana na podstawie tego drzewa tablica routingu, będzie posiadała następujące wpisy opisujące lokalne obiekty docelowe:

Obiekt docelowy	Następny skok	Odległość
N1	RT3	10
N2	RT3	10
N3	RT3	7
N4	RT3	8
Ib	*	7
Ia	RT10	12
N6	RT10	8
N7	RT10	12
N8	RT10	10
N9	RT10	11
N10	RT10	13
N11	RT10	14
H1	RT10	21
RT5	RT5	6
RT7	RT10	8

Tabela 33) Tablica routingu routera RT6

Po stworzeniu drzewa badane są informacje o trasach prowadzących do innych systemów autonomicznych. Informacje takie mogą pochodzić od innych protokołów routingu, takich jak BGP, lub też być wprowadzone statycznie[31]. Również trasy domyślne mogą być częścią informacji o routingu zewnętrznym. Trasy dotyczące routingu zewnętrznego są rozsyłane niezmienione w obrębie całego systemu autonomicznego. Tak więc, w przykładowej topologii wszystkie routery posiadają informacje, że router RT7 ma dwie trasy zewnętrze o metrykach 2 i 9[31].

W dotychczasowych rozważaniach przyjmowaliśmy, że trasy do odległych obiektów docelowych zawsze prowadzą przez router brzegowy systemu autonomicznego, który uczestniczy w wymianie informacji OSPF. Nie zawsze jednak jest to najlepsze rozwiązanie[31]. Aby zobrazować sytuacje, w której takie rozwiązanie nie jest najefektywniejsze przyjmijmy, że w topologii sieci istnieje dodatkowy router RTX, przyłączony do sieci N6. Przyjmijmy dalej, że router ten nie uczestniczy w wymianie informacji protokołu OSPF, ale używa protokołu BGP i za jego pomocą kontaktuje się z brzegowym routerem RT7. Router RT7 ogłasza trasy nauczone od RTX poprzez protokół OSPF i w ten sposób staje się pośrednikiem w procesie przesyłania pakietów. Jest to więc dodatkowy, niepotrzebny skok, ponieważ pakiety mogłyby być kierowane za pośrednictwem sieci N6 bezpośrednio do routera RTX. Aby zapobiegać takim sytuacjom routery brzegowe systemu autonomicznego mają możliwość określenia w przesyłanych pakietach LSA następnego routera na trasie. Wykorzystując to router RT7, powiadomi innych uczestników systemu autonomicznego o możliwości bezpośredniego kierowania pakietów do routera RTX.

Umieszczanie informacji o następnym routerze na ścieżce może służyć jeszcze jednemu celowi. Umożliwia to działanie routerów znajdujących się wewnątrz systemu autonomicznego, jako „serwerów tras”[31]. W naszej przykładowej sieci serwerem tras mógłby stać się router RT6. Informacje o odległych obiektach docelowych mogłyby być zbierane bądź to

ze statycznych wpisów, bądź przy użyciu zewnętrznych protokołów routingu. Tak zebrane informacje rozsyłane byłyby w postaci pakietów LSA, w których umieszczone zostałyby jednocześnie dane na temat routera, do którego pakiety powinny być kierowane.

10.6 Dzielenie systemu autonomicznego na obszary

Protokół OSPF umożliwia grupowanie przyległych sieci i hostów w obszary[31]. OSPF jest protokołem bezklasowym (patrz też rozdziały 7 i 8), co oznacza, że wraz z adresem IP przesyłana jest maska sieci, co pozwala opis grupy adresów za pomocą jednego wpisu w bazie danych, a w konsekwencji grupowanie sieci. Routery posiadające interfejsy do zgrupowanych sieci wchodzą również w skład obszaru. Każdy z takich obszarów posługuje się oddzielnym protokołem stanu łącza. Oznacza to, że każdy z obszarów posiada własną bazę stanów łączy i własny graf opisujący topologię sieci wewnątrz obszaru. Szczegóły tej topologii nie są widoczne na zewnątrz obszaru. Z drugiej strony natomiast, szczegóły dotyczące topologii na zewnątrz obszaru nie są znane routerom należącym do wewnętrznej struktury obszaru[31]. Taka izolacja wiadomości umożliwia znacząco zmniejszyć ruch pakietów związany z wymianą informacji między routerami. W związku z podziałem systemu autonomicznego na obszary nie jest prawdą, że wszystkie routery systemu autonomicznego posiadają identyczną bazę opisującą topologię sieci. Prawdą jest natomiast, że dwa routery należące do tego samego obszaru posiadają identyczną bazę stanów łączy. Ze względu na podział systemu autonomicznego na obszary występują dwa rodzaju trasowania w obrębie systemu. Jeśli nadawca i odbiorcą należą do tego samego obszaru, to jest to trasowanie wewnątrzobszarowe, a jeśli należą do różnych obszarów, to nazywamy to trasowaniem międzyobszarowym.

Na routerach Cisco istnieje możliwość skonfigurowania obszarów i ich parametrów[14]. Możliwe jest włączenie uwierzytelnienie dla obszaru, zdefiniowanie obszaru jako obszaru końcowego, określenie kosztu trasy prowadzącej poza obszar końcowy. Konfiguracji podlegają również parametry mające wpływ na przebieg procesu agregacji tras.

10.6.1 Szkielet systemu autonomicznego

Szkieletem systemu autonomicznego nazywamy specjalny obszar o numerze 0[11]. Do szkieletu należą wszystkie brzegowe routery obszarów w systemie. Zadaniem szkieletu jest dystrybucja informacji trasowania pomiędzy obszarami. Każdy obszar musi mieć połączenie ze szkieletem systemu autonomicznego[14]. Jeśli szkielet uległ przerwaniu, lub został celowo podzielony konieczne jest ustanowienie połączeń wirtualnych. Połączenie takie tworzone jest pomiędzy brzegowymi routerami obszarów. Wirtualne połączenie nie może zostać utworzone między obszarami końcowymi.

10.6.2 Trasowanie międzyobszarowe i podział routerów

Jak wspomniano wcześniej trasowanie międzyobszarowe występuje, gdy nadawca i odbiorca należą do różnych obszarów[31]. W takim przypadku do przepływu informacji między obszarami używany jest szkielet systemu autonomicznego. Trasę, jaką przebywają pakiety można podzielić na trzy przylegające do siebie części: trasę wewnątrz obszaru - od nadawcy do routera brzegowego obszaru, trasę wewnątrz szkieletu - łączącą obszary i trasę wewnątrz docelowego obszaru - od routera brzegowego obszaru do końcowego odbiorcy. Algo-

rytm trasowania jest odpowiedzialny za znalezienie takiego zestawu tras, których całkowity koszt jest najmniejszy. Stosowanie w obrębie szkieletu połączeń wirtualnych daje administratorowi systemu pewną kontrolę nad wyborem przez algorytm tras łączących obszary.

Wybór odpowiedniego routera brzegowego obszaru jest dokonywany dokładnie w ten sam sposób, co wybór routera ogłaszającego trasy zewnętrzne w systemie. Każdy router brzegowy obszaru ogłasza skrócone informacje na temat sieci znajdujących się poza obszarem. Po zbudowaniu drzewa najkrótszych ścieżek dla obszaru, trasy do wszystkich obiektów leżących poza obszarem są obliczane na podstawie analizy skróconych informacji o sieciach otrzymanych od routerów brzegowych obszaru.

Podział systemu autonomicznego na obszary oraz cały mechanizm przepływu informacji powoduje, że routery spełniają różne funkcje w zależności od położenia w topologii[31]. Ze względu na funkcjonalność wyróżniamy następujące rodzaju routerów:

- routery wewnętrzne – posiadają interfejsy do sieci należących do tego samego obszaru. Routery te posługują się pojedynczym protokołem routingu.
- routery brzegowe obszaru – routery te przyłączone są do kilku obszarów i dla każdego z nich posiadają uruchomioną oddzielną kopię algorytmu trasowania. Routery brzegowe kondensują informacje na temat przyłączonych obszarów i przekazują te informacje do szkieletu systemu autonomicznego. Z kolei szkielet zajmuje się dystrybucją tak skondensowanych danych do innych obszarów.
- routery szkieletu – są to te routery, które posiadają interfejs do obszaru 0, czyli do szkieletu systemu. Do tej grupy routerów zaliczamy wszystkie routery posiadające interfejsy do więcej niż jednego obszaru. Wynika z tego, że routery brzegowe obszarów są routerami szkieletu. Jednakże, routery szkieletu nie muszą być routerami brzegowymi obszaru. Dopuszczalna jest bowiem sytuacja, w której router posiada interfejsy tylko i wyłącznie do obszaru 0.
- routery brzegowe systemu autonomicznego – routery wymieniające informacje trasowania z routerami należącymi do innego systemu autonomicznego. Routery tego typu ogłaszają informacje o zewnętrznym routingu w obrębie całego systemu autonomicznego, do którego należą i z tego względu każdy router należący do systemu posiada informacje o ścieżkach prowadzących do wszystkich routerów brzegowych systemu autonomicznego. Routerem brzegowym systemu autonomicznego może być router wewnętrzny lub router brzegowy obszaru. Może (ale nie musi) być częścią szkieletu systemu autonomicznego.

Wymiana informacji może odbywać się również w kooperacji z routerami z zaimplementowanymi innymi niż OSPF protokołami routingu dynamicznego[14]. Routery Cisco dają nam taką możliwość. W obrębie systemu autonomicznego oznacza to importowanie tras od takich protokołów jak IGRP, RIP, IS-IS. Możliwa jest również sytuacja odwrotna, w której trasy OSPF eksportowane są do IGRP, RIP, IS-IS. Na poziomie routingu między systemami autonomicznymi implementacja Cisco wspomaga wymianę między OSPF a takimi protokołami jak EGP i BGP.

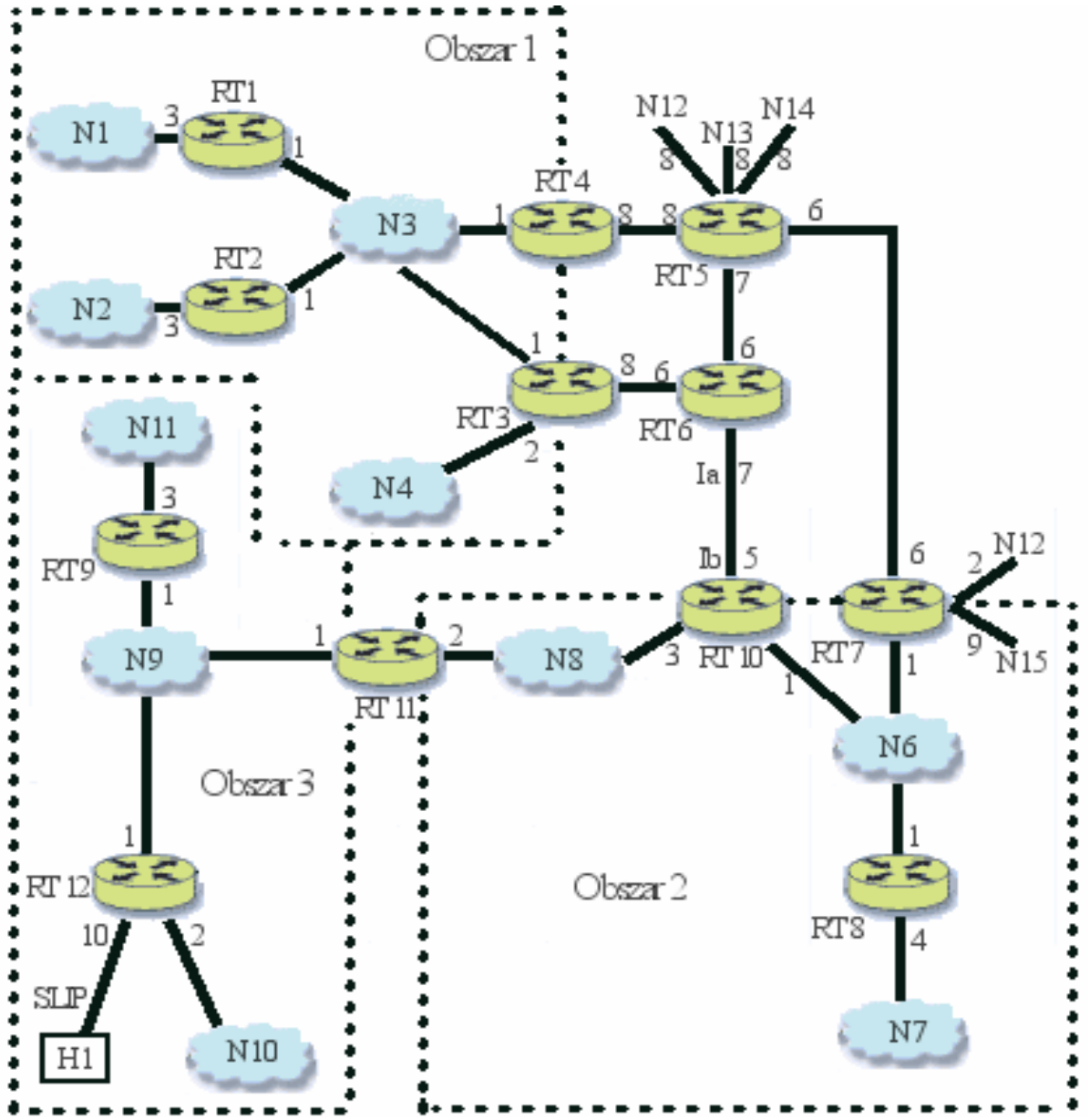
10.6.3 Przykład podziału systemu autonomicznego na obszary

Ponownie, w celu lepszego zobrazowania sytuacji, jakie mogą wystąpić w topologii, posłużymy się przykładem[31]. Rysunek 8 przedstawia przykładową konfigurację obszarów.

Pierwszy obszar składa się z sieci N1-N4 i przyłączonych do nich routerów N1-N4. Obszar drugi to sieci N6-N8 i routery RT7, RT8, RT10 i RT11. W skład trzeciego obszaru wchodzą sieci N9-N11 i host H1 i routerów RT9, RT11 i RT12. Trzeci obszar został tak skon-

figurowany, że sieci N9-N11 i host H1 będą kondensowane do pojedynczego wpisu w ogłoszeniach kierowanych na zewnątrz obszaru.

W przykładowej topologii routerami wewnętrznymi są routery: RT1, RT2, RT5, RT6, RT8, RT9 i RT12, natomiast routerami brzegowymi obszarów są routery: RT3, RT4, RT7, RT10 i RT11. Routery RT5(wewnętrzny) i RT7(brzegowy obszaru) są dodatkowo routerami brzegowymi systemu autonomicznego.



Rysunek 8) Podział systemu autonomicznego na obszary w OSPF

Baza stanów łączy dla obszaru 1 będzie miała postać jak pokazano w tabeli 34.

Zadaniem routerów brzegowych RT3 i RT4 jest dostarczenie do obszaru 1 informacji o obiektach leżących poza tym obszarem, jak i informacji o umiejscowieniu brzegowych routerów systemu autonomicznego RT5 i RT7. Informacja o umiejscowieniu routerów brzegowych systemu jest konieczna, ponieważ routery te ogłaszają trasy zewnętrzne w całym systemie autonomicznym, więc również w obszarze 1. Te trasy zewnętrzne są zawarte w bazie obszaru i opisują dostęp do sieci N12-N15.

Routery RT3 i RT4 tworzą skrót informacji dotyczących obszaru 1 i zajmują się dystrybucją tych danych do szkieletu w postaci pakietów LSA. W tabeli 35 przedstawione są ogłaszane przez RT3 i RT4 informacje, w których skład wchodzą sieci znajdujące się w obszarze 1 i koszt dotarcia do nich. Koszt ten to koszt dotarcia od routera brzegowego obszaru do danej sieci.

	RT1	RT2	RT3	RT4	RT5	RT7	N3
RT1							0
RT2							0
RT3							0
RT4							0
RT5			14	8			
RT7			20	14			
N1	3						
N2		3					
N3	1	1	1	1			
N4			2				
Ia, Ib			20	27			
N6			16	15			
N7			20	19			
N8			18	18			
N9-N11, H1			29	36			
N12					8	2	
N13					8		
N14					8		
N15						9	

Tabela 34) Baza stanów łączy dla obszaru 1

Sieć	Ogłoszenia RT3	Ogłoszenia RT4
N1	4	4
N2	4	4
N3	1	1
N4	2	3

Tabela 35) Ogłoszenia sieci w obszarze 1

Baza danych dla szkieletu systemu autonomicznego pokazana jest w tabeli 36. Przedstawione w tabeli routery należą do szkieletu systemu. Router R11 także należy do szkieletu, bo posiada interfejsy do dwóch obszarów. Aby szkielet systemu był połączony, między routerami RT10 i RT11 zostało skonfigurowane połączenie wirtualne.

Szkielet systemu umożliwia wymianę skróconych informacji o sieciach między obszarami[11]. Każdy router brzegowy obszaru, na podstawie informacji przesyłanych od innych routerów brzegowych, tworzy obraz sieci na zewnątrz obszaru. W naszej topologii routery RT3 i RT4 tworzą drzewo najkrótszych ścieżek dla szkieletu systemu. Na podstawie tego drzewa mogą wyznaczyć trasy do wszystkich routerów brzegowych obszarów w systemie. W drzewie umieszczone są również informacje dotyczące sieci (Ia i Ib) oraz brzegowych route-

rów systemu (RT5 i RT7) należących do szkieletu. Trasy do routerów w systemie wyliczone na podstawie drzewa są pokazane w tabeli 37. Trasy do obiektów docelowych wyliczone na podstawie drzewa są następnie ogłaszane wewnątrz obszaru, do którego należą RT3 i RT4. Zestaw tych ogłoszeń przedstawia tabela 38.

	RT3	RT4	RT5	RT6	RT7	RT10	RT11
RT3				6			
RT4			8				
RT5		8		6	6		
RT6	8		7			5	
RT7			6				
RT10				7			2
RT11						3	
N1	4	4					
N2	4	4					
N3	1	1					
N4	2	3					
Ia						5	
Ib				7			
N6					1	1	3
N7					5	5	7
N8					4	3	2
N9-N11, H1							11
N12			8		2		
N13			8				
N14			8				
N15					9		

Tabela 36) Baza danych szkieletu systemu autonomicznego

	Odległość od RT3	Odległość od RT4
Do RT3	*	21
Do RT4	22	*
Do RT7	20	14
Do RT10	15	22
Do RT11	18	25
Do Ia	20	27
Do Ib	15	22
Do RT5	14	8
Do RT7	20	14

Tabela 37) Wyliczone na podstawie drzewa trasy do routerów w systemie autonomicznym

Obiekt	Ogłoszenia RT3	Ogłoszenia RT4
Ia, Ib	20	27
N6	16	15
N7	20	19
N8	18	18
N9-N11, H1	29	36
RT5	14	8
RT7	20	14

Tabela 38) Zestaw ogłoszeń od routerów RT3 i RT4

Informacje rozesłane do obszaru 1 pozwalają routerom wewnętrznym, takim jak RT1, na inteligentny wybór routera brzegowego obszaru do przesyłania danych na zewnątrz obszaru. Tak więc RT1 wybierze router RT4 do przysyłania pakietów, których odbiorca znajduje się w sieci N6, a RT3 dla pakietów skierowanych do sieci N10. Ruch do sieci N8 będzie dzielony między routery RT3 i RT4. W ten sam sposób RT1 może decydować, który router wybrać do przesyłania pakietów na zewnątrz systemu (RT5 lub RT7).

10.6.4 Obszary końcowe

W pewnych systemach autonomicznych większą część bazy stanów łączy mogą wypełniać informacje dotyczące tras zewnętrznych. Ogłoszenia o trasach zewnętrznych są rozgłaszane w całym systemie autonomicznym. Jednakże, OSPF umożliwia określanie pewnych obszarów jako obszary końcowe. Pakiety LSA z zewnętrznymi trasami nie są rozgłaszane w obrębie tych obszarów, a routing do obiektów znajdujących się poza systemem autonomicznym jest zorganizowany na zasadzie tras domyślnych. To znacznie redukuje rozmiar bazy danych, a tym samym rozmiar potrzebnej pamięci, routerów wewnątrz obszarów końcowych[31].

W celu używania tras domyślnych w obszarze końcowym jeden, lub więcej, routerów brzegowych obszaru końcowego musi ogłaszać trasy domyślne wewnątrz obszaru. Takie trasy domyślne będą użyte dla każdego obiektu, który nie może być osiągany ani poprzez trasę wewnątrz obszaru, ani trasę między obszarami (na przykład obiekty poza systemem autonomicznym).

Obszar może przykładowo być zdefiniowany jako końcowy, gdy istnieje tylko jeden punkt wyjściowy z obszaru. Obszar 3 na rysunku 8 może zostać zdefiniowany jako końcowy, ponieważ cały ruch z tego obszaru przechodzi przez brzegowy router obszaru RT11. Jeśli obszar 3 zdefiniowany byłby jako końcowy, router RT11 ogłaszałby trasę domyślną wewnątrz obszaru 3, zamiast rozsyłać pakiety LSA zawierające informacje o trasach zewnętrznych (trasy do sieci N12-N15).

OSPF zapewnia, że wszystkie routery należące do obszaru końcowego zgadzają się działać w tym trybie, co jest konieczne w celu uniknięcia nieprawidłowości związanych z możliwością jednoczesnego ogłaszaniem tras domyślnych przez jeden router i pakietów LSA o trasach zewnętrznych przez inny router[31]. Istnieją dodatkowe restrykcje co do końcowych obszarów: połączenia wirtualne nie mogą być konfigurowane przez obszar końcowy, brzegowy router systemu autonomicznego nie może być umiejscowiony wewnątrz obszaru końcowego.

10.7 Routery przyległe (adjacent routers)

Routery przyległe, to sąsiadujące ze sobą routery, połączone ze sobą logicznie w celu wymiany informacji o trasowaniu[11]. Nie wszystkie routery sąsiadujące stają się routerami przyległymi.

Za nawiązanie i utrzymanie połączenia między przyległymi routerami odpowiedzialny jest protokół Hello. Zapewnia on komunikacje dwukierunkową między tymi routerami. Pakiety protokołu Hello są rozsyłane cyklicznie poprzez wszystkie interfejsy routera. Jeśli router otrzymujący pakiet Hello odnajdzie siebie samego na liście routerów ogłoszonych w pakiecie, to oznacza to, że istnieje dwukierunkowe połączenie z nadawcą tego pakiety. W sieciach rozgłoszeniowych i NBMA protokół Hello służy również wybraniu routera desygnowanego (Designed Router).

W zależności od rodzaju połączenia między routerami, to znaczy rodzaju zastosowanej technologii sieciowej, protokół Hello działa odmiennie[31]. W sieciach rozgłoszeniowych, każdy router ogłasza samego siebie używając multiemisji do rozsyłania pakietów Hello. To pozwala na dynamiczne odnajdowanie sąsiadujących routerów. Pakiety Hello, w tym przypadku, zawierają informacje o routerze desygnowanym i listę routerów, od których ostatnio otrzymane zostały pakiety Hello.

W sieciach NBMA każdy router, który może potencjalnie stać się routerem desygnowanym (ma przypisany odpowiedni priorytet) posiada listę wszystkich routerów przyłączonych do sieci. Gdy interfejs do tejże sieci zostaje włączony, router wysyła pakiety Hello do wszystkich innych routerów będących w stanie stać się routerami desygnowanymi. Jest to próbą znalezienia routera desygnowanego dla sieci. W przypadku nie odnalezienia routera desygnowanego, router sam staje się routerem desygnowanym i rozpoczyna on rozsyłać pakiety Hello do wszystkich routerów w sieci[31].

W sieciach rozgłoszeniowych i NBMA router desygnowany spełnia dwie podstawowe funkcje. Po pierwsze, jest źródłem pakietów LSA opisujących sieć. Te pakiety LSA zawierają listę wszystkich routerów aktualnie przyłączonych do sieci. Identyfikatorem tych pakietów jest adres interfejsu routera desygnowanego. Po drugie, router desygnowany staje się routerem przyległym dla wszystkich routerów w sieci i tym samym staje się centralnym punktem synchronizacji baz stanów łączy.

Dodatkowo, oprócz podstawowego routera desygnowanego, istnieje desygnowany router rezerwowy. Router ten jest również przyległy do wszystkich innych routerów w sieci i w przypadku awarii routera podstawowego zajmuje jego miejsce.

Po odnalezieniu sąsiadującego routera i zapewnieniu dwukierunkowego połączenia a w sieciach rozgłoszeniowych i NBMA dodatkowo wyborze routera desygnowanego, następują dalsze czynności prowadzące do ustanowienia pełnej przyległości routerów[31].

Pierwszym krokiem jest ustalenie, który z routerów jest routerem nadrzędnym oraz ustalenie początkowego numeru DD używanego do numeracji przesyłanych pakietów opisu bazy danych[11]. Po tych czynnościach routery przesyłają do siebie pakiety zawierające skrócony opis całej zawartości ich baz stanów łączy. Każdy pakiet zawiera sekwencyjny numer DD, a wysłanie następnego pakietu jest uzależnione od otrzymania potwierdzenia dotarcia pakietu do odbiorcy. Router nadrzędny wysyła pakiet jako pierwszy, natomiast router podrzędny potwierdza odbiór przesyłki poprzez odesłanie pakietu z opisem swojej bazy opatrzonego tym samym numerem sekwencyjnym. Numer sekwencyjny jest zwiększany przez router nadrzędny przy wysyłaniu kolejnego pakietu. Jeśli router nadrzędny nie otrzyma potwierdzenia w określonym czasie następuje retransmisja pakietu. Retransmitować pakiet może tylko router nadrzędny.

Cała wymiana pakietów ma na celu synchronizację baz obu przyległych routerów. Jeśli router otrzyma pakiet i zauważy, że sąsiedni router posiada świeższe informacje na temat któregoś z elementów topologii sieci, to zażąda przesłania dokładnych, aktualnych danych na temat tegoż elementu. Każdy pakiet posiada pole M-bit określające, czy po pakiecie nastąpią dalsze pakiety opisujące bazę stanu łączy. Proces wymiany pakietów między przyległymi routerami kończy się, gdy router wysłał i otrzymał pakiet z wyzerowanym polem M-bit.

Po zakończeniu wymiany routery oznaczone zostają jako całkowicie przyległe. Od tego czasu przyległość routerów jest rozgłaszana w wysyłanych przez nie pakietach LSA.

Typowa implementacja OSPF wymaga skoordynowania pracy routerów wewnętrznych, routerów brzegowych obszarów i routerów brzegowych systemu autonomicznego[14]. Minimalne parametry protokołu mogą być skonfigurowane poprzez włączenie ustawień domyślnych, bez uwierzytelniania, z interfejsami przypisanymi do obszarów. Implementacja Cisco pozwala na wprowadzania zmian w ustawieniach interfejsów, jeśli jest to konieczne. Zmiany ustawień jednego routera mogą wymusić zmiany parametrów konfiguracyjnych innych routerów w systemie. Należy pamiętać, że pewne parametry muszą przyjmować taką samą wartość we wszystkich komunikujących się routerach dołączonych do sieci.

10.8 Stabilność

Algorytm wyznaczania tras stosowany w OSPF jest zupełnie inny od tego stosowanego w protokołach wektora odległości. Budowa drzewa najkrótszych ścieżek i na jego podstawie wyznaczanie tras w naturalny sposób likwiduje możliwość powstawania pętli w sieci. Nie są więc tu wymagane mechanizmy zapobiegania pętlom[31]. Po zmianie stanu łącza router generuje nowy pakiet LSA, który rozsyłany jest od routera do routera, a każdy router otrzymujący ten pakiet musi przeliczyć od nowa drzewo najkrótszych ścieżek i na jego podstawie zaktualizować tabelę routingu. Dzięki takiemu rozwiązaniu zapewniona jest stabilność trasowania[31].

11 Protokół NLSP

11.1 Wstęp

NLSP (NetWare Link-Services Protocol) jest protokołem trasowania dynamicznego przede wszystkim w średnich i dużych sieciach informatycznych z łączami WAN.

Podstawowe właściwości, jakimi cechuje się protokół NLSP to[1]:

- połączeniowy charakter trasowania, co oznacza, że routery NLSP gwarantują doręczenie pakietów,
- przechowywanie identycznego obrazu całej sieci (grafu sieci) w bazach danych wszystkich routerów NLSP,
- zarządzanie SNMP (Simple Network Management Protocol),
- kompresowanie nagłówka IPX (Internetwork Packet eXchange), zmniejszające rozmiar pakietów (szczególnie ważne w ruchu WAN),
- oszczędny ruch wywołany aktualizacją grafu sieci i tabel przyległości routerów,
- szybka reakcja na zerwanie połączenia i przełączenie na alternatywną trasę, połączone z uaktualnieniem topologii sieci w tabelach wszystkich routerów,
- uzależnienie metodyki aktualizacji tabel routerów od własności sieci (LAN, łącza WAN),
- wsteczna kompatybilność z protokołem RIP,
- równoważenie obciążeń łączy równoległych o tym samym koszcie,
- adresowanie hierarchiczne,
- rozsyłanie informacji do grupy routerów przy wykorzystaniu multiemisji (multicasting) lub rozgłaszanie (broadcasting)

11.2 Pakiety przesyłające wiadomości

W tym rozdziale przedstawione zostaną formaty pakietów Hello. O wykorzystaniu pakietów mowa jest w dalszej części pracy.

Pakiety WAN i LAN Hello są bardzo podobne do siebie. Pakietom WAN Hello nie są potrzebne pola: *No Multicast*, *Priority* i *LAN IDentifier*, a pakietom LAN Hello pola: *State* i *Local WAN Circuit ID*.

11.2.1 Format ramki LAN Hello

Identyfikator			Wskaźnik długości	
Mniejsza wersja			Zarezerwowane	
Zarezerwowane				Typ pakieu
Większa wersja			Zarezerwowane	
Zarezerwowane			NM	Zarezerwowane	CT
ID źródła				
Czas utrzymania				
Długość pakietu				
Zarezerwowane	Priorytet		LAN ID	
LAN ID				

Tabela 39) Format ramki LAN Hello

Opis :

- Identyfikator – identyfikator poziomu trasowania
- Wskaźnik długości – wskaźnik liczby bajtów nagłówka
- Typ pakietu – pole to zajmuje 5 bitów i określa typ przesyłanego pakietu
- NM (no multicasting) – 1 bitowy przełącznik multicast/broadcast, ustawiany przez nadawcę. Jeśli sterownik karty sieciowej nie akceptuje trybu grupowego, nadawca ustawia NM=1 i pakiety adresowane do niego będą rozgłaszane.
- CT (circuit type) – 2 bitowe pole określające typ łącza powiązanego z danym portem. Może przyjmować następujące wartości:
  0 – wartość zarezerwowana. Powoduje pominięcie całego pakietu przy odbiorze,
  1 – tylko trasowanie poziomu 1,
  2 – tylko trasowanie poziomu 2 (router – nadawca używa tego łącza do trasowania na poziomie 2),
  3 – trasowanie poziomów 1 i 2 (nadawcą jest router poziomu 2, wykorzystuje to łącze dla ruchu na poziomie 1 i 2).
- ID źródła – identyfikator routera wysyłającego pakiet,
- Czas utrzymania (hold time) – czas utrzymania rekordu w tabeli przyległości mierzony w sekundach. Po upływie tego czasu połączenie zostanie wykreślone z tabeli przyległości.
- Długość pakietu – długość całego pakietu w bajtach, łącznie z nagłówkiem NLSP;
- Priorytet – 5-bitowy priorytet routera na określonym porcie. Domyślna wartość przypisywana temu polu wynosi 44. Maksymalna wartość priorytetu to 127. Router wybrany na router desygnowany na jednym porcie nie musi być nim na innym porcie.
- LAN ID – identyfikator LAN poziomu 1. Składa się z 6-bajtowego systemu nazewniczego routera desygnowanego poziomu 1 i 1-bajtowego identyfikatora nadawanego przez router.

11.2.2 Format ramki WAN Hello

Identyfikator	Wskaźnik długości	
Mniejsza wersja	Zarezerwowane	
Zarezerwowane		Typ pakietu
Większa wersja	Zarezerwowane	
Zarezerwowane	Stan	CT
ID źródła		
Czas utrzymania		
Długość pakietu		
ID obwodu		

Tabela 40) Format ramki WAN Hello

Opis:

- ID obwodu - Pole to określa unikatowy identyfikator obwodu, nadawany przez router ustalający swoje przyległości WAN,
- Stan – 2 bitowe pole stanu łącza WAN, wykorzystywane przez routery ustalające swoje przyległości WAN; 0 – nieczynne, 1 – inicjalizacja, 2 – aktywne.

11.3 Maska i adresowanie hierarchiczne

NLSP został przystosowany do schematu adresowania hierarchicznego. Każdy obszar trasowania jest rozpoznawany po dwu 32-bitowych numerach, z których jeden stanowi adres sieci (network address), na przykład 00112200 w zapisie szesnastkowym, a drugi maskę (mask), jak FFFFFF00 w tym samym zapisie. Para takich numerów nosi wspólną nazwę adresu obszaru (area address). FFFFFF identyfikuje obszar trasowania (tu 001122), a 00 (dwójkowo 00000000) - indywidualne numery sieci w tym obszarze: A1, czyli 10100001, C2 (11000010) czy 1B (00011011) itd. Adresy kolejnych sieci w obszarze przybiorą ostatecznie formy w rodzaju: 001122A1, 001122C2 itp.

W podobny sposób dochodzi się aż do adresu sieci globalnej. Odpowiednia maska wskaże wtedy, jaka część takiego adresu przypadnie na numer domeny, numer obszaru w tej domenie i wreszcie na numer sieci w obszarze.

Trasowanie obszaru może wykorzystywać aż 3 różne adresy obszaru, każdy z inną maską. Tak pokaźny i elastyczny system adresowania umożliwia organizowanie trasowania obszaru bez przerywania innych operacji sieciowych. W określonej domenie może być wykorzystana dowolna kombinacja takiego adresu obszaru.

11.4 Bazy danych

NLSP jest protokołem trasowania dynamicznego należącym do grupy protokołów stanu łącza (link state). Oznacza to, że każdy router NLSP przechowuje bazę stanów wszystkich łączy i buduje graf połączeń na podstawie, której podejmowane są decyzje o routingu. Baza stanów łączy nie jest nigdy przesyłana w całości. Jest natomiast uaktualniana albo co 2 godzi-

ny, albo po każdej zmianie konfiguracji czy usług lub po awarii. Wszystkie routery muszą dysponować bazą o identycznej zawartości. Skompletowanie danych następuje przy użyciu specjalnych pakiety LSP (Link State Packets).

W bazie danych routera znajduje się także relatywnie niewielka tabela jego bezpośredniego sąsiedztwa, nazywana tabelą przyległości (Adjacency Database). W myśl protokołu NLSP router A jest sąsiadem routera B, jeśli może się z nim komunikować bez pośrednictwa innego routera[6]. Router należy do tylu przyległości, ile ma czynnych portów. Zawartości tabel przyległości, w przeciwieństwie do zawartości bazy stanów łączy, są różne w każdym routerze. Tabela określonego routera zawiera dane o sąsiednich routerach oraz opisy stanu połączeń z każdym z nich. Routery aktualizują swoje tabele na podstawie odpowiedzi uzyskiwanych po wyemitowaniu pakietów LAN Hello lub WAN Hello, zwykle co 20 sekund. Router w każdej przyległości reprezentuje jeden z nich, nazywany routerem desygnowanym DR (Designated Router).

Po skompletowaniu tabeli stanów łączy routery są gotowe do wypełniania swoich unikatowych tabel wyboru najlepszej trasy w grafie (Forwarding Database), przechowywanych w pamięci RAM[34]. Procedurą wyboru steruje algorytm Dijkstry, a więc wyliczane trasy będą się charakteryzowały najniższym kosztem przesyłania danych[15]. Router najpierw wyliczy koszt fragmentów tras prowadzących do węzłów, a następnie najniższy koszt całej trasy. A kiedy do tego samego celu prowadzi kilka tras o jednakowym koszcie, router zrównoważy ich obciążenia, kierując pakiety na każdą z nich. Jeśli najlepsza droga z A do E prowadzi przez B, C i D, to najlepsza droga z B do E prowadzi również przez C i D. Tak przekłada się na trasowanie spójność tabel stanów łączy i wspólny algorytm wyboru dróg.

Tabela wyboru najlepszej drogi składa się z rekordów zawierających numery sieci dostępnych z określonego routera, kolejnego skoku, kosztu przejścia przez sieć, interfejsu, na który zostanie skierowany pakiet, itp. Modyfikacja tabeli następuje po każdej zmianie w tabeli stanów łączy[34].

Wieloskładnikowa sieć globalna w ujęciu NLSP składa się z domen, domeny z obszarów, a obszary z sieci lokalnych lub segmentów. Obraz takiej sieci globalnej w postaci grafu powstaje ze złożenia przyległości wszystkich routerów. Z tych względów protokół NLSP został wyposażony w kilka mechanizmów umożliwiających precyzyjne ustalanie i modyfikowanie sąsiedztwa przez łącza LAN i WAN oraz rozpropagowanie tych wszystkich obrazów przyległości po całej sieci.

11.4.1 Tabela przyległości i pakiety Hello

Tabela przyległości routera NLSP zawiera dane o routerach z bezpośredniego sąsiedztwa oraz opisy ich połączeń (stanów). Taką tabelę kompletuje każdy router na podstawie odpowiedzi, jakie otrzymuje od swoich sąsiadów po rozesłaniu do nich specjalnych pakietów Hello. Z uwagi na duże różnice charakterystyk łączy LAN i WAN - protokół NLSP definiuje dwa rodzaje wspomnianych pakietów: LAN Hello i WAN Hello.

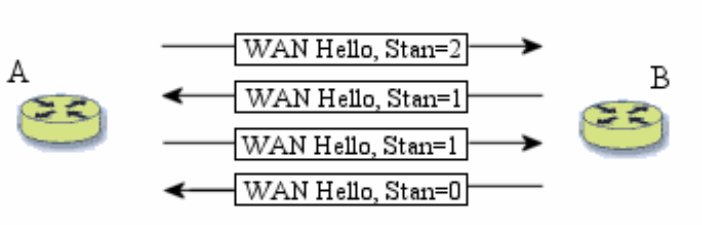
Kompletowanie tabel przyległości za pośrednictwem pakietów LAN Hello jest bardzo proste, gdyż sprowadza się do zidentyfikowania sąsiednich połączeń i routerów poziomu 1, funkcjonujących w tym samym obszarze trasowania.

Zalecanym trybem emisji pakietów LAN Hello jest zawsze multicasting, czyli adresowanie grupowe, obejmujące routery z przyległości[6]. Jeśli jednak sterowniki kart sieciowych nie akceptują takiego trybu, pakiety będą rozgłaszane (broadcasting).

Pakiety WAN Hello umożliwiają routerom odnalezienie wzajemnych powiązań, rozstrzygnięcie przynależności do tego samego poziomu trasowania i sprawdzenie, czy inne routery i łącza są jeszcze aktywne. Routery generują pakiety WAN Hello w następujących sytu-

acjach: kiedy określają swoje bezpośrednie sąsiedztwo po raz pierwszy lub kiedy upłynął określony czas, albo też zawartość następnego Hello różni się od zawartości poprzedniego. Pakiety WAN Hello są emitowane cyklicznie tak długo, jak długo istnieje określone łącze.

Pierwsze ustalanie przyległości przez łącza WAN jest trudniejsze niż każde inne. Routery muszą najpierw poznać charakterystyki łącza zależne od medium. W tym celu muszą się posłużyć protokołem IPXWAN v.2 (Internetwork Packet eXchange Wide Area Network). Z kolei pakiety WAN Hello będą wykorzystywane podczas uaktualniania tabel przyległości. Podstawowym parametrem dialogu będzie wówczas dwubitowa zawartość pola Circuit Type. Jeśli łącze między dwoma routerami jest czynne, zapis takiego sąsiedztwa do tabeli przyległości nastąpi po czterostopniowym dialogu (rysunek 9)



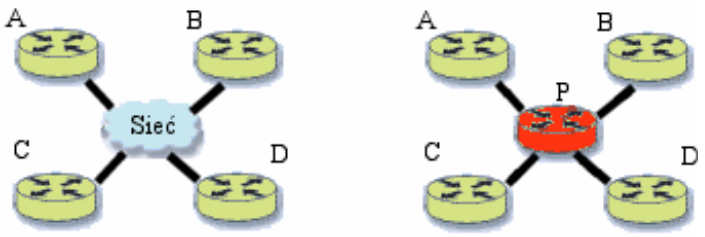
Rysunek 9) Czterostopniowy dialog routerów w NLSP

W pakietach WAN Hello przy stanie równym 2 połączenie jest nieczynne, gdy stan osiągnie wartość 1, oznacza to, że trwa proces łączenia. Stan 0 oznacza, że routery nawiązały łączność.

Router ustalający swoje przyległości WAN spodziewa się potwierdzania swojego pakietu Hello w precyzyjnie ustalonym czasie (Holding Time), kontrolowanym przez licznik (holding timer). W razie braku potwierdzenia w przewidzianym czasie router wykreśla odpowiedni rekord w swojej tabeli przyległości i rozsyła pakiety LSP, informujące środowisko o zmianie topologii[34].

11.4.2 Węzeł umowny i router desygnowany

Sieci w NLSP reprezentowane są przez pseudowęzły. Dzięki tej oryginalnej koncepcji Novella router A (rysunek 10) zawęzi swoje sąsiedztwo tylko do jednego węzła umownego - pseudowęzła PN. Podobnie węzły B, C i D. Tabele przyległości tych routerów będą teraz zawierały tylko po jednym rekordzie, charakteryzującym stany ich połączeń z pseudowęzłem. Ponieważ jednak każdy węzeł musi aktywnie uczestniczyć w normalnym ruchu pakietów, jak też w ruchu wywołanym aktualizacją tabel - pseudowęzeł zostaje zastąpiony w tych wszystkich funkcjach przez router desygnowany.



Rysunek 10) Router desygnowany

Routerem desygnowanym w określonej przyległości zostaje router o najwyższym priorytecie, a przy równych priorytetach - router o najwyższym adresie MAC. Router po ta-

kiej elekcji dodaje jeszcze do swojego priorytetu liczbę 20. Administrator sieci może podwyższyć priorytet do 100 lub nieco wyżej, ale nie więcej niż do 127, gdyż pole priorytetu każdego routera NLSP jest siedmiobitowe. Wszystkie routery po zainstalowaniu do sieci mają początkowo takie same priorytety - 44.

Do innych zadań routera desygnowanego należą jeszcze:

- komunikacja z węzłami korzystającymi z protokołów: trasowania,
- rozsyłanie pakietów kontrolnych CSNP (Complete Sequence Number Packets) do routerów sieci lokalnych, umożliwiających wykrycie zmian obrazu sieci przechowywanego przez te routery oraz DR,
- wysyłanie pakietów LAN Hello dwukrotnie częściej niż inne routery.

11.4.3 Przetwarzanie informacji trasowania

NLSP jest protokołem trasowania hierarchicznego, trzystopniowego, w którym trasowanie odbywa się w obrębie:

- obszaru, rozumianego jako zbiór połączonych ze sobą sieci, charakteryzujących się jednakowymi adresami obszaru,
- systemu autonomicznego, stanowiącego zbiór obszarów, administrowanych przez tę samą organizację, oraz
- wieloskładnikowej sieci globalnej, tworzonej przez systemy autonomiczne, które są powiązane ze sobą łączami sieciowymi, ale zarządzają nimi różne organizacje.

Trasowanie odpowiadające tej hierarchii przeprowadza się na trzech poziomach (Level 1-3 routing):

1 - komunikacja między sieciami wewnątrz wspólnego obszaru,
2 - komunikacja między obszarami oraz z routerem poziomu 1 wewnątrz jego obszaru,
3 - komunikacja między systemami autonomicznymi oraz z routerem poziomu 2 wewnątrz jego domeny.

Trasowanie hierarchiczne upraszcza i przyspiesza dystrybucję informacji o routingu, minimalizując ruch oraz zmniejszając ilość informacji niezbędnych do przesłania pakietów pod określonym adresem[34]. Na przykład router poziomu 1 powinien znać jedynie szczegółowe dane o łączach i routerach ze swojego najbliższego sąsiedztwa. Kiedy przesyła pakiety do innego obszaru - musi tylko znaleźć najbliższy router poziomu 2. Z kolei w komunikacji między obszarami routery poziomu 2 ogłaszają swoje adresy w akceptowanych obszarach, a nie w całej sieci. Podobnie funkcjonują routery poziomu 3, odpowiadającego trasowaniu między domenami. Dokładny opis trasowania hierarchicznego został omówiony w trakcie opisu protokołu OSPR w rodziale 10.6. Opis trasowania hierarchicznego można bezpośrednio odnieść do trasowania stosowanego w NLSP przy założeniu, że routery wewnętrzne OSPF to routery poziomu 1 NLSP, routery brzegowe obszarów to routery poziomu 2 NLSP, a routery brzegowe systemu autonomicznego to routery poziomu 3 NLSP.

11.4.4 Stabilność

W sieci z routerami NLSP, oprócz pakietów Hello, pojawiają się jeszcze trzy inne pakiety:

- LSP (*Link State Packet*),
- CSNP (*Complete Sequence Number Packet*),
- PSNP (*Partial Sequence Number Packet*).

Wszystkie te pakiety służą jednemu celowi: utrzymywaniu spójnej, jednakowej tabeli stanów łączy we wszystkich routerach NLSP. Pakiety LSP emitowane są przez wszystkie routery co 2 godziny lub przez router bezpośrednio po wykryciu w jego otoczeniu zmiany topologii czy usługi. Pakiety LSP wysyłane do łącza WAN są potwierdzane, gdyż nadmierny ruch może przeciążyć węzły na krańcach takiego łącza. W sieci lokalnej obowiązują inne procedury kontroli spójności tabeli stanów łączy, oparte na pakietach CSNP i PSNP.
Pakiety LSP przenoszą kompletne informacje o przyległościach routera, stanie jego połączeń i usługach. Każdy pakiet musi zawierać unikatowy identyfikator LSP ID (LSP IDentifier), złożony z numeru routera, numeru portu (circuit ID), z którego pochodzi pakiet, oraz z numeru fragmentu ID (fragment ID), kiedy cała informacja zajmuje więcej niż jeden pakiet. Takie pakiety przesyła router do wszystkich innych routerów przyległych, które uaktualniają swoje tabele stanów łączy i przesyłają je dalej, do własnych routerów przyległych itd. Sieć zalewają wtedy pakiety LSP (LSP flooding), gdyż każdy router musi otrzymać komplet pakietów, przenoszących w sumie pełny obraz sieci - tabelę stanów łączy. Taki proces zostanie zainicjowany ponownie po upływie 2 godzin.

W okresach między kolejnymi rozsyłaniami pakietów LSP może wystąpić wiele zdarzeń. Router po wykryciu zmiany w swoim otoczeniu emituje niezwłocznie pakiety LSP opatrzone identyfikatorem LSP ID i numerem sekwencyjnym (sequence number), po którym routery rozpoznają aktualność informacji i podejmują decyzję o modyfikowaniu swoich tabel. Pakiety kierowane do łącza WAN są potwierdzane.

Synchronizacją obrazu sieci we wszystkich routerach NLSP zajmują się routery desygnowane DR. Co około 30 s wysyłają one do routerów ze wszystkich swoich przyległości pakiety CSNP, przenoszące bardzo okrojony wypis z tabeli stanów łączy, obejmujący: identyfikator LSP ID, numer sekwencyjny oraz sumę kontrolną każdej informacji LSP. Informacje wysyłane przez DR mogą być - w odniesieniu do informacji stanów łączy adresatów - nowsze, identyczne lub starsze. Identyczne są pomijane, pozostałe wywołują różne reakcje.

Kiedy informacje z DR są nowsze, router z sąsiedztwa tworzy wtedy pakiet PSNP, który jest żądaniem skierowanym do DR o przesłanie szczegółowych informacji LSP opatrzonych tym samym identyfikatorem LSP ID, ale nowszym numerem sekwencyjnym. W przeciwnej sytuacji, a więc kiedy informacje DR są starsze, router, do którego był skierowany pakiet routera desygnowanego, zacznie rozpowszechniać swoje nowsze informacje w trybie multicast LSP. W taki sposób wszystkie routery będą miały jednakowy obraz topologii sieci, niezależnie od sytuacji.

12 Protokół IS-IS

12.1 Wstęp

Dla zestawu protokołów OSI (Open Systems Interconnection) opracowano w ramach organizacji ISO (International Standard Organization) kompletny zbiór protokołów routingu. Są to: ES-IS (End System-to-Intermediate System), IS-IS (Intermediate System-to-Intermediate System) i IDRP (Interdomain Routing Protocol).

Protokół routingu IS-IS opracowany przez OSI umożliwia wymianę informacji o routingu i dynamiczny routing zarówno pakietów TCP/IP jak i pakietów OSI[28]. Istnienie dwóch zestawów protokołów w sieci (TCP/IP i OSI) nie oznacza rozdzielenia sieci i ruchu pakietów. Wręcz przeciwnie, sieć musi być spójna, a pakiety muszą być przesyłane od odbiorcy do nadawcy bez względu na rodzaj używanych w sieciach pośrednich protokołów. Routery, które są odpowiedzialne za przesyłanie pakietów z informacjami muszą sprostać zadaniu trasowania nawet w przypadku istnienia różnych zestawów protokołów w sieciach, do których jest bezpośrednio podłączony. Istnieją dwa główne sposoby rozwiązania tego problemu. Pierwszy sposób, to implementacja na routerze niezależnych protokołów, po jednym dla każdego zestawu protokołów. Drugim sposobem rozwiązania problemu jest zastosowanie zintegrowanego protokołu IS-IS.

Zintegrowany IS-IS to pojedynczy protokół routingu zapewniający skuteczny i równoczesny routing dla TCP/IP i dla OSI[28]. Protokół dostarcza możliwość obsługi podsieci IP, podsieci o zmiennej długości maski, możliwość routingu opartego o TOS, oraz trasowanie na zewnątrz systemu autonomicznego. IS-IS przewiduje również możliwość uwierzytelniania informacji, bądź to przy użyciu hasła, bądź przy użyciu bardziej zaawansowanych metod uwierzytelniania. Protokół pozwala na współpracę routerów pracujących w trybie adresowania tylko IP (routery IP), adresowania tylko OSI (routery OSI) i adresowania podwójnego (routery mieszane posługujące się równocześnie i adresami IP i adresami OSI). Pozwala to na stosowanie protokołu w sieciach posługujących się wyłącznie protokołami TCP/IP, w sieciach posługujących się wyłącznie protokołami OSI, lub w sieciach posługujących się równocześnie protokołami TCP/IP i OSI[28].

12.2 Pakiety przesyłające wiadomości

W IS-IS wyróżniamy trzy główne rodzaje pakietów wymienianych między routerami: pakiety Hello, pakiety LSP (link state packets) i pakiety SNP (sequence number packets).

Pakiety Hello są wykorzystywane od tworzenia i utrzymywania połączenia między routerami przyległymi (więcej o routerach przyległych w rozdziale 10.7 opisującym protokół OSPF). Dodatkowo pakiety te dzielą się na:

- pakiety Hello LAN poziomu 1 – używane przez routery poziomu 1 w sieciach LAN rozgłoszeniowych,
- pakiety Hello LAN poziomu 2 – używane przez routery poziomu 2 w sieciach LAN rozgłoszeniowych,
- pakiety Hello punkt-punkt – używane w sieciach innych niż rozgłoszeniowe, przykładowo przy połączeniu routerów typu punkt-punkt.

Pakiety LSP służą do wymiany informacji o stanie połączeń w topologii między routerami. Jak już wspomniano wcześniej wyróżniamy dwa rodzaje tych pakietów: poziomu 1 i poziomu 2.

Pakiety SNP, czyli pakiety sekwencji, służą do numeracji pakietów LSP i mają na celu utrzymanie w całym obszarze jednakowego obrazu topologii w bazach routerów. Dzięki nim routery mogą decydować o tym, które dane są bardziej aktualne. Istnieją cztery rodzaje pakietów SNP : a) kompletne pakiety SNP poziomu 1, b) kompletne pakiety SNP poziomu 2, c) częściowe pakiety SNP poziomu 1 i d) częściowe pakiety SNP poziomu 2. Częściowe pakiety SNP zawierają listę aktualnych numerów sekwencyjnych pakietów LSP i używane są jako potwierdzenia pakietów LSP. Od standardowych pakietów potwierdzających różnią się tym, że jednocześnie mogą one potwierdzać wiele pakietów LSP[28]. Kompletne pakiety SNP zawierają listę wszystkich aktualnych numerów sekwencyjnych zgromadzonych w bazie danych routera. Używane są do synchronizacji zawartości baz danych przyległych routerów zarówno podczas okresowego sprawdzania identyczności baz jak i podczas początkowego nawiązywania połączenia między przyległymi routerami.

12.3 Metryka

Podobnie jak w przypadku protokołu OSPF w IS-IS wyróżniamy dwa rodzaje metryki: zewnętrzną i wewnętrzną[28]. Całkowity koszt trasy od nadawcy do odbiorcy obliczany jest jako suma wartości metryk poszczególnych odcinków składających się na całą ścieżkę. Wartość metryki nie może przekroczyć 63. Jeśli wynik sumowania jest większy od 63, to przyjmuje się, że metryka jest równa 63. Z ogłaszanymi trasami, przesyłanymi w pakietach LSP, może być związanych kilka metryk, określających stopień preferencji łącza w zależności od różnorakich parametrów, takich jak opóźnienie czy pewność. Dodatkowe metryki stosuje się w odniesieniu do TOS (type of service). Metryki te są opcjonalne i nie muszą wystąpić w ogłoszeniach LSP, w odróżnieniu od metryki domyślnej, która jest obowiązkowym elementem każdego ogłoszenia trasy[28].

Z trasami zewnętrznymi może być związana bądź to metryka zewnętrzna, bądź wewnętrzna. Jeśli z trasą są związane dwa rodzaje metryki to przy obliczeniach kosztu trasy brana jest wartość metryki wewnętrznej.

12.4 Tablica routingu

Rolę tablicy routingu spełnia w IS-IS baza danych zawierająca stany połączeń w sieci[28]. Na podstawie tej bazy budowane jest graf skierowany służący do wyboru najlepszej trasy. Najlepszą, czyli taką, która posiada najmniejszą wartość metryki. Jeśli cała ścieżka prowadząca od nadawcy do odbiorcy wspiera TOS (type of service), to wykorzystana może być w procesie wyboru trasy metryka odpowiednia danemu TOS. Ze względu na stosowanie tego samego algorytmu (Dijkstry) przez protokoły IS-IS i OSPF proces budowy drzewa i wyboru trasy jest analogiczny (patrz rozdział 10.5).

12.5 Przetwarzanie informacji trasowania

Wymiana informacji o routingu miedzy routerami dokonywana jest przy użyciu pakietów LSP (link state packet). Wyróżniamy dwa rodzaje tych pakietów: poziomu 1 i 2.

Routery poziomu 2 umieszczają w ogłaszanych pakietach LSP poziomu 2 listę obiektów docelowych osiągalnych w jego obszarze[28]. Lista ta składa się z rekordów zawierających: adres IP, maskę podsieci i metrykę. Router poziomu 2 może czerpać informacje o obiektach w obrębie obszaru z pakietów LSP poziomu 1 otrzymanych od wszystkich routerów w obszarze. Jest wysoko pożądane, by router poziomu 2 dokonał skrótu otrzymanych informacji o obszarze i dopiero wtedy przekazał obraz obszaru do innych routerów. Dokonuje się tego poprzez ręczną konfiguracje adresów skróconych. Każdy router poziomu 2 może posiadać jeden lub więcej tak skonfigurowanych adresów postaci [adres IP, maska podsieci, metryka]. Zestaw osiągalnych adresów otrzymany w LSP poziomu 1 jest porównywany z adresami skonfigurowanymi. Powtarzające się trasy nie są umieszczane w pakietach LSP poziomu 2. Generalnie rzecz biorąc, adresy skonfigurowane są adresami mało specyficznymi, co oznacza, że będą one zawierać liczne adresy otrzymane od routerów poziomu pierwszego. Adresy skonfigurowane pojawiają się tylko w LSP poziomu 2. Adres taki zostanie umieszczony w pakiecie LSP, jeśli co najmniej jeden adres jest osiągalny poprzez router ogłaszający. Dla ręcznie konfigurowanych adresów poziomu drugiego określana jest również ręcznie metryka.

Każdy adres uzyskany od routera poziomu 1, który nie zawiera się w adresach skonfigurowanych dołączany jest do listy adresów pakietu LSP poziomu 2. W takim wypadku wartość metryki obliczana jest jako suma wartości metryki uzyskanej od routera poziomu 1, plus koszt trasy łączącej router poziomu 1 i router poziomu 2, który będzie ogłaszał trasę. Wartość metryki nie może przekroczyć 63[22]. Jeśli wynik sumowania jest większy od 63, to przyjmuje się, że metryka jest równa 63. Metryki odnoszące się do różnych rodzajów usług (TOS) będą zawarte w LSP, jeśli:

- router poziomu 2 obsługuje dany rodzaj usługi (TOS),
- ścieżka łącząca router poziomu 2 z odpowiednim routerem poziomu 1 jest utworzona z połączeń wspierających dany rodzaj usługi (TOS),
- router poziomu 1, który potrafi osiągnąć bezpośrednio adres odbiorcy również obsługuje dany rodzaj usługi; router poziomu 1 powiadamia o możliwości wspierania danej usługi w swoich pakietach LSP poziomu 1.

W przypadku otrzymania przez router poziomu 2 tych samych tras od różnych routerów poziomu 1, ogłaszana jest tylko jedna kopia trasy w pakietach LSP poziomu 2 (przyjmując, że adres nie zawiera się w puli adresów skonfigurowanych ręcznie). Metryka ogłaszana z daną trasą jest obliczana jako minimum metryk otrzymanych od routerów poziomu 1[28].

Adresy osiągane bezpośrednio przez routery poziomu drugiego są traktowane dokładnie tak samo jak adresy uzyskane od routerów poziomu 1.

12.5.1 Adresacja routerów

Protokół IS-IS wymaga, aby adresy routerów były adresami OSI[28]. Oznacza to, że adresy pojawiające się w przesyłanych pakietach mają adresy routerów w notacji adresowej OSI. Przykładowo, adresy te wykorzystywane są do określania przynależności routera do danego obszaru. Każdy router musi mieć przypisany adres OSI. Dla routerów IP, adres ten będzie służył wyłącznie dla potrzeb protokołu IS-IS. Adresu routerów OSI i routerów mieszanych są przyznawane przez upoważnioną do tego instytucje i są unikalne w obrębie sieci globalnej.

Adresy OSI dla routerów IP mogą być uzyskane w dwojaki sposób:

- dla sieci, w których występują adresy OSI, lub adresy takie mogą pojawić się w przyszłości, przyznanie adresu OSI routerowi odbywa się w trybie formalnego uzyskania

adresu od odpowiedniego organu nadającego adresy OSI; ten sposób uzyskiwania adresów routerów jest rekomendowany, nawet jeśli sieć jest typową siecią opartą o adresy IP,

- w przypadku sieci używających tylko protokołów TCP/IP może nie być koniecznym uzyskanie adresu od organu nadającego adresy OSI; w zamian stosowany jest odpowiedni algorytm nadawania adresów OSI routerom na podstawie istniejących adresów IP.

12.6 Podejmowanie decyzji o trasowaniu

12.6.1 IS-IS dla OSI

Protokół IS-IS był początkowo zaprojektowany przez ISO do obsługi wyłącznie otoczenia posługującego się pakietami OSI[28]. W tym rozdziale przedstawiony zostanie sposób, w jaki protokół działa w sieciach posługujących się tylko adresami OSI.

IS-IS postrzega całą sieć jako sieć podzieloną na systemu autonomiczne. Protokół działa wewnątrz systemu autonomicznego, a połączenie z resztą sieci realizowane jest poprzez trasy oznaczone jako trasy zewnętrze. Jeśli trasa oznaczona jest jako prowadząca poza system autonomiczny, to nie są tą trasą wysyłane żadne wiadomości obsługi protokołu IS-IS. Informacje zewnętrze są wprowadzane statycznie w postaci numerów sieci, które są osiągalne poprzez daną trasę.

System autonomiczny podzielony jest na obszary, co jest konsekwencją trasowania hierarchicznego, dwupoziomowego stosowanego w IS-IS[22]. Routery poziomu 1 znają topologie wewnątrz swojego obszaru, włączając wszystkie routery i sieci końcowe w obrębie obszaru. Routery te nie posiadają obrazu topologii sieci na zewnątrz obszaru, a cały ruch do odbiorców znajdujących się poza obszarem kierowany jest do routerów poziomu 2 należących do tego samego obszaru. Podobnie routery poziomu 2 znają topologię poziomu drugiego i wiedzą, które adresy są osiągalne poprzez routery poziomu 2. Nie muszą jednak znać wewnętrznego obrazu sieci wewnątrz obszarów. Wyjątkiem jest sytuacja, w której router jest równocześnie routerem poziomu 1 wewnątrz danego obszaru. Routery poziomu 2 są jedynymi routerami uprawnionymi do możliwości wymiany pakietów i informacji o routingu z routerami zlokalizowanymi poza lokalnym systemem autonomicznym[28].

IDP	DSP

AFI	IDI	HO-DSP	ID	SEL

Tabela 41) Adres ISO

Jak przedstawia tabela 41 adresy ISO podzielone są na części: IDP (Initial Domain Part), i DSP (Domain Specific Part). Część IDP jest numerem nadawanym przez organizację OSI i określa format i władze odpowiedzialne za określanie dalszej postaci adresu. Pole DSP jest określane przez władze do tego powołane, identyfikowane numerem IDP. DSP jest dalej podzielone na HO-DSP (High Order Part of DSP), identyfikator systemu (ID), i selektor NSAP – SEL. Para (IDP, HO-DSP) określa jednoznacznie system autonomiczny i obszar w obrębie systemu. Z tego względu omawiana para może być uznana jako adres obszaru.

Przeważnie, wszystkie routery w obszarze posiadają pojedynczy, taki sam adres obszaru[28]. Jednakże, w pewnych sytuacjach pożądane jest posiadanie przez obszar kilku adresów. Oto przykłady takich sytuacji:

- Zmiana adresu A obszaru na adres B. W takim przypadku najefektywniej będzie początkowo korzystać z obu adresów A i B, a następnie, gdy już wszystkie routery w obrębie obszaru dowiedzą się o istnieniu dwóch adresów, wycofać adres A.
- Połączenie dwóch obszarów A i B w jeden obszar. Dokonuje się tego poprzez dodanie wiedzy o adresie B do obszaru A i odwrotnie.
- Podzielenie obszaru C na dwa obszary A i B. Można tego dokonać przypisując do żądanej części routerów adres A, a do pozostałych routerów dodatkowy adres B. Po pewnym czasie wycofujemy z użycia adres C, co powoduje utworzenie dwóch oddzielnych obszarów, o adresach A i B.

Ze względu na omówiony podział systemu autonomicznego na obszary i adresacje tych obszarów jest bardzo łatwo zdecydować routerom poziomu 1, które pakiety nie należą do lokalnego obszaru i powinny być przekazane routerom poziomu drugiego.

Router poziomu 1 posiada przypisany przez administratora adres obszaru. W przypadku otrzymania propozycji od innego routera, aby stać się jego routerem sąsiednim, to propozycja taka zostanie przyjęta tylko pod warunkiem, że adresy obszarów pokrywają się.

Routery poziomu 2 natomiast akceptują routery jako sąsiednie bez względu na adres obszaru. W przypadku, gdy adresy obszarów nie pokrywają się połączenie między routerami jest połączeniem wyłącznie poziomu 2 i wymieniane będą między tymi routerami wiadomości tylko poziomu 2.

IS-IS umożliwia naprawę podzielonych obszarów[28]. W przypadku podziału (poprzez zerwanie połączenia między routerami) obszaru, jeśli funkcja naprawcza jest zaimplementowana, połączenie między podzielonymi częściami obszaru następuje poprzez routery poziomu 2. Protokół wymaga jednak, by routery poziomu 2 były połączone ze sobą. W przypadku utraty łączności między routerami poziomu 2 następuje podział szkieletu systemu autonomicznego, a w konsekwencji brak łączności między podzielonymi częściami systemu.

Podział systemu autonomicznego na obszary został dokładnie omówiony podczas omawiania protokołu OSPF (rozdział 10.6). Cały mechanizm współdziałania routerów w obrębie systemu autonomicznego podzielonego na obszary jest w IS-IS bardzo podobny. W przypadku, gdy pojedynczy router poziomu 2 utraci połączenie ze szkieletem systemu router ten ogłasza zaistniały fakt w wiadomościach LSP poziomu 1. Router poziomu 1 analizuje otrzymaną wiadomość i kieruje pakiety do innego routera poziomu 2, z którym ma połączenie.

Specjalnie traktowane są sieci rozgłoszeniowe, przez co możliwe jest uniknięcie dwóch rodzajów sytuacji: każdy router w sieci ogłasza połączenie do każdego innego routera w sieci, co powoduje n^2 ogłoszeń połączeń; każdy router zgłasza identyczną listę systemów końcowych w sieci. Te problemy rozwiązuje się poprzez wprowadzenie pseudowęzłów reprezentujących sieć rozgłoszeniową. Zasady działania pseudowęzłów omówione zostały podczas opisu protokołu NLSP (patrz rozdział 11.4.2).

12.6.2 Zintegrowany IS-IS

Zintegrowany protokół IS-IS, jak już wspomniano wcześniej, pozwala na trasowanie pakietów IP oraz OSI[28]. Oznacza to, że ten sam poziom 2 będzie używany w odniesieniu do IP i OSI. Każdy obszar będzie zdefiniowany jako: obszar tylko z ruchem pakietów IP, obszar tylko z ruchem pakietów OSI, lub obszar mieszany z ruchem pakietów IP i OSI. Taki podział

obszarów nie pozwala na zachodzenie na siebie obszarów używających tylko IP i obszarów z pakietami OSI. Oznacza to, że routery nie mogą należeć równocześnie do dwóch takich obszarów.

W obrębie systemu autonomicznego istnieje jeden szkielet, wspólny dla wszystkich obszarów.

Ruch pakietów oraz przekazywanie informacji o trasowaniu między routerami poziomu pierwszego i drugiego w obszarach IP i mieszanych przebiega identycznie jak w obszarach OSI, omówionych wcześniej.

Struktura adresów IP pozwala dzielić sieci na podsieci. Nie jest pożądane, by wymagać od adresów podsieci IP jakichkolwiek specyficznych powiązań z adresami obszarów OSI. Dla przykładu, w wielu przypadkach mieszane routery mieszane mogą być instalowane o otoczeniu, gdzie adresy IP i/lub OSI są już przypisane. Z tego względu specyfikacja IS-IS nie wymaga istnienia relacji między numerem sieci IP a strukturą obszaru. Oznacza to, że adresy IP mogą być przypisane zupełnie niezależnie od adresacji OSI i rozmiaru obszaru. Niemniej jednak, wyższą skuteczność trasowania uzyskuje się, gdy adresy IP obszaru są związane ze strukturą obszaru, na przykład, gdy cały obszar posiada ten sam numer sieci IP.

Zintegrowany IS-IS wyróżnia trzy rodzaje routingu: IP, OSI i mieszany. W systemach autonomicznych z routingiem IP wszystkie routery muszą być w stanie przekazywać ruch pakietów IP. Routery IP mogą się komunikować z routerami mieszanymi. Specyficzne pola odnoszące się do ruchu OSI mogą być dodawane przez mieszane routery w komunikatach do routerów IP. Routery IP ignorują te pola. W takich systemach trasowany będzie jedynie ruch pakietów IP, a pakiety OSI będą odrzucane.

W systemach OSI wszystkie routery muszą obsługiwać ruch pakietów OSI. Routery OSI mogą nawiązywać połączenia z routerami mieszanymi[28]. Routery mogą dodawać w ogłaszanych komunikatach pewne specyficzne pola dla ruchu IP, które to pola są ignorowane przez routery OSI. Tylko ruch pakietów OSI będzie obsługiwany. Pakiety IP będą odrzucane.

Systemy autonomiczne mieszane mogą składać się z routerów IP, OSI i mieszanych.

Podobnie, każdy obszar może być zdefiniowany jako obszar z trasowaniem: IP, OSI lub mieszanym. W obszarach IP routery IP mogą komunikować się z routerami mieszanymi, a ruch przekazywany przez routery poziomu 1 ogranicza się do ruchu pakietów IP. W obszarach OSI routery OSI mogą nawiązywać kontakt z routerami mieszanymi, a routery poziomu 1 trasują tylko pakiety OSI. W obszarach mieszanych trasowany jest ruch pakietów IP i OSI. W przypadku, gdy w obrębie systemu autonomicznego mieszanego mają być trasowane pakiety IP i OSI, to wszystkie routery poziomy 2 muszą być routerami mieszanymi[28].

Podejmowanie decyzji o wyborze najlepszej ścieżki prowadzącej do danego obiektu docelowego odbywa się na podstawie drzewa opisującego topologie sieci. Drzewo połączeń w sieci tworzone jest przy pomocy algorytmu Dijkstr'y. Algorytm ten nie uzależnia swojego działania od rodzaju stosowanych adresów, dlatego może być stosowany zarówno w odniesieniu do notacji OSI jak i notacji TCP/IP. Szczegóły dotyczące budowy drzewa i podejmowania decyzji na jego podstawie zostały szczegółowo omówione w rozdziale poświęconym protokołowi OSPF.

13 Protokół BGP

13.1 Wstęp

Protokół BGP (Border Gateway Protocol) wykonuje routing międzydomenowy w sieciach pracujących z protokołem TCP/IP[30]. Należy do klasy protokołów zewnętrznych, bezklasowych. Wykonuje routing pomiędzy wieloma systemami autonomicznymi (domenami) i wymienia informacje o routingu i dostępności z innymi systemami posługującymi się protokołem BGP. Protokół BGP efektywnie rozwiązuje problemy związane z routingiem międzydomenowym oraz skalowaniem sieci Internet. BGP jest protokołem wektora ścieżki (path vector), co oznacza, że wraz z przesyłaną informacją umieszczane (dopisywane) są również identyfikatory poszczególnych systemów autonomicznych na ścieżce dystrybucji trasy. Pozwala to w prosty sposób zapobiegać powstawaniu pętli w tablicach.

Protokół BGP wykonuje trzy typy routingu:

- wewnątrz systemów autonomicznych - między dwoma lub większą liczbą routerów BGP zlokalizowanych w jednym systemie autonomicznym, na przykład w przedsiębiorstwie, uczelni lub u jednego dostawcy usług internetowych,
- na zewnątrz systemów autonomicznych - między dwoma lub większą liczbą routerów w różnych systemach autonomicznych,
- przez systemy autonomiczne - między dwoma lub większą liczbą routerów BGP, które wymieniają ruch przez system autonomiczny, nie obsługujący protokołu BGP.

Podobnie jak każdy protokół routingu, BGP utrzymuje tablice routingu, przesyła uaktualnienia routingu i podejmuje decyzje o trasie kierowania ruchu, opierając się na miarach routingu. Główną funkcją systemu BGP jest wymiana z innymi systemami BGP informacji o dostępności sieci. Jednym z elementów opisujących trasę prowadzącą do danej sieci jest informacja o systemach autonomicznych, przez które trasa przechodzi[14]. Informacja ta jest niezbędna do konstrukcji grafu połączeń systemów autonomicznych, z którego można eliminować pętle i wprowadzać w życie strategiczne decyzje z poziomu systemów autonomicznych

Routery Cisco obsługują protokół routingu BGP w wersjach 2, 3 i 4 [14].

13.2 Format przesyłanych wiadomości

Za przesyłanie wiadomości jest odpowiedzialny protokół TCP, co gwarantuje niezawodność połączenia. Przetwarzanie wiadomości rozpoczyna się dopiero po całkowitym odebraniu przesyłki. Maksymalny rozmiar przesyłki to 4096 oktetów. Rozmiar ten musi być przestrzegany przez wszystkie implementacje BGP. Najmniejszy pakiet może składać się jedynie z samego nagłówka.

13.2.1 Format nagłówka

Każda wiadomość posiada stałej długości nagłówek, za którym mogą (lecz nie muszą) wystąpić dodatkowe dane, w zależności od typu wiadomości. Format nagłówka przedstawia tabela 42.

Znacznik (16 oktetów)	
Długość (2 oktety)	Typ (1 oktet)

Tabela 42) Format nagłówka wiadomości protokołu BGP

Znacznik (marker) - jest polem o długości 16 oktetów, którego zawartość odbiorca jest w stanie przewidzieć. Jeśli wiadomość jest typu 'otwarcie' (patrz dalej) to pole posiada ustawione bity na wszystkich pozycjach. W przeciwnym wypadku wartość tego pola jest z góry znana odbiorcy, na podstawie odpowiedniego algorytmu, i jest częścią mechanizmu uwierzytelniania. Pole znacznika może być również wykorzystywane do detekcji utraty synchronizacji między parą routerów BGP.

Długość - pole to określa całkowitą długość wiadomości włącznie z nagłówkiem. Długość podana jest w oktetach. Długość ta musi się zawierać w przedziale od 19 do 4096 oktetów. Interpretacja długości wiadomości jest zależna od rodzaju wiadomości.

Typ - to jednooktetowe pole wskazuje na rodzaj przesyłanej wiadomości. Zdefiniowane są następujące typy wiadomości:

1 - OTWARCIE (open)
2 - UAKTUALNIENIE (update)
3 - ZAWIADOMIENIE (notification)
4 - KeepAlive

13.2.2 Format wiadomości OTWARCIE

Po zainicjowaniu i ustaleniu połączenia przez protokół TCP, pierwszą wiadomością wysyłaną przez obie strony jest wiadomość typu OTWARCIE[30]. Gdy wiadomość ta zostanie przyjęta i zaakceptowana, w odpowiedzi odsyłana jest wiadomość typu KeepAlive. Po otrzymaniu potwierdzenia KeepAlive mogą być wymieniane przez strony połączenia wiadomości pozostałych typów.
Wiadomość ta składa się z nagłówka oraz występujących po nim dodatkowych informacji przedstawionych w tabeli 43.

Wersja (1 oktet)	
Numer AS (2 oktety)	Czas wstrzymania (2 oktety)
Identyfikator BGP (4 oktety)	
Dł. param. (1 oktet)	
Parametry opcjonalne (zmienna długość)	

Tabela 43) Format wiadomości OTWARCIE

Wersja - to jednooktetowe pole wskazuje na numer wersji protokołu BGP. Aktualny numer wersji to 4.

Numer AS - określa numer systemu autonomicznego nadawcy wiadomości.

Czas wstrzymania (hold time)- jest to proponowany przez nadawcę wiadomości czas wstrzymania. Podczas przetwarzania wiadomości typu OTWARCIE odbiorca musi określić czas wstrzymania dla połączenia. Czas ten jest brany jako minimum dwóch wartości: czasu lokalnie zdefiniowanego i czasu otrzymanego w wiadomości OTWARCIE. Czas ten musi przyjmować albo wartość zero, albo co najmniej wartość trzech sekund. Określony czas wstrzymania jest maksymalną ilością sekund, jaka może upłynąć pomiędzy otrzymaniem od nadawcy prawidłowych wiadomości typu KeepAlive i/lub UAKTUALNIENIE[30].

Identyfikator BGP - jest adres IP przypisany do nadawcy wiadomości. Wartość identyfikatora jest ustalana przy starcie routera i jest taka sama dla każdego interfejsu[9].

Długość parametrów - jest to całkowita liczba dodatnia określająca w oktetach całkowitą długość pola parametrów opcjonalnych. Jeśli przyjmuje wartość zero, to oznacza to, że parametry opcjonalne nie występują.

Parametry opcjonalne - pole to może zawierać listę parametrów opcjonalnych przesyłanych w wiadomości. Każdy parametr opisują trzy pola: typ, długość i wartości parametru. Pole typu to jednooktetowe pole, które jednoznacznie określa rodzaj parametru. Pole długości, również jednooktetowe, wskazuje na długość pola parametrów. Pole parametrów natomiast to zmiennej długości pole zawierające dane interpretowane w odniesieniu do typu parametru opcjonalnego[30]. Parametrem opcjonalnym może być przykładowo uwierzytelnianie.

13.2.3 Format wiadomości UAKTUALNIENIE

Wiadomości tego typu służą przesyłaniu między połączonymi routerami informacji trasowania. Informacje dostarczone przez UAKTUALNIENIE są wykorzystywane do konstrukcji grafu opisującego relacje między systemami autonomicznymi. Dzięki zastosowaniu odpowiednich reguł przy konstrukcji grafu wykrywane są pętle i inne anomalie, które następnie są usuwane, aby nie miały negatywnego wpływu na routing między systemami autonomicznymi. W wiadomościach tego typu przesyłane są informacje o użytecznych trasach, lub też dane informujące, że dana trasa nie jest już aktualna i powinna być usunięta z tablicy routingu[30]. Wiadomość UAKTUALNIENIE zawiera nagłówek stałej długości, po którym mogą wystąpić dodatkowe, specyficzne dla tego typu wiadomości pola. Układ pól przedstawia tabela 44.

Długość tras nieużytecznych (2 oktety)
Trasy nieużyteczne (zmienna długość)
Długość atrybutów (2 oktety)
Atrybuty (zmienna długość)
Informacje dostępności sieci (zmienna dł)

Tabela 44) Format wiadomości UAKTUALNIENIE

Długość tras nieużytecznych - dwuoktetowe pole określające w oktetach całkowitą długość pola tras nieużytecznych. Wartość tego pola musi pozwalać na prawidłowe określenie długości pola opisującego osiągalną sieć (patrz dalej).
Jeśli wartość tego pola to 0, to żadna trasa nie jest zawarta w sekcji tras nieużytecznych.

Trasy nieużyteczne - to zmiennej długości pole zawiera listę rekordów opisujących obiekty, które są już nieaktualne i nie powinny być brane pod uwagę w procesie trasowania.

Rekord skład się z dwóch pól opisujących obiekt. Ponieważ obiektami są sieci, tak więc dane, jakie potrzebujemy do określenia obiektu jest numer sieci. Pierwsze pole rekordu to długość, która określa liczbę bitów numeru sieci. Drugie pole to numer sieci. Jeśli długość numeru sieci przyjmuje wartość zero, to oznacza to trasę domyślną.

Długość atrybutów - jest to całkowita długość pola atrybutów ścieżki określona w oktetach. Wartość zero wskazuje na nieobecność w wiadomości informacji dotyczących dostępnej trasy.

Atrybuty - jest to lista rekordów opisujących trasę. Każdy rekord posiada trzy składowe: typ, długość i wartość atrybutu. Jak widać ze składowych, rekord jest zmiennej długości.

Typ atrybutu to dwuoktetowe pole składające się z flag atrybutu i kodu atrybutu (tabela 45)

Bity:	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Flagi atrybutu (1 oktet)	Kod atrybutu (1 oktet)

Tabela 45) Format pola „typ atrybutu”

Opis:

- Bit 0 pola flag określa czy atrybut jest opcjonalny (wartość 1) czy też obowiązkowy (wartość 0).
- Bit 1 pola flag jest bitem przechodniości. Dla opcjonalnych atrybutów określa on czy atrybut powinien być przekazany dalej (ustawiony na 1) czy też nie powinien być dalej przekazywany (ustawiony na 0). Dla atrybutów obowiązkowych, bit ten musi mieć wartość 1.
- Bit 2 definiuje, czy informacja zawarta w opcjonalnym, przechodnim atrybucie jest kompletna (0) czy też nie (1). Dla obowiązkowych atrybutów i atrybutów nieprzechodnich bit ten musi być wyzerowany.
- Bit 3 jest bitem rozszerzenia długości. Jeśli ustawiony oznacza to, że pole długości atrybutu zajmuje dwa oktety, jeśli bit przyjmuje wartość zero to pole długości atrybutu ma długość jednego oktetu. Bit może przyjąć wartość 1 tylko wtedy, gdy długość atrybutu przekracza 255 oktetów.
- Cztery najmniej znaczące bity pola flag są nieużywane i muszą być wyzerowane.
- Kody atrybutów opisane są w dalszej części pracy (patrz 13.3).

Po typie atrybutu występuje pole długości zajmujące jeden lub dwa oktety (w zależności od bitu 3).

Pozostałe pola atrybutu trasy zawierają dane specyficzne dla danego atrybutu, jego kodu i ustawienia flag.

Informacje dostępności sieci - to zmiennej długości pole zawiera listę numerów sieci IP[30]. Długość tego pola nie jest oddzielnie kodowana, lecz może zostać wyznaczona na podstawie wzoru:

długość wiadomości - 23 - długość tras nieużytecznych - długość atrybutów

Informacja o dostępności sieci jest w postaci jednej lub wielu par złożonych z pola długości i pola numeru sieci IP.

13.2.4 Format KeepAlive

BGP nie używa mechanizmu opartego na warstwie transportowej do określania czy router sąsiedni jest osiągalny[30]. Zamiast tego używane są wiadomości typu KeepAlive. Wymiana tych wiadomości pomiędzy sąsiadującymi routerami odbywa się na tyle często, że nie powoduje wygaśnięcia liczników wstrzymania związanych z sąsiadami. Przyjmuje się, że maksymalnym czasem wysyłania wiadomości KeepAlive do sąsiadów jest czas będący jedną trzecią czasu wstrzymania. Wiadomości KeepAlive nie mogą być jednak wysyłane częściej niż raz na sekundę. W przypadku, gdy wynegocjowany podczas zestawiania połączenia czas wstrzymania ma wartość zero, wiadomości KeepAlive nie mogą być wysyłane do tego routera.

Wiadomości KeepAlive składają się tylko i wyłącznie z nagłówka i mają długość 19 oktetów.

13.2.5 Format wiadomości ZAWIADOMIENIE

Wiadomość ZAWIADOMIENIE jest wysyłana w przypadku wystąpienia błędu. Po wysłaniu ZAWIADOMIENIA połączenie BGP jest zamykane.
Po nagłówku następują pola:

Kod błędu (1 oktet)	Szczegółowy kod (1)
Dane (zmienna długość)	

Tabela 46) Format wiadomości ZAWIADOMIENIE

Kod błędu - to jednooktetowe pole wskazujące na rodzaj błędu, który został wykryty.

Szczegółowy kod błędy - to bardziej dokładny opis błędy zajmujący jeden oktet. Interpretacja pola zależy od rodzaju błędu.
Dane – interpretacja pola zależy od rodzaju błędu i kodu błędu szczegółowego.

13.3 Atrybuty ścieżek

Rozdział ten opisuje atrybuty ścieżek stosowane w wiadomościach UAKTUALNIENIE. Zawarte w wiadomościach trasy posiadają przypisane atrybuty, na podstawie których można określić najlepszą trasę spośród wielu tras prowadzących do tego samego obiektu docelowego[30]. Atrybuty te można podzielić na cztery kategorie:

1 - Powszechne obowiązkowe (well-known mandatory)
2 - Powszechne dowolne (well-known discretionary)
3 - Opcjonalne przechodnie (optional transitive)
4 - Opcjonalne nieprzechodnie (optional non-transitive)

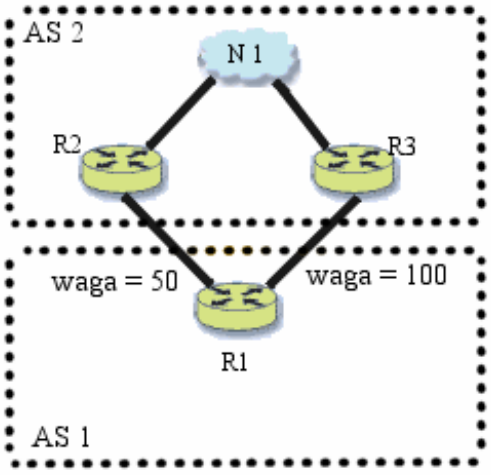
Atrybuty powszechne muszą być rozpoznawalne przez wszystkie implementacje BGP, a te z nich, które są obowiązkowe, muszą znaleźć się w wiadomościach UAKTUALNIENIE. Dodatkowo, atrybuty powszechne muszą być przekazywane przez router (po odpowiedniej zmianie, jeśli to konieczne) do sąsiednich routerów BGP.

W wiadomościach UAKTUALNIENIE mogą znaleźć się również atrybuty opcjonalne. Nie jest przy tym wymagane, by wszystkie implementacje obsługiwały te atrybuty. Sposób postępowania z nierozpoznanymi, opcjonalnymi atrybutami zależy od ustawienia bitu przechodniości w oktecie flag atrybutu[9].

Ścieżka z nierozpoznanym, przechodnim, opcjonalnym atrybutem powinna być zaakceptowana. Jeśli ścieżka taka została zaakceptowana i przekazana do sąsiedniego routera BGP, to wraz ze ścieżką musi być przekazany nierozpoznany, przechodni, opcjonalny atrybut i dodatkowo musi zostać ustawiony bit drugi pola flagi. Jeśli akceptowana jest ścieżka z rozpoznanym przechodnim, opcjonalnym atrybutem, to przekazując tą ścieżkę do sąsiedniego routera BGP przekazuje również wartość bitu 2 flagi. Oznacza to, że jeśli bit ten był ustawiony na wartość 1 przez któryś z poprzednich routerów na ścieżce, to nie jest on ustawiany ponownie na zero przez bieżący system autonomiczny. Nierozpoznane nieprzechodnie, opcjonalne atrybuty muszą być zignorowane i nie są przekazywane innym routerom BGP.
Aktualna implementacja Cisco nie generuje atrybutów przechodnich opcjonalnych [14].

13.3.1 Atrybut WEIGHT

Jest to atrybut wagi nie ujęty w dokumentacji protokołu BGP, ale wprowadzony w implementacji Cisco protokołu[7]. Atrybut ten używany jest lokalnie przez router i nie jest ogłaszany do sąsiednich routerów. Atrybut ten jest przypisywany do trasy w momencie otrzymania trasy od sąsiedniego routera. Jeśli istnieje wiele tras prowadzących do tego samego obiektu docelowego, to wybrana zostanie trasa o największej wartości tego atrybutu. Na rysunku 11 router R1 otrzymuje ogłoszenie o sieci N1 od routera R2 i R3. W momencie otrzymania ogłoszenia o trasie od routera R2, przypisywana jest waga równa 50. Trasie otrzymanej od routera R3 natomiast przypisywana jest waga równa 100. Obie trasy, prowadzące do sieci N1, zostaną umieszczone w bazach danych, z przypisanymi im wagami. W bazie lokalnej zostanie zainstalowana trasa o większej wadze.



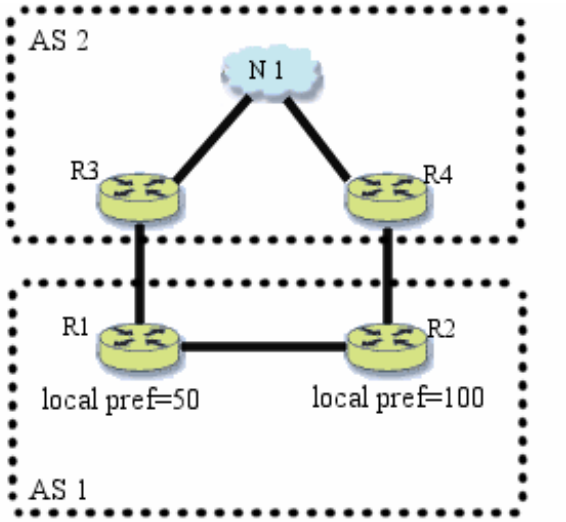
Rysunek 11) Atrybut WEIGHT

13.3.2 Atrybut LOCAL_PREF

Ten atrybut zaliczamy do drugiej kategorii[30]. Powinien się on znaleźć w UAKTUALNIENIACH, które wysyłane są do routerów znajdujących się w tym samym systemie autonomicznym co nadawca wiadomości. Router ogłaszający trasy do swoich sąsiadów znajdujących się w tym samym systemie autonomicznym powinien oszacować stopień użyteczności tras zewnętrznych (na zewnątrz systemu autonomicznego) i zawrzeć wynik oszacowania w

polu omawianego atrybutu. Router BGP powinien wykorzystać dane uzyskane z tego pola w procesie decyzji (patrz rozdział 13.7.3).

Inaczej mówiąc, jest to atrybut określający preferowany punkt wyjścia z lokalnego systemu autonomicznego[7]. Gdy istnieje kilka punktów wyjścia z systemu autonomicznego, to wartość tego atrybutu jest decydująca w procesie wyboru drogi dla pakietów kierowanych poza lokalny system. Na rysunku 12, system autonomiczny AS1 otrzymuje dwa ogłoszenia o sieci N1 od systemu autonomicznego AS2. Gdy router R1 otrzymuje ogłoszenie o tejże sieci, odpowiadająca wartość atrybutu trasy prowadzącej do tej sieci jest równa 50. W przypadku, gdy router R2 otrzymuje informacje o trasie prowadzącej do sieci N1, to przypisany tej trasie jest atrybut LOCAL_PREF o wartości 100. Otrzymane trasy i przypisane im wartości atrybutu zostaną wymienione o obrębie systemu autonomicznego AS1, czyli między routerami R1 i R2. Ponieważ trasy prowadzące przez R2 posiadać będą większą wartość omawianego atrybutu, to ruch do sieci N1 (znajdującej się poza lokalnym systemem autonomicznym) będzie kierowany przez router R2[7].

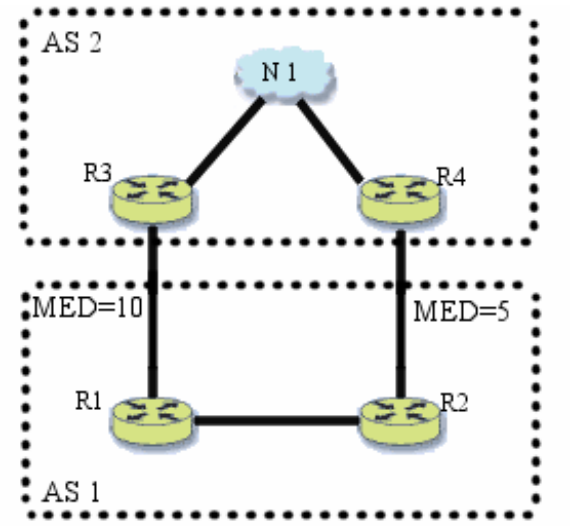


Rysunek 12) Atrybut LOCAL_PREF

Atrybut ten nie powinien się zawrzeć w wiadomościach ogłaszanych do sąsiednich systemów autonomicznych[7]. W przypadku otrzymania wiadomości z tym atrybutem od routera znajdującego się w sąsiednim systemie, atrybut powinien być zignorowany.

13.3.3 Atrybut MULTI_EXIT_DISC (MED)

Pole tego atrybutu przechowuje czterooktetową liczbę bez znaku będącą metryką[30]. W przypadku wielu tras prowadzących do tego samego obiektu, wartość tego atrybutu może mieć wpływ na wybór jednej z tras. Atrybut ten może, lecz nie musi zostać użyty do wyboru najlepszej trasy, gdyż jednoznaczne wyznaczenie najbardziej preferowanej trasy może nastąpić na podstawie innych atrybutów. Wartość tego atrybutu ogłaszana jest w całym systemie autonomicznym[7]. Na rysunku 13, router R3 ogłasza sieć N1 z metryką równą 10, podczas gdy router R4 ogłasza tą samą sieć z metryką równą 5. Preferowana jest trasa o mniejszej metryce, dlatego system autonomiczny AS1 będzie wybierał router R4 do przesyłania pakietów, których odbiorca znajduje się w sieci N1 systemu autonomicznego AS2.



Rysunek 13) Atrybut MULTI_EXIT_DISC

Wartość atrybutu MED jest przypisywana w obrębie systemu autonomicznego zgodnie z ustalonymi regułami polityki. Atrybut ten informuje, który punkt wejściowy do systemu autonomicznego jest preferowany przez tenże system. Ponieważ wartość tego atrybutu zależy tylko od lokalnie zdefiniowanych reguł, to porównywanie atrybutów MED pochodzących z różnych systemów autonomicznych nie ma sensu[30].

13.3.4 Atrybut ORIGIN

Jest to powszechny obowiązkowy atrybut, który określa pochodzenie informacji o trasie[30]. Atrybut ten generowany jest przez system autonomiczny generujący pakiet UAKTUALNIENIA. Wartość tego atrybutu jest brana pod uwagę w czasie podejmowania decyzji o wyborze trasy [7]. Pole tego atrybutu może przyjmować trzy wartości:

IGP – trasa jest trasą wewnętrzną w stosunku do ogłaszającego ją systemu autonomicznego. Wartość ta jest używana, gdy trasa zostaje włączona do bazy informacji BGP przy pomocy polecenia konfiguracyjnego routera.

BGP – trasa otrzymana została poprzez protokół BGP

Nieznany – trasa otrzymana została z nieznanego źródła lub w inny sposób. Taka wartość przypisywana jest w przypadku otrzymania informacji o trasie od innych protokołów routingu.

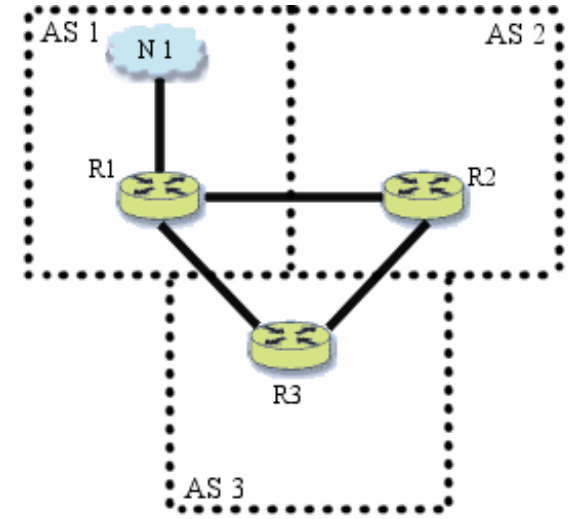
13.3.5 Atrybut AS_PATH

Atrybut należy do kategorii atrybutów powszechnych obowiązkowych [30]. Określa on systemy autonomiczne, przez które informacja zawarta w wiadomości była przesyłana. Router BGP przesyłający wiadomość do innych routerów jest zobowiązany do modyfikacji tego atrybutu w zależności od lokalizacji odbiorcy.

Gdy ogłoszenie o trasie przechodzi przez systemy autonomiczne, każdorazowo dodawany jest do uporządkowanej listy numer systemu autonomicznego[7]. Aby lepiej zobrazować przebieg wypełniania atrybutu AS_PATH posłużmy się przykładem. Rysunek 14 pokazuje sytuacje, w której informacja o trasie przechodzi przez trzy systemy autonomiczne. System autonomiczny AS1 jest twórcą ogłoszenia o trasie prowadzącej do sieci N1 i ogłoszenie

to kierowane jest do AS2 i AS3. W tym momencie lista atrybutu AS_PATH zawiera tylko jedną wartość {1}, czyli numer systemu ogłaszającego. System autonomiczny AS3 ogłosi trasę z powrotem do AS1 z atrybutem zawierającym już dwie wartości {3, 1}. Również AS2 w analogiczny sposób ogłosi trasę do AS1 z atrybutem {2, 1}. System autonomiczny AS1 odrzuci te trasy, gdyż numer przypisany temu systemowi autonomicznemu (czyli numer 1) pojawia się na liście numerów atrybutu AS_PATH. W ten sposób BGP radzi sobie z problemem powstawania pętli w topologii[7].

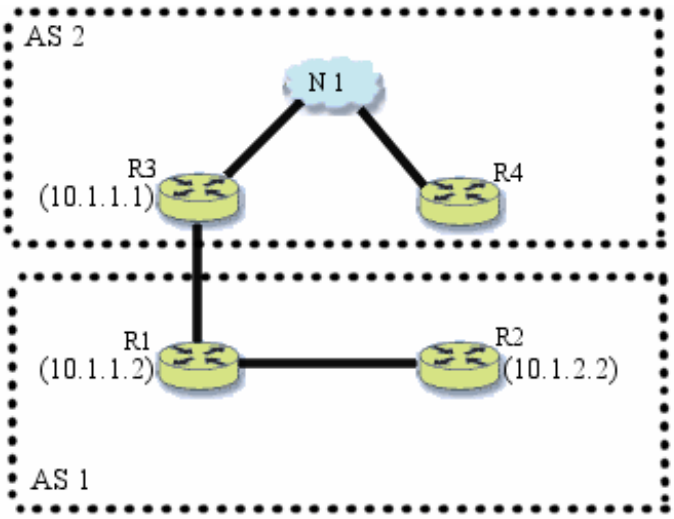
Systemy autonomiczne AS2 i AS3 ogłaszają trasę również pomiędzy sobą, odpowiednio modyfikując listę atrybutu AS_PATH. Trasy te jednak nie zostaną zainstalowane w bazach lokalnych (ale w bazach wejściowych oczywiście tak), ponieważ każdy z routerów posiada już w swojej bazie trasy (uzyskane bezpośrednio od AS1) o krótszej liście atrybutu AS_PATH[7].



Rysunek 14) Atrybut AS_PATH

13.3.6 Atrybut NEXT_HOP

NEXT_HOP jest atrybutem powszechnym obowiązkowym, który określa adres IP routera brzegowego, który powinien być użyty jako następny skok na trasie do obiektu docelowego zamieszczonego w wiadomości[30]. Oznacza to, że jest to adres routera w sąsiednim systemie autonomicznym. Atrybut ten nie jest zmieniany, gdy ogłaszana jest trasa do routera wewnątrz lokalnego systemu autonomicznego. Router BGP nie może nigdy ogłaszać tras, których atrybut NEXT_HOP wskazuje na router będący odbiorcą wiadomości. Nie wolno również routerowi umieszczać w tablicy trasowania ścieżek, których atrybut NEXT_HOP wskazuje na siebie samego[7]. Poniższy rysunek obrazuje użycie atrybutu.



Rysunek 15) Atrybut NEXT_HOP

Router R3 ogłasza trasę do sieci N1 z atrybutem następnego skoku ustawionym na wartość 10.1.1.1. Gdy router R1 ogłasza tą trasę wewnątrz swojego lokalnego systemu autonomicznego, to wartość atrybutu NEXT_HOP się nie zmienia. Jeśli router R2 nie posiada w swojej bazie informacji pozwalającej mu dostać się do podanego adresu 10.1.1.1, to trasa zostaje odrzucona[7].

13.3.7 Atrybut ATOMIC_AGGREGATE

Jest to atrybut kategorii drugiej[30]. Atrybut ten jest dołączany do listy atrybutów w sytuacji, gdy router wybiera spośród możliwych tras prowadzących do obiektu docelowego, trasę, która nie posiada największego stopnia preferencji. Atrybut ten nie powinien być usuwany podczas przekazywania wiadomości do dalszych routerów[7]. Router poprzez otrzymanie atrybutu informowany może być również o fakcie, że wiadomość przeszła przez systemy autonomiczne nie wyszczególnione w AS_PATH.

13.4 Bazy informacji o trasach

Na podstawie informacji o trasach zgromadzonych w bazie routera podejmowane są wszelkie czynności związane z trasowaniem [30]. Cała baza jest po prostu tablicą routingu routera. Baza składa się z trzech oddzielnych części:

- wejściowej - w tej bazie przechowywane są informacje o routingu, które zawarte były w przybyłych wiadomościach UAKTUALNIENIA. Informacje te służą jako dane wejściowe dla procesu decyzji (patrz 13.7.3).
- lokalna - zawiera lokalne, czyli stosowane przez router, informacje o routingu, które zostały wybrane spośród informacji zawartych w bazie wejściowej poprzez zastosowanie odpowiednich dla routera reguł.
- wyjściowa - w tej bazie zawarte są te trasy, które zostały wybrane przez router do ogłoszenia do sąsiednich routerów. Trasy te będą przekazane do sąsiednich routerów za pomocą wiadomości UAKTUALNIENIA.

Mimo tego, że specyfikacja protokołu wyróżnia trzy części bazy, nie wymusza to na implementacji protokołu utrzymywania trzech oddzielnych kopii informacji o routingu. Wy-

bór rozwiązania, jakie zostanie zastosowane (na przykład, trzy kopie informacji lub jedna kopia ze wskaźnikami) nie jest opisane protokołem.

13.5 Analiza atrybutów ścieżek

Router może posiadać w swoich bazach wejściowych wiele tras prowadzących do jednego obiektu docelowego[7]. Spośród tych tras musi zostać wybrana jedna trasa o największym stopniu preferencji. Wyboru tej jednej trasy dokonuje się o oparciu o wartości atrybutów. Implementacja protokołu BGP firmy Cisco stosuje następujące kroki przy wyborze trasy [14] :

- Jeśli atrybut NEXT_HOP wskazuje na adres nieosiągalny, to odrzuć trasę.
- Wybierz trasę o największej wadze.
- Jeśli wagi są równe, wybierz trasę o największej wartości atrybutu LOCAL_PREF.
- Jeśli wartości atrybutu LOCAL_PREF są równe, wybierz trasę, która została wygenerowana przez protokół BGP uruchomiony na tym routerze.
- Wybierz trasę o najkrótszej liście atrybutu AS_PATH.
- Wybierz trasę o najmniejszej wartości parametru ORIGIN (gdzie wartość IGP jest mniejsza od EGP, a EGP jest mniejsza od Nieznany).
- Jeśli nadal istnieje kilka tras, wybierz trasę posiadającą mniejszą wartość atrybutu MULTI_EXIT_DISC.
- Wybierz trasę zewnętrzną.
- Jeśli wciąż nie ma jednoznaczności, wybierz trasę przechodzącą przez najbliższy router sąsiedni IGP.
- Wybierz trasę ogłoszoną przez router o najmniejszym adresie IP.

Implementacja Cisco pozwala na zmianę wielu ustawień routera, w tym również parametrów dotyczących atrybutów ścieżek przypisywanych do tras [14]. Należy pamiętać, że po dokonaniu zmian konieczne jest ponowne ustanowienie połączeń z sąsiednimi routerami. Dopiero wtedy, zmiany zostaną uwzględnione.

13.6 Liczniki i stałe czasowe

Podobnie jak poprzednio opisane protokoły, również BGP posiada zestaw parametrów czasowych, stosowanych w procesie wymiany informacji o trasowaniu [30]. Oto parametry używane przez protokół BGP:

i) czas oczekiwania na połączenie (ConnectRetry)
ii) czas wstrzymania (hold time)
iii) minimalny odstęp ogłaszania tras (MinRouteAdvertisementInterval)
iv) minimalny odstęp lokalnego systemu autonomicznego (MinASOriginationInterval)
v) odstęp wysyłania wiadomości KeepAlive

Ad i) Używany podczas początkowego nawiązywania połączenia pomiędzy routerami przy pomocy protokołu TCP. Jeśli w tym czasie nie nastąpi nawiązanie połączenia router próbuje nawiązać połączenie z innym routerem. Sugerowana wartość parametru to 120 sekund[2].

Ad ii) Patrz rozdział 13.2. Sugerowana wartość parametru to 90 sekund.

Ad iii) Parametr określa minimalną ilość czasu, jaka musi upłynąć pomiędzy ogłoszeniami tras prowadzących do określonej grupy obiektów pochodzących od pojedynczego routera BGP. Inaczej mówiąc, dwie wiadomości AKTUALIZACYJNE, pochodzące od tego samego routera, znajdującego się w sąsiednim systemie autonomicznym, ogłaszającego ważne trasy odnoszące się do wspólnego zestawu obiektów docelowych, muszą być oddzielone od siebie odstępem czasowym co najmniej równym wartości tego parametru czasowego. Takie zadanie może być spełnione, gdy z każdym zestawem obiektów docelowych związany jest licznik czasowy[30].
W związku z potrzebą szybkiej zbieżności w obrębie systemu autonomicznego, licznik ten nie ma zastosowania do tras otrzymanych od routerów znajdujących się w tym samym systemie autonomicznym. Aby uniknąć długo pozostających w systemie dziur, opisana procedura nie odnosi się również do wycofywanych tras nieważnych (tych znajdujących się w polu tras nieważnych wiadomości AKTUALIZACYJNYCH). Sugerowana wartość parametru to 30 sekund[2].

Ad iv) Ten parametr czasowy wskazuje minimalny odstęp czasu, który mysi być zachowany pomiędzy prawidłowym ogłoszeniem dwóch wiadomości AKTUALIZACYJNYCH ogłaszających zmiany w obrębie lokalnego systemu autonomicznego. Sugerowana wartość parametru to 15 sekund.

Ad v) patrz opis formatu KeepAlive (rozdział 13.2.4). Sugerowana wartość parametru to 30 sekund[30].

Wymienione wartości domyślne odnoszą się do implementacji protokołu na routerach Cisco [14]. Routery firmy Cisco zapewniają możliwość modyfikacji wartości tych parametrów. Jeśli routery posiadają odmienne wartości, to podczas nawiązywania połączenia wartości odpowiednich parametrów są negocjowane.

13.7 Przetwarzanie informacji trasowania

Początkową fazą wymiany informacji przez dwa systemu jest utworzenie połączenia zapewniającego transport danych między tymi systemami [30]. Pierwsze wymieniane wiadomości służą ustaleniu parametrów połączenia. Po negocjacji parametrów i udanym zestawieniu połączenia następuje przesłanie całej zawartości tablicy routingu BGP. Protokół nie wymaga okresowego odświeżania całej zawartości tablicy, ale każdy router ogłaszający musi pamiętać aktualną wersję tablicy BGP dla każdego ustanowionego połączenia z innym routerem. Tablica odpowiadająca danemu sąsiadowi jest utrzymywana w pamięci przez cały czas trwania połączenia. W celu upewnienia się, że połączenie jest aktualne, periodycznie wysyłane są wiadomości KeepAlive. W razie wystąpienia błędu lub w przypadku zaistnienia pewnych specjalnych okoliczności wysyłane są komunikaty zawiadomienia[30]. Jeśli zauważony zostanie błąd w połączeniu, wysłane zostanie zawiadomienie, a samo połączenie zostanie zamknięte.

W implementacji Cisco protokołu istnieje pewne zabezpieczenie, w postaci listy routerów mogących stać się routerami sąsiednimi[14]. Jeśli router sąsiedni próbuje nawiązać połączenie, to połączenie takie zostanie zaakceptowane tylko wtedy, gdy jego adres znajduje się na liście. Jeśli zabezpieczenie to jest wyłączone, to połączenia od wszystkich routerów są akceptowane.

Zmiany topologii sieci są ogłaszane między parami routerów BGP w postaci wiadomości UAKTUALNIENIE[30]. W wiadomościach tych przesyłane są informacje o obiekcie

docelowych i o ścieżce prowadzącej do tego obiektu. Router otrzymujący taką wiadomość analizuje zawarte w niej informacje i na ich podstawie dokonuje zmian w swoich bazach informacji o routingu. Jeśli router ten zdecyduje się na przesłanie UAKTUALNIENIA do innych routerów, to ma on możliwość dodania lub modyfikacji pewnych atrybutów ścieżki przed wysłaniem wiadomości.

BGP dostarcza również mechanizmów, dzięki którym router ogłaszający ma możliwość poinformowania sąsiadów o fakcie, że wcześniej ogłaszana trasa nie jest już aktualna. Istnieją trzy przypadki, w których router uzna trasę za nieprzydatną[30]:

- numer sieci IP opisujący obiekty docelowe znajdzie się w polu tras nieużytecznych wiadomości UAKTUALNIENIE
- trasa zostanie zastąpiona trasą nowszą,
- połączenie z routerem ogłaszającym trasę zostanie zamknięte, co spowoduje usunięcie z bazy informacji o trasach wszystkich tras nauczonych od tego routera.

13.7.1 Analiza wiadomości UAKTUALNIENIE

Pierwszym krokiem w procesie analizy wiadomości UAKTUALNIENIE jest sprawdzenie poprawności poszczególnych pól wiadomości[30]. Następnie, na podstawie zawartych informacji, podejmowane są odpowiednie działania.

Jeśli w wiadomości występuje opcjonalny atrybut nieprzechodni to jest on ignorowany, jeśli natomiast atrybut jest przechodni, to ustawiany jest 3 bit flagi (patrz rozdział 13.2), a sam atrybut jest pozostawiany do przekazania innym routerom BGP.
Jeśli opcjonalny atrybut jest rozpoznany przez router i jego wartość jest prawidłowa to w zależności od rodzaju atrybutu podejmowane są odpowiednie działania. Atrybut jest przetwarzany, pozostawiany, a jeśli istnieje taka potrzeba to również modyfikowany[30].

W przypadku, gdy wiadomość zawiera niepuste pole tras nieużytecznych, to trasy uprzednio wprowadzone do bazy wejściowej powinny być usunięte. Po usunięciu tras z bazy wejściowej router uruchamia proces decyzji (patrz rozdział 13.7.3) w celu weryfikacji nowej zawartości bazy.

Gdy UAKTUALNIENIE zawiera informacje o dostępności sieci, czyli użyteczną trasę, to powinna ona zostać umieszczona w odpowiedniej bazie wejściowej i dodatkowo podjęte powinny być następujące działania[30]:

i) jeśli pole informujące o dostępności sieci, czyli pole zawierające numer sieci IP jest identyczne z odpowiadającym mu polem trasy znajdującej się już w bazie wejściowej, to nowa trasa powinna zastąpić trasę starą. Wobec zmiany w bazie powinien zostać uruchomiony proces decyzji.

ii) jeśli nowa trasa zachodzi na trasę (patrz 13.7.3.1) zawartą już w bazie wejściowej, to router powinien uruchomić proces decyzji ponieważ bardziej specyficzna (patrz 13.7.3) ścieżka powoduje nieużyteczność części ścieżki mniej specyficznej (ścieżka mniej specyficzna zawiera w sobie pulę adresów ścieżki bardziej specyficznej).

iii) gdy nowa trasa, posiadająca identyczne atrybuty co ścieżka już zawarta w bazie, jest bardziej specyficzna nie potrzebne są dodatkowe czynności

iv) jeśli trasa zawiera numer sieci IP, który nie występuje w bazie, to trasa powinna zostać wprowadzona do bazy wejściowej, po czym powinien zostać uruchomiony proces decyzji.
v) jeśli nowa trasa zachodzi na istniejącą trasę mniej specyficzną, to powinien być uruchomiony proces decyzji odnoszący się tylko do obiektów opisanych przez trasę mniej specyficzną.

Implementacja BGP na routerach Cisco zapewnia możliwość wymiany informacji o trasowaniu nie tylko pomiędzy routerami z uruchomionym procesem protokołu BGP, ale również wymianę danych pomiędzy routerami z zaimplementowanymi różnymi protokołami routingu[14]. Istnieje więc możliwość importowania i eksportowania tras.

13.7.2 Polityka routingu

Polityka routingu to zbiór reguł, które służą podejmowaniu decyzji o wyborze tras[27]. Odpowiednia baza przechowuje na routerze wszystkie reguły. Modyfikacja tej bazy pozwala administratorowi systemu autonomicznego na pełną kontrolę trasowania i przepływu informacji. Wymagane jest, aby wszystkie routery w obrębie systemu autonomicznego posługiwały się tymi samymi regułami. Baza reguł składa się z trzech sekcji:

- preferencji tras – reguły te wykorzystywane są w procesie decyzji do oszacowania przydatności trasy; mówią o tym, które z tras baz wejściowych powinny się znaleźć w bazie lokalnej.
- agregacji tras – reguły wyboru tras mających ulec agregacji, jak również sposób agregacji tras. Na podstawie tej sekcji reguł wybierane są te trasy spośród bazy lokalnej, które mają zostać przesłane do sąsiednich routerów w postaci zagregowanej. Odnoszą się te reguły zarówno do tras ogłaszanych na zewnątrz systemu autonomicznego, jak i wewnątrz lokalnego systemu autonomicznego.
- dystrybucji tras – reguły modyfikacji i selekcji tras, które zostaną przesłane do sąsiednich routerów. Kontrolują one ruch w systemie autonomicznym poprzez określenie, które trasy z bazy lokalnej mają się znaleźć w bazach wyjściowych routera.

Reguły buduje się na podstawie z góry określonej składni i semantyki. Budowa reguł nie jest częścią protokołów routingu[18]. Zainteresowanych dokładną metodą budowy reguł odsyłam do[20].

13.7.3 Proces decyzji

Proces decyzji wybiera trasy spośród tras zgromadzonych w bazie wejściowej stosując lokalne dla routera reguły[30]. Rezultatem działania procesu decyzji jest zestaw tras, które będą ogłaszane do wszystkich sąsiadujących routerów, oraz tras lokalnie używanych do trasowania pakietów. Trasy te gromadzone są w bazie lokalnej i bazach wyjściowych.

Ważnym elementem procesu decyzji jest wyznaczenie stopnia preferencji danej trasy. Odpowiedzialna za to zadanie funkcja przyjmuje jako parametry wyceny trasy atrybuty trasy, a na wyjściu zwraca szacowany stopień użyteczności trasy. Podczas szacowania użyteczności trasy nie mogą być brane pod uwagę takie czynniki jak: istnienie innych tras, nieobecność innych tras czy parametry innych tras[30]. Wybór trasy odbywa się na podstawie zwracanym wartości preferencji danej trasy. Spośród wielu tras, wybierana jest ta o największym stopniu preferencji.

Jak już wspomniano proces decyzji operuje na danych zebranych w bazie wejściowej. Proces jest odpowiedzialny za wybór tras ogłaszanych do sąsiednich routerów znajdujących się w tym samym co router ogłaszający systemie autonomicznym, za wybór tras ogłaszanych do routerów zlokalizowanych w sąsiednich systemach autonomicznych, za agregacje tras i redukcje informacji[30].

Na proces decyzji składają się trzy fazy:

Faza 1 odpowiedzialna jest za oszacowanie stopnia preferencji każdej trasy otrzymanej od routera BGP zlokalizowanego w sąsiednim systemie autonomicznym i ogłoszenie do routerów w lokalnym systemie autonomicznym tras o największym stopniu preferencji dla każdego obiektu docelowego.

Router BGP powinien określić stopień preferencji dla każdej nowo otrzymanej trasy[30]. Jeśli trasa ta otrzymana została od routera lokalnego systemu autonomicznego, to stopień preferencji jest albo brany jako wartość otrzymana w polu LOCAL_PREF, albo szacowany na podstawie wcześniej zdefiniowanych reguł.

Faza 2 uruchamiana jest w momencie zakończenia się fazy 1. Jej zadaniem jest wybór najlepszej trasy (spośród wszystkich dostępnych znajdujących się we wszystkich bazach wejściowych) dla każdego obiektu docelowego i umieszczenie jej w bazie lokalnej. Jeśli atrybut NEXT_HOP trasy wskazuje na adres, do którego lokalny router BGP nie posiada trasy w bazie lokalnej, to taka trasa powinna być wykluczona ze zbioru tras rozważanych w fazie 2. Dla każdego zestawu tras (bierzemy pod uwagę trasy ze wszystkich baz wejściowych) prowadzących do danego obiektu docelowego lokalny router powinien wybrać tą, która[30]:

- posiada najwyższy stopień preferencji, lub
- jest jedyną trasą prowadzącą do obiektu, lub
- zostaje wybrana spośród tras posiadających ten sam stopień preferencji (patrz dalej).

Po wyborze trasy, router powinien wprowadzić tą trasę do lokalnej bazy, zastępując dowolną trasę znajdującą się już w bazie, a prowadzącą do tego samego obiektu docelowego. Dla wprowadzanej do lokalnej bazy trasy router musi określić wartość atrybutu NEXT_HOP. Wartość ta to bezpośredni, następny skok na trasie.

Trasy nieużyteczne powinny zostać usunięte z lokalnej bazy informacji.

Jeśli wśród rozważanych tras znajdą się dwie (lub więcej) trasy opisujące ten sam obiekt docelowy posiadające ten sam stopień preferencji router może spośród nich wybrać tylko jedną, która znajdzie się w bazie lokalnej. Trasy nauczone od routerów lokalnego i sąsiedniego systemu autonomicznego są traktowane w taki sam sposób[30]. Wybór trasy odbywa się według następujących zasad:

- jeśli router jest skonfigurowany do obsługi atrybutu MULTI_EXIT_DISC i trasy mają różne wartości tego parametru, to wybór pada na trasę o mniejszej wartości atrybutu.
- w przeciwnym wypadku, wybiera się trasę posiadającą mniejszy koszt dotarcia do adresu wymienionego w polu NEXT_HOP. Jeśli koszt ten jest równy dla obu tras, to:
    - jeśli przynajmniej jedna z tras kandydujących była ogłoszona przez router w sąsiednim systemie autonomicznym, to router lokalny wybierze tą trasę, która została ogłoszona przez router o najmniejszej wartości identyfikatora.
    - w przeciwnym razie, router wybiera się trasę ogłoszoną przez router o najmniejszym identyfikatorze.

Faza 3 rozpoczyna działanie po modyfikacji lokalnej bazy i jest odpowiedzialna za ogłoszenie tras zgromadzonych w lokalnej bazie do routerów znajdujących się w sąsiednich systemach autonomicznych. Ogłaszanie tras jest uzależnione do lokalnych reguł. W tej fazie może następować także agregacja tras i redukcja informacji.

13.7.3.1 Pokrywanie się tras

Pokrywanie się tras występuje w sytuacji, gdy ten sam zbiór obiektów docelowych opisany jest przez kilka różnych tras[30]. Zbiór obiektów docelowych opisywany jest jako numer sieci, czyli inaczej mówiąc bardziej znaczącą część adresu. Pokrywanie się tras, będzie związane z pojawianiem się tras o bardziej szczegółowym opisie obiektów docelowych i tras opisujących ten sam zbiór obiektów docelowych, ale w sposób ogólny. Mówimy, że trasa opisująca mniejszą ilość obiektów docelowych (dłuższy prefiks adresu) jest trasą bardziej specyficzną od trasy opisującej większą ilość obiektów (krótszy prefiks).

W przypadku pokrycia się tras następuje rozłożenie trasy mniej specyficznej na dwie części opisujące:

- zbiór obiektów opisanych tylko trasą mniej specyficzną i
- zbiór obiektów opisanych nachodzącymi na siebie trasami mniej i bardziej specyficznymi.

Gdy w bazie wejściowej zdarzy się pokrycie tras, to pierwszeństwo, czyli wyższy stopień preferencji powinna mieć trasa bardziej specyficzna. Pokrycie powoduje, że mniej specyficzna trasa opisująca część obiektów docelowych jest trasą prawidłową, lecz aktualnie nieużywaną (w procesie trasowania do tychże obiektów). Jeśli jednak trasa bardziej specyficzna zostanie wycofana z bazy, to obiekty opisane tą trasą będą nadal dostępne poprzez wykorzystanie trasy mniej specyficznej, która znajduje się cały czas w bazie[30].

13.7.4 Proces wysyłania uaktualnień

Proces wysyłania uaktualnień jest odpowiedzialny za tworzenie i wysyłanie wiadomości AKTUALIZACYJNYCH do sąsiednich routerów[30]. Ogłaszane są trasy przygotowane przez proces decyzji.

13.7.5 Uaktualnienia wewnętrzne

Jest to proces dystrybucji informacji o trasowaniu w obrębie lokalnego systemu autonomicznego[30].

Gdy router A otrzyma wiadomość AKTUALIZACYJNĄ od routera B zlokalizowanego w tym samym systemie autonomicznym, to router A nie powinien rozgłaszać otrzymanej informacji trasowania do pozostałych routerów lokalnego systemu. Natomiast w przypadku, gdy otrzymana wiadomość pochodzi od routera znajdującego się w sąsiednim systemie autonomicznym, to nowo nauczona trasa powinna zostać ogłoszona wszystkim routerom lokalnego systemu autonomicznego, jeśli zajdzie któryś z warunków:

- stopień preferencji przypisany przez router A nowej trasie jest większy niż stopień preferencji wszystkich innych tras otrzymanych wcześniej, lub
- nie ma innych tras otrzymanych od routerów sąsiednich systemów autonomicznych.

Jeśli wiadomość AKTUALIZACYJNA, która dotarła do routera A posiada niepuste pole tras nieużytecznych, to wszystkie trasy prowadzące do obiektów wymienionych w tym polu (jako numery sieci IP) powinny zostać usunięte z tablicy wejściowej routera A[30]. Jeśli usunięte trasy nie były wcześniej ogłaszane przez router A, to żadne dodatkowe akcje nie są wymagane. Jeśli natomiast trasy te były ogłaszane, to:

- gdy wybrana została nowa trasa na miejsce trasy usuniętej, to powinna ta nowa trasa zostać ogłoszona
- gdy nie ma trasy, która mogłaby zastąpić trasę usuniętą, to router A powinien utworzyć wiadomość AKTUALIZACYJNĄ, w której umieszczone zostaną nieosiągalne obiekty (w polu tras nieużytecznych). Wiadomość ta powinna zostać wysłana do wszystkich routerów, którym wcześniej były ogłaszane usunięte trasy.

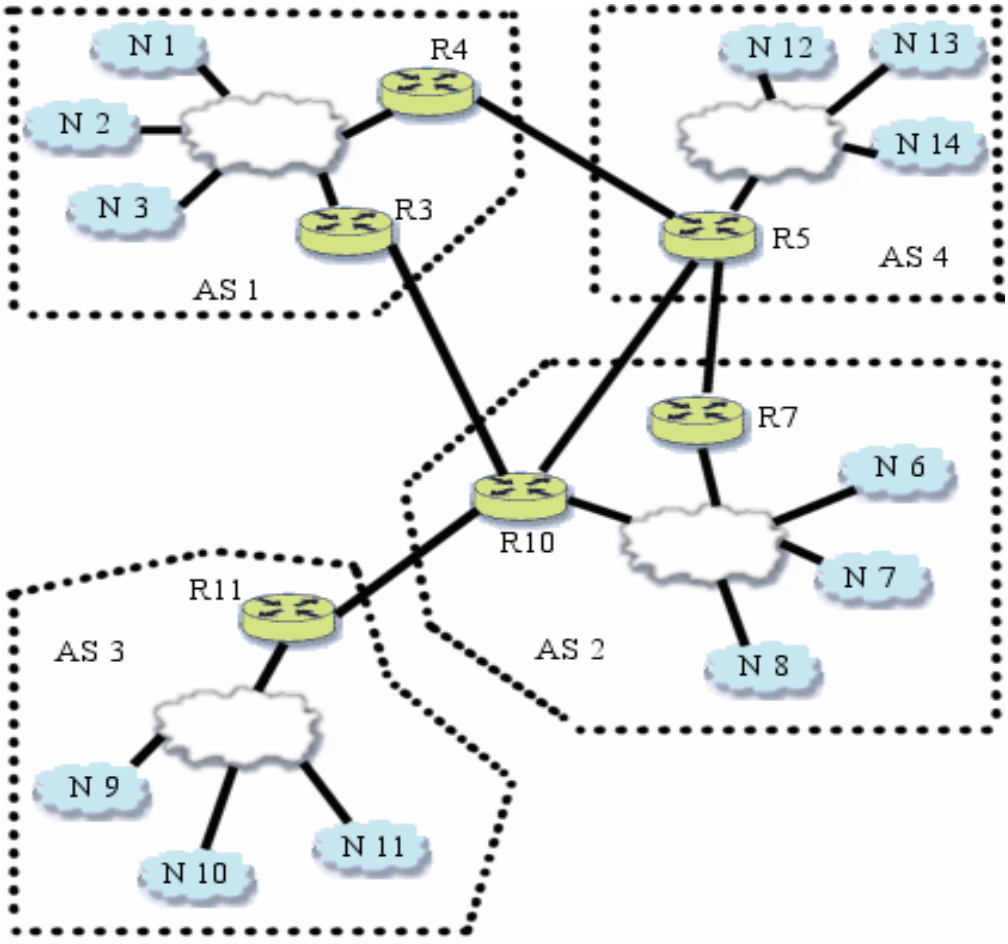
Wszystkie prawidłowe, aktualne trasy, które zostały ogłoszone powinny zostać umieszczone w odpowiednich bazach wyjściowych, a ogłoszone trasy nieużyteczne powinny z bazy wyjściowej zostać usunięte.

13.8 Dzielenie systemów autonomicznych i ich łączenie w konfederacje

Protokół BGP może być stosowany jako protokół wymiany informacji o trasowaniu w obrębie systemu autonomicznego[14]. W takim przypadku stawiane jest wymaganie, aby połączenia między routerami miały charakter „każdy z każdym”. Takie wymaganie bardzo obciąża zasoby routerów. W celu redukcji obciążenia możliwy jest podział systemu autonomicznego na kilka mniejszych systemów autonomicznych i uznanie ich jako systemów autonomicznych będących pod wspólną administracją.

Możliwy jest również podział systemu autonomicznego na kilka mniejszych, a następnie zgrupowanie ich w konfederacje[14]. Każdy z tych systemów ma wewnętrznie zapewnione połączenia „każdy z każdym” i posiada połączenia z innymi systemami autonomicznymi wchodzącymi w skład konfederacji. Na zewnątrz konfederacja widziana jest jako pojedynczy system autonomiczny (patrz również opis konfederacji w protokole IDRP).

13.9 Przykład topologii



Rysunek 16) Przykład topologii dla BGP

Routery w tej topologii przesyłają wzajemnie wiadomości, w których opisują trasy do poszczególnych sieci. Wiadomości te po dotarciu do routera, będącego odbiorcą, są umieszczane w bazach wejściowych. Na początek zobaczmy trasy i przypisane im atrybuty, które pojawiają się w tablicach wejściowych routera R3.

Sieć	Next_hop	Local_pref	Path	Origin	Weight	MED
N6	R4	200	AS {2,4}	BGP	75	-
N7	R4	200	AS {2,4}	BGP	75	-
N8	R4	200	AS {2,4}	BGP	75	-
N9	R4	200	AS {3,2,4}	BGP	75	-
N10	R4	200	AS {3,2,4}	BGP	75	-
N11	R4	200	AS {3,2,4}	BGP	75	-
N12	R4	200	AS {4}	BGP	75	-
N13	R4	200	AS {4}	BGP	75	-
N14	R4	200	AS {4}	BGP	75	-

Tabela 47) Baza wejściowa nr 1 (dane od routera R4)

Sieć	Next_hop	Local_pref	Path	Origin	Weight	MED
N6	R10	150	AS {2}	BGP	75	-
N7	R10	150	AS {2}	BGP	75	-
N8	R10	150	AS {2}	BGP	75	-
N9	R10	150	AS {3,2}	BGP	75	-
N10	R10	150	AS {3,2}	BGP	75	-
N11	R10	150	AS {3,2}	BGP	75	-

Tabela 48) Baza wejściowa nr 2 (dane od routera R10)

Podczas procesu decyzji zostaną wybrane trasy o najlepszych wartościach atrybutów. W bazie lokalnej będą następujące wpisy :

Sieć	Next_hop	Local_pref	Path	Origin	Weight	MED
N1	-	-	-	IGP	-	-
N2	-	-	-	IGP	-	-
N3	-	-	-	IGP	-	-
N6	R4	200	AS {2,4}	BGP	75	-
N7	R4	200	AS {2,4}	BGP	75	-
N8	R4	200	AS {2,4}	BGP	75	-
N9	R4	200	AS {3,2,4}	BGP	75	-
N10	R4	200	AS {3,2,4}	BGP	75	-
N11	R4	200	AS {3,2,4}	BGP	75	-
N12	R4	200	AS {4}	BGP	75	-
N13	R4	200	AS {4}	BGP	75	-
N14	R4	200	AS {4}	BGP	75	-

Tabela 49) Baza lokalna

Ponieważ wagi tras zarówno od routera R4, jak i od R10 mają wartość 75, dlatego decydującym atrybutem przy instalowaniu tras w bazie lokalnej ma wartość atrybutu Local_pref. W takiej sytuacji wybierane są trasy pochodzące od routera R4.

Dla przykładu rozważmy również część wpisów jakie pojawią się w bazach wejściowych routera R10.

Sieć	Next_hop	Local_pref	Path	Origin	Weight	MED
N9	R11	-	AS3	IGP	100	-
N10	R11	-	AS3	IGP	100	-
N11	R11	-	AS3	IGP	100	-
. . .						

Tabela 50) Baza wejściowa nr 1 (dane od routera R11)

Sieć	Next_hop	Local_pref	Path	Origin	Weight	MED
N1	R3	-	AS {1}	IGP	75	-
N2	R3	-	AS {1}	IGP	75	-
N3	R3	-	AS {1}	IGP	75	-
N6	R3	-	AS {2,4,1}	BGP	75	100
N7	R3	-	AS {2,4,1}	BGP	75	100
N8	R3	-	AS {2,4,1}	BGP	75	100
N9	R3	-	AS{3,2,4,1}	BGP	75	100
N10	R3	-	AS{3,2,4,1}	BGP	75	100
N11	R3	-	AS{3,2,4,1}	BGP	75	100
N12	R3	-	AS {4,1}	BGP	75	100
N13	R3	-	AS {4,1}	BGP	75	100
N14	R3	-	AS {4,1}	BGP	75	100

Tabela 51) Baza wejściowa nr 2 (dane od routera R3)

Sieć	Next_hop	Local_pref	Path	Origin	Weight	MED
N1	R5	-	AS {1,4}	BGP	100	-
N2	R5	-	AS {1,4}	BGP	100	-
N3	R5	-	AS {1,4}	BGP	100	-
N6	R5	-	AS {2,4}	BGP	100	-
N7	R5	-	AS {2,4}	BGP	100	-
N8	R5	-	AS {2,4}	BGP	100	-
N9	R5	-	AS {3,2,4}	BGP	100	-
N10	R5	-	AS {3,2,4}	BGP	100	-
N11	R5	-	AS {3,2,4}	BGP	100	-
N12	R5	-	AS {4,1}	BGP	100	-
N13	R5	-	AS {4,1}	BGP	100	-
N14	R5	-	AS {4,1}	BGP	100	-

Tabela 52) Baza wejściowa nr 3 (dane od routera R5)

Podczas procesu decyzji zostaną wybrane trasy o najlepszych wartościach atrybutów. W bazie lokalnej routera R10 będą następujące wpisy :

Sieć	Next_hop	Local_pref	Path	Origin	Weight	MED
N1	R5	-	AS {1,4}	BGP	100	-
N2	R5	-	AS {1,4}	BGP	100	-
N3	R5	-	AS {1,4}	BGP	100	-
N6	-	-	-	IGP	-	-
N7	-	-	-	IGP	-	-
N8	-	-	-	IGP	-	-
N9	R11	-	AS{3}	IGP	100	-
N10	R11	-	AS{3}	IGP	100	-
N11	R11	-	AS{3}	IGP	100	-
N12	R5	-	AS {4,1}	BGP	100	-
N13	R5	-	AS {4,1}	BGP	100	-
N14	R5	-	AS {4,1}	BGP	100	-

Tabela 53) Baza lokalna routera R10

Przy wyborze tras decydującym atrybutem był atrybut Weight, ale również bardzo ważną rolę spełniał atrybut Path, gdyż niektóre trasy musiały zostać odrzucone właśnie z jego powodu. Działo się tak, aby zapobiec powstawaniu pętli w sieci. Przykładem może być trasa otrzymana od routera R3 prowadząca do sieci N9. Na liście systemów autonomicznych przez które informacja ta została przekazana znalazł się system AS2, czyli lokalny system autonomiczny dla routera R10. W takim wypadku akceptacja takiej trasy oznaczałoby zapętlenie w procesie kierowania pakietów. Do takiej sytuacji oczywiście nie wolno dopuścić, dlatego trasa ta na samym początku została odrzucona.

14 Protokół IDRP

14.1 Wstęp

Protokół IDRP (Interdomain Routing Protocol) jest protokołem OSI definiującym sposób wymiany informacji pomiędzy systemami autonomicznymi[18]. Protokół IDRP jest oparty na protokole BGP i posiada wiele jego cech. Każdy router (w terminologii IDRP, Border Intermediate System – BIS) z zaimplementowanym protokołem IDRP ogłasza do swoich sąsiadów (BISs) obiekty docelowe dostępne poprzez ten router. Podobnie jak w BGP, do ogłaszanych tras przypisane są pewne atrybuty opisujące daną trasę i mające wpływ na podejmowanie decyzji o trasowaniu pakietów. Trasy otrzymane od sąsiednich routerów przetrzymywane są w bazach danych routera. Spośród tych tras wybierane są te o najlepszych parametrach transmisji i instalowane w osobnej tablicy routingu. IDRP zapewnia mechanizm niezawodnego przekazywania pakietów między routerami[18]. Działanie protokołu opiera się o cztery procesy: proces odbierania pakietów BISPDU protokołu, proces wysyłania pakietów BISPDU, proces decyzji i proces przesyłania.

14.2 Format przesyłanych wiadomości

14.2.1 Format nagłówka

RPI (1 oktet)	Długość (2 oktety)	Typ (1 oktet)
Numer sekwencyjny wiadomości(4 oktety)		
Potwierdzenie (4 oktety)		
Odbiór (1 oktet)	Nadanie (1 oktet	
Test (16 oktetów)		

Tabela 54) Format nagłówka wiadomości protokołu IDRP

RPI – jest to identyfikator wskazujący, że jest to pakiet protokołu IDRP[18]

Długość – określa długość całego pakietu włączając nagłówek

Typ – informuje, jakiego typu wiadomość następuje po nagłówku; wyróżniamy 6 typów (cztery takie jak w BGP i dwa dodatkowe):

1 - OTWARCIE (open)
2 - UAKTUALNIENIE (update)
3 - ZAWIADOMIENIE (error)
4 – KeepAlive
5 – USTANIE (cease)
6 – ODŚWIEŻENIE (refresh)

Potwierdzenie – jest to numer sekwencyjny ostatnio otrzymanego poprawnie pakietu z wiadomością.

Odbiór – wskazuje na ilość wiadomości, które nadawca tego pakietu jest w stanie przyjąć od routera sąsiedniego; pole to wykorzystywane jest do kontroli przepływu.

Nadanie - wskazuje na ilość wiadomości, które nadawca tego pakietu jest w stanie wysłać od routera sąsiedniego; pole to wykorzystywane jest do kontroli przepływu.

Test – służy do testowania poprawności zawartości pakietu; może też służyć procedurom uwierzytelniania.

14.2.2 Format wiadomości UAKTUALNIENIE

Wiadomości UAKTUALNIENIE są używane w celu ogłaszania tras dostępnych lub wycofywania tras, które stały się niedostępne, do sąsiednich routerów[20]. Pojedyncza wiadomość może równocześnie ogłaszać kilka dostępnych tras i wycofywać więcej niż jedną trasę z użycia. Wiadomość ma postać bardzo podobną do wiadomości UAKTUALNIENIA protokołu BGP (patrz 13.2.3). Wiadomości różnią się jedynie tym, że w IDRP bit 3 pola flagi atrybutów jest nieużywany i musi przyjmować wartość zero.

14.3 Atrybuty ścieżek

Wiadomości uaktualnień przenoszą trasy w postaci pary składającej się z opisu obiektu docelowego i ścieżki prowadzącej do tego obiektu. Ścieżka zdefiniowana jest jako zbiór atrybutów[18]. Podział atrybutów jest dokładnie taki sam jak w protokole BGP. Również zasady przekazywania atrybutów między routerami nie zmieniły się. W IDRP dodatkowo pewne atrybuty, z kategorii powszechnym dowolnych, mogą być oznaczone jako atrybuty rozróżnienia. Jak już wspomniano wcześniej atrybuty te identyfikują jednoznacznie bazy danych, w których powinna się znaleźć dana trasa. Atrybuty te służą rozróżnieniu tras prowadzących do tego samego obiektu docelowego, a posiadające różne miary użyteczności trasy. Atrybuty rozróżnienia mogą być dodawane do zbioru atrybutów przekazywanych z trasą tylko przez router, który generuje informację trasowania. Mogą być jednak modyfikowane przez routery przekazujące informację o trasie do innych routerów.

Do każdej trasy może być przypisana pewna kombinacja atrybutów rozróżnienia. Dozwolone zestawy atrybutów mogą zawierać:

- atrybut SECURITY,
- jeden z atrybutów : RESIDUAL_ERROR, TRANSIT DELAY, EXPANSE lub LOCALLY DEFINED QOS,
- atrybut PRIORITY.

Wynika stąd, że w skład zestawu atrybutów mogą wchodzić maksymalnie trzy atrybuty. Możliwy jest również pusty zestaw atrybutów rozróżnienia, który jest zestawem domyślnym[18].

Dodatkowo, atrybuty rozróżnienia dzielą się na dwa rodzaje: a) atrybuty typu, które są identyfikowane jednoznacznie przez typ (do tej grupy należą: RESIDUAL_ERROR, TRANSIT DELAY, EXPANSE i PRIORITY) oraz b) atrybuty typu i wartości, których rozróżnienie wymaga podania typu i wartości atrybutu (SECURITY i LOCALLY DEFINED QOS). Podział ten ma znaczenie przy porównywaniu atrybutów. Dwa atrybuty uznane zostaną za równoważne, gdy:

- należą do grupy atrybutów typu i mają ten sam typ,
- należą do grupy atrybutów typu i wartości, i oba mają ten sam typ i tą samą wartość.

14.3.1 Atrybut ROUTE_SEPARATOR

Atrybut ten, to atrybut separacji, należący do kategorii atrybutów powszechnych obowiązkowych[18]. Długość, jaką zajmuje w wiadomości to 5 oktetów. Dla każdej trasy ogłaszanej w komunikacie powinien być obecny jeden taki atrybut. W komunikacie może pojawić się kilka atrybutów tego rodzaju, po jednym dla każdej ogłaszanej trasy. Atrybut składa się z dwóch pól: identyfikatora trasy (ROUTE_ID) zajmującego 4 oktety i jednooktetowego lokalnego stopnia preferencji trasy (LOCAL-PREF).

Identyfikator trasy to unikalny w obrębie połączenia dwóch sąsiednich routerów numer trasy, pozwalający jednoznacznie określić trasę w bazie wejściowej (routera otrzymującego wiadomość) i wyjściowej (routera ogłaszającego wiadomość)[18]. Zaznaczyć należy, że ten sam numer może być przypisany do tras posiadających różne atrybuty rozróżnienia. Stąd, aby zidentyfikować trasę należy brać pod uwagę identyfikator trasy i związane z trasą atrybuty rozróżnienia. Inaczej mówiąc, atrybuty rozróżnienia decydują, w której bazie wejściowej/wyjściowej znajduje się trasa, a identyfikator wskazuje, która to trasa w bazie[20].

Atrybut rozgranicza zestawy atrybutów rozróżnienia przypisanych do trasy. Zestaw atrybutów rozróżnienia decyduje, do których baz zostanie wpisana trasa. Zauważmy, że dla każdego zestawu tras, może być przypisana różna wartość lokalnego stopnia preferencji trasy używanego w procesie decyzji.

Wartość pola preferencji trasy powinna przyjmować wartość zero, jeśli trasa jest ogłaszana do routera znajdującego się w innym systemie autonomicznym[18]. Pole to jest ignorowane, jeśli informacja pochodzi od routera znajdującego się w sąsiednim systemie autonomicznym.

Wszystkie atrybuty nie będące atrybutami rozróżnienia odnoszą się do wszystkich ogłaszanych tras, niezależnie od położenia w stosunku do atrybutów separacji.

14.3.2 Atrybut EXT_INFO

Jest to atrybut z grupy powszechnych dowolnych nie posiadający żadnych pól z danymi[18]. Jego obecność wskazuje, że kilka (lub wszystkie) atrybutów lub część tras (lub wszystkie) ogłaszanych w polu informacji dostępności sieci pochodzi ze źródeł nie będących częścią protokołu IDRP. Nieobecność tego atrybutu oznacza więc, że wszystkie przekazywane informacje uzyskane zostały metodami opisanymi protokołem IDRP. Atrybut powinien być generowany przez router będący źródłem informacji o trasie. Jeśli atrybut raz został przypisany do trasy, to powinien on być przekazywany wraz z tą trasą przez wszystkie routery propagujące informacje o trasie w sieci. Ponieważ informacje o trasie posiadającej ten atrybut pochodzą z innych źródeł niż IDRP, to może to prowadzić do powstawania pętli w sieci[18]. Dlatego też powinny być podjęte szczególne środki ostrożności w przypadku propagowania takich tras.

14.3.3 Atrybut RD_PATH

Jest to atrybut powszechny obowiązkowy złożony z serii segmentów opisujących systemy autonomiczne. Każdy segment składa się z trzech pól: typu segmentu, długości segmentu i wartości segmentu. Typ segmentu jest jednooktetowym polem mogącym przyjmować jedną z czterech wartości: RD_SET, RD_SEQ, ENTRY_SEQ lub ENTRY_SET. Segmenty typu RD_SEQ i ENTRY_SEQ tworzą listę identyfikatorów systemów autonomicznych, przez

które kolejno przeszła informacja trasowania. Segmenty typu RD_SET i ENTRY_SET tworzą natomiast nieuporządkowaną listę identyfikującą systemy autonomiczny, przy czym informacja trasowania niekoniecznie musiała przejść przez wszystkie z wymienionych systemów[18].

Długość segmentu zajmuje dwa oktety i wyraża w oktetach długość pola wartości segmentu.

Wartość segmentu to jedno lub więcej par złożonych z: długości i identyfikatora systemu autonomicznego. Długość określa rozmiar pola zawierającego identyfikator systemu autonomicznego. Identyfikator jest zakodowany zgodnie ze specyfikacją ISO 8348/Add.2.

14.3.4 Atrybut NEXT_HOP

Znaczenie tego atrybutu jest analogiczne jak w protokole BGP (patrz 13.3.6)

14.3.5 Atrybut DIST_LIST_INCL

Ten atrybut należy do grupy atrybutów powszechnych dowolnych. Jeśli jest obecny, atrybut zawiera listę identyfikatorów określających systemy autonomiczne, do których dana trasa może być ogłaszana.

14.3.6 Atrybut MULTI_EXIT_DISC

Znaczenie tego atrybutu jest analogiczne jak w protokole BGP (patrz 13.3.3)

14.3.7 Atrybut TRANSIT DELAY, RESIDUAL ERROR, EXPENSE

Są to atrybuty powszechne dowolne. Każdy z tych atrybutów jest zawierany w wiadomościach uaktualnień w celu wskazania, że parametry przekazywane z trasą zostały wyznaczone z naciskiem położonym na opóźnienie/pewność/koszt towarzyszący przesyłaniu danych tą trasą[18]. Router ogłaszający ten atrybut oznajmia tym samym, że posiada oddzielne bazy informacji (wejściowe, wyjściowe i lokalne) przechowujące dane o trasach, z którymi związane są liczby wyrażają stopień opóźnień/pewności/kosztów występujących na trasie.

14.4 Bazy danych

Baza danych (Routing Information Base – RIB), na podstawie, której router opiera swoją wiedzę dotyczącą topologii sieci, składa się z trzech oddzielnych części[20]:

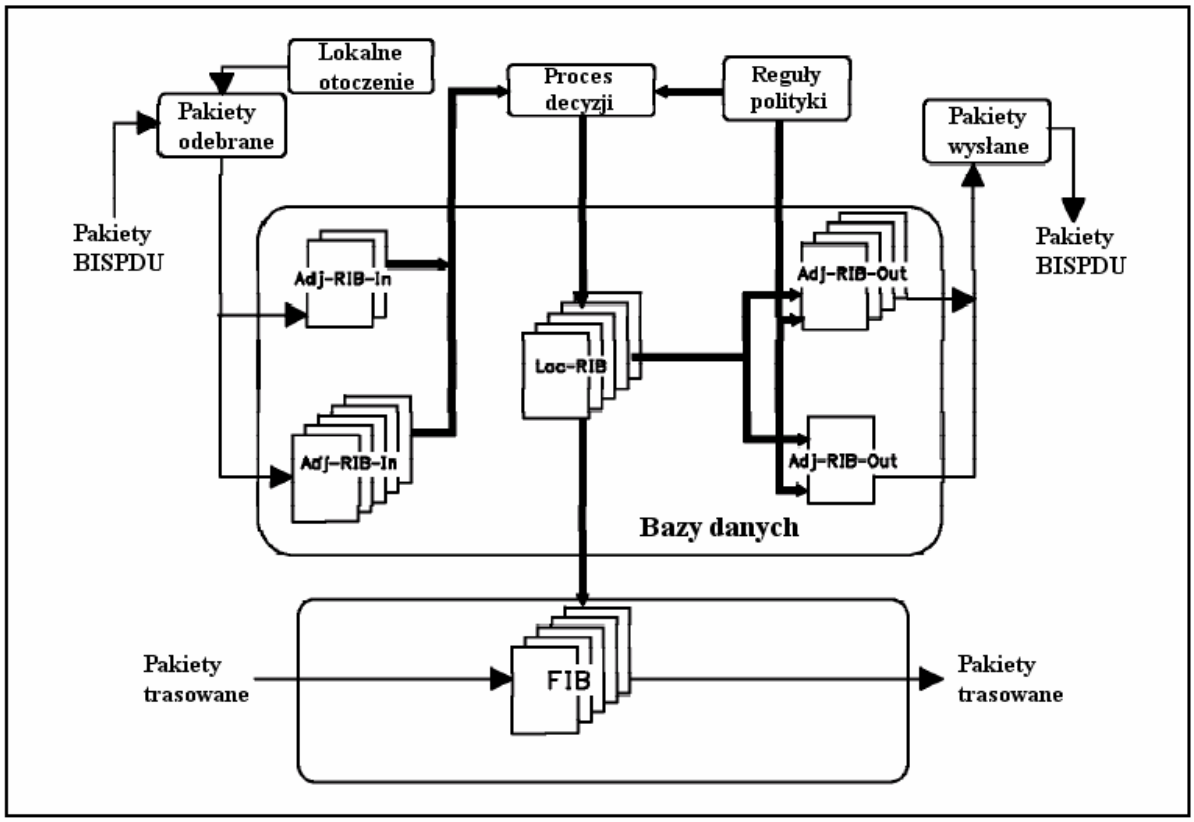
- baza wejściowa (Adj-RIBs-In) – w tej bazie przechowywane są informacje o trasach, które zostały uzyskane od sąsiednich routerów. Zawartość tej bazy reprezentuje zbiór tras będących danymi wejściowymi dla procesu decyzji. Każdy router musi dysponować przynajmniej jedną bazą wejściową dla każdego z sąsiadów. Opcjonalnie router może utrzymywać kilka baz wejściowych dla tego samego routera sąsiedniego. W takim przypadku rozróżnienie baz następuje na podstawie atrybutów (RIB-Att) przypi-

sanych do każdej z baz (patrz dalej). Oznacza to, że dwie bazy, odnoszące się do tego samego routera sąsiedniego, nie mogą posiadać identycznych wartości atrybutów.

- baza lokalna (Loc-RIBs) – baza zawiera lokalne informacje trasowania, które zostały wybrane spośród tras zgromadzonych w bazie wejściowej, poprzez zastosowanie pewnych reguł selekcji tras (proces decyzji). Pojedynczy router (z zaimplementowanym pojedynczym protokołem IDRP) może utrzymywać kilka baz lokalnych, rozróżnialnych na podstawie przypisane do nich atrybuty baz.
- baza wyjściowa (Adj-RIBs-Out) – w tej bazie przechowywane są te trasy, które zostały wybrane do ogłoszenia sąsiednim routerom. Dla każdego sąsiada istnieje przynajmniej jedna baza wyjściowa. Opcjonalnie może istnieć ich kilka[18]. W obrębie zbioru baz wyjściowych, przypisanych do pojedynczego routera sąsiedniego, żadna para baz nie może posiadać takich samych atrybutów. Informacje o routingu zgromadzone w tej bazie będą rozsyłane do sąsiednich routerów w postaci pakietów uaktualnień.

Widać tu podobieństwo do protokołu BGP, gdzie również baza danych podzielona jest na trzy części. Protokół IDRP wprowadza dodatkowo możliwość istnienia kilku wersji baz rozróżnialnych na podstawie przypisanych atrybutów (atrybuty rozróżnienia)[18].

Dodatkowo router utrzymuje jeszcze jedną bazę, która jest bezpośrednio wykorzystywana w procesie trasowania pakietów nie będących pakietami protokołu IDRP. Baza ta różni się ilością przechowywanych informacji o trasach i sposobem indeksacji. Mowa tu o bazie FIB (Forwarding Information Base). W bazie FIB przechowywane są tylko te atrybuty tras, które są bezpośrednio wykorzystywane podczas trasowania pakietów, czyli na przykład atrybuty RESIDUAL ERROR, TRANSIT DELAY[18]. Baza przechowuje dla każdej trasy lokalny interfejs, na który powinien być skierowany pakiet oraz adres następnego routera na trasie. Podobnie jak bazy informacji o sieci, również i ta baza ma kilka wersji w zależności od przypisanych do tras atrybutów rozróżnienia. W każdej z baz FIB trasy indeksowane są w oparciu o adres obiektu docelowego. Ten sposób indeksacji różni się od tego, który występuje w bazach wejściowych i wyjściowych, gdzie trasy są porządkowane na podstawie numeru przypisanego do danej trasy (Route-ID). Rysunek 17 przedstawia graficznie drogę jaką przebywa informacja pomiędzy bazami.



Rysunek 17) Graficzna droga przepływu informacji pomiędzy bazami w IDRP

14.4.1 Atrybuty baz

Każda pojedyncza baza informacji (pojedyncza baza wejściowa, pojedyncza baza lokalna, lub pojedyncza baza wyjściowa) posiada jeden zbiór atrybutów (atrybutów rozróżnienia) jednoznacznie identyfikujący daną bazę[18]. W zbiorze tym mogą się znaleźć tylko te atrybuty, które są rozpoznawalne przez router. Zbiór może składać się z jednego atrybutu, lub też ich dozwolonej kombinacji. Możliwe atrybuty i ich kombinacje są z góry ustalone i są to takie kombinacje, które router jest w stanie obsłużyć. Wszystkie routery w obrębie systemu autonomicznego powinny posiadać możliwość obsługi tych samych atrybutów i ich kombinacji[20]. Weryfikacja atrybutów, które obsługuje router następuje podczas początkowego łączenia się sąsiednich routerów. W przesyłanym komunikacie otwarcia połączenia zawarty jest zbiór możliwych kombinacji atrybutów akceptowalnych przez dany router. Router odbierający taki komunikat porównuje zestaw atrybutów otrzymanych ze swoim własnym zestawem i w przypadku, gdy występują jakieś niezgodności zgłaszany jest odpowiedni błąd. Każdy router zobowiązany jest obsługiwać domyślne bazy danych, z którymi związany jest pusty zestaw atrybutów opisujących bazę[18].

14.4.2 Detekcja błędów w bazach

Ponieważ na podstawie danych zawartych w bazach routera podejmowane są wszelkie czynności dotyczące trasowania pakietów, bardzo ważne jest zapewnienie ochrony tym danym przed pojawieniem się błędów, a w razie ich wystąpienia jak najszybsze wykrycie i usunięcie błędów z baz[20]. Biorąc pod uwagę fakt, że błędy, które pojawiły się w bazie routera

mogą być rozesłane do sąsiednich routerów i przedstawiać błędny obraz topologii, problem ten staje się na tyle ważny, że wprowadzono w IDRP mechanizmy sprawdzania poprawności zawartości baz danych. Mechanizm ten polega na okresowym sprawdzaniu sumy kontrolnej danych zawartych w bazach. Istnieje specjalny licznik czasowy służący wywoływaniu algorytmu sprawdzania integralności. Po upływie czasu określonego przez stałą sprawdzania sumy kontrolnej (maxRIBIntegrityCheck) wykonywane są kolejne kroki[18]:

- ponowne obliczenie sumy kontrolnej każdej z baz wejściowych. W razie niezgodności sumy kontrolnej, oznaczającej pojawienie się błędu w bazie, baza zostaje usunięta z pamięci.
- niezależnie od wyniku sprawdzania sumy kontrolnej uruchamiany jest ponownie proces decyzji. W wyniku działania procesu decyzji bazy lokalne i bazy wyjściowe są wypełniane nowo wyznaczonymi trasami. Z tego względu zbyteczne jest sprawdzania sumy kontrolnej dla tych baz.
- ustawienie wartości licznika wywołującego sprawdzanie integralności.

W celu uzupełnienia danych, w którejś z baz wejściowych router może wysłać do odpowiedniego routera sąsiedniego specjalny pakiet żądania odświeżenia danych (RIB-REFRESCH PDU). W przesyłanym żądaniu znajdują się także atrybuty bazy, której przesłanie ma nastąpić. Router sąsiedni otrzymując takie żądanie, przesyła w pakietach uaktualnień zawartość swojej bazy wyjściowej związanej z routerem, od którego nadeszło żądanie. Przesyłanie pakietów uaktualnienia poprzedza wysłanie pakietu rozpoczęcia przesyłania odświeżanych danych, a kończy wysłaniem pakietu zakończenia odświeżania danych[18].

14.5 Przetwarzanie informacji trasowania

14.5.1 Polityka routingu

O polityce routingu była już mowa przy okazji omawiania protokołu BGP. Wszystkie zasady odnoszące się do zakresu stosowania reguł polityki pozostają niezmienione[18].

Wymiana reguł między routerami nie jest opisana protokołem. IDRP odzwierciedla wynik działania reguł w ogłaszanych wiadomościach aktualizacyjnych, w których przesyłane są wybrane trasy. Ze względu na przymus używania takich samych, stałych reguł w obrębie całego systemu autonomicznego, IDRP udostępnia metody wykrywania reguł nie spełniających tego wymagania[18].

Każdy system autonomiczny ustala swoją politykę i kontroluje jej przestrzeganie. Lokalnie reguły składowane są w bazie informacji o polityce PIB (Policy Information Base) oddzielonej od baz informacji o trasach.

14.5.2 Proces decyzji

Podobnie jak to miało miejsce w BGP wyboru tras, którymi posługuje się protokół, dokonuje się w czasie procesu decyzji. Proces ten przebiega analogicznie do tego, jaki jest stosowany w BGP.

14.5.3 Trasowanie pakietów

Pierwszą czynnością dokonywaną przez router z zaimplementowanym protokołem IDRP po otrzymaniu pakietu jest rozpoznanie adresu odbiorcy. Po weryfikacji adresu i decyzji o potrzebie przesłania pakietu dalej następują czynności mające na celu znalezienie kolejnego routera na trasie. Decyzje o kierowaniu pakietów przychodzących do routera podejmowane są na podstawie zawartości tablicy FIB[18].

Po rozpoznaniu adresu odbiorcy podejmowane są następne czynności. Jeśli odbiorca znajduje się w obrębie lokalnego systemu autonomicznego routera, to pakiet przekazywany jest do jednego z routerów znajdującego się na wewnętrznie utrzymywanej przez router liście routerów wewnątrz systemu autonomicznego. Są to routery posługujące się protokołem wewnątrzdomenowym i to one odpowiedzialne są za dostarczenia pakietu do końcowego odbiorcy.

Gdy odbiorca wiadomości należy do innego systemu autonomicznego to na podstawie danych zawartych w nagłówku pakietu wyznaczane są parametry, jakie powinny być brane pod uwagę w procesie trasowania[18]. Chodzi tu o takie własności trasy jak: bezpieczeństwo, priorytet i jakość usług (Quality of Service). Inaczej mówiąc, określane są atrybuty rozróżnienia dla trasy, dzięki którym będzie można zidentyfikować odpowiednią bazę danych, a na podstawie tej bazy dokonać trasowania pakietu. W przypadku, gdy określone parametry rozróżnienia dla trasy zostaną dopasowane do którejś z baz FIB, to odnajdywana jest najbardziej specyficzna trasa do odbiorcy i pakiet zostaje przekazany do routera określonego w atrybutach trasy. Jeśli router nie posiada bazy o parametrach rozróżnienia pasujących do tych określonych dla przybyłego pakietu to generowany jest odpowiedni komunikat błędu[18].

14.5.4 Proces wysyłania uaktualnień

Proces wysyłania pakietów uaktualnień jest odpowiedzialny za ogłaszanie wewnętrznych wiadomości protokołu IDRP[18]. Przykładowo, ogłasza trasy wybrane w procesie decyzji do sąsiednich routerów zlokalizowanych tak w lokalnym systemie autonomicznym (uaktualnienia wewnętrzne), jak i w sąsiednim systemie autonomicznym (uaktualnienia zewnętrzne).

14.5.4.1 Uaktualnienia wewnętrzne

Proces uaktualnień wewnętrznych jest rozważany jako proces ogłaszania tras do routerów znajdujących się w tym samym systemie autonomicznym co router ogłaszający. Każdy router wybiera spośród tras otrzymanych od routerów z sąsiedniego systemu autonomicznego trasy najbardziej preferowane i ogłasza te trasy do wszystkich routerów w lokalnym systemie autonomicznym [18].

Kolejne czynności powinny być podjęte dla każdego zestawu atrybutów rozróżnienia obsługiwanych przez router:

- jeśli wiadomość aktualizacyjna otrzymana została od routera lokalnego systemu autonomicznego, to router odbierający nie powinien ogłaszać otrzymanych informacji do innych routerów swojego systemu autonomicznego.
- gdy wiadomość aktualizacyjna nadeszła od routera z sąsiedniego systemu autonomicznego, to wiadomość powinna zostać przekazana dalej do wszystkich routerów systemu lokalnego, jeśli:

    - stopień preferencji przypisany nowej trasie jest większy niż stopień preferencji wszystkich innych tras otrzymanych wcześniej, lub
    - nie ma innych tras otrzymanych od routerów sąsiednich systemów autonomicznych.

- jeśli wiadomość AKTUALIZACYJNA, która dotarła do routera posiada niepuste pole tras nieużytecznych, to wszystkie trasy prowadzące do obiektów wymienionych w tym polu (jako numery sieci IP) powinny zostać usunięte z tablicy wejściowej routera[18]. Jeśli usunięte trasy nie były wcześniej ogłaszane przez router, to żadne dodatkowe akcje nie są wymagane. Jeśli natomiast trasy te były ogłaszane, to:
    - gdy wybrana została nowa trasa na miejsce trasy usuniętej, to powinna ta nowa trasa zostać ogłoszona
    - gdy nie ma trasy, która mogłaby zastąpić trasę usuniętą, to router powinien utworzyć wiadomość AKTUALIZACYJNĄ, w której umieszczone zostaną nieosiągalne obiekty (w polu tras nieużytecznych). Wiadomość ta powinna zostać wysłana do wszystkich routerów, którym wcześniej były ogłaszane usunięte trasy.

W przypadku, gdy router posiada połączenie z kilkoma routerami sąsiednimi utrzymuje, dla każdego z tych sąsiadów oddzielną bazę wejściową. W bazach tych może być zawartych wiele tras prowadzących do tego samego obiektu docelowego, o tych samych atrybutach rozróżnienia i równym stopniu preferencji. Router musi jednoznacznie wybrać jedną z tras stosując się do zasad[18]:

- jeśli wszystkie z tras zawierają atrybut MULTI_EXIT_DISC (zawiera wartość metryki), a trasy różnią się tylko wartością atrybutów NEXT_HOP i MULTI_EXIT_DISC, to wybierana jest trasa o najmniejszej wartości atrybutu MULTI_EXIT_DISC. Jeśli w rezultacie nie udało się wyznaczyć pojedynczej trasy, to wybierana jest trasa, która została ogłoszona od routera o najmniejszym adresie.
- We wszystkich innych przypadkach wybierana jest trasa, która została ogłoszona od routera o najmniejszym adresie.

14.5.4.2 Uaktualnienia zewnętrzne

Są to uaktualnienia dotyczące przesyłania informacji do routerów w sąsiednich systemach autonomicznych. Podczas procesu decyzji wypełniane są bazy wyjściowe i bazy FIB. Wszystkie trasy nowo zainstalowane w tych bazach, lub trasy usunięte, do których nie ma tras zastępczych, powinny zostać ogłoszone do routerów IDRP w przyległych systemach autonomicznych[18].

Routery nie powinny przesyłać wiadomości aktualizacyjnych zawierających zestawy atrybutów rozróżnienia, które to zestawy nie były wynegocjowane podczas łączenia się routerów. W przypadku ogłoszenia takich wiadomości połączenie zostanie zerwane przez router odbierający.

14.6 Routing hierarchiczny

Protokół IDRP umożliwia organizowanie systemów autonomicznych w konfederacje. Dzięki takiemu rozwiązaniu może być tworzona hierarchicznie zorganizowana struktura poprawiająca skuteczność trasowania[18].

Tworzenie konfederacji jest strawą prywatną uczestniczących w konfederacji systemów autonomicznych i nie wymaga podejmowania specjalnych zmian i czynności na zewnątrz tworzonej konfederacji. Z zewnątrz, konfederacja widziana jest jako pojedynczy system autonomiczny (przykładowo konfederacja posiada numer identyfikujący systemy autonomiczne)[18]. Konfederacje mogą być zagnieżdżone, rozłączne lub nachodzić na siebie.

Każdy składnik konfederacji może posiadać własny zbiór reguł polityki.

Jako, że konfederacja jest na zewnątrz postrzegana jako pojedynczy system autonomiczny, to mechanizmy zapobiegania powstawaniu pętli w topologii wykryją sytuacje, w której trasa opuszcza konfederacje, a następnie do niej powraca. Oznacza to, że trasa pomiędzy dwoma obiektami znajdującymi się w tej samej konfederacji nie może przebiegać przez system autonomiczny nie należący do konfederacji[18].

15 Protokoły multiemisji

15.1 Wstęp

Multiemisja jest techniką pozwalającą na przesyłanie ruchu IP pochodzącego od jednego nadawcy, a kierowanego do wielu odbiorców. Zamiast przesyłać oddzielnie pakiety do każdego z indywidualnych odbiorców, wysyłany jest pojedynczy pakiet do grupy multiemisyjnej, która to grupa identyfikowana jest poprzez pojedynczy adres IP[21]. Multiemisja została wprowadzona w celu zapewnienia lepszej skuteczności w działaniu nowych, wymagających aplikacji, do których użycie metod rozgłaszania, czy metod regularnego przesyłania pakietów IP nie było wystarczające.

15.2 MOSPF

Protokół MOSPF (Multicast Open Shortest Path First) nie został jak dotąd zaimplementowany na routerach Cisco. Jest on jednak ważnym protokołem rodziny protokołów multiemisyjnych, dlatego jego opis jest całkowicie uzasadniony[14].

Protokół MOSPF jest rozszerzeniem protokołu OSPF wersji 2 wprowadzonym w celu umożliwienia routingu multiemisyjnego[21]. MOSPF jest w pełni kompatybilny z protokołem OSPF, co oznacza, że routery posługujące się tym protokołem będą współpracować z protokołem OSPF w przypadku przesyłanie regularnego (unicast) ruchu IP.

MOSPF dostarcza mechanizmy przesyłania pakietów multiemisyjnych z jednej sieci IP do innej sieci IP[29]. Ustalanie drogi pakietów odbywa się na podstawie zarówno umiejscowienia odbiorcy jak i umiejscowienia nadawcy informacji. Protokół MOSPF posługuje się bazą danych identyczną jak protokół OSPF, z tym jednak wyjątkiem, że w bazie pojawiają się pakiety LSA nowego rodzaju – opisujące przynależność do grup multiemisyjnych. Dzięki tym pakietom położenie wszystkich członków grup multiemisyjnych jest opisane w bazie MOSPF. Droga, jaką przebędzie pakiet multiemisyjny wyznaczana jest na podstawie drzewa najkrótszych ścieżek, w którym korzeniem jest nadawca pakietu[29]. Gałęzie drzewa nie zawierające członków danej grupy emisyjnej są usuwane. Cały proces budowy drzewa jest uruchamiany w momencie przybycia pierwszego pakietu. Rezultaty wyliczania najkrótszych ścieżek są przechowywane, w celu użycia ich ponownie w razie konieczności trasowania pakietu multiemisyjnego o tych samych adresach: źródłowym i docelowym[21].

OSPF umożliwia podział systemu autonomicznego na obszary (patrz rozdział 10.6). Podział taki powoduje utratę pewnych wiadomości o topologii systemu. Inaczej mówiąc, wiedza o sieci nie jest pełna. W przypadku przesyłania pakietów multiemisyjnych pomiędzy obszarami budowane drzewo najkrótszych ścieżek nie jest kompletne. Może to prowadzić do pewnej nieskuteczności routingu[29]. Analogiczna sytuacja zachodzi, gdy źródło pakietów multiemisyjnych znajduje się w innym systemie autonomicznym. W obu tych przypadkach bezpośrednie sąsiedztwo źródła nie jest znane.

Routery z zaimplementowanym protokołem MOSPF mogą współpracować z routerami posługującymi się protokołem OSPF w przesyłaniu pakietów regularnego ruchu IP[29]. Przesyłanie ruchu multiemisyjnego jest uzależnione od ilości routerów posługujących się MOSPF i połączeń między nimi. Tunelowanie pakietów multiemisyjnych przez routery nie obsługujące multiemisji nie jest dostępne w MOSPF.

Trasowanie pakietów od nadawcy do odbiorców oznacza się następującymi cechami:

- droga, którą przesyłany jest pakiet uzależniona jest od położenia nadawcy i od położenia członków grupy multiemisyjnej (jest to odróżnienie od protokołów takich jak OSPF, w których decyzja o trasowaniu odbywa się wyłącznie na podstawie położenia odbiorcy).
- ścieżka łącząca nadawcę z pojedynczym członkiem grupy multiemisyjnej posiada najniższy koszt spośród wszystkich dostępnych tras. Koszt trasy wyrażony jest w ten sam sposób jak ma to miejsce w OSPF.
- dla danego pakietu multiemisyjnego, wszystkie routery wyznaczają identyczne drzewo najkrótszych ścieżek i wybierana jest tylko jedna droga łącząca nadawcę z pojedynczym odbiorcą.

Protokół MOSPF najbardziej efektywnie działa w sieciach, w których jest stosunkowo mało aktywnych jednocześnie par: źródło-grupa odbiorców[21]. Działanie protokołu znacznie traci na efektywności w sieciach o dużej liczbie aktywnych źródeł oraz w przypadku występowania częstych zmian topologii sieci.

15.3 DVMRP

DVMRP (Distance Vector Multicast Routing Protocol) jest protokołem należącym do klasy protokołów wektora odległości. Mechanizmy trasowania stosowane w DVMRP są bardzo podobne jak w protokole RIP[16]. Podobnie jak w RIP podejmowanie decyzji o routingu odbywa się na podstawie liczby skoków, z tą jednak różnicą, że DVMRP wyznacza trasy na podstawie drzewa budowanego dla danej grupy multiemisyjnej.

W momencie, gdy nadawca rozpoczyna multiemisję, przyległy router do każdego ze swoich sąsiadujących routerów wysyła pakiet multiemisyjny. Ten proces trwa do momentu, gdy wszyscy członkowie grupy otrzymają pakiet[21].

Gdy router otrzymuje pakiet multiemisyjny, sprawdza w tablicy routingu, jaka trasa jest najkrótsza do nadawcy pakietu. Jeśli trasa ta prowadzi poprzez ten sam interfejs, przez który otrzymany został pakiet multiemisyjny, to router włącza informacje o grupie multiemisyjnej do swojej wewnętrznej bazy i przesyła pakiet do przyległych routerów poprzez wszystkie interfejny, z wyjątkiem interfejsu, z którego otrzymany został pakiet. Jest to proces zwany RPF (Reverse Path Forwarding) zapewniający, że w drzewie nie ma pętli, a używane trasy są trasami najkrótszymi [21].

Routery posługujące się DVMRP utrzymują wewnętrzną tablicę, w której przechowywane są informacje o członkach grup multiemisyjnych[16]. Z tego względu możliwe jest ograniczenie ruchu poprzez zaprzestanie przekazywania pakietów multiemisyjnych do sieci, w których nie ma członków danej grupy[21].

Okresowo DVMRP rozsyła pakiety multiemisyjne do wszystkich stacji. Jest to robione w celu odnajdywania nowych członków grup emisyjnych[16]. Stacja, która jest zainteresowana otrzymywaniem danych w drodze multiemisji wysyła co pewien czas do routera stosowny komunikat. Dzięki takiej komunikacji router ma możliwość określenia potrzeby wysyłania pakietów multiemisyjnych poprzez wszystkie interfejsy. Jeśli router uzna, że jego uczestnictwo w procesie przesyłania pakietów jest zbędne, wysyła do routera, od którego otrzymuje pakiety odpowiedni komunikat. Jest to mechanizm zawężania (czyszczenia) drzewa multiemisji. Mechanizm ten jest szczególnie skuteczny w przypadku dużej koncentracji stacji odbiorczych[21].

Do głównych wad DVMRP należy zaliczyć problemy ze skalowalnością[16]. Przyczynia się do tego:

- konieczność utrzymywania dużej ilości informacji wykorzystywanych w procesie routingu,
- utrzymywanie informacji o trasowaniu dla każdej grupy multiemisyjnej,
- okresowe uczestniczenie w procesie odnajdywania nowych członków grup.

Omawiany protokół nie jest implementowany na routerach firmy Cisco[14]. Niemniej jednak routery Cisco z zaimplementowanym protokołem PIM (patrz dalej) potrafią się komunikować i wymieniać informacje o trasowaniu z routerami używającymi protokołu DVMRP[14].

15.4 PIM

Protokół ten jest używany przez routery Cisco[14].

Jest to protokół uniwersalny, współpracujący ze wszystkimi protokołami routingu[21]. Protokół ten może działać w dwóch trybach: SM (sparse mode) i DM (dense mode) w zależności od warunków i natężenia ruchu pakietów multiemisyjnych.

Tryb DM protokołu PIM (Protocol-Independent Multicast) jest najbardziej skuteczny w sytuacjach, gdy[21]:

- nadawca i odbiorcy są w bliskim sąsiedztwie,
- jest mało stacji nadawczych i dużo stacji odbiorczych,
- ruch powodowany multiemisją jest duży
- strumień pakietów jest stały.

Tryb DM jest bardzo podobny w działaniu do protokołu DVMRP. Podobnie jak DVMRP używany jest tutaj mechanizm RPF. Najbardziej znaczącą różnicą pomiędzy DVMRP i trybem DM protokołu PIM jest to, że PIM nie wymaga szczególnego protokołu do wyznaczania trasy prowadzącej do źródła emisji pakietów[23]. Określanie tejże trasy odbywa się na podstawie dowolnego protokołu routingu.

Drugi z trybów, w których może pracować PIM, to tryb SM[21]. Najbardziej efektywny jest, gdy:

- jest mało odbiorców w grupie multiemisyjnej,
- nadawcy i odbiorcy są rozdzieleni łączami WAN,
- ruch pakietów jest sporadyczny lub o zróżnicowanym natężeniu.

Ze względu na podane powyżej okoliczności stosowania protokołu PIM-SM, używanie techniki RPF byłoby niekorzystne, ze względu na zbyt dużo obciążanie sieci[19]. Dlatego w trybie SM definiuje się tak zwany punkt spotkań (rendezvous point). Nadawca chcąc wysłać dane, wysyła je najpierw do punktu spotkań. Gdy natomiast odbiorca jest zainteresowany w uczestniczeniu w wymianie informacji, musi on się najpierw zarejestrować w punkcie spotkań[32]. W momencie rozpoczęcia przepływu pakietów od nadawcy do odbiorców za pośrednictwem punktu spotkań, routery znajdujące się na ścieżkach będą optymalizować ruch tychże pakietów, w celu jak najefektywniejszego ich dostarczenia. W trybie SM stosowane jest założenie, że pakiety są kierowane tylko i wyłącznie do tych stacji, które zgłoszą chęć otrzymywania pakietów multiemisyjnych danej grupy.

Protokół PIM ma możliwość równoczesnej pracy w obu trybach, stosując tryb SM dla części grup multiemisyjnych, a tryb DM dla pozostałych grup[21].

16 Protokół PNNI

PNNI (Private Network-to-Network Interface) jest protokołem trasowania z dynamicznym badaniem stanu połączenia, który służy do tworzenia sieci typu ATM[24]. Typowa sieć ATM składa się z wielu grup przełączników ATM różnego typu. Przełączniki te posiadają połączenia lokalne i w poszczególnych grupach komunikują się ze sobą przy pomocy interfejsu NNI (Network-to-Netword Interface).

Protokół PNNI umożliwia rozsyłanie informacji o topologii sieci pomiędzy grupami przełączników ATM[24]. Interfejs PNNI jest standardowym protokołem sygnałowym, który umożliwia wymianę informacji o topologii sieci pomiędzy urządzeniami różnych producentów. Informacje te wykorzystywane są następnie przez przełączniki do znajdowania najlepszej drogi łączącej dwa urządzenia końcowe. Po określeniu jej przebiegu, w sieci tworzone jest komutowane łącze wirtualne. Ponieważ sieć ATM może składać się z tysięcy wzajemnie połączonych przełączników pochodzących od różnych producentów, istnienie standardowego protokołu (np. PNNI) ma zasadnicze znaczenie dla współdziałania tego sprzętu[24].

Protokół routingu PNNI należy do grupy protokołów stanu łącza i działanie tego protokołu bardzo podobne jest do opisanego wcześniej protokołu OSPF (patrz rozdział 10). PNNI wspiera Quality of Service (QoS), jak również umożliwia tworzenie hierarchicznej struktury trasowania, co pozwala na skalowalność sieci o dużych rozmiarach. W przeciwieństwie do strategii internetowych, gdzie używane są różne protokoły na różnych poziomach (OSPF, BGB itp.), PNNI używa jednego protokołu routingu dla całej sieci. Protokół w znaczny sposób redukuje również ilość koniecznych czynności konfiguracyjnych związanych z wieloma poziomami routingu.

Każdy węzeł (przełącznik) w sieci posiada przypisany identyfikator (node ID). Identyfikator przełącznika oparty jest na adresie przełączników ATM, i dlatego adres ten jest automatycznie generowany przez implementacje (nie jest wymagana konfiguracja)[24].

PNNI grupuje przełączniki w hierarchiczną strukturę, dzięki czemu zmniejsza się ilość informacji przesyłanych w celu określenia topologii sieci[24]. W dużych sieciach ATM przesyłanie informacji o całej sieci i każdym jej węźle nie miałoby sensu. Zamiast tego tworzy się grupy przełączników ATM o podobnej strukturze adresów. Grupy przełączników (peer group) również identyfikowane są przez identyfikator. Wszystkie zgrupowane węzły muszą mieć ten sam identyfikator grupy. Identyfikatory grup wywodzą się z adresów przełączników w ATM, dlatego też są one automatycznie generowane i nie ma potrzeby ręcznej konfiguracji[24].

Do odnajdywania sąsiednich węzłów i utrzymywania z nimi kontaktu PNNI używa specjalnego rodzaju pakietów – pakietów Hello. Dzięki cyklicznej wymianie pakietów Hello z sąsiadującymi punktami węzłowymi przełącznik ma możliwość śledzenia aktywności węzłów. Pakiety Hello przenoszą identyfikator grupy przełączników, do której należy nadawca pakietu, dzięki czemu dwa sąsiadujące punkty węzłowe mogą określić czy znajdują się w tej samej grupie węzłów.

Podczas nawiązywania łączności, węzły pytają inne węzły, czy posiadają zasoby umożliwiające stworzenie określonego połączenia (Quality of Service). Na przykład przekaz video wymagać będzie dużej prędkości transmisji. Węzeł docelowy oszacuje swój własny udział w wymianie danych i możliwość obsługi jeszcze jednego połączenia. W przypadku negatywnej odpowiedzi węzeł proszący o połączenie wycofa się i zacznie szukać innego połączenia[24].

Grupy przełączników można połączyć w większą grupę, która w hierarchii będzie zajmowała wyższy poziom i wymieniała informacje z innymi grupami tego samego poziomu.

Protokół PNNI umożliwia obsługę maksymalnie 100 poziomów tak zbudowanej hierarchii[24].

Grupy, które wymieniają między sobą informacje o topologii sieci, zwą się grupami równoprawnymi. Każda z nich ma przypisany identyfikator. W każdej grupie znajduje się co najmniej jeden węzeł graniczny, który wymienia identyfikatory oraz informacje o topologii sieci z węzłami granicznymi znajdującymi się w innych grupach równoprawnych. Węzły graniczne wymieniają między sobą dane opisujące topologię ich domeny, tj. logicznej grupy węzłów posiadających ten sam identyfikator. Informacje te wykorzystywane są do znajdowania tras w sieci.

Routing w PNNI oparty jest na algorytmie stanu łącza. Każdy punkt węzłowy tworzy rekord, w którym opisuje swoją lokalną topologie, samego siebie, każde wychodzące połączenie z punktu węzłowego oraz każdy prefiks adresu, który punkt węzłowy jest w stanie opisać[24]. W specyfikacji PNNI rekordy te nazywane są PNNI Topology State Elements, w skrócie PTSE. Jeżeli punkt węzłowy posiada wszystkie rekordy PTSE dla wszystkich punktów węzłowych w grupie (peer group), to ma kompletną topologie, i może wyznaczać ścieżki dla każdego adresu w grupie przełączników.

Aby możliwy był routing przez każdy punk węzłowy w sieci, informacja o topologii (PTSEy) musi być rozprowadzana do każdego punktu węzłowego. Ponadto, gdy zmienią się parametry QoS (takie jak: połączenia, przepustowość i opóźnienia), informacja o topologii musi być zaktualizowana tak, aby routing odbywał się po rzeczywistej topologii[24]. PNNI przewiduje dwie metody dystrybuowania PTSE-ów: podczas włączania się do pracy nowego węzła oraz w razie zmiany topologii sieci.

W momencie włączenia się do pracy punktu węzłowego, nie zna on topologii sieci i nie może brać udziału w wyznaczaniu połączeń. Informacje o topologii sieci czerpane są od węzłów sąsiednich[24]. Każdy sąsiedni punkt węzłowy wysyła swoje streszczenie rekordów PTSE, jakie posiada w bazie danych. Włączający się punkt węzłowy może zażądać brakującej informacji od sąsiada. W ten sposób świeżo aktywowany punkt węzłowy może zgromadzić pełną informacje na temat topologii sieci od swoich sąsiednich punktów węzłowych, które mają kompletny zbiór PTSE-ów.

Informacja na temat topologii sieci ulega zmianie, kiedy aktywują się nowe połączenia, lub, gdy połączenia ulegają uszkodzeniu. Również parametry QoS mogą się zmieniać w zależności od bieżącego obciążenia sieci. Gdy w jakiejkolwiek części topologii sieci zajdzie znacząca zmiana to w konsekwencji punkt węzłowy stworzy nowy PTSE (w przypadku nowego połączenia), lub zaktualizuje istniejące PTSE-y (w przypadku zmiany parametrów QOS)[24]. Następnie, nowy PTSE jest rozsyłany do wszystkich sąsiadujących punktów węzłowych. Wszystkie punkty węzłowe, które otrzymają PTSE-a zachowują zawarte w nim informacje i dalej rozsyłają go do swoich sąsiadów. W ten sposób nowa informacja szybko zostanie rozesłana w obrębie całej grupy. Otrzymane PTSE-y są potwierdzane. Jeżeli PTSE nie zostanie potwierdzony jest on retransmitowany.

Sygnalizacja PNNI oparta jest na routingu źródłowym, co oznacza, że wejściowy przełącznik ATM oblicza drogę poprzez sieć ATM do przełącznika wyjściowego, który ogłosił spójność dla wywołanego przyjęcia. Trasa źródłowa, która jest listą punktów węzłowych po drodze (od jednego przełącznika do drugiego) nazywana jest Designated Transit List (DTL)[24]. Lista ta wchodzi w skład wiadomości sygnalizacyjnej nawiązywania połączeń. Wewnętrzne punkty węzłowe w sieci nie podejmują decyzji dotyczących wyznaczania trasy, ale podają dalej ustawienia zgodnie z listą DTL. Routing źródłowy zwiększa wydajność sieci, gdyż decyzje wyznaczania trasy nie są wymagane w każdym punkcie węzłowym pomiędzy przełącznikiem wejściowymi i wyjściowym[24]. Routing źródłowy zwiększa również ela-

styczność sieci, ponieważ obliczanie drogi może odbywać się za pomocą różnych strategii wykorzystywanych przez różne przełączniki. Nie ma również obawy, że wystąpią zapętlenia[24]. Ponieważ algorytm wyznaczania drogi w sieci nie jest częścią specyfikacji PNNI, implementacja daje możliwość wyboru przez administratora różnych strategii najbardziej odpowiednich dla danych konfiguracji.

17 Zakończenie

Analizując poszczególne protokoły trasowania zauważamy, że każdy z tych protokołów ma pewne wady i zalety. Każdy z nich ma swoje miejsce wykorzystania i nie można powiedzieć jednoznacznie, który z nich jest najlepszy. Szybki rozwój Internetu i decentralizacja kontroli jego struktur, wymuszała wprowadzanie nowych technik routingu oraz stosowania nowych mechanizmów i algorytmów. Mimo wprowadzania coraz to nowych protokół trasowania, stare protokoły nadal znajdują zastosowanie w pewnych sytuacjach.

I tak, protokół RIP nadaje się do przeprowadzania routingu w małych sieciach o jednolitej technologii i w takich sytuacjach sprawdza się całkiem dobrze. Jeszcze kilkanaście lat temu protokół taki jak RIP był wystarczający do obsługi większości rzeczywistych sieci. Niestety, jednym z jego podstawowych problemów jest zbieżność, przez co zerwanie łącza zostaje odzwierciedlone w tabelach routingu poszczególnych routerów dopiero po pewnym czasie. Jest to do zaakceptowania w małych sieciach, ale w pewnym momencie, gdy sieci bardzo się rozrosły, RIP stał się mało wydajny i w dużych sieciach korporacyjnych zastąpiony został protokołem IGRP. Wprowadzone mechanizmy uaktualnień wymuszonych i wstrzymywania uaktualnień miały na celu zmniejszenie czasu zbieżności. Nowością w IGRP jest również obsługa TOS oraz kierowanie pakietów wieloma trasami. Oczywiście żadne narzędzie nie rozwiązuje wszystkich problemów i wkrótce pojawił się kolejny protokół trasowania, wprowadzający nowe rozwiązania i algorytmy obliczeń. Protokołem tym był EIGRP. Stosowanie skutecznych mechanizmów trasowania pociągało za sobą konieczność zwiększania mocy obliczeniowej procesorów oraz pamięci operacyjnej routerów, co oczywiście wiązało się z kosztami.

Wadą protokołów wektora odległości jest generowanie dodatkowego ruchu w sieci poprzez cykliczne rozgłaszanie pełnych tabel routingu, nawet wówczas, gdy w topologii sieci nie zachodzą żadne zmiany. Ta wada wyeliminowana została przez opracowanie protokołu stanu łącza OSPF, który nie rozsyła cyklicznych ogłoszeń, a dodatkowy ruch generuje tylko przy zmianie stanu łącza. Bardzo ważną zaletą OSPF jest szybkie reagowanie na zmiany w topologii sieci. OSPF w celu optymalizacji wprowadza podział systemu autonomicznego na obszary, co wpływa na to, że OSPF przeznaczony jest do obsługi znacznie większych sieci niż protokoły wektora odległości. Ale i ten protokół posiada pewne wady, takie jak: zwiększone zapotrzebowanie na pasmo transmisji w początkowej fazie ich działania, zwiększone wymagania dotyczące procesora i pamięci operacyjnej routera.

Rozwój protokołów routingu zaobserwować można również w obrębie protokołów międzydomenowych. Początkowo stosowany EGP, zastąpiony został protokołem BGP, a ten z kolei ewaluował w kolejnych wersjach, aż do aktualnej wersji BGP 4.

Niektóre z powstałych protokołów są bardzo podobne, szczególnie jeśli chodzi o ich działanie i stosowane algorytmy. Oczywiście każdy z protokołów oferuje pewne nowe możliwości i definiuje obszar zastosowania. I tak opracowany przez OSI zestaw protokołów (IS-IS, IDRP) umożliwia skuteczne trasowanie dla sieci stosujących zarówno adresację TCP/IP jak i OSI. Zasady działania tych protokołów są niemal identyczne jak protokołów OSPF (jeśli chodzi o IS-IS) i BGP (jeśli chodzi o IDRP). Wprowadzono tylko nieznaczne zmiany.

Obserwując kolejno powstające protokoły można powiedzieć, że protokoły te wyraźnie zmierzają w kierunku stworzenia wielopoziomowej struktury identyfikacji adresata informacji zbliżonej do hierarchicznego systemu numeracji telefonicznej. Odbiega to znacznie od pierwotnego pomysłu na całkowicie płaski system adresacji w Internecie. Wielkość sieci uniemożliwiła praktyczne stosowanie tej metody.

Analizując poszczególne protokoły, rozważając ich wady i zalety można dojść do wniosku, że nie ma możliwości zaprojektowania algorytmu wyznaczania trasy optymalnego

dla każdego rodzaju fizycznej topologii sieci. Należy więc pamiętać, że niektóre protokoły są efektywniejsze w ściśle określonych strukturach sieci. Analogicznie można stwierdzić, że nie ma najlepszego protokołu, a każdy ma dobre i złe cechy, zależnie od konkretnej sytuacji.

Szybki wzrost wydajności urządzeń powoduje potrzebe wprowadzania nowych protokołów i nowych standardów. Takim właśnie standardem stał się zatwierdzony przez ATM Forum Multiprotocol Over ATM (MPOA). Architektura MPOA umożliwia przełączanie pakietów typu unicast w warstwie 3 przez urządzenia brzegowe z ominięciem routerów, tradycyjnie odpowiedzialnych za tą funkcjonalność. Inną, nową techniką jest MPLS (Multi Protocol Label Switching), w której zasadniczej zmianie ulega sposób spojrzenia na sieć i jej możliwości. Zasada działania MPLS polega na zdolności do interpretacji etykiet określających kierunek przesyłania informacji.

Wynika stąd, że temat protokołów routingu jest nadal tematem otwartym i ciągle zmieniającym się. Nadal opracowywane są nowe techniki mające na celu jak najlepsze wykorzystanie łączy. Równocześnie kładzie się duży nacisk na wymagania, które stawiają użytkownicy coraz to szybszych sieci. Dynamika rozwoju sieci i oprogramowania, pojawiające się coraz to nowe techniki i rozwiązania pokazują, że temat trasowania w sieciach jest nadal tematem otwartym.

18 Literatura

1. Chris Lewis „Routing Cisco TCP/IP dla profesjonalisty”
   tł. z ang. - Wydawnictwo PLJ, Warszawa 1999, ISBN: 83-7101-428-7
2. Innokenty Rudenko „Routery Cisco. Czarna księga”
   tł. z ang. - Wydawnictwo Helion, Gliwice 2001, ISBN: 83-7197-286-5
3. Mark Sportack „Routing IP”
   tł. z ang. - Wydawnictwo MIKOM, Warszawa 2000, ISBN: 83-7158-230-7
4. Mark Sportack „Sieci komputerowe. Księga eksperta”
   tł. z and. – Wydawnictwo Helion, Gliwice 1999, ISBN: 83-7197-076-5
5. Mark Tripod „Routery Cisco”
   tł. z ang. - Wydawnictwo Robomatic, Wrocław 2000, ISBN: 83-87150-95-9
6. Ravi Malhotra „IP Routing”
   Wydawnictwo O’Reilly, UK 2000, ISBN: 0-596-00275-0
7. Border Gateway Protocol (BGP).htm
   http://www.cisco.com/univercd/cc/td/doc/cisintwk/ito_doc/bgp.htm
8. Cisco - An Introduction to IGRP.htm
   http://www.cisco.com/warp/public/103/5.html
9. Cisco - Border Gateway Protocol (BGP).htm
   http://dis.eafit.edu.co/cursos/st080/material/libros/ito/bgp.htm
10. Cisco - Introduction to EIGRP.htm
    http://www.cisco.com/warp/public/459/7.html
11. Cisco - OSPF Design Guide - Section 1.htm
    http://www.cisco.com/warp/public/104/2.html
12. Cisco - White Paper EIGRP.htm
    http://www.cisco.com/warp/public/103/eigrp1.html
13. Cisco - Standards Supported in Cisco IOS Software Release 12_1-12_1T.htm
    http://www.cisco.com/warp/public/cc/general/bulletin/software/general/1189_pp.htm
14. Configuring IP Routing Protocols.html
    http://www.cisco.com/univercd/cc/td/doc/product/software/ios11/cbook/ciproute.htm
15. Dijkstry algorytm.html
    http://www.algorytm.cad.pl/Algorithms/11-20/algorithm16.html
16. Distance Vector Multicast Routing Protocol.htm
    http://www.voxtechnologies.com/enterasys_files/distance_vector_multicast_routing_protocol.htm
17. Forda-Bellmana algorytm.html
    http://www.algorytm.cad.pl/Algorithms/11-20/algorithm15.htm
18. IDRP.pdf
    http://www.helios-is.com/downloads/iso/idrp-fnl.pdf
19. Internet Protocol Multicast.pdf
    http://www.pulsewan.com/data101/pdfs/ipmulti.pdf
20. ISO 10747 - IDRP.htm
    http://www.acm.org/sigcomm/standards/iso_stds/IDRP/10747.TXT
21. Multicast Routing.htm
    http://cio.cisco.com/warp/public/614/17.html
22. NETWORLD - Protokoły routingu(1).htm
    http://www.networld.pl/artykuly/20910.html

23. PIM Overview.htm
http://www.juniper.net/techpubs/software/junos52/swconfig52-multicast/html/pim-overview.html
24. PNNI Specifikation.pdf
25. Protokoly routingu dynamicznego.htm
http://www.pckurier.pl/archiwum/art0.asp?ID=5011
26. RFC 1058 - RIP.htm
http://www.faqs.org/rfcs/rfc1058.html
27. RFC 1104 - Models of policy based routing_ H_W_ Braun.htm
http://rfc.sunsite.dk/rfc/rfc1104.html
28. RFC 1195 - IS-IS.htm
http://www.cis.ohio-state.edu/cgi-bin/rfc/rfc1195.html
29. RFC 1584 - Multicast Extensions to OSPF.htm
http://www.cis.ohio-state.edu/cgi-bin/rfc/rfc1584.html
30. RFC 1771 - BGP4.htm
http://www.cis.ohio-state.edu/cgi-bin/rfc/rfc1771.html
31. RFC 2178 - OSPF Version 2.htm
http://www.kblabs.com/lab/lib/rfcs/2100/rfc2178.txt.html
32. RFC 2362 - PIM-SM.txt
http://www.isi.edu/in-notes/rfc2362.txt
33. RFC 2453 - RIP Version 2.htm
http://www.faqs.org/rfcs/rfc2453.html
34. Wide Routing Protocols - Networld .htm
http://www.networld.pl/artykuly/20910_9.html