Notatki z sieci komputerowych

Adrian Jałoszewski

Streszczenie

Częściowo stworzone z nudów, częściowo dla pozbierania wszystkich sensownych informacji w jednym miejscu (aby nie były porozrzucane w losowych miejscach na prezentacjach). Ogólnie bez konfiguracji specyficznej pod Cisco, rozłożone na ładnych stronach A4.

Spis treści

1 NAT + IPv6

1.1 NAT

NAT może dokonywać translacji w trybie:

- Jeden adres ip ↔ jeden adres
- Wiele adresów IP (pula IP) ↔ jeden adres
- Wiele adresów IP ↔ wiele adresów IP

Konwersja nie musi być wiązana z przejściem datagramu przez bramkę – możliwe jest użycie tzw. rutera na patyku (router on a stick), gdzie datagram po konwersji wraca do tej samej sieci

Z reguły NAT operuje na trzech blokach adresów przypisanych do tzw. sieci prywatnych. Znaczna większość ruterów nie posiada jedneak ograniczenia NAD do adresacji sieci prywatnej. Bloki adresów traktowanych jako przynależne do sieci prywatnych:

- Klasa A: 10.0.0.0 – 10.255.255.255
- Klasa B: 172.16.0.0 – 172.31.255.255
- Klasa C: 192.168.0.0 – 192.168.255.255

1.1.1 Overloading

Overloading umożliwia stworzenie relacji typu „jeden do wielu" pomiędzy adresami IP podczas translacji (NAT). NAT bez trybu overloading pozwala tylko na translację kolejnych adresów IP na kolejne inne.

W większości ruterów funkcjonalność NAT jest konfigurowalna w oparciu o tzw. pule adresów IP (w ten sposób określa się zakres adresów). Kontrola procesu translacji jest regulowana przez standardowe listy kontrolne (ACL)

Wyróżniane są dodatkowe interfejsy pomiędzy którymi następuje translacja:

- inside
- outside

W trybie overloading możliwe jest zdefiniowanie puli adresów – lecz wtedy tylko jedno-adresowej i przypisanie jej do interfejsu inside (z tego interfejsu odbierane są datagramy od różnych nadawców – więc z różnymi adresami IP do konwersji na „adres nadawcy" w outside).

1.1.2 Rodzaje adresów

Podział ze względu na występowanie:

- Adres local – występujący w datagramach sieci inside (prywatnej)
- Adres global – występujący w datagramach IP w sieci outside

Z zachowaniem tego podziału w NAT występują 4 rodzaje adresów:

- inside local – adres hosta w sieci prywatnej (inside)
- inside global – adres interfejsu IP w sieci prywatnej, który jest nadawcą datagramów do sieci publicznej po konwersji (występuje w imieniu hostów z sieci prywatnej). Gdy wybieramy opcję overloading adres ten będzie adresem interfejsu outside w ruterze NAT. W przeciwnym wypadku – interfejsu inside w tym ruterze.
- outside global – rzeczywisty adres hosta w sieci publicznej (outside)
- outside local – adres hosta z sieci publicznej, pod którym występuje on w sieci prywatnej (przeważnie taki sam jak outside global)

1.2 Adresowanie IPv6

1.2.1 Adresacja

IPv6 posiada adresy 128 bitowe. Maska jest zapisywana na końcu po znaku /. Ogólnie są zapisywane jako pełna notacja szesnastkowa z dwukropkami co 16 bitów, możliwe jest pomijanie zer (6A2E:0:0:0:0:0:B6:A25E = 6A2E::B6:A25E).

Adres IPv6 składa się z następujących części:

- 3 bity Format (prefix)
- 13 bitów TLA (Top Level Aggregation) ID – określają identyfikator dostawcy pierwszego poziomu (Top Level Aggregator) – odpowiadają wpisom w globalnej tablicy rutowania, może ich być 8192, ilość tę można w przyszłości powiększyć biorąc bity z 8 zarezerwowanych
- 8 bitów – zarezerwowane
- 24 bity NLA (Next Level Aggregation) ID – identyfikator dostawcy drugiego poziomu (Next Level Aggregator) – powinny określać cel w ramach jednego TLA, typowo jeden identyfikator może być przyznawany jednej instytucji

- 16 bitów SLA (Site Level Aggregation) ID – lokalny identyfikator sieci – pozwala na określenie podsieci lokalnych, może ich być 65535
- 64 bity – identyfikator interfejsu – jest kojarzony z adresem Ethernet (MAC), ale przeznaczono na niego 65, a nie 48 bitów. Często konstruuje się go z MAC wstawiając po 3-cim bajcie MAC dodatkowe 16 bitów o wartoci 0xFFFE.

1.2.2 EUI-64

Notacja EUI-64 (64-bIT eXTENDED Unique Identifier) – gdy istnieje interface id (64-bitowa najmłodsza część adresu IPv6) zawiera MAC address (48 bitów) uzupełniony pośrodku wartością 0xFFFE (np. dla Cisco 5555:1111:1111:1111::/64 eui-64)interfejs otrzyma adres będący połączeniem powyższej konstrukcji i MAC (dodatkowo 7-my najstarszy bit MAC jest ustawiany na wartość 1).

- 2000::/3 – Global Unicast
- fc00::/7 – Unique Local Unicast
- fe80::/10 - Link Local Unicast

1.2.3 Adresy specjalne i prefiksy

Przeznaczenie adresu IP określa jego prefiks (określany jako pewn ilość najstarszych bitów przy odpowiedniej masce), np.:

- ::/128 – adres nie wyspecyfikowany (odpowiednik 0.0.0.0/32 w IPv4) – używany do określania dozwolonych klientów dla połączeń
- ::/0 – używany do określenia default routes przy rutowaniu (odpowiednik 0.0.0.0/0 w IPv4)
- ::1 - odpowiednik 127.0.0.1 w IPv4
- fc80:prefix::/10 – unique local address, służący do (unikatowego także w przypadku wydostania się poza bieżącą sieć) adresowania hostów lokalnych. Nierutowalny,. Często uzupełniany przez EUI-64. Prefix to 41+16 bitów (subnet+ link ID).
- fe80::/10 – link-local prefix (odpowiednik adresu auto-konfiguracji 169.254.0.0/16 w IPv4), każdy interfejs go posiada awaryjnie, często uzupełniony przez EUI-64.

- fec0::/7 – unikatowe adresy lokalne (site local), nie przeznaczone do rutowania (lub rutowalne w bardzo ograniczonych sytuacjach w grupie kilku wyróżnionych węzłów sieci)
- 2001::/?, 2002::? – adresy przeznaczone do tunelowania (różne technologie, przy maskach '?')
- ff01::?, ff02::?, ff05?=::? - adresy multicast przeznaczone do wspierania funkcji specjalnych (cyfra zmienna oznacza liczbę, która identyfikuje kolejną funkcję). Przykładowe funkcje to ruch związany z procesami rutowania (np. identyfikacja ruterów, protokoły RIPng, OSPF, EIGRP dla IPv6), DHCP, propagowanie nazwenictwa sieci i łącz, DNS, NTP i inne
- ff::0/8 są zarezerwowane

1.2.4 Cechy interfejsów IPv6

Interfejs może mieć wiele adresów IPv6, np: kilka link-local address, global address itp. Adresy mogą być trwałe lub tymczasowe (mogę być użyte tylko dla konkretnego połączenia wychodzącego identyfikując klient + usługę na bazie treści adresu IPv6).

Istnieją konfigurowalne „tablice preferencji" łączące każdy prefiks danego adresu (routing prefix) z tzw. precedence level (liczbę będącą priorytetem). W przypadku posiadania (np. chwilowego) danego adresu możliwe jest podjęcie decyzji, którego adresu użyć np. jako źródłowego dla połączenia wychodzącego.

Każdy adres IPv6 ma czas życia (domyślnie skonfigurowany jako nieskończony). Ruter konfigurujący adresy zdalnych interfejsów IPv6 może wymieniać ich adresy podając interfejsom dodatkową wartość lifetime adresu. Przeterminowany adres IPv6 przechodzi w interfejsie ze stanu preferred do deprecated (dalej może być używany lecz bez nawiązywania nowych połączeń). Po dalszym czasie przechodzi do stanu invalid i może być przypisany innemu interfejsowi.

1.3 IPv6 multicast

Zawiera:

- 8-bitowy prefiks o wartości 0xFF
- 4-bitowe pole flag (3 bity używane)

- 4-bitowe pole scope field identyfikujące unikatowość adresu i (częściowo) przynależność do konkretnej usługi oraz przede wszystkim zasięg w jakim znajdować się mogą odbiorcy datagramu
- 112-bitowe pole identyfikatora grupy multicast

Adresy IPv6 multicast są powszechnie użytkowane do wspierania protokołów Internetu zarządzających ruchem nad IPv6.

Istotne wartości pola scope decydującego o klasyfikacji ruchu IPv6 multicast (x to dowolna liczba z przedziału od 1 do 8)

- ff00::/128-ff0f::/128 – zarezerwowane
- ffx1::/16 – Intrface-local – datagramy nie wychodzą poza localhost (możliwe jest rozgłaszanie w obrębie jednego hosta z prezkazaniem ruchu do wielu aplikacji)
- ffx2::16 - Link-local – z przeznaczeniem dla lokalnego segmentu sieci, datagramy nie będą rutowane (odpowiednik IPv4 224.0.0.0/24)
- ffx5::16 – Site-local – z przeznaczeniem dla lokalnej sieci fizycznej – w tym przypadku dodatkowo (zależnie od specyfiki urządzeń) – ruch może być blokowany w urządzeniach konwertujących medium (np. w Access Point WiFi)
- ffxe::/16 - Global scope – zwykłe i rutowalne grupy IPv6 multicast

1.3.1 Rutowanie i maski

W przypadku hostów nie stosuje się przypisań masek do interfejsów. Natywne rozgraniczeni prefiksu sieci następuje w połowie długości adresu (młodsze 64 bity to identyfikator interfejsu). Inne niż /64 maski definiowane są w ruterach dla sieci bezpośrednio podłączonyc. W praktyce tylko dla tych, które nie zawierają hostów docelowych.

Host posiada tablicę rutowania. Domyślna reguła powinna kierować ruch na bramkę (uwage – dotyczy ona wszystkich adresów nie link–local). W zależności od wybranego adresu różne jest zachowanie hosta-nadawcy (inny adres źródłowy w datagramie IP, inna interpretacja lokalnej tablicy rutowania). Np. gdy wyślemy pakiet multicast typu global scope zostanie użyty adres nadawcy global. Gdy local scope – adres nadawcy unicast local scope.

1.3.2 Mapowanie IPv4

Użytkowanie IPv6 jest obecnie w dużej mierze oparte na tunelach wykorzystujących poprzednią wersję protokołu. Technika ta określana jest skrótem **6to4**. Mapowanie adresów IPv4:

$$2002 : xxxx : xxxx :: /16$$

gdzie $xxxx : xxxx$ to adres IPv4. Istnieją także adresy IPv6 mapowane z IPv4:

$$:: FFFF : x.x.x.x/96$$

lub

$$:: x.x.x.x/96$$

gdzie $x.x.x.x$ to adres IPv4 (stosuje się tu notację dziesiętną, która jest z automatu zamieniana na hex).

1.3.3 Tunelowanie IPv6 – mechanizm 6in4

Mechanizm polega na umieszczeniu datagramów IPv6 w tunelach punkt–punkt tworzonych w sieciach IPv4 (następuje enkapsulacja). Do oznaczenia protokołu tunelowanie 6in4 w datagramach IP wykorzystywana jest wartość 41. Tunele są zestawiane manualnie przez odpowiednią konfigurację ruterów-bramek. Istnieje możliwość zestawienia tuneli „dynamicznych" ('proto-41heartbeat' tunnels) – gdzie przeciwległy koniec tunelu może migrować pomiędzy kilkoma hostami (ruterami). Nowy adres tunelu jest przekazywany przez komunikat heartbeat.

Mechanizm ten nie wymaga ręcznego konfigurowania tuneli pomiędzy sieciami IPv6. Bazuje na stosowaniu indywidualnych hostów 6to4 lub ruterów brzegowych 6to4 ekranujących całe sieci IPv6 (tzn. wyspy IPv6). W obydwu przypadkach wymagane jest posiadanie adresu IPv4 w sieci globaclnej. Emulowany jest wówczas pseudo-interfejs IPv6 o adresie: 2002:xxxx:xxxx::/48 gdzie xxxx:xxxx to globalny adres IP. Ruter 6to4 automatycznie turneluje datagramy IPv6 gdy posiadają one prefiks 20002:.

Tunele także korzystają z protokołu tunelowania z wykorzystywaniem wartości identyfikatora protokołu: 41, lecz adresy końców tuneli są generowane na podstawie treści IPv6.

1.3.4 Użytkowanie adresów IPv6 w systemie operacyjnym

W przypadku typowania URL stosujemy nawias [ ] – należy go wpisać do UR, np:

- http://[adres IPv6]/

- https://[adres IPv6]:443/

Ścieżka UNC wykorzystująca IPv6 wygląda następująco (kreski zamiast :, oraz przyrostek DNS typujący adres IPv6):
\\1111-1-1-1-1-1-1-1111.ipv6-literal.net

W DNS do konwersji (name resolving) służy rekord AAAA, analogicznie do rekordu A w przypadku IPv4.

Istnieje protokół ICMPv6 - analog do ICMP w IPv4

1.3.5 NAT w IPv6

NAT z uwagi na szeroki zakres adresacyjny nie jest z nie jest technologią bezpośrednio z IPv6 powiązaną (powstał jako odpowiedź na braki adresów w IPv4). Implementowane jest prowadzenie kowersji host-host (bez overloading) – tehchnologia nosi nazwę NAT66. Jest to tak zwany wariant stateless.

NAT stosowany jest tdość często do prowadzenia konwersji IPv4 $\leftrightarrow$ IPb6 – gdy hosty nie posiadające przeciwległych adresów potrzebują się komunikować – technologia nosi nazwę NAT65. W jej przypadku konieczna jest translacja wszystkiego, co napływa do interfejsu IPv6 (nie są definiowane pule adresów IP). Po stronie IPv6 technika używa pseudo-interfejsów IPv6 (generuje adresy IPv6 nadawcy). Po stronie IPv4 używany jest adres urządzenia, które dokonuje translacji rutera.

2 Rutowanie IPv4, BGP

2.1 Border Gatway Protocol

2.1.1 Ogólne wiadomości

BGP – jest to *Border Gateway Protocol*. Protokół ten pracuje przy pomocy TCP (port 179) i dzieli się na protokół typu exterior (EGP) oraz interior (IGP). Wykorzystuje wielowarstwowe sesje między ruterami (peering) – tworzy relacje pomiędzy ruterami, na których pracuje:

- Wewnętrzne, internal (iBGP) – pomiędzy ruterami w tym samym Systemie Autonomicznym
- Zewnętrzne, external (eBGP) – pomiędzy ruterami w różnych systemach autonomicznych

Różnice między eBGP, a iBGP:

- domyślna wartość TTL w datagramach IP eBGP to 1 (tylko to najbliższego rutera w innym AS), dla iBGP jest to najczęściej 64
- Administrative Distance w tablicy rutowania wynosi 20 dla eBGP i 200 dla iBGP (Cisco)
- iBGP przesyła dodatkowe dane dotyczące LOCAL-PREFERENCE

2.1.2 Internal BGP (iBGP)

Ten wariant protokołu operuje w jednym systemie autonomicznym (jako IGP). Zaletą rozwiązania jest brak konieczności konwertowania komunikatów o trasach pomiędzy AS (jak to miało miejsce np. w OSPF).

Rutery iBGP mogą:

- przekazywać między sobą informacje o prefiksach z innych AS
- przekazywać informacje o prefiksach do innych AS
- odbierać informacje o prefiksach z innych AS

Rutery iBGP nie mogą przekazywać informacji o prefiksach z innych ruterów w swoim AS między sobą (zapętlenia). Dlatego – w przypadku iBGP (aby ruter posiadał informacje o komplecie prefiksów) – konieczne jest utrzymywanie secji pomiędzy ruterami iBGP w trybie „każdy z każdym” przez TCP – full mesh. Generuje to jednak duży ruch – ruch ten niekoniecznie przebiega tylko po segmencie

sieci współdzielonej przez dwa rutery uczestniczące w sesji iBGP (TCP może korzystać z wielu innych segmentów).

Dwie techniki skalowania iBGP:

- Konfederacje Systemów Autonomicznych (AS Confederations) – stworzenie kilku fikcyjnych systemów autonomicznych w istniejącym już AS i skonfigurowaniu ruterów do wymiana ruterów pomiędzy nimi (uwaga – trzeba kontrolować możliwość zapętleń)
- Route Reflector (RR) – zdefiniowanie rutera ze zmodyfikowaną implementacją iBGP wyjątkowo pozwalającą na przekazywanie informacji o prefiksach w bieżącym AS do innych wyróżnionych ruterów iBGP (zwanych RRC – Route REflector Client)
  - dzieli obszar działania na części – radykalnie zmniejsza liczbę sesji
  - w jednym AS może występować wiele RR
  - RR przekazuje komunikaty od klientów RR do innych ruterów iBGP
  - RR nie przekazuje komunikatów pomiędzy innymi ruterami iBGP (nieklientami)
  - RR nie przyjmuje komunikatów od innych RR (mogą być jednak klientami)

2.1.3 Ogólne zasady komunikacji w BGP

Przestrzeń działania BGP (Internet) dzielona jest na systemy autonomiczne (AS - Autonomic System). Są one identyfikowane 16-bitowymi liczbami (powyżej 65 000 zarezerwowane do testów). Komunikacja odbywa się w oparciu o wiadomości:

- OPEN (otwarcie połączenia) – pierwsza wiadomość wysyłana przez obie strony, jak przyjęta i zaakceptowana, to w zwrocie jest KEEPALIVE
- KEEPALIVE (potwierdzenie aktualności) – wysyłane ciągle z wynegocjowaną podczas otwierania sesji BGP częstotliwością
- UPDATE (wprowadzenie zmian) – przesyłanie między połączonymi routerami informacji o prefiksach oraz o ich dezaktualizacji
- NOTIFICATION (obsługa awarii) – wysyłanie w przypadku błędu, powoduje zamknięcie sesji między ruterami BGP

2.1.4 Prefiksy i ich atrybuty

Prefiks to ogólnie zakres adresów IP określony adresem sieci i maską (dowolną – VLSM). W BGP jest on opisywany wieloma dalszymi cechami. Informacja jest przesyłana w komunikatach BGP. Inne znaczenie prefiksu – informacja o zdalnej sieci posiadana przez dany ruter, powiązana z ID docelowego systemu autonomicznego, w którym sieci opisane prefiksem się znajdują.

Klasy atrybutów w prefiksach:

- Powszechne Obowiązkowe (Well-known Mandatory)
- Powszechne Dowolne (Well-known Discretionary)
- Opcjonalne Przechodnie (Optional Transitive)
- Opcjonalne Nieprzechodnie (Optional Non-transitive)

Atrybuty Well-known muszą być rozpoznawalne przez wszystkie implementacje BGP i obowiązkowo przetwarzane. Atrybuty przechodnie przekazywane są dalej do następnych ruterów.

Lista parametrów:

- ORIGIN - źródło informacji o prefiksie – atrybut określa, czy prefiks pochodzi iBGP, eBGP lub jest tworzony przez redystrybucję (wtedy wartość: *incomplete*)
- AS-PATH – kolejka systemów autonomicznych, przez które trzeba przejść aby dotrzeć do prefiksu
- NEXT-HOP – któryś z adresów na ścieżce do którego trzeba kierować ruch w stronę prefiksu (niekoniecznie najbliższy)
- Parametry dodatkowe – pochodzenie ich to przynależność do default-communities o nazwach:

  No-Export – zakaz rozgłaszania do innych sąsiadów eBGP

  No-Advertise – zakaz rozgłaszania gdziekolwiek

  Local-as – zakaz wysyłania poza lokalny AS w konfederacji

  Internet – zezwalaj na wysyłanie gdziekolwiek

- WEIGHT – parametr definiowany lokalnie dla rutera – priorytet zapisany w bazie danych określonego rutera, na podstawie którego będzie wybierana trasa, nie jest przekazywany poza bieżący ruter – implementacja zależy od producenta
- MED (Multi Exit Discriminator) – określa preferowane wyjście, gdy dwa AS mają wiele połączeń i nie wiadomo przez które powinna prowadzić droga wskazywana przez AS-PATH

2.1.5 Procedura wyboru trasy

Każdy ruter korzysta z informacji zawartej w bazie prefiksów na podstawie zoptymalizowanego przez producenta algorytmu, który dokonuje wyboru trasy na podstawie porównania kolejnych atrybutów. Dla Cisco kolejność brana pod uwagę to:

- WEIGHT - najwyższa
- LOCAL-PREFERENCE –najwyższa
- trasy zgłoszone ręcznie przez komendę network w BGP
- ORIGIN –preferowane są trasy z iBGP
- AS-PATH – jak najmniej AS do odwiedzenia po drodze
- MED – najniższa
- starsza (dłużej istniejąca) ścieżka
- ścieżka do rutera o niższym Router_id
- ścieżka do rutera o niższym IP

Reguły przy wyborze tras:

- Nie wybieraj trasy, dla której NEXT-HOP jest nieosiągalny
- Nie bierz pod uwagę trasy iBGP, jeśli sesja pomiędzy procesami nie jest aktywna (np. została zamknięta)
- Następnie stosuj kolejno kryteria wyboru trasy zgodnie z listą preferencji
- Najlepsza trasa jest także wysyłana do następnych ruterów w innych AS
- Możliwy jest multithoming – czyli definiowanie kilku dobrych tras z dopuszczeniem podziału ruchu

2.1.6 Atrybut AS-PATH

Ogólnie jest to podobnie jak z regexami:

- . – dokładnie jeden znak w ścieżce
- * – dowolna liczba znaków
- ˆ- początek ścieżki
- $ – koniec ścieżki
- _ – dowolny fragment ścieżki

Przykłady:

- .* – cokolwiek
- ˆ$ – trasy lokalne dla tego AS
- _65002$ – trasy stworzone w AS 65002
- ˆ65002_– trasy z AS 65002 (jako pierwszego na ścieżce)
- _65002_ – trasy przez AS 65002
- _65005_65004_ - trasy przez AS 65004, a później przez AS 65005

2.1.7 Zabezpiecnia przeciw zapętleniom

Aby zabezpieczyć przed zapętleniami eBGP korzysta z AS-PATH. iBGP nie przekazuje informacji o trasach wewnątrz AS (więc nie ma mowy o zapętleniu).

W zamian sesje iBGP utrzymywane są ciągle tworząc Full mesh (każdy z każdym utrzymuje sesję). Nie jest przy tym wymagane fizyczne połączenie każdego z każdym, gdyż TTL wynosi dla iBGP 64.

Rute Reflector (dla iBGP) – przy dużych AS liczba sesji jest znaczna, wprowadza się dlatego jeden ruter (Route REflector), który utrzymuje i przekazuje dalej informacje z innych. Inne rutery (Rourte Reflector Client) są skonfigurowane tak aby przekazywać informację do RR.

2.1.8 Redystrybucja protokołów rutowania

W ruterze obsługującym kilka protokołów rutowania dynamicznego jednocześnie istnieje możliwość przenoszenia inforamacji o trasach pozyskanych z użyciem jednego z nich do struktur danych utrzymywanych przez procesy związane z innym. Zabieg ten jest nazywany redystrybucją (danych o sieciach). W związku z tym informacje o trasach opuszczają ruter, ale tylko w kierunku sąsiadów powiązanych innym protokołem rutowania dynamicznego.

Oprócz tego możliwa jest także redystrybucja danych ze statycznej tablicy rutowania IP.

Jeżeli dane dodatkowe dotyczące trasy nie są generowane przez protokół źródłowy, to należy je uzupełnić (np. wartość liczby przeskoków przy redystrybucji do RIP, czy wartość metryki przy redystrybucji do OSPF).

2.1.9 RIPE

Jednostką rejestrującą informację o adresach IP i o systemach autonomicznych BGP w Europie jest RIPE (Reseaux IP Europeens) – https://www.ripe.net, strona dysponuje wyszukiwarką rejestrów https://apps.db.ripe.net/search/query/html. Na świecie istnieją ogólnie cztery inne odpowiedniki RIPE. RIS – Routing Information Service - system informacyjny RIPE o Systemach Autonomicznych BGB i trasach między nimi.