

Politechnika Warszawska - Wydział Elektroniki i Technik  
Informacyjnych

# **Bezpieczeństwo i skalowalność protokołu BGP w wersji 4.**

Robert Święcki

11 lutego 2005

Opiekun pracy: dr inż. Andrzej Bąk

---

## **Abstract**

BGP is one of the most important protocols in the Internet. Simple, accidental configuration error can disrupt Internet connectivity all over the world. The consequences of BGP process misconfiguration or direct attack on the protocol are not well recognized. In this thesis the study of the security and scalability issues of BGP protocol is presented. The aim of this work is to help network engineers and protocol designers to create secure and scalable routing procedures. In the practical part of this work the extension of the existing BGP protocol to support 32 bit ASN is presented and discussed.

---

## **Podziękowania**

Chciałbym podziękować mojemu promotorowi, dr inż. Andrzejowi Bąkowi za pomoc i opiekę nad redakcją pracy.

Podziękowania należą się także członkom Zespołu Eksploatacji Sieci (NOC) ATMAN, za niezwykle kompetentne rozmowy na tematy opisywane w tej pracy.

## Spis treści

<b>1. Znaczenie Internetu</b>	<b>7</b>
<b>I. Wprowadzenie do BGP w wersji 4.</b>	<b>10</b>
<b>2. Informacje podstawowe</b>	<b>10</b>
2.1. Wstęp . . . . .	10
2.2. Słowniczek pojęć . . . . .	10
2.3. Standaryzacja BGP . . . . .	10
<b>3. Opis protokołu</b>	<b>13</b>
3.1. Protokoły podkładowe. . . . .	13
3.2. Obsługiwane protokoły . . . . .	13
3.3. Zestawianie sesji BGP . . . . .	13
3.4. External BGP . . . . .	14
3.5. Internal BGP . . . . .	14
3.5.1. Route-reflectors . . . . .	15
3.5.2. Konfederacje AS . . . . .	16
3.6. CIDR . . . . .	16
3.7. Nagłówek BGP . . . . .	17
3.8. Typy wiadomości BGP . . . . .	17
3.9. Atrybuty . . . . .	18
3.10. Strategie routingu . . . . .	20
<b>II. Analiza aspektu bezpieczeństwa protokołu BGP</b>	<b>22</b>
<b>4. Uwarunkowania historyczne bezpieczeństwa protokołu</b>	<b>22</b>
<b>5. Rozgłaszanie prefiksów</b>	<b>24</b>
5.1. Incydent 7007 (AS 7007 Incident) . . . . .	24
5.1.1. Opis . . . . .	24
5.1.2. Analiza przypadku . . . . .	25
5.1.3. Podsumowanie . . . . .	29
5.2. 128/9 „disaster” . . . . .	30
5.2.1. Opis . . . . .	30
5.2.2. Analiza przypadku . . . . .	31
5.2.3. Podsumowanie . . . . .	32

---

5.3. Wnioski . . . . .	32
5.4. Proponowane rozwiązanie . . . . .	33
5.4.1. Wnioski . . . . .	33
<b>6. Błędy konfiguracyjne</b>	<b>35</b>
6.1. Origin misconfiguration . . . . .	35
6.2. Export misconfiguration . . . . .	36
6.3. Wnioski . . . . .	37
<b>7. Hijacking</b>	<b>38</b>
7.1. Hijacking adresów i Spam . . . . .	38
7.1.1. Teoretyczny przebieg ataku . . . . .	39
7.1.2. Wnioski . . . . .	40
<b>8. Podsumowanie analizy aspektu bezpieczeństwa protokołu BGP</b>	<b>41</b>
<b>III. Analiza aspektu skalowalności protokołu BGP</b>	<b>42</b>
<b>9. Filtracja BGP</b>	<b>42</b>
9.1. Koordynacja pomiędzy operatorami . . . . .	42
9.2. Bogon filters . . . . .	49
<b>10. Konfederacje AS</b>	<b>52</b>
10.1. Ocena skuteczności . . . . .	53
<b>11. Objętość informacji routingowej</b>	<b>54</b>
11.1. Zarys problemu . . . . .	54
11.2. Możliwe rozwiązania . . . . .	55
<b>12. Ograniczony zasób - ASN</b>	<b>57</b>
12.1. Zarys problemu . . . . .	57
12.2. Proponowane działania . . . . .	58
12.2.1. Weryfikacja przyznaných zasobów . . . . .	58
12.2.2. 32bit ASN . . . . .	58
<b>13. Modyfikacja demona Quagga</b>	<b>60</b>
13.1. Zakres modyfikacji . . . . .	60
13.2. Zestawienie sesji BGP . . . . .	62
13.3. Rozgłaszanie prefiksów . . . . .	64

13.4. Wnioski . . . . .	64
<b>14. Podsumowanie analizy aspektu skalowalności protokołu BGP</b>	<b>65</b>

# Wstęp

## 1. Znaczenie Internetu

Wraz z popularyzacją Internetu stopniowo nasilają zjawiska, które wpływają negatywnie na jego funkcjonowanie. Z jednej strony mogą być to działania celowe - inicjowane przez ludzi lub organizacje - z drugiej ograniczenia związane ze wzrostem liczby jego użytkowników i pierwotnymi założeniami projektowymi.

Z chwilą, gdy Internet wydostał się z kręgów akademickich, jego wpływ na procesy gospodarcze wzrósł do ogromnych rozmiarów. Od poprawnego działania tej sieci zależą wielkie korporacje, banki a coraz częściej także udostępniające za jego pomocą swe zasoby instytucje użyteczności publicznej. Internetowe usługi stały się ważnym sektorem gospodarki, generującym ogromne zyski a rozrywka internetowa jest postrzegana przez potentatów tego rynku za bardzo atrakcyjne miejsce dla ekspansji. Warto wspomnieć chociażby tzw. „bankowość wirtualną” (mBank, Intelligo), czy też sieciowe sklepy z muzyką (iTunes, IPlay), które z roku na rok zwiększają swe zyski i obroty a także zdobywają coraz więcej klientów. Większość z „internautów” jest gotowa zapłacić za luksus rezygnacji z kolejek na rzecz internetowych portali, obsługiwanych z wygodnych foteli, przy użyciu przeglądarki internetowej zainstalowanej na domowym PeCecie.

Dzięki Internetowi oddziały dużych korporacji, wykorzystując logiczne połączenia, tzw. VPN - Virtual Private Networks, mogą komunikować się z centralami w celu wymiany danych, w szczególności przy użyciu komunikacji głosowej i wizualnej (telekonferencje). Usługi typu VoIP coraz śmielej poszerzają swój udział w rynku transmisji głosu. Z Internetu korzysta na wielką skalę sektor reklamowy, przeprowadzane są przy jego udziale sondaże opinii publicznej. Przesyłanie wiadomości elektronicznych (e-mail) jest istotną, a co ważne, relatywnie taną formą kontaktu z klientami dla wielu przedsiębiorstw. Z dobrodziejstw usług sieciowych coraz częściej korzystają sami pracownicy, którzy wolą wykonywać swe zadania w miłej, domowej atmosferze przy użyciu stałego łącza sieciowego. Ta forma wykonywania swych obowiązków służbowych, nazywana telepracą, jest chętnie wykorzystywana przez pracodawców, kuszonych wizją obniżenia kosztów.

Sprzedaż usług internetowych jest obecnie koniecznością dla każdego opera-

tora telefonii i telewizji kablowej. Proponowane przez nich ceny, pozwalające na dołączenie swojego urządzenia sieciowego do globalnej sieci, spadły drastycznie na przestrzeni ostatniego dziesięciolecia. Można przypuszczać, że w ciągu kilku następnych lat domowe połączenie z Internetem przestanie być postrzegane jako luksus, a zacznie być koniecznością wynikającą z faktu przenoszenia wielu ważnych usług na platformy korzystające z globalnej sieci. W niektórych krajach, jak np. Estonia, Internet stał się ważną formą kontaktu „państwa” z jego obywatelami. W Polsce także można zauważyć coraz większe zainteresowanie administracji publicznej Internetem. Wszystkie polskie ministerstwa i ważniejsze centralne urzędy posiadają swoje strony WWW, wiele instytucji samorządowych udostępnia na własnych witrynach użyteczne dla potencjalnych interesantów zbiory danych w postaci Biuletynów Informacji Publicznej. Zainteresowaniem cieszy się wizja tzw. „internetowych wyborów”. Wprawdzie jeszcze nierealizowalna, ale jej przedsmak możemy oglądać na stronach Państwowej Komisji Wyborczej (<http://www.pkw.gov.pl>), udostępniających, niemal natychmiastową, wizualizację wyników organizowanych głosowań.

Skoro od Internetu zależy tak wiele, to czy jego infrastruktura jest odpowiednio chroniona? To pytanie zaczyna być coraz częściej podnoszone w środowiskach odpowiedzialnych za funkcjonowanie poszczególnych części tej globalnej sieci. A jest ono niezwykle ważne, ponieważ Internet i oferowane przez niego usługi stały się niezwykle istotną częścią światowej gospodarki. Kolejne wątpliwości przynosi pytanie o skalowalność Internetu: „Czy, jeżeli protokoły Internetowe były projektowane z myślą o znacznie mniejszych sieciach niż Internet, to jak dadzą sobie radę w przyszłości, gdy liczba jego użytkowników i oferowanych usług będzie dalej szybko się powiększać?”.

Ważnym elementem internetowej infrastruktury są protokoły routingu. Pozwalają one dobierać efektywne trasy dla pakietów IP, a w przypadku awarii znajdować trasy zastępcze. Najważniejszym protokołem routingu IP w wymiarze globalnym jest BGP (Border Gateway Protocol), obecnie stosowany w wersji 4. Dzięki niemu Internet działa; łączność pomiędzy odległymi zakątkami globu nie sprawia - zazwyczaj - większego problemu.

Postawione powyżej pytania odnoszą się w dużej mierze właśnie do BGPv4. Od jego architektury i konkretnych implementacji zależy najbardziej skalowalność i stabilność dzisiejszego i przyszłego Internetu.

Chociaż BGPv4 jest protokołem zaawansowanym, posiada on nałożone przez twórców, a wynikające z niedoszacowania roli Internetu, ograniczenia. Nieuwzględnienie nadchodzących zmian w globalnej sieci w chwili jego projek-

towania jest obecnie problemem zarówno pod względem skalowalności jak i bezpieczeństwa. Niniejszej praca stara się przeanalizować te problemy i zaproponować rozwiązania, które mogą pomóc w uniknięciu związanych z nimi zagrożeń.

# Część I.

## Wprowadzenie do BGP w wersji 4.

### 2. Informacje podstawowe

#### 2.1. Wstęp

BGP (*ang. Border Gateway Protocol*) jest protokołem routingu dynamicznego stosowanym w celu wymiany informacji routingowych w Internecie. Pozwala na aktywne tworzenie strategii routingu (polityka routingowa) oraz utrzymywanie wielu połączeń (linków) z różnymi ISP<sup>1</sup> w celu uzyskania redundancji połączeń jak i load-balancingu (*ang. równoważenia obciążenia*) przepływu pakietów pomiędzy łączami. BGP na podstawie zdefiniowanych wcześniej reguł pozwala na automatyczne dobieranie najbardziej efektywnych ścieżek routingowych. Pozwala to na optymalne wykorzystanie infrastruktury sieciowej używanej w Internecie.

BGP jest obecnie jedynym protokołem używanym globalnie w Internecie. Jego poprzednikiem był EGP (*ang. Exterior Gateway Protocol*), którego ograniczenia były tak duże w odniesieniu do skalowalności, że zdecydowano się na stworzenie nowego protokołu, któremu nadano nazwę BGP.

#### 2.2. Słowniczek pojęć

Na stronie 66 znajduje się słowniczek podstawowych pojęć używanych w niniejszej pracy, odnoszących się do BGP.

#### 2.3. Standaryzacja BGP

Pierwsza wersja BGP została udostępniona w 1989r.. Opisano ją w RFC<sup>2</sup>1105. Czwarta wersja BGP została opisana w RFC 1771 oraz RFC 1772. Lista ważniejszych RFC dotyczących protokołu BGPv4 znajduje się w załączonej tabeli:

---

<sup>1</sup>ISP - Internet Services Provider (*ang. Dostawca Usług Internetowych*)

<sup>2</sup>Request For Comments - IETF

RFC	Opis
RFC 1771 1995	<i>A Border Gateway Protocol 4 (BGP-4)</i> Dokument definiuje protokół BGP w wersji 4
RFC 1772 1995	<i>Application of the Border Gateway Protocol in the Internet</i> Dokument wraz z RFC1771 definiuje sposoby użycia BGPv4 w Internecie
RFC 1773 1995	<i>Experience with the BGP-4 protocol</i> Dokument poddaje analizie zgodność stworzonego standardu założeniami zawartymi w Drafcie stworzonym przez IESG (Internet Engineering Steering Group)
RFC 1774 1995	<i>BGP-4 Protocol Analysis</i> Analiza protokołu BGPv4 pod kątem zgodności z Drafetm IESG. Szczególny nacisk jest kładziony na aspekt skalowalności protokołu
RFC 1930 1996	<i>Guidelines for creation, selection, and registration of an Autonomous System (AS)</i> Dokument zawiera zalecenia dotyczące tworzenia i rejestracji Systemów Autonomicznych (AS)
RFC 1966 1996	<i>BGP Route Reflection: An alternative to full mesh iBGP</i> Dokument opisuje technikę 'BGP Route Reflection' pozwalającą na uniknięcie topologii full-mesh dla BGP w ramach jednego Systemu Autonomicznego
RFC 1997 1996	<i>BGP Communities Attribute</i> Dokument opisuje dodatkowe atrybuty ścieżek przesyłane poprzez BGP
RFC 1998 1996	<i>An Application of the BGP Community Attribute in Multi-home Routing</i> Dokument przedstawia możliwości stosowania atrybutów typu community w celu uproszczenia konfiguracji i zarządzania sieciami Internetowymi

RFC 2270 1998	<i>Using a Dedicated AS for Sites Homed to a Single Provider</i> Zalecenia służące ograniczeniu tempa przyrostu przyznawanych numerów systemów AS
RFC 2385 1998	<i>Protection of BGP Sessions via the TCP MD5 Signature Option</i> Dokument opisuje sposób użycia opcji TCP, mogącej przesyłać skrót payloadu protokołu BGP do celów uwierzytelniania i integralności zawartych tam danych
RFC 2858 2000	<i>Multiprotocol Extensions for BGP-4 Inter-Domain Routing</i> Dokument opisuje sposoby przesyłania informacji routingowych dla protokołów innych niż IPv4
RFC 3065 2001	<i>Autonomous System Confederations for BGP</i> Dokument opisuje jedną z technik pozwalających na uniknięcie topologii full-mesh routerów iBGP.
RFC 3562 2003	<i>Key Management Considerations for the TCP MD5 Signature Option</i> Dokument analizuje sposoby generowania kluczy służących do zabezpieczenia payloadu pakietów BGP przy użyciu opcji MD5 protokołu TCP

## 3. Opis protokołu

Wszystkie odniesienia do BGP w niniejszej pracy oznaczają BGP w wersji 4 o ile bezpośrednio w tekście nie napisano inaczej.

### 3.1. Protokoły podkładowe.

Pakiety BGPv4 przesyłane są poprzez protokół TCP<sup>3</sup>, który korzysta z protokołu IP. Standardowym portem dla usługi BGP jest 179. Sesja BGP zestawiana jest pomiędzy *peerami* BGP. Urządzenia uczestniczące w sesji BGP nazywane są także BGP neighbors.

### 3.2. Obsługiwane protokoły

BGP został stworzony by obsługiwać IPv4, lecz z biegiem czasu uzupełniono go o możliwość przenoszenia informacji routingowych opisujących inne protokoły. To rozszerzenie zostało nazwane Multi-Protocol Extension. Dzięki niemu BGP w wersji 4 sprawdza się nie tylko w sieciach IPv4, ale także w obecnie już produkcyjnie wdrażanych sieciach IPv6 i innych (np. przenoszenie etykiet MPLS).

### 3.3. Zestawianie sesji BGP

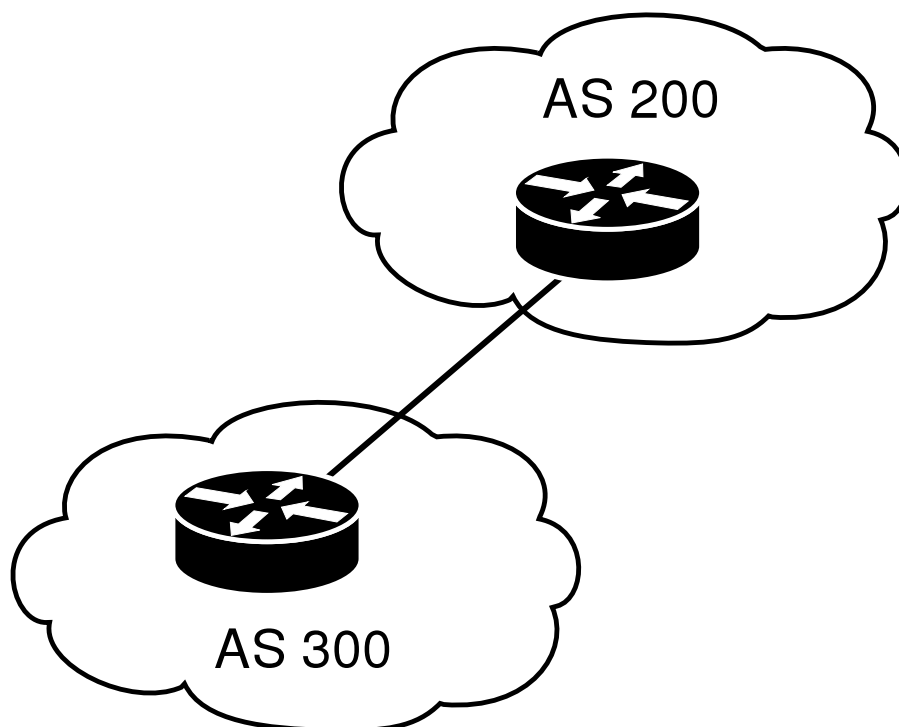
Sesja BGP musi zostać skonfigurowana pomiędzy routerami manualnie na urządzeniach. Zazwyczaj w konfiguracji procesu routingu umieszcza się adres IP peera BGP oraz numer jego systemu autonomicznego. Peer BGP może zostać opisany także innymi parametrami, dotyczącymi m.in. filtracji rozgłaszanych do niego i przyjmowanych od niego informacji routingowych. Możliwe jest także zdefiniowanie niektórych atrybutów jak np. globalny i lokalny next-hop<sup>4</sup> dla celów obsługi routingu IPv6.

Dla celów zestawiania sesji BGP konieczna jest obustronna widoczność urządzeń za pomocą protokołu IP. Nie jest wymagane ich połączenie za pomocą wspólnego linku warstwy drugiej ISO/OSI.

---

<sup>3</sup>Transmission Control Protocol

<sup>4</sup>Adres IP następnego routera dla wysyłanego pakietu IP. Adres ten służy jedynie do odnalezienia adresu warstwy 2giej ISO/OSI tego urządzenia oraz lokalnego interfejsu wyjściowego.



Rysunek 1: Sesja eBGP

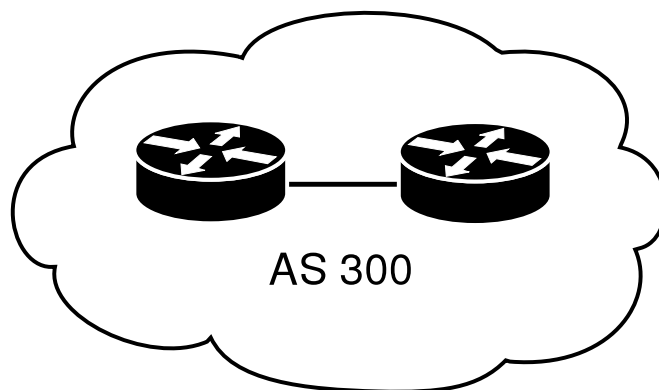
### 3.4. External BGP

Sesja BGP zestawiona pomiędzy dwoma *peerami* należącymi do różnych AS (Autonomous System) nazywana jest eBGP - External BGP. External BGP stosowany jest na styku operator-operator lub klient-operator. Przykład sesji External BGP przedstawiony jest na Rysunku 1.

### 3.5. Internal BGP

Sesja BGP zestawiona pomiędzy routerami należącymi do tego samego AS nazywana jest iBGP - Internal BGP. Sesje takie zestawiane są wtedy, gdy operator posiada peeringi BGP na różnych routerach należących do jego sieci (w obrębie jednego systemu autonomicznego - operatorzy mogą korzystać z wielu AS w danym momencie). Dzięki temu eliminowana jest konieczność terminowania peeringu BGP klientów i innych operatorów na jednym, wydzielonym routerze systemu autonomicznego. Schemat sesji Internal BGP przedstawiony jest na Rysunku 2.

W przypadku korzystania z iBGP w obrębie jednego systemu autonomicznego na wielu urządzeniach sieciowych wymagane jest, by wszystkie urządzenia uczestniczące w routingu BGP były połączone sesjami iBGP, tworząc topolo-



Rysunek 2: Sesja iBGP

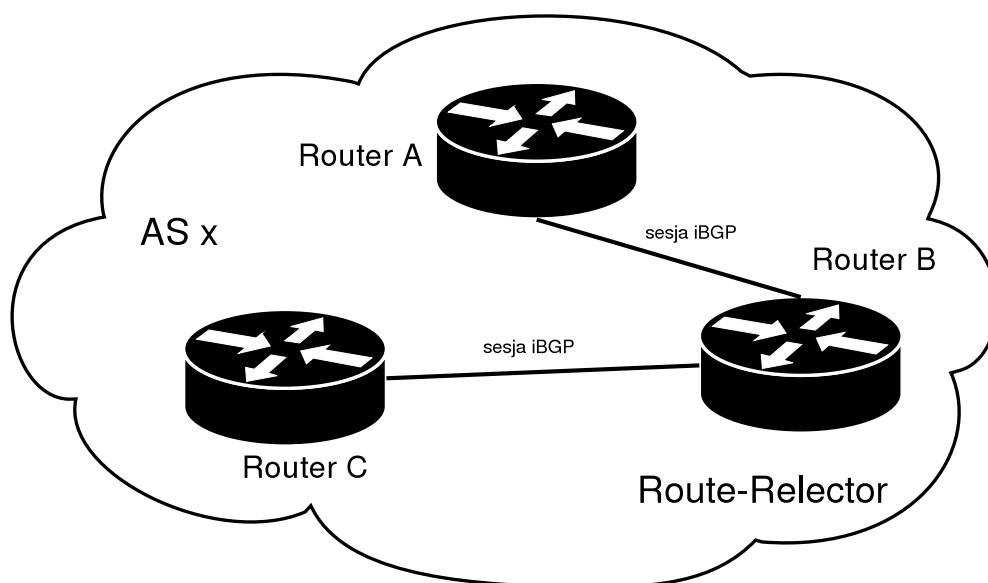
gię połączeń typu full-mesh. Jest to przyczyna wielu niedogodności w fazie projektowania i eksploatacji w odniesieniu do sieci, na którą nakłada się takie wymagania. Wymaganie to bierze się z założenia, że routery iBGP domyślnie nie redystrybuują informacji routingowych pozyskanych od peerów iBGP do pozostałych peerów w tym samym AS.

Uniknięcie tych problemów jest możliwe dzięki zastosowaniu tzw. „reflektorów BGP” oraz konfederacji AS.

### 3.5.1. Route-reflectors

Reflektor BGP jest urządzeniem obsługującym sesję iBGP, które jest skonfigurowane tak, by przekazywać informacje pozyskane podczas sesji iBGP do pozostałych peerów iBGP. Dzięki temu zmniejsza się liczba linków koniecznych do zestawienia pomiędzy urządzeniami sieciowymi, co jest ważne w przypadku posiadania dużej liczby routerów w ramach jednego AS. W efekcie zmniejsza to ilość informacji routingowych przesyłanych w systemie autonomicznym. Zmniejsza to zatem ryzyko degradacji jakości ruchu IP w sieci z powodu awarii linków na których zestawiane są sesje iBGP, a w konsekwencji zmian w trasach, którymi przesyłane są pakiety IP. Mniejsze jest także zapotrzebowanie na zasoby CPU i pamięć operacyjną urządzeń obsługujących BGP z powodu mniejszej liczby peeringów. Mniejsza liczba sesji BGP, to także mniej prefiksów i tras do obsłużenia na pojedynczym urządzeniu sieciowym.

Przykład konfiguracji sieci korzystającej z reflektora BGP przedstawiony jest na Rysunku 3.



Rysunek 3: Route-reflector

### 3.5.2. Konfederacje AS

Konfederacja AS jako przykład techniki wydatnie zwiększającej skalowalność protokołu BGP opisana jest na stronie 52.

## 3.6. CIDR

CIDR jest nowym schematem adresacji w Internecie, pozwalającym na bardziej efektywne przyznawanie zakresów IP. Przed wprowadzeniem CIDR, pula adresów IP dzieliła się, ze względu na długość maski sieci, na zakresy oznaczane literami A,B,C,D,E z których trzy pierwsze definiowały adresację IP ogólnego przeznaczenia (D to adresy zarezerwowane dla transmisji multicastowych. E to pula testowa, zarezerwowana przez IANA). Przypisane do nich maski sieci:

- A - 8 bitów
- B - 16 bitów
- C - 24 bity

sprawiły, że zazwyczaj przyznane adresy IP nie były do końca wykorzystywane. Działo się to szczególnie w przypadku klas A oraz B, które zawierają odpowiednio 16777214 oraz 65534 użytecznych adresów IP. Dzięki wprowadzeniu CIDR, przyznawane klasy mogą mieć dowolne rozmiary masek sieci. Z tej funkcjonalności korzysta BGP, który nie tylko przekazuje adresy wraz z

ich maskami ale także potrafi łączyć podzakresy w większe całości i w takiej postaci przekazywać je do swoich peerów. Dzięki temu komunikaty routingowe mogą być mniejsze a przetwarzanie tablic routingowych szybsze. Oszczędza to pasmo sieciowe oraz czas i moc procesorów zainstalowanych w routerach. Taka forma operacji na adresacji IP nazywana jest agregacją adresów.

Wartym zauważenia jest fakt, że prefiksy posiadające maski sieci dłuższe niż 24 bity są filtrowane na routerach części operatorów w celu ograniczenia wpisów routingowych w urządzeniach routujących pakiety IP. Stąd też niepisany standardem jest rozgłaszanie poprzez BGP jedynie prefiksów o długości masek mniejszej od 24 bitów. Prefiksy o dłuższych długościach masek mogą nie być widoczne w poszczególnych częściach Internetu w zależności od konfiguracji filtrów routerów obsługujących protokół BGP.

### 3.7. Nagłówek BGP

Każda z wiadomości BGP składa się z nagłówka, który zawiera:

1. Marker - zawierający informacje uwierzytelniające, które może przewidzieć odbiorca wiadomości,
2. Długość - całkowita długość pakietu BGP podana w bajtach,
3. Typ wiadomości BGP

### 3.8. Typy wiadomości BGP

**Open** Wiadomość otwierająca sesję BGP, pierwsza po nawiązaniu sesji TCP.

Podczas niej przesyłana jest informacja o ASN peera transmitującego wiadomość. Wiadomość Open zawiera następujące informacje:

1. Wersję protokołu BGP,
2. Lokalny ASN,
3. Proponowany hold-time, wartość, która określa czas nieaktywności (brak komunikatów od peera), po której sesja BGP jest zrywana,
4. Identyfikator BGP - zazwyczaj adres IP lokalnego interfejsu sieciowego,
5. Opcjonalne parametry.

**Update** dzięki której routery przekazują informacje o dostępności prefiksów wraz z ich atrybutami. Wiadomość Update zawiera pola:

1. Długość wektora opisującego prefiksy IP oznaczone jako niedostępne,
2. Usunięte (niedostępne) prefiksy adresów IP,
3. Rozmiar wektora opisującego atrybuty ścieżek
4. Ścieżki AS wraz z ich atrybutami, informacjami o preferencjach, communities, konfederacjach AS itp.
5. Prefiksy adresów IP oznaczonych jako dostępne.

**Notification** Wiadomość ta informuje peera BGP o błędzie. Składa się ona z kodu błędu, subkodu oraz danych związanych z zaistniałą sytuacją awaryjną. Po transmisji tej wiadomości sesja BGP jest kończona. Wiadomość ta może dotyczyć w szczególności błędów w protokole lub przekroczenia wartości timeoutu nieaktywnej sesji.

**Keep-alive** Wiadomość składająca się jedynie z nagłówka BGP. Jest wysyłana w celu podtrzymania sesji. Częstotliwość jej wysyłania jest większa niż wynegocjowany parametr Hold-Time, co w przypadku poprawnie działającego łącza uniemożliwia zerwanie sesji z powodu timeoutów.

### 3.9. Atrybuty

Wraz z rozgłaszanymi przez peera BGP prefiksami przekazywane są dodatkowe parametry pozwalające na nadanie konkretnym wpisom pewnych wartości, od których może zależeć zachowanie routerów odbierających i przetwarzających wiadomości routingu BGP. Najczęściej mówią one o wadze danej ścieżki routingowej i wspomagają proces decyzyjny wyboru najlepszych tras. Te wartości, przypisane do konkretnych prefiksów sieciowych, nazywane są atrybutami. Standardowe atrybuty to:

1. Weigth (WEIGHT) - atrybut lokalny dla routera. Nie jest eksportowany do peerów BGP. Oznacza wagę konkretnej ścieżki AS (AS path), na podstawie której wybierane są najlepsze trasy, dzięki którym można osiągnąć prefiksy,
2. Local preference (LOCAL\_PREF) - atrybut podobny do Weigth, lecz rozgłaszany w obrębie AS. Wyższa wartość local preference rozgłaszana z danego routera, oznacza, że to właśnie przez niego będą „wychodziły” pakiety skierowane do konkretnego prefiksu,

3. Multi-exit discriminator (MED) - atrybut, który sugeruje przyległemu (*ang. adjacent*) systemowi autonomicznemu (takiemu, z którym lokalny system posiada zestawioną sesję BGP) o sugerowanej ścieżce pakietów przychodzących lub tranzytowanych przez lokalny system autonomiczny,
4. Origin (ORIGIN) - atrybut opisujący pochodzenie danej wpisu routinowego, rozgłaszanego przez BGP. Może przyjmować wartości
  - IGP - ścieżka redystrybuowana do BGP z protokołu typu IGP lub wynikająca z wewnętrznej konfiguracji ścieżek routingu,
  - EGP - ścieżka uzyskana poprzez protokół eBGP,
  - Incomplete - wartość opisująca nieznanne pochodzenie ścieżki routinowej.
5. AS Path (AS\_PATH) - ścieżka AS. Atrybut ten opisuje systemy autonomiczne, przez które musi przejść pakiet by osiągnąć docelowy prefix,
6. Next-Hop (NEXT\_HOP) - atrybut definiujący adres IP, który powinien być wykorzystany jako Hop dla pakietów, jeżeli wpis routingowy opisywany przez ten atrybut zostanie wybrany.
7. COMMUNITIES - sposób oznaczenia prefiksów na podstawie którego mogą być podejmowane akcje związane z routingiem. Dobrze znanymi przykładami communities są:
  - NO-EXPORT - zabrania rozgłaszania oznaczonej w ten sposób trasy do innych peerów eBGP
  - NO-ADVERTISE - zabrania rozgłaszania do jakiegokolwiek peera BGP (iBGP oraz eBGP)
  - NO-EXPORT-SUBCONFED - zabrania rozgłaszania tras do peerów eBGP oraz do innych sub-AS w ramach konfederacji.

Atrybuty dzielą się na te dobrze określone (*well-known*) oraz opcjonalne (*optional*). Do tych pierwszych należą:

- ORIGIN, AS\_PATH, NEXT\_HOP, LOCAL\_PREF

Do opcjonalnych zaś należą m.in.:

- MED, COMMUNITIES

Atrybuty typu well-known muszą być obsługiwane przez wszystkie urządzenia obsługujące BGP. Atrybuty opcjonalne mogą być ignorowane.

Kolejny podział jest związany ze sposobem propagacji atrybutów. Te, które nie są redystrybuowane przez urządzenia BGP, są nazywane *non-transitive attrs* (atrybuty nieprzechodnie). Należą do nich:

- LOCAL\_PREF (tylko w ramach iBGP, nieobsługiwany na peeringach eBGP), MED

Te, które mogą być rozgłaszane do innych peerów, nazywane są atrybutami przechodnimi (*transitive attrs*)

- ORIGIN, COMMUNITIES, AGGREGATOR

### 3.10. Strategie routingu

Operator stosujący BGP na brzegu swojej sieci może aktywnie tworzyć politykę routingu. Może być ona zależna od wielu parametrów, lecz najpopularniejsze strategie to:

- użycie otrzymywanych prefiksów w komunikatach BGP do podejmowania decyzji o włączeniu trasy do tablic routingu, ewentualnie o eksportowaniu ich do konkretnych peerów BGP.
- filtrowanie prefiksów w zależności od ścieżki AS. Najpopularniejsze w tym przypadku jest użycie tzw. wyrażeń regularnych mogących służyć jako wzorzec dopasowania do danych typów ścieżek. Np. wzorzec `_12464_1887_` oznacza ścieżkę, które zawiera przyległe (*ang. adjacent*) ASy o numerach (ASN) 12464 oraz 1887. Wzorce umożliwiają konfigurację wielu typowych konfiguracji BGP. Najprostszym przykładem jest tzw. *leaf AS*, czyli AS nietranzytowy, posiadający połączenia do dwóch lub więcej peerów BGP. W takiej konfiguracji najczęściej zezwala się na rozgłaszanie ścieżek pasujących do pustego wzorca, czyli do `^$`. Taki wzorzec pasuje do lokalnego numeru AS, który będzie jako jedyny rozgłaszany do peerów BGP.
- dokonywanie filtrowania prefiksów na podstawie communities.

Dzięki możliwości filtrowania prefiksów i ścieżek w BGP na podstawie podanych wyżej parametrów możliwe jest uproszczenie zarządzania dużymi sieciami

IP. Wybrane obszary sieci - najczęściej dzielone na tzw. peering, kraj, świat - mogą być ogłaszane peerom BGP, a prefiksy IP pochodzące od nich, w procesie filtrowania, mogą być rozgłaszane innym, przyłączonym sieciom.

## Część II.

# Analiza aspektu bezpieczeństwa protokołu BGP

## 4. Uwarunkowania historyczne bezpieczeństwa protokołu

Protokół BGP rozwijany był w warunkach akademicko-inżynierskich dla potrzeb wtedy już (początek lat 90-ych) licznej, lecz dobrze znającej się grupy organizacji stanowiących trzon ówczesnego Internetu. Zaufanie, którym ta grupa się darzyła było bardzo duże, a w porównaniu do dzisiejszej rzeczywistości, relatywnie mała liczność grupy nie pozwalała na zachowania niezgodne z ogólnie przyjętymi normami. Było to wymuszone obawą natychmiastowego napiętnowania i wykluczenia z grupy tworzącej innowacyjny projekt. W takich warunkach zaczęto dyskutować o nowym protokole, który uprościłby zarządzanie tablicą routingu w Internecie.

W dotychczasowej, scentralizowanej formule istniał tzw. Internet Backbone - szkielet Internetu. ARPANET kończył swoją działalność zgodnie z decyzją amerykańskiego Department of Defense - Departament Obrony. Ostatni węzeł ARPANETu został wyłączony w 1989r..

Centralny routing w Internecie zaczął być obsługiwany przez NFSnet - sieć National Science Foundation - Narodowa Fundacja Nauki, cywilną agencję Stanów Zjednoczonych. Ograniczenia z tym związane, czyli niekomercyjność „ruchu” w sieci były przeszkodą dla firm chcących świadczyć płatne usługi na bazie powstałego Internetu. Zaczęto tworzyć połączenia pomiędzy powstającymi wtedy szybko ISP (Internet Service Provider). Taki kierunek rozwoju Internetu przyspieszył upadek NSFnet i wprowadzenie zdecentralizowanego routingu opartego o BGP.

30 kwietnia 1995 NFSnet jako szkielet Internetu przestał istnieć. Sieci zaczęły organizować połączenia pomiędzy sobą za pomocą tzw. NAP (Network Access Points), które przekształciły się w dzisiejsze IXP (Internet eXchange Points).

BGP zaczął być podstawowym protokołem routingu między-operatorского w Internecie. I jak czas pokazał, z jednej strony protokołem niezwykle elastycz-

nym, z drugiej „nieświadomym” nadchodzących zagrożeń.

## 5. Rozgłaszanie prefiksów

Największym zagrożeniem dla stosowanego obecnie w Internecie BGP wydaje się być rozgłaszanie prefiksów IP przez nieuprawnione speakery BGP. Rozgłaszanie jednego prefiksu przez wiele speakerów BGP, chociaż dopuszczalne, w przypadku gdy jest to nieustalone pomiędzy operatorami, może spowodować duże problemy związane z routowaniem w Internecie.

Jeżeli urządzenie odbierające informacje routingowe BGP uzna na podstawie zaimplementowanego kryterium (zazwyczaj długości ścieżki BGP), że pakiety IP do konkretnego prefiksu należy kierować poprzez ścieżkę prowadzącą do nieuprawnionego AS, wtedy pakiet taki najprawdopodobniej zostanie odrzucony w fazie filtracji IP u operatora docelowego (rozgłaszającego prefiks pod nieuprawnionym AS). Rozgłoszenie dużej liczby prefiksów może doprowadzić do koncentracji ruchu na wybranych łączach (prowadzących do operatora rozgłaszającego prefiksy) a w konsekwencji do przeciążenia routera operatora oraz routerów operatorów upstreamowych.

Historia routingu BGP notuje kilka przypadków takich sytuacji, z których dwa najpoważniejsze przedstawione są poniżej.

### 5.1. Incydent 7007 (AS 7007 Incident)

#### 5.1.1. Opis

25 kwietnia 1997 r. na listę NANOG (The North American Network Operators' Group) wysłany został e-mail o następującej treści (tłumaczenie z ang.):

From: Stephen A Misel  
Date: Fri Apr 25 13:20:40 1997  
Subject: Wow, AS7007!

*Byłem zalogowany na routerze 7505 dzisiejszego popołudnia, kiedy nagle zniknął cały Internet. Z początku pomyślałem, że to moja wina, lecz po chwili poszukiwań okazało się, że AS o numerze 7007 rozgłasza należące do MNIE prefiksy. Nie była to wprawdzie należąca do mojej firmy cała przestrzeń adresowa - posiadamy bowiem kilka klas /18, a wydaje się, że jedynie pierwsza klasa /24 z każdej z klas /18 była rozgłaszana.*

*Gdy znaleźliśmy maszynę na końcu klasy /18, wykonaliśmy szybki whois dla AS 7007 - Florida Internet Exchange - i zadzwoniliśmy*

*do nich.*

*Florida Internet Exchange stwierdziło, że to ich klient rozgłasza błędne prefiksy i natychmiast odcięło ich router. Kilka chwil później wszystko ustabilizowało się i zaczęliśmy widzieć prawdziwe trasy.*

*Poprawcie mnie, jeżeli się mylę, ale:*

*(1) Przeczytamy niedługo w każdym czasopiśmie komputerowym, gazecie i usłyszymy w telewizji o „końcu internetu”,*

*(2) Listy dostępu na operatorach sieci szkieletowych powinny być temu zapobiec,*

*(3) Rejestratorzy internetowi powinni czy też nie powinni byli temu zapobiec?*

*Mam nadzieję, że ta luka w bezpieczeństwie zostanie naprawdę szybko załatwana.*

*Steve*

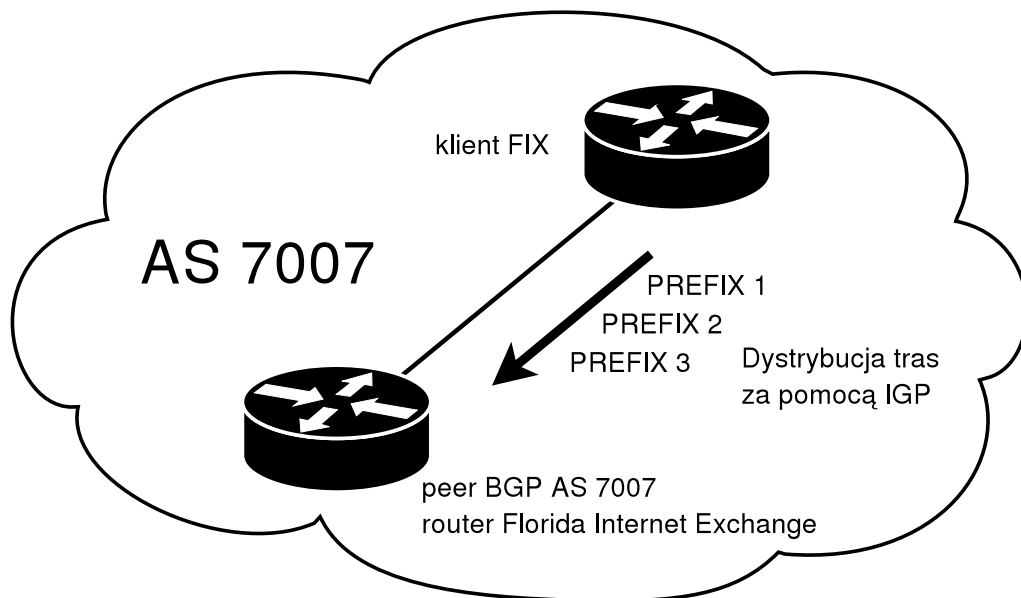
Powyższy list opisuje sytuację, w której operator internetowy, posiadający własny ASN rozgłasza prefiksy nie należące do niego. Taka sytuacja była (i dalej jest) możliwa, ponieważ nie jest stosowany na szerszą skalę mechanizm pozwalający na autoryzację AS do rozgłaszania konkretnych prefiksów. Wprawdzie rejestratorzy internetowi (RIR) prowadzą takie bazy (whois), lecz nie istnieją mechanizmy służące importowi tych baz i korzystania z nich w procesie weryfikacji poprawności rozgłaszanych tras na routerach.

### **5.1.2. Analiza przypadku**

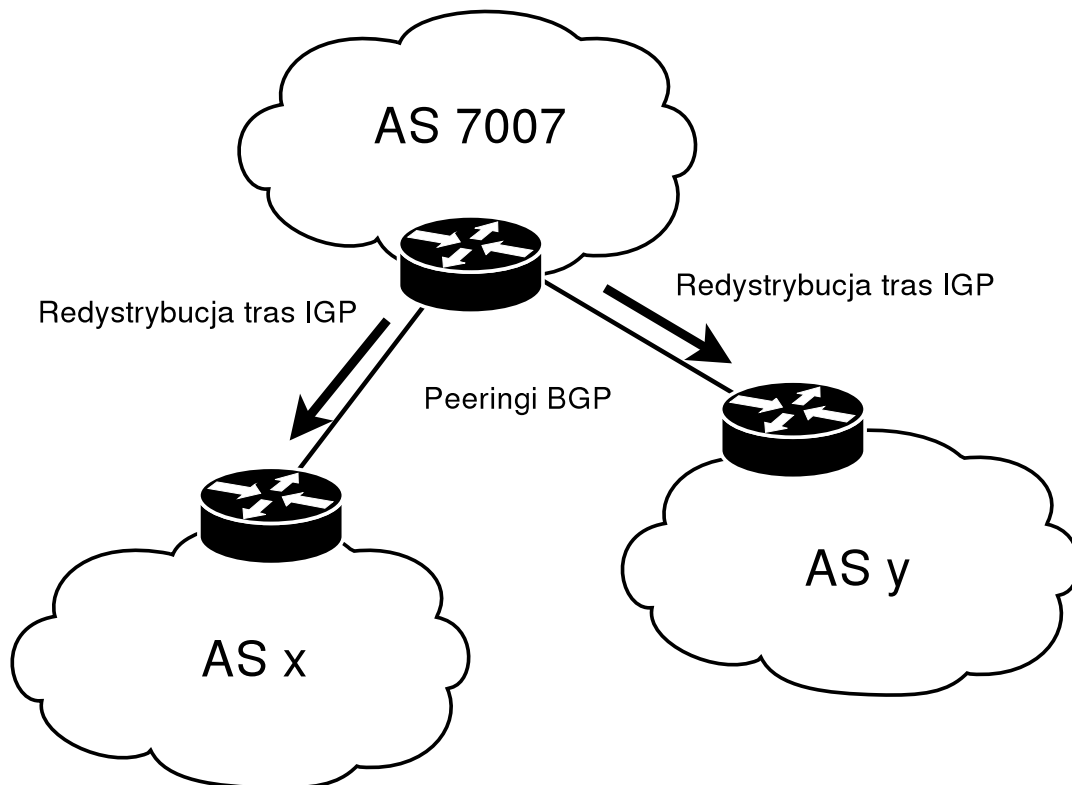
W opisanym przypadku AS7007 otrzymał od swojego klienta trasy za pomocą (najprawdopodobniej) jednego z protokołów IGP (IGRP, EIGRP, OSPF, RIP) lub prywatnego BGP. Nazwałem to Fazą 1 incydentu. Fazę tę przedstawia Rysunek 4. Router kliencki uzupełnia tablicę dostępnych prefiksów routera FLIX o kolejne prefiksy - oznaczone tutaj jako PREFIX n. Router należący do FLIX po aktualizacji swych tablic rozpoczyna proces redystrybucji tras routingowych. Trasy uzyskane za pomocą protokołu IGP są umieszczane w kolejnych pakietach update protokołu BGP.

Proces rozgłaszania prefiksów za pomocą BGP, oznaczony tutaj jako Faza 2 incydentu, przedstawia Rysunek 5.

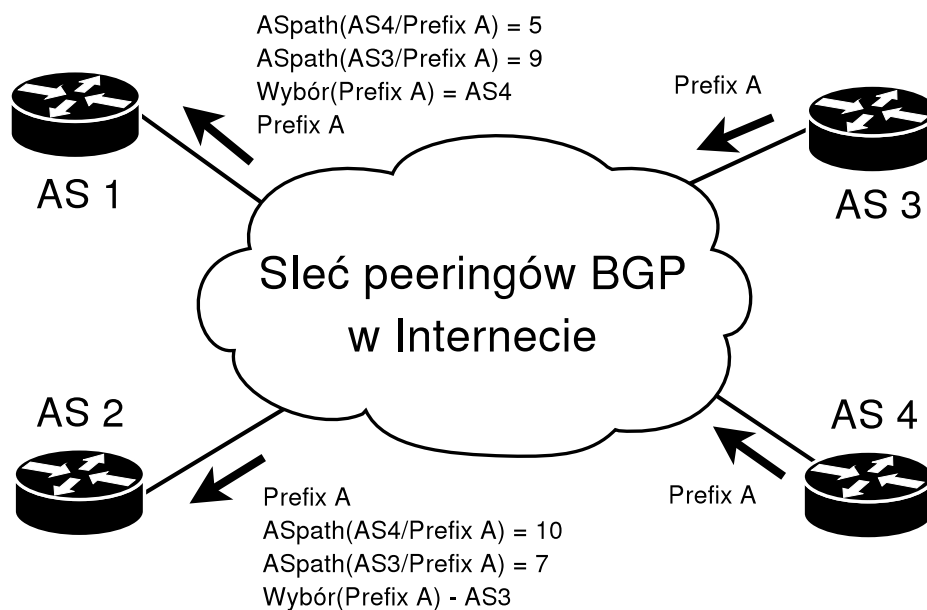
Rozgłoszone do providerów upstreamowych trasy zostają ogłoszone do kolejnych AS. Trasy do rozgłoszonych prefiksów powoli zaczynają się zmieniać.



Rysunek 4: Incydent 7007. Faza 1



Rysunek 5: Incydent 7007. Faza 2



Rysunek 6: Incydent 7007. Wybór tras

Proces zmiany jest zależny od kilku czynników. Najważniejszym jest długość ścieżki AS (AS path). Jeżeli ścieżka do prefiksu osiągalnego przez AS7007 jest krótsza od tej, którą można osiągnąć za pomocą uprawnionego AS, wtedy wpis w tablicy routingu FIB<sup>5</sup> przełącza się na trasę lepszą według zadanych reguł, czyli tę osiągalną poprzez krótszą ścieżkę AS. Schemat wyboru ostatecznej trasy, oparty na długości ścieżki AS, czyli podstawowej metodzie wyboru tras w BGP przedstawia Rysunek 6.

Routery obsługujące AS3 oraz AS4 rozgłaszają ten sam Prefix A. Jeżeli AS1 oraz AS2 otrzymują dwie trasy do Prefiksu A o docelowych ASN 3 oraz 4, wtedy rozpoczyna się proces wyboru trasy.

Router o ASN równym 1 otrzymuje ścieżkę w do Prefiksu A w formacie:

```
Prefix A: Ścieżka ASa, ASb,
          ASc, ASd, AS4, ASPathLength = 5
Prefix A: Ścieżka ASe, ASf, ASg, ASH, ASi,
          ASj, ASk, ASl, AS3, ASPathLength = 9
```

Działając na podstawie kryterium najkrótszej ścieżki przy wyborze optymalnej trasy w tablicy FIB routera obsługującego AS1 znajdzie się wpis, by pakiety z adresem docelowym IP zawierającym się w Prefiksie A zostały skierowane ścieżką, której docelowym ASN jest 4.

Analogicznie dla routera obsługującego AS2.

<sup>5</sup>Forwarding Information Base - tablica routingu routera

Warto zauważyć, że sam proces rozgłaszania jednego prefiksu pod dwoma ASN jest dopuszczalny, lecz wymaga zgody administratorów obu systemów autonomicznych jak i właściciela prefiksu. Dzięki temu możliwy jest do zrealizowania „tani” multihoming. Multihoming, czyli łączność do co najmniej dwóch operatorów udostępniających zasoby internetu jest tanią i prostą formą zabezpieczenia się przed utratą łączności do jednego z operatorów jak. Zaletą tak realizowanego multihomingu jest także load-balancing na poziomie ścieżek AS. Prefix docelowy jest osiągany przez krótszą ścieżkę AS.

26 kwietnia operatorzy sieci FLIX wysłali na listę NANOG wyjaśnienie zaistniałej sytuacji (fragmenty tłumaczenia z ang.):

Subject: 7007 Explanation and Apology

From: Vincent J. Bono

Date: Sat Apr 26 19:42:16 1997

*Witajcie,*

*(...)*

*25 kwietnia 1997 o godz 11:30 nasz router brzegowy zaczął otrzymywać od naszego klienta zestaw informacji routingowych obejmujących całą tablicę routingu (właściwie zestaw zawierający 23000 tras).*

*Nie posiadaliśmy list dystrybucyjnych zainstalowanych na łączu do tego klienta.*

*(...)*

*Telefon ciągle dzwonił a my zauważyliśmy, że trasy do 7007 zaczęły być osiągalne przez nasze peeringi BGP. Wpadliśmy w panikę, i o 12:15 odłączyliśmy nasze urządzenia sieciowe od prądu. Wtedy o 12:25 otrzymaliśmy telefon od Zespołu Zarządzania Siecią Sprint, w którym poinformowano nas, że nasze upstreamowe połączenie DS-3 zostało wyłączone. Oni ciągle widzieli te trasy. Powiedzieliśmy im, że nie ma problemu, jako, że nasze routery były i tak wyłączone.*

*(...)*

*Telefony od ISP z całego świata dzwoniły do ok 16:45.*

*W rozmowie z Zespołem Zarządzania Siecią Sprint, dowiedziałem się, że nie mogli usunąć tras z routerów, ponieważ ciągle się pojawiały.*

*Wydaje się także, że jeden z dużych operatorów zaczął rozgłaszać te trasy na zachodnim wybrzeżu i trwało to do ok 19:00.*

*Byli u nas inżynierowie od producenta naszego sprzętu sieciowego i do ok 13:00 dzisiaj badali router, starając się wyjaśnić to wszystko niewłaściwą konfiguracją. W tej chwili mamy założone odpowiednie listy dystrybucyjne na każdym z naszych łącz.*

*(...)*

*Nie zrobiliśmy tego specjalnie. Nie jestem pewien, czy potrafiłbym powtórzyć to zachowanie routera BGP.*

*(...)*

*Chciałbym także podziękować AT&T WorldNet, NASA Sciences Institute oraz Net Access Corporation. Oni nie zadzwonili po wyjaśnienia, ale po to, by zaoferować pomoc.*

*Sincerely, MANAGEMENT ANALYSIS, INCORPORATED*

*Vincent J. Bono Director Network Services*

Przytoczony list ukazuje ciekawe zjawisko, czyli duży czas propagacji szczątkowych informacji routingowych. Pomimo natychmiastowej reakcji, błędne wpisy routingowe były widoczne przez kilka godzin po odłączeniu sprzętu sieciowego, co uwidoczniło się w czasie, po którym odbierano telefony od zdezorientowanych operatorów sieci.

Zjawisko to nie zostało w toku dyskusji wyjaśnione. Można zatem podejrzewać, że propagacja takich informacji zależy w dużym stopniu od rozległości sieci peeringów BGP, ponieważ teoretyczne modele BGP wykluczają takie zachowanie sieci złożonych z niewielkiej liczby speakerów BGP. Informacje o niedostępnych prefiksach powinny zostać rozpropagowane w wiadomości Update protokołu. Możliwym wyjaśnieniem jest też błędna implementacja protokołu w ówczesnie stosowanych urządzeniach sieciowych.

### **5.1.3. Podsumowanie**

Prawdopodobna, błędna konfiguracja jednego z urządzeń sieciowych, niekoniecznie uczestniczącego w procesie światowego BGP miała fatalne skutki dla światowego Internetu. Rozgłoszone, nieautoryzowane prefiksy znajdujące się na końcu błędnej ścieżki AS spowodowały brak wzajemnej widoczności większości hostów internetowych.

Jeżeli proces wybierający najlepsze trasy do danego prefiksu uznał (np. na podstawie krótszej ścieżki AS), że należy wysłać pakiety IP w kierunku błędnego AS (AS7007), wtedy komunikacja była niemożliwa, a pakiety najczęściej krążyły aż do odrzucenia pakietu na podstawie pola TTL lub odrzucenia pakietu z powodu niedostępności routera odbierającego pakiety dla AS (po wyłączeniu routera przez operatorów LIX).

Wydarzenie spowodowało długą dyskusję na listach NANOG i wśród operatorów internetowych na świecie. Dyskusja ta doprowadziła do stworzenia kilku projektów, które w założeniach miały eliminować tego typu zjawiska w przyszłości.

## 5.2. 128/9 „disaster”

### 5.2.1. Opis

7 października 1997r. na listę NANOG został wysłany kolejny e-mail informujący o dziwnym stanie tablic BGP na świecie. Harold Willison w liście zatytułowanym „*UUnet routes from hell..*” napisał:

*Tutaj przedstawiam to, co widać z sieci AGIS*

```
BGP table version is 35293027, main routing table
version 35293027
76174 network entries (103312/230110 paths) using
15856492 bytes of memory
11155 BGP path attribute entries using 1438184 bytes
of memory
3682 BGP route-map cache entries using 58912 bytes
of memory
9872 BGP filter-list cache entries using 157952 bytes
of memory
Dampening enabled. 18898 history paths, 1516
dampened paths.
```

*Wydaje mi się, czy czy też wszyscy zauważyliście, że AS701 rozgłasza dodatkowo 30000+ prefiksów?*

*A to widać z sieci Digex:*

```
BGP table version is 20811375, main routing table
version 20811375
```

79158 network entries (173777/371074 paths) using  
20851336 bytes of memory 24653 BGP path attribute  
entries using 3466820 bytes of memory  
7506 BGP route-map cache entries using 120096 bytes  
of memory  
0 BGP filter-list cache entries using 0 bytes of  
memory  
Dampening enabled. 8747 history paths, 7465  
dampened paths

*UUnet<sup>6</sup> stwierdził „ojoj” i obiecał to naprawić do 9:15.*

Następnie na liście pojawiły się kolejne informacje potwierdzające ten stan.

### 5.2.2. Analiza przypadku

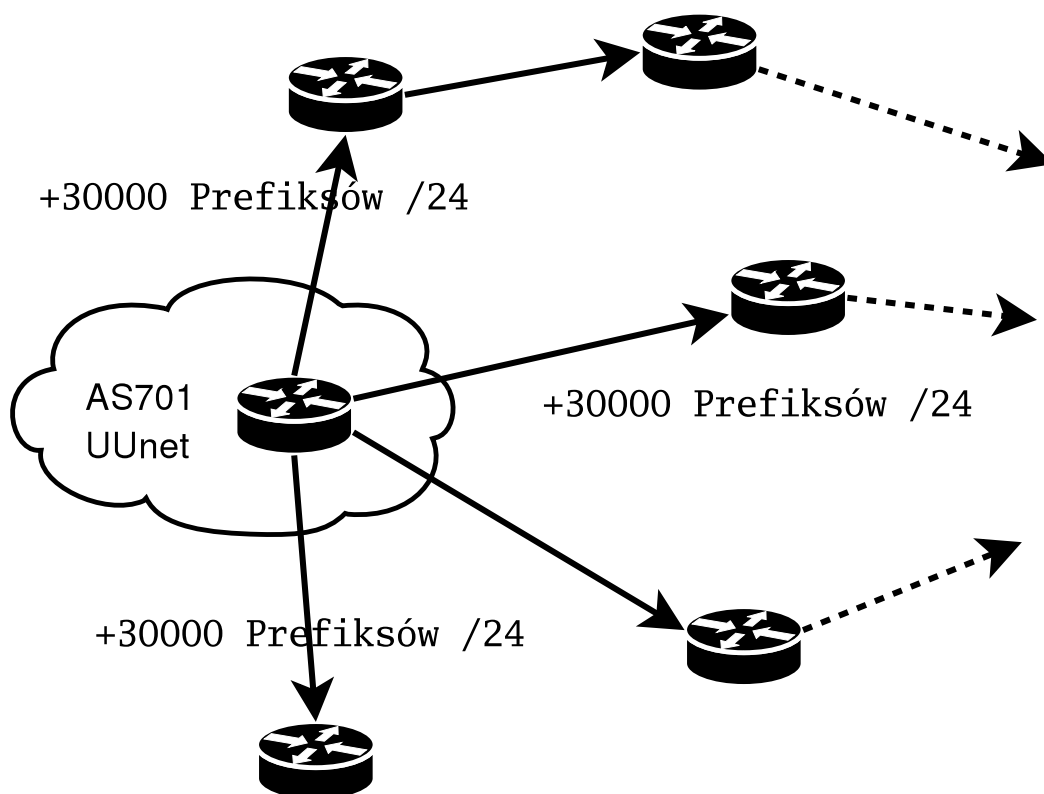
Kolejność zdarzeń:

- AS o numerze 701 (UUnet) wieczorem 7 października z powodu błędu w konfiguracji lub też błędnego oprogramowania zainstalowanego na routerach zaczyna rozgłaszać prefiksy /24 należące do klasy 128.0.0.0/9
- Operatorzy internetowi notują wzrost listy tras zainstalowanych na routerach. Dotychczasowa liczba tras, znajdująca się poniżej 50000 wzrosła do ponad 75000.
- Operatorzy różnych AS zaczynają kontaktować się z operatorami sieci UUnet informując o problemie
- Operatorzy sieci UUnet wyłączają swoje sesje BGP z innymi sieciami
- Sytuacja stabilizuje się

„Wpływ na Internet”:

- Brak widoczności prefiksów z klasy 128.0.0.0/9 o ile algorytm decyzyjny routerów uznał trasę prowadzącą do AS701 za najoptymalniejszą
- Odmowa usługi w odniesieniu do routerów, którym zwiększona o prawie 50% liczba otrzymanych prefiksów przepełniła pamięć operacyjną lub przetwarzanie tras routingowych wysyciło dostępne zasoby CPU. W konsekwencji spowodowało to brak łączności z zasobami internetowymi wielu odbiorców usług.

<sup>6</sup>UUnet posiada ASN o numerze 701



Rysunek 7: 128/9 disaster. Dystrybucja prefixów

Na Rysunku 7 znajduje się schemat rozgłaszania nieautoryzowanych prefixów do swoich peerów BGP.

### 5.2.3. Podsumowanie

Błąd konfiguracyjny lub też błąd w oprogramowaniu routera obsługującego sesję BGP w sieci UUnet spowodował rozgłoszenie bardzo dużej ilości dodatkowych informacji routingowych. Spowodowało to ograniczoną widoczność sieci umiejscowionych w klas, konsternację wśród operatorów sieci oraz kłopoty sprzętowe u ISP, którzy nie przewidzieli podobnych sytuacji.

Przypadek ten zapoczątkował poważną dyskusję na listach operatorów sieci (szczególnie na NANOG) na temat zaufania wobec peerów BGP, w aspekcie ich polityki routingowej. Dyskutowane były różne sposoby filtrowania informacji routingowych w celu niedopuszczenia do podobnych sytuacji w przyszłości.

## 5.3. Wnioski

Wspomniane powyżej dwa przypadki dużych awarii, dotyczących najbardziej krytycznej infrastruktury Internetu, czyli routingu BGP, uświadomiły wielu

operatorom zmiany, które zaszły w tej dziedzinie.

Liczba uczestników routingu BGP wzrosła z niewielkiej grupy na początku jego działania, do kilkunastu tysięcy ASN pod koniec lat 90-tych ubiegłego wieku. Wzajemna kontrola wprowadzanych do sieci routingowych stała się bardzo ograniczona. Łatwo zauważalne stały się jedynie przypadki drastyczne, w których tracona była kontrola nad dużymi partiami sieci, zaś pomniejsze błędy są niezauważane lub toną w statystykach sobie podobnych.

## 5.4. Proponowane rozwiązanie

W celu uniknięcia rozgłaszania prefiksów przez nieautoryzowane systemy autonomiczne można posłużyć się systemami kryptograficznymi. Aby tego dokonać jedna z jednostek koordynujących pracę internetu, np: IANA lub InterNIC, powinna zostać CA dla systemu kryptograficznego.

Para AS źródłowy-Prefix podpisywana byłaby kluczem certyfikowanym przez CA. Posiadaczami certyfikowanych kluczy byłyby jednostki RIR odpowiedzialne za przydziały klas adresowych oraz ASN.

Router wspierający uwierzytelnianie prefiksów posiadałby zaimportowany klucz publiczny CA oraz zaimplementowany w stosie BGP kod służący weryfikacji podpisu elektronicznego (np. X.509).

W ciągu każdej sesji BGP speaker wysyłałby do peera wszystkie posiadane klucze publiczne RIR opatrzone identyfikatorami. Peer BGP sprawdzałby poprawność certyfikacji tych kluczy za pomocą zaimportowanego wcześniej klucza publicznego CA.

Wraz z atrybutem AS\_PATH przesyłany byłby identyfikator klucza RIR oraz podpis złożony kluczem prywatnym odpowiadającym identyfikatorowi. Speaker BGP otrzymujący atrybut AS\_PATH sprawdzałby poprawność podpisu elektronicznego za pomocą klucza o identyfikatorze odpowiadającym identyfikatorowi klucza podpisem którego opatrzona jest para AS źródłowy-Prefiks. W przypadku niezgodności podpisu uaktualnienie informacji routingowej nie byłoby akceptowane.

### 5.4.1. Wnioski

Relatywnie niewielkie przeróbki w oprogramowaniu obsługującym BGP na urządzeniach sieciowych oraz samym protokole (atrybut w którym przesyłany byłby podpis dla ścieżki AS oraz dodany protokół lub typ pakietu definiujący protokół wymiany kluczy RIR).

Dzięki wprowadzeniu powyższego schematu uwierzytelniania par AS źródłowy-Prefiks można wykluczyć przypadki typu „Incydent 7007” lub „129/8 disaster” a także celowe próby zdeorganizowania działania BGP w Internecie poprzez nieprawdziwe informacje routingowe rozprzestrzeniane w sieci BGP.

## 6. Błędy konfiguracyjne

Największe dotychczas badania nad zakresem i wpływem błędnie skonfigurowanych procesów routingu BGP zostały przeprowadzone przez Ratula Mahajana, Davida Wetheralla oraz Toma Andersona z Computer Science and Engineering University of Washington Seattle. W swoim opracowaniu<sup>7</sup> szacują że 0.2% do 1% rozgłaszanych codziennie tras jest oparta na błędnych konfiguracjach procesów routingowych.

Dzielią oni błędy konfiguracyjne na te dotyczące błędnego rozgłaszania informacji źródłowych (origin misconfiguration) oraz te związane z tranzytem tych informacji (export misconfiguration).

### 6.1. Origin misconfiguration

Praca wskazuje, że błędy typu origin misconfiguration zawierają w sobie takie zjawiska jak:

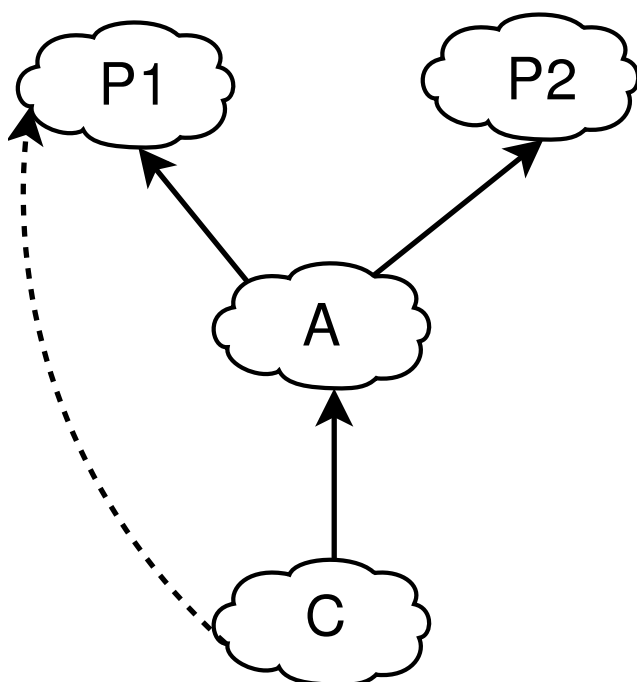
1. Błędna sumaryzacja i agregacja tras. Wynikiem jest większa ilość informacji routingowych przekazywanych do Internetu.
2. Kradzież klas adresowych (address hijacking). Błąd ten wynika z rozgłaszania pod niewłaściwym ASN prefiksów, które rozgłaszane w ten sposób być nie powinny.
3. Propagacja prefiksów, które nie powinny być rozgłaszane poza obrębem danego systemu autonomicznego. Najczęściej dotyczy to klas adresowych przeznaczonych do użytku prywatnego (10.0.0.0/8 itp), adresów pętli zwrotnej (127.0.0.0/8), adresów multicastowych i nieprzyznaných przez żadnego z RIR.

Wynikiem błędów tego typu jest najczęściej:

- Zbędne obciążenie routerów obsługujących BGP, co jest szczególnie ważne w momencie, gdy rozmiary tablic BGP zwiększają się bardzo szybko
- Problem z łącznością. Błędnie rozgłaszane trasy mogą prowadzić do braku widoczności klas adresowych i zakłóceń w łączności w Internecie.

---

<sup>7</sup>Understanding BGP Misconfiguration - <http://www.cs.washington.edu/homes/ratul/bgp/>



Rysunek 8: Prefix based configuration

- Naruszenie polityki routingowej, prefiksy mogą być rozgłaszane na niewłaściwych łączach, co w konsekwencji może prowadzić do niewłaściwej widoczności prefiksów w Internecie, włączając w to wybór mniej optymalnych ścieżek do prefiksów.

## 6.2. Export misconfiguration

Błędy typu export misconfiguration pojawiają się wtedy, gdy trasy są eksportowane niezgodnie z założoną polityką routingową lub gdy eksportowana ścieżka zawierająca listę ASów jest niespójna lub błędna.

Praca rozróżnia błędy związane z:

- Prefix based configuration - przykład konfiguracji widoczny jest na Rysunku 8.

Sieci *P1* oraz *P2* są dostawcami usług internetowych dla *A*. Z kolei sieć *A* dostarcza usługi internetowe dla sieci *C*. Sieć *C* posiada także połączenie z *P1* nieprzechodzące przez *A*.

Gdy łącza działają bez uszkodzeń, polityka routingu jest nienaruszona, lecz w przypadku awarii łącza *A-C* polityka ta może zostać naruszona. Jeżeli łącze to ulegnie awarii, wtedy sieć *P1* stanie się Asem tranzytowym do sieci *C* dla sieci *A*. Ostatecznie sieć *A* zostanie Asem tranzytowym dla sieci *P2* w

przypadku ruchu skierowanego z sieci *P2* do sieci *C*. W części przypadków jest to nieakceptowalne zachowanie, które może wpłynąć na znacznie pogorszenie jakości świadczonych usług internetowych. Dzieje się tak wtedy, gdy łącza są zestawiane tylko dla celów backupowych (linia przerywana na Rysunku 8). Ruch pakietów z sieci *P2* poprzez sieci *A* oraz *P1* jest uznawany w tym przypadku za rzecz niepożądaną, degradującą jakość usług sieciowych na łączach backupowych.

### 6.3. Wnioski

Błędy konfiguracyjne związane z procesem routingu BGP mogą być bardzo groźne. Wspomniane we wstępie zależności od Internetu są bardzo kosztowne w przypadku problemów z łącznością. Z powodu braku połączenia z bankiem internetowym, użytkownicy mogą nie wykonać zaplanowanych operacji finansowych. Informacje o finansach spółek, obecnie transmitowane często za pomocą sieci IP, mogą nie dotrzeć do inwestorów.

Duże problemy w przypadku rozległej awarii związanej z BGP mogą mieć pracownicy pracujący w domu (telepraca). Także inne sektory gospodarki mogą znacznie ucierpieć wskutek niedostępności dużych obszarów sieci internetowej. Warto zauważyć, że do spowodowania takiej awarii wystarczy zaledwie jedno urządzenie sieciowe działające w infrastrukturze BGP. A tych urządzeń jest ogromna ilość, przekraczająca znacznie liczbę obecnie używanych ASN. Często są one obsługiwane przez niekompetentnych administratorów lub używają wątpliwej jakości lub przestarzałego oprogramowania.

Pomyłkę starają się minimalizować filtry pomiędzy peerami BGP. Jednak nie wszyscy operatorzy jest stosują ufając swoim peerom na tyle, by akceptować od nich dowolne informacje routingowe.

Jak pokazuje historia bezpieczeństwa BGP oraz obecne badania na tym polu taktyka taka potrafi spowodować ogromne zakłócenia w Internecie przekładające się na wymierne straty finansowe i niezadowolenie klientów.

## 7. Hijacking

Hijacking odnosi się do praktyk kradzieży zasobów w Internecie. IP address hijacking jest celową kradzieżą klas adresowych, należących formalnie do innego użytkownika sieci.

Powodów, dla którego używana jest ta technika, może być kilka, lecz zazwyczaj są to przyczyny finansowe lub związane z przechwytywaniem informacji.

### 7.1. Hijacking adresów i Spam

Spam, wg definicji z najbardziej znanej polskiej witryny dotyczącej tej tematyki - <http://nospam-pl.net>, to:

*„Spam, określany również 'UCE' unsolicited commercial email (niezamawiany komercyjny email) lub 'UBE' unsolicited bulk email (niezamawiany wielokrotny email) - to takie informacje lub przesyłki, których odbiorca sobie nie zażyczył ani wcześniej na nie się nie zgodził. Najczęściej do niczego mu niepotrzebne, powodujące nieekonomiczne wykorzystanie użytych do przesyłki zasobów, często wywołujące irytację”*

Przesyłki te w początkach swego istnienia były przesyłane bezpośrednio z hostów osób je tworzących lub organizacji, która była reklamowana w przesyłce. Z czasem na większych serwerach pocztowych zaczęto blacklistować<sup>8</sup> adresy IP należące do osób i organizacji spamujących - spamerów.

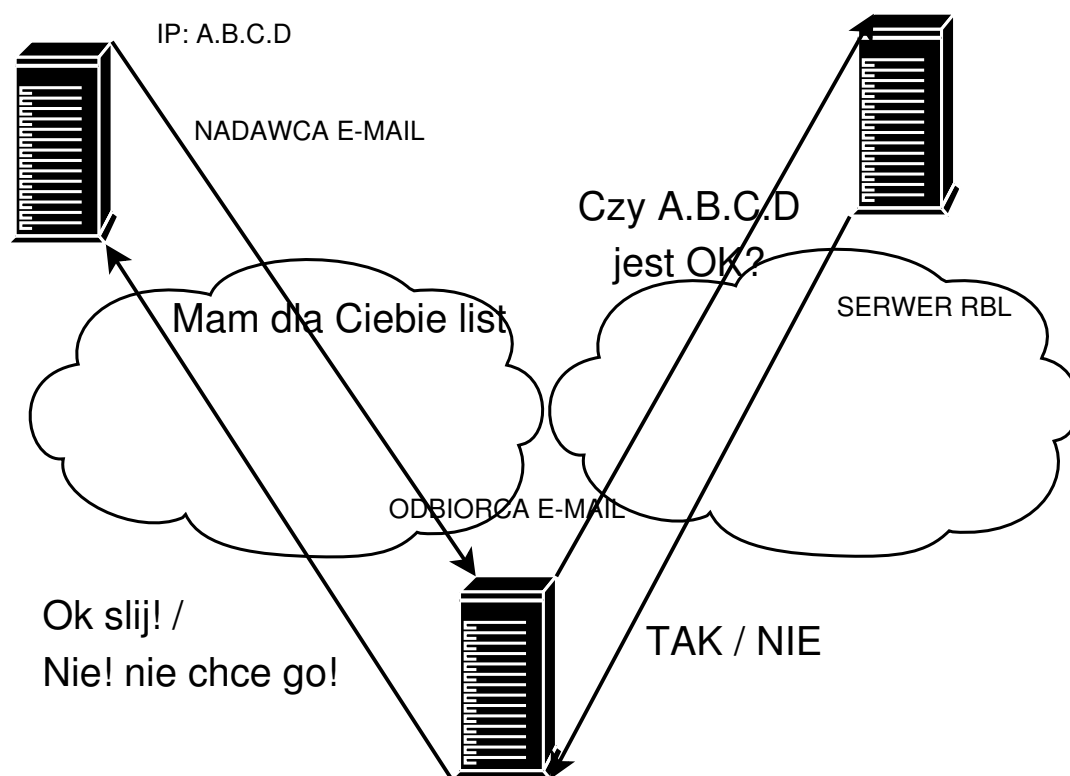
W tym celu stworzono kilka rodzajów baz przechowujących dane IP hostów, z których często otrzymywany jest Spam. Bazy te nazwano RBL - Relay Block List. Oparte są one o serwis DNS, który na odpowiednio zadane zapytanie potrafi odpowiedzieć kodem, mówiącym, czy dane IP jest listowane w danym RBL. Schemat działania usługi RBL jest przedstawiony na Rysunku 9.

Większość serwerów RBL jest publicznie dostępna. Odnośniki do większości światowych i wszystkich polskich serwerów RBL wraz ze szczegółowym opisem zjawiska spamu można znaleźć na stronach Nospam-PL <http://nospam-pl.net>.

Jeden z kilku polskich serwerów RBL - [bl.student.pw.edu.pl](http://bl.student.pw.edu.pl) - strona domowa projektu <http://www.bl.student.pw.edu.pl> - jest zlokalizowany na Politechnice Warszawskiej i administruje nim autor niniejszego opracowania.

---

<sup>8</sup>przechowywać w bazach i podejmować działania na ich podstawie, np. odrzucać pocztę pochodzącą z listowanego IP



Rysunek 9: Schemat działania serwera RBL

Gdy serwer wspierający usługę RBL otrzymuje połączenie ze zdalnego hosta na swój port SMTP może odpytać dowolny serwer RBL o status łączącego się IP. Jeżeli odpowiedź jest pozytywna, wtedy akcja zależy od serwera. Najczęściej stosowane jest rozłączenie sesji SMTP za pomocą kodu 500 na polecenie EHLO/HELO wraz z podaniem opisu błędu. Opis błędu jest zazwyczaj pobierany także z serwera RBL jako pole TXT zapytania DNS.

Inną akcją może być oznaczenie przyjmowanej przesyłki. Najczęściej jest to dodatkowy nagłówek ustawiany w przesyłce lub zmiana tematu wiadomości.

Obecne bazy RBL zawierają bardzo szczegółowe dane na temat tzw. „spamerских hostów”. Celem spamerów jest dostarczenie przesyłki bez odrzucenia jej lub oznaczenia jako potencjalny Spam. Do tego potrzebne są nieznane w bazach RBL adresy IP. W celu uzyskania dostępu do takich adresów można posłużyć się hijackingiem klas adresowych.

### 7.1.1. Teoretyczny przebieg ataku

- Uzyskanie od RIR numeru AS oraz ewentualnie klasy adresowej PI - usługi bezpłatne

- Uzyskanie dostępu do peeringu BGP - wymaganiem jest by peer BGP nie filtrował otrzymywanych od lokalnego peera BGP tras.

Dostęp do peeringu można zdobyć u komercyjnego operatora. Dla celów jednokrotnej zmiany tras BGP, wystarczy chętnie udostępniana przez ISP, usługa testowania połączenia. Taki peering nie będzie już w przyszłości potrzebny, więc wystarczający jest jednokrotny dostęp do usługi.

Innym wariantem jest przejęcie routera z sesją BGP, wymaga to jednak dużej wiedzy oraz jest jednak mało prawdopodobne, ze względu na dobre zabezpieczenia (szczególnie za pomocą list dostępu) większości routerów z sesjami BGP.

- Zestawienie sesji BGP

Do zestawienia sesji BGP będzie potrzebny router z taką możliwością. Wystarczający jest dowolny system typu UNIX wraz z odpowiednim oprogramowaniem (pakiety zebra/quagga).

- Rozgłoszenie prefiksów

Konieczne jest skonfigurowanie systemów routingu tak, by rozgłaszały ważne dla nas prefiksy sieciowe. „Dobrą praktyką” byłoby rozgłaszanie prefiksów, które są od nas odległe w sensie długości ścieżki AS. Wtedy na większych obszarach sieci routing do wskazanych klas zostanie skierowany do lokalnego (spammerskiego) AS. Dalszy AS (zautoryzowany do rozgłaszania prefiksów) oznacza, że więcej systemów autonomicznych w światowym internecie uzna rozgłaszane przez nas prefiksy za lepsze w sensie długości ścieżki AS.

- Ustawienie odpowiednich IP na interfejsach sieciowych hostów i rozesłanie spamu z tych adresów.

### 7.1.2. Wnioski

Przedstawiony wariant ataku na adresację IP może posiadać różne odmiany. Chwilowe rozgłoszenie jednego prefiksu pod przypadkowym Asem (nie zautoryzowanym przez RIR do takiego rozgłoszenia) może nawet nie zostać zauważone w natłoku informacji pochodzących z procesów routingu BGP. Dochody ze Spamów są ogromne, a zainteresowanie tą formą „reklamy” przejawiają nawet zorganizowane grupy przestępcze, które mogą pozwolić sobie na zakup odpowiedniego sprzętu i usług od ISP.

## 8. Podsumowanie analizy aspektu bezpieczeństwa protokołu BGP

W przeprowadzonej analizie nie ujęto problemu bezpieczeństwa BGP w kontekście protokołów warstw niższych.

Szeroko znane są ataki zarówno no warstwę IP jak i TCP. Szczególnie niebezpieczny może być atak TCP Reset polegający na wysyłaniu spoofowanych pakietów TCP z ustawioną flagą RST, co przy spełnieniu kilku dodatkowych warunków może skutkować zerwaniem sesji BGP. Opis tych ataków można znaleźć w Internecie i na grupach dyskusyjnych zajmujących się śledzeniem niedoskonałości systemów informatycznych (BugTraq, VulnWatch itp.) lub w opracowaniach zajmujących się tymi zagadnieniami,

Przedstawione w analizie przykłady dezorganizacji routingu w Internecie spowodowały duże zmiany w środowisku operatorów sieci internetowych. Choć protokół nie został uzupełniony o funkcjonalności, które mogłyby zapobiec takim błędom lub celowym atakom, to podejście administratorów do zagadnień konfigurowania procesów BGP zmieniło się znacznie. Obecnie bardzo duży odsetek ISP stosuje filtry na swoich peeringach BGP. Takie podejście cechuje szczególnie operatorów sieci międzynarodowych. Wynika to z zaufania jakim darzą ich operatorzy będący klientami takich sieci. Jako, że sieć operatora międzynarodowego dostarcza ISP większość prefiksów bardzo trudnym byłoby ustawienie filtrowania tras na takim łączu. Wymagałoby to codziennej rekonfiguracji filtrów BGP u ISP. Dlatego też powszechną praktyką jest stosowanie filtrów na łączach z klientami i małymi sieciami. Na łączach na których otrzymywana jest bogata i zróżnicowana informacji routingowa stosowanie filtrów jest mało prawdopodobne.

Pozostaje mieć nadzieję, że operatorzy wielkich międzynarodowych sieci IP stosują rozsądną i bezpieczną politykę filtrowania BGP na swoich łączach.

## Część III.

# Analiza aspektu skalowalności protokołu BGP

Problem skalowalności BGP w wersji 4 jest zagadnieniem niezwykle ważnym dla rozwoju Internetu. Znaczenie poprawnego routingu jest ogromne, a dynamiczny rozwój globalnej sieci stawia obsługującemu ten proces protokołowi wysokie wymagania.

Protokołowi w głównej mierze zagraża lawinowy wzrost uczestniczących w procesie globalnego routingu urządzeń. W połączeniu z ograniczonymi zasobami od których zależy BGP istnieje potrzeba szybkiego podjęcia działań, które zapobiegną ograniczeniom, które mogłyby spowolnić rozwój Internetu w przyszłości.

## 9. Filtracja BGP

Dużym problemem w utrzymaniu spójności rozproszonego systemu routingowego, jakim jest BGP, jest koordynacja działań w sferze administracyjnej pomiędzy operatorami, których urządzenia wykorzystują protokół BGP.

### 9.1. Koordynacja pomiędzy operatorami

Duża liczba użytkowników protokołu BGP stosuje filtry ramach procesów BGP, które kontrolują zachowanie się protokołu, szczególnie podczas wymiany informacji routingowych w innymi operatorami.

Wymagania biznesowe nakładają na administratorów urządzeń stosowanie, niekiedy bardzo wyrafinowanych, konfiguracji służących pożądanemu rodzajowi dystrybucji i przyjmowania informacji routingowych. Dla przykładu dostęp do jednych sieci może być dla operatora bardziej kosztowny niż do innych, co wpływa na ostateczną ofertę przedstawianą klientowi operatora.

Przykładami sieci o niskich kosztach dostępu są peerigi w ramach IXP - Internet Exchange Points, które są tworzone właśnie w tych celach. Innym przykładem taniej usługi jest dostęp wyłącznie do zasobów operatora.

Dużymi kosztami są jest obłożony dostęp do zasobów zagranicznych oraz innych operatorów, których polityka zakłada płatny dostęp. Zasoby zagraniczne

są droższe, ponieważ wymaga to najczęściej dużych nakładów na infrastrukturę lub środków pieniężnych na dzierżawę infrastruktury innego operatora. Najczęstszy model zakłada jednak kupno pasma sieciowego od operatorów zajmujących się jedynie tranzytem w ujęciu światowym. Przykładami takich operatorów są np: Telia, DTAG, Level3, OpenTransit.

Płatny dostęp do zasobów operatora zazwyczaj nie jest praktykowany, lecz zdarza się tam, gdzie znajdują się operatorzy dominujący na rynku, posiadający znaczny udział z usługach internetowych. Takim przykładem jest Telekomunikacja Polska, która za dostęp (akceptowanie i przyjmowanie prefiksów i ścieżek AS w ramach BGP) żąda od pozostałych operatorów opłat, nierzadko większych niż standardowe opłaty za dostęp do sieci zagranicznych.

Wspomniane warunki działalności operatorów wymagają więc rozróżnienia w usługach sprzedawanych swoim klientom, to zaś przekłada się na odpowiednie konfiguracje filtrów BGP.

W przypadku sprzedaży usługi tranzytu do wydzielonych sieci najczęstszym schematem działania jest

1. Uzgodnienie typu tranzytu (łącze peeringowe, tranzyt ruchu z/do świata, tranzyt ruchu z/do wydzielonych sieci)
2. Zestawienie peeringu BGP pomiędzy klientem a operatorem. Wymaga to zestawienia kanału logicznego, uzgodnienia adresacji IP, numerów AS
3. Uaktualnienie filtrów prefiksów i ścieżek AS na routerach obsługujących BGP operatora.
4. Informacja do operatorów o rozgłaszaniu nowych prefiksów i/lub ścieżek na łączach operatora. Przykład listu informującego operatora międzynarodowego o rozgłaszaniu nowego prefiksu przedstawiony jest poniżej

From: noc@operator.pl

To: bgp-team@inln-operator.com

Subject: Filter update request

*Hello ASxxx-BGP4 Administration team*

*Customer name: Operator S.A.*

*Technical customer name: Operator.*

*Network name: OPERATOR\_NETWORK*

*Router name: WAW-BB3-A2*

*Port: AT/0/4.572*

*We start to advertise new AS and prefix:*

*AS1234 123.45.67.89/24*

*Please, upgrade the filters for BGP session with AS 4321.*

Odpowiedź od operatora zazwyczaj jest podobna do przytoczonej poniżej:

From: noc@intl-operator.com

To: noc@operator.pl

Subject: Re: Filter update request

*Dear customer,*

*our filters have been adjusted.*

Podobna informacja jest wysyłana do wszystkich operatorów posiadających z którymi zestawiona jest sesja BGP4 oraz o których wiadomo, że stosują filtry odnoszące się do protokołu BGP.

W skrajnych przypadkach, dla bardzo dużych operatorów liczba peeringów na których filtrowane są prefiksy/ścieżki AS może być duża.

Najczęstszym sposobem rozpowszechnianie takich informacji jest poczta elektroniczna. Brak potwierdzeń otrzymania przesyłki może skutkować zaginięciem listu pomiędzy kolejnymi mail exchangerami. Taka przesyłka może być także pominięta wśród natłoku innych. Skutkiem tego mogą być problemy z rozgłaszaniem informacji routingowych.

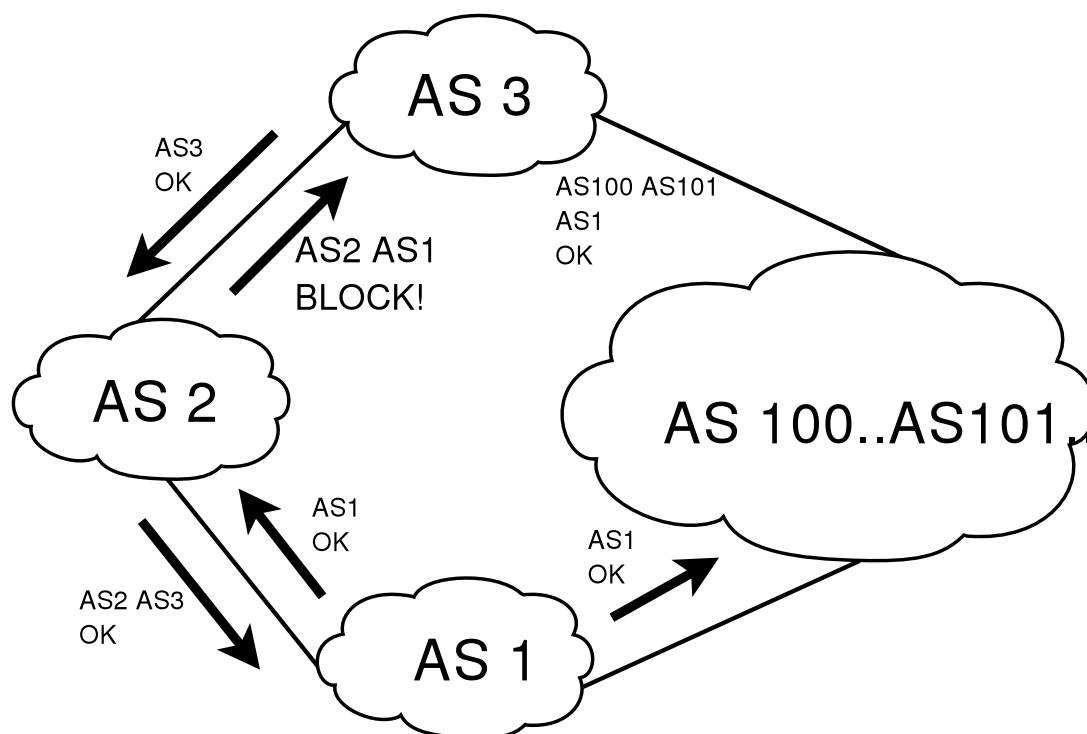
Innym przypadkiem jest brak informacji o stosowaniu filtrów na peeringach BGP. Jako, że taka informacja nie jest najczęściej dostępna publicznie, zarówno operator jak i klient może nie być świadomy przez długi okres czasu tego typu problemów.

Częstą przyczyną braku koordynacji we wprowadzaniu filtrów BGP jest asymetria routingu.

Rysunek 10 przedstawia rozgłaszanie i widoczność podsieci AS1 u operatora AS3.

AS2 będąc operatorem udostępniającym sieć AS3 dla swoich klientów, w szczególności dla AS2, rozgłasza prefiksy ze ścieżką AS2..AS3 do AS1. AS1 akceptuje te ścieżki i umieszcza w swoich tablicach routingu. Jest to best path do sieci AS3.

Problemy w uzgodnieniu filtrów na styku AS2-AS3 powodują, że AS3 nie akceptuje prefiksów na ścieżce AS2..AS1 na swoim łączu do AS2. Akceptuje



Rysunek 10: Filtry BGP

je jednak na łączu do operatora międzynarodowego, na styku z którym nie posiada żadnych filtrów BGP.

Wynikiem braku dostatecznej wymiany informacji pomiędzy operatorami AS2 oraz AS3 jest brak akceptacji sieci AS1 na łączu o krótszej ścieżce AS.

Powstaje asymetryczność routingu, która może być przyczyną wielu problemów.

Przykład hipotetycznej trasy z AS1 do AS2 mógłby wyglądać następująco

```

1. 1.2.3.4 rtr.as1.net 0 0 0
2. 5.6.7.8 rtr.as2.net 1 0 1
3. 9.0.1.2 rtr.as3.net 2 0 3
4. 3.4.5.6 host.as3.net 3 3 4

```

Zaś trasy z AS3 do AS1

```

1. 9.0.1.2 rtr.as3.net 0 0 1
2. 11.0.1.2 rtr.as100.net 15 15 16
3. 12.2.3.4 rtr.as101.net 20 21 18
4. 1.2.3.4 rtr.as1.net 100 100 101

```

Taki przykład obrazuje zwiększenie opóźnień pomiędzy AS1 oraz AS3.

1. Asymetryczność routingu utrudnia diagnostykę. Ścieżka przepływu pakietów odnaleziona z jednego AS nie odpowiada ścieżce odnajdywanej w odwrotnej sytuacji. Brak sprawdzenia tras pakietów z obu urządzeń jest popularnym błędem popełnianym podczas diagnostyki sieci IP. W powyższym przykładzie tras AS1-AS3 odtwarzanie ścieżki pakietów z AS1 do AS3 nie wykazuje problemów, który leży w konfiguracji trasy powrotnej pakietów.
2. Asymetryczność ruchu narusza politykę routingową operatorów. W niektórych przypadkach wynikiem asymetryczności routingu jest wysyłanie pakietów przez łącza „gorsze” w ocenie operatora. Może to oznaczać zarówno większe koszty ponoszone na obsługę ruchu IP (konieczność kupna większych „przepływności” na łączach lub nakładów związanych z rozbudową posiadanej infrastruktury) jak i degradację jakości usług IP (większe opóźnienia oraz mniejsze przepływności).

Stosowane obecnie przez operatorów Internetowych techniki pozwalają na „fizyczne” kształtowanie polityk routingu BGP. Taka polityka względem klienta i innych operatorów może się opierać na filtrowaniu prefiksów rozgłaszanych na różnych łączach. Brak koordynacji oraz nieuwaga podczas dokonywania poprawek w filtrach BGP może być przyczyną nieprawidłowości, która mogą objawić się zarówno jako brak widoczności jednych sieci w obszarach innych oraz w degradacji jakości usług IP.

Rozwiązaniem może być nakłonienie operatorów do utrzymywania aktualnych wpisów o akceptowanych prefiksach na swoich łączach w bazach whois. Bazy whois prowadzone przez RIR, dla Europy jest to RIPE, zawierają wiele informacji związanych z procesami routingu, w tym listę akceptowanych prefiksów na różnych interfejsach. Przykład dla sieci PW-NET - AS12464.

```
[jagger@ares jagger]$ whois AS12464
% This is the RIPE Whois query server #1.
% The objects are in RPSL format.
%
% Rights restricted by copyright.
% See http://www.ripe.net/db/copyright.html

as-block: AS12288 - AS13311
descr: RIPE NCC ASN block
remarks: These AS numbers are further assigned by RIPE NCC
```

```
remarks: to LIRs and end-users in the RIPE NCC region
remarks: Please refer to these documents
remarks: <http://www.ripe.net/ripe/docs/ir-policies-procedures.html>
<http://www.ripe.net/ripe/docs/asnrequestform.html>
<http://www.ripe.net/ripe/docs/asnsupport.html>
org: ORG-NCC1-RIPE
admin-c: CREW-RIPE
tech-c: OPS4-RIPE
mnt-by: RIPE-NCC-HM-MNT
mnt-lower: RIPE-NCC-HM-MNT
changed: hostmaster@ripe.net 20010423
changed: hostmaster@ripe.net 20011024
changed: hostmaster@ripe.net 20011120
changed: hostmaster@ripe.net 20020408
changed: ripe-dbm@ripe.net 20040421
source: RIPE
```

```
aut-num: AS12464
as-name: UNSPECIFIED
descr: WARSAW U-TECH Campus Network
descr: WARSAW UNIVERSITY OF TECHNOLOGY
import: from AS1887
        action pref=100;
        accept ANY
import: from AS8890
        action pref=50;
        accept AS8890
export: to AS1887
        announce AS12464
export: to AS8890
        announce AS12464
default: to AS1887
        action pref=100;
        networks ANY
admin-c: SLAN1-RIPE
tech-c: PJ1343-RIPE
mnt-by: AS12464-MNT
```

```
changed: P.Jankowski@coi.pw.edu.pl 20021219
source: RIPE
```

Z listingu można odczytać informacje na temat posiadanych połączeń, oraz oferowanych jak i rozgłaszanych prefiksach.

Jednak wielokrotnie informacje znajdujące się w tych bazach są przestarzałe i nie odzwierciedlają aktualnego stanu. Nie jest to więc wiarygodne źródło aktualnych informacji o filtrach stosowanych na stykach BGP. Częstokroć w celu uzgodnienia polityki filtrowania na peeringach BGP stosowane są najprostsze metody komunikacji telefonicznej lub rozsyłanie poczty elektronicznej.

Istnieją też operatorzy, którzy dbają o informacje routingowe, w tym o communities, które umożliwiają „oznakowywanie” prefiksów i podejmowanie akcji w zależności od zastosowanego atrybutu community.

Fragment wpisu w bazach whois dla sieci INTERNET-TECHNOLOGIES-POLSKA-AS - AS8246 może być wzorem dla innych operatorów, ukazującym właściwą politykę informacyjną odnośnie procesów routingu BGP:

```
[jagger@ares jagger]$ whois AS8246
/.../
aut-num: AS8246
as-name: INTERNET-TECHNOLOGIES-POLSKA-AS
descr: GTS Internet Partners
descr: al. Niepodleglosci 69
descr: 02-626 Warsaw, Poland
import: from AS5617 action pref=300; accept ANY
import: from AS1239 action pref=200; accept ANY
import: from AS6714 action pref=200; accept AS6714
import: from AS8308 action pref=200; accept AS-NASK
import: from AS8664 action pref=100; accept AS8664 AS8890
import: from AS12324 action pref=100; accept AS12324
import: from AS12346 action pref=100; accept AS12346
/.../
export: to AS5617 announce AS8246:AS-TPNET
export: to AS1239 announce AS-IPARTNERS
export: to AS8308 announce AS8246:AS-NASK
export: to AS8664 announce AS8246 AS6714
export: to AS6714 announce ANY
export: to AS12324 announce ANY
```

```

export: to AS12346 announce ANY
export: to AS12423 announce ANY
/.../
remarks: Internet Partners BGP community support
remarks: e-mail contact: <bgp4@ipartners.pl>
remarks: -----
remarks: Communities to control traffic (settable by peers):
remarks:
remarks: 8246:2000 Do not announce to GTS CE (AS5588)
remarks: 8246:2001 Prepend +1 when announcing to GTS CE
remarks: 8246:2002 Prepend +2 when announcing to GTS CE
remarks: 8246:2003 Prepend +3 when announcing to GTS CE
/.../

```

## 9.2. Bogon filters

Bogon filters są przykładem filtrów mających na celu wyeliminowanie z procesów routingu tras do prefiksów, które uważane są za niepożądane w danych AS. Najczęściej są to klasy adresowe nieprzyznane przez żadnego z RIR.

Brak aktualizacji prefiksów znajdujących się w filtrach typu bogon mogą prowadzić do poważnych utrudnień w działaniu protokołu IP. Na stronach internetowych <http://www.cymru.com/Documents/secure-bgp-template.html> znajduje się przykładowa konfiguracja procesu BGP na routerach systemach operacyjnych Cisco IOS. Zawiera ona *prefix-list* o nazwie *bogons*. Wycinek definicji tej listy prefiksów:

```

ip prefix-list bogons description Bogon networks we won't accept.
ip prefix-list bogons seq 5 deny 0.0.0.0/8 le 32
ip prefix-list bogons seq 10 deny 1.0.0.0/8 le 32
ip prefix-list bogons seq 15 deny 2.0.0.0/8 le 32
ip prefix-list bogons seq 20 deny 5.0.0.0/8 le 32
ip prefix-list bogons seq 25 deny 7.0.0.0/8 le 32
ip prefix-list bogons seq 30 deny 10.0.0.0/8 le 32
ip prefix-list bogons seq 35 deny 23.0.0.0/8 le 32
ip prefix-list bogons seq 40 deny 27.0.0.0/8 le 32
ip prefix-list bogons seq 45 deny 31.0.0.0/8 le 32
ip prefix-list bogons seq 50 deny 36.0.0.0/8 le 32

```

```
ip prefix-list bogons seq 55 deny 37.0.0.0/8 le 32
ip prefix-list bogons seq 60 deny 39.0.0.0/8 le 32
ip prefix-list bogons seq 65 deny 41.0.0.0/8 le 32
ip prefix-list bogons seq 70 deny 42.0.0.0/8 le 32
/../
```

Prefiksy należące do klas znajdujących się w tej liście prefiksów ze słowem kluczowym deny nie są akceptowane przez proces obsługujący protokół BGP i nie pojawią się w tablicy routingu routera (o ile nie doda go wpis statyczny lub protokół IGP).

Cel, który można osiągnąć dzięki zastosowaniu takiego filtru, czyli eliminacja akceptacji klas adresowych, które nie zostały oficjalnie dopuszczone do użytku, jest bardzo ważny, ogranicza bowiem propagację niewłaściwych informacji routingowych.

Niestety, brak aktualizacji tego typu wpisów na urządzeniach obsługujących BGP jest przyczyną wielu problemów z brakiem widoczności klas adresowych w Internecie.

Na początku 2004 problem ten dotknął nowej klasy adresowej 83/8, której część (83/11) została przydzielona sieci TPNET.

W efekcie zaistniałych problemów RIPE (europejski RIR) stworzył nowy model wprowadzania zarezerwowanych klas do użytku. Został stworzony tzw. draft-document, który analizuje zaistniały problem. Dokument dostępny jest pod adresem: <http://www.ripe.net/ripe/draft-documents/deboganising-draft.html>. Pierwsze zdanie rozdziału „Problem statement” doskonale opisuje problem:

*„Filtering of unallocated address space (a.k.a. bogon filtering) is becoming more prolific. This is a good thing. However when those filters are not kept up-to-date they can quickly become too much of a good thing.”*

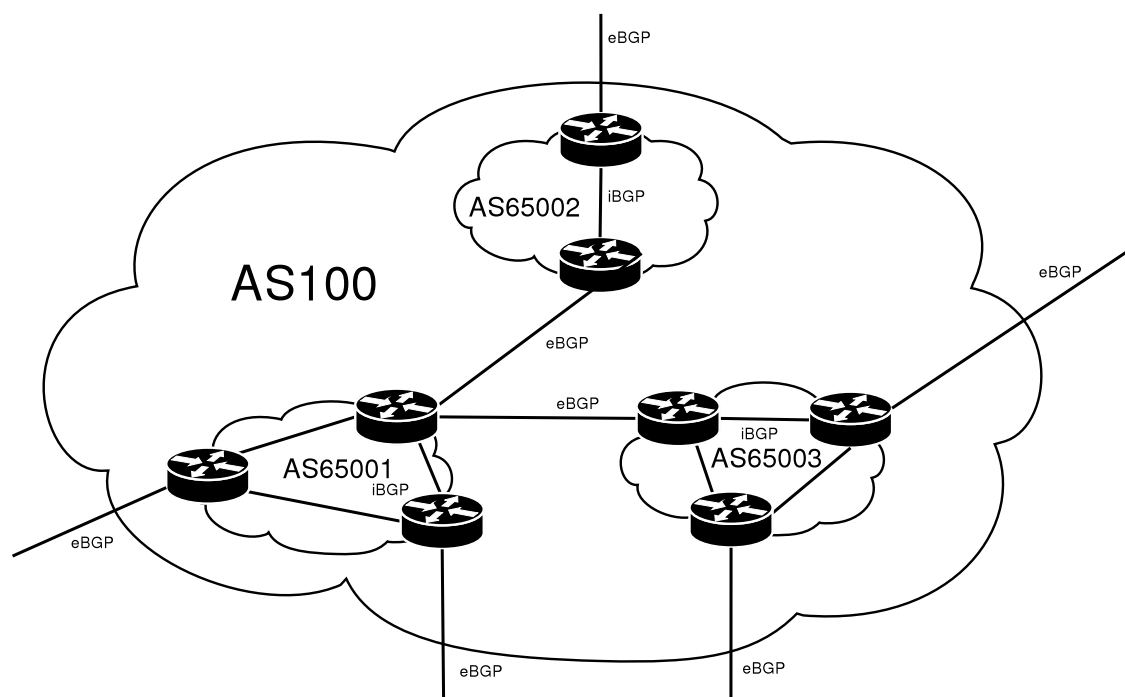
Dokument wspomina skutki problemów związanych z filtrowaniem nowo-wprowadzonych klas:

- frustracja użytkowników i ISP związana z dostępnością usług
- krytyka RIR (RIPE) za umożliwienie zaistnienia wspomnianej sytuacji

Przedstawione są też środki zaradcze w odniesieniu do nowych klas przyznawanych ISP. RIPE zobowiązało się do bardziej intensywnego informowania społeczności internetowej o nowo-rozgłaszanych klasach, w tym do indywidualnego

informowania ISP, którzy nie zareagują na pierwsze informacje o rozgłaszaniu nowych klas.

Dodatkowo klasy, które są przeznaczone do wprowadzenia do użytku będą przez pewien okres czasu rozgłaszane przez RIR (RIPE) pod tzw. *beacon AS 12654*. Wg RIPE powinno to doprowadzić do zminimalizowania w przyszłości problemów z akceptacją klas, które są obecnie umieszczane w filtrach typu bogon.



Rysunek 11: Konfederacja AS

## 10. Konfederacje AS

Konfederacje AS rozszerzają skalowalność BGP w obrębie systemów autonomicznych. Głównym celem wprowadzania konfederacji w ramach systemu autonomicznego jest wyeliminowanie konieczności stosowania topologii full-mesh pomiędzy speakerami BGP.

Urządzenia będące speakerami BGP w ramach systemu autonomicznego dzielone są w grupy nazywane subAS, którym przyznaje się numer systemu autonomicznego. Przyjęło się, że są to numery z puli prywatnej, z zakresu powyżej 65000.

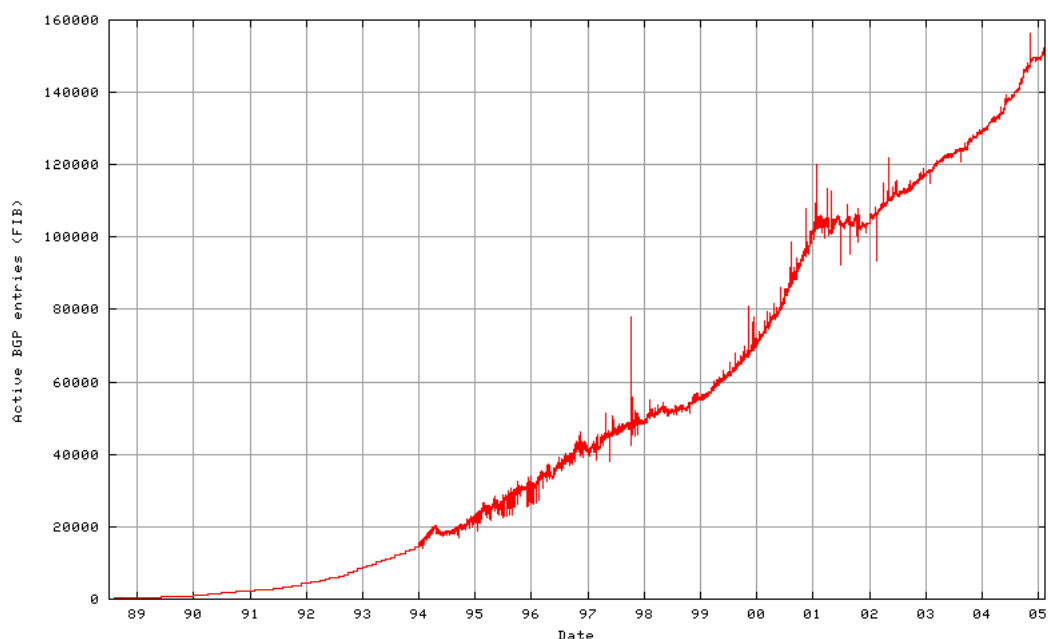
Dla każdego subAS musi zostać zachowany warunek topologii full-mesh speakerów BGP, lecz w ramach całego AS subASy mogą być połączone w dowolny sposób (zapewniający jednak, że istnieje ścieżka z dowolnego subAS do każdego innego subAS). Schemat przykładowego AS wraz z subAS przedstawiony jest na Rysunku 11.

Dzięki konfederacjom system autonomiczny podzielony na subAS jest widoczny dla innych sieci jako jeden AS.

## 10.1. Ocena skuteczności

Użycie konfederacji systemów autonomicznych dla BGP może przyczynić się do łatwiejszej implementacji i zarządzania protokołem BGP w sieciach o dużej liczbie speakerów BGP. Eliminowana jest konieczność stosowania topologii full-mesh pomiędzy routerami a podział na subAS może przyspieszyć diagnostykę sytuacji awaryjnych związanych z protokołem routingowym (mniejsze jednostki, to szybsze wyszukiwanie źródła problemu).

Z drugiej strony podział sieci na wewnętrzne jednostki logiczne może stać się przyczyną degradacji ruchu w obrębie sieci. Nierozważnie obrany podział może spowodować zwiększenie liczby hopów w obrębie AS, co może być przyczyną zwiększonych opóźnień wynikających z większej liczby przełączających urządzeń. Warto także pamiętać, że rozszerzenie atrybutu AS\_PATH o segmenty AS\_CONFED\_SET oraz AS\_CONFED\_SEQUENCE nie należy do kategorii well-known a więc nie musi być implementowane w ramach standardowej implementacji BGP na urządzeniach sieciowych.



Rysunek 12: Wzrost liczby rozgłaszanych prefiksów

## 11. Objętość informacji routingowej

### 11.1. Zarys problemu

Wzrost liczby operatorów internetowych (ISP) korzystających z BGP wzrasta bardzo szybko. W 1994r tablica prefiksów wg danych zbieranych przez administratorów AS1221 <http://bgp.potaroo.net/> wynosiła ok 15tys. wpisów, w chwili obecnej router otrzymujący z jednego źródła pełną tablicę światowego bgp charakteryzuje się następującymi parametrami:

```
r9-waw2> show ip bgp summary | inc path entries
154407 path entries using 7411536 bytes of memory
```

Oznacza to ponad 150tys. otrzymywanych wpisów routingowych. Od 1994r. liczba wpisów wzrosła 10-krotnie. Rysunek 12 przedstawia w formie graficznej tę statystykę.

Tak szybki wzrost informacji routingowych, które przetwarzają urządzenia obsługujące BGP. Na wykresie można zauważyć kilka tendencji. Od roku 1994 do roku 1999 wzrost miał charakterystykę liniową. Od roku 1999 można zaobserwować gwałtowny wzrost liczby rozgłaszanych prefiksów. Taką tendencję można wiązać z ówczesnym tzw. „boomem internetowym”, gdy każda forma działalności gospodarczej związanej z usługami internetowymi przynosiła ogromne zyski. W roku 2001 wraz z tzw. „pęknięciem internetowej bańki”

następuje gwałtowne wyhamowanie a nawet spadek liczby rozgłaszanych prefiksów. lecz po krótkim czasie znów przybiera tendencję liniową. ze wzrostem rzędu 20tys prefiksów/rok.

Tak szybki wzrost informacji routingowej nie pozostaje bez wpływu na procesy zarządzania i routingu w sieciach internetowych.

Pierwszym zagadnieniem jest ilość informacji, które może przeanalizować administrator obsługujący router BGP. W obecnej rzeczywistości nikt się już nie „przejmuje” brakiem widoczności pojedynczych prefiksów a problem pojawia się dopiero wtedy, gdy z routingu BGP znikają duże partie sieci (vide „AS7007 incident” oraz „128/9 disaster”).

Powoduje to, że bardzo częste są przypadki znikania pojedynczych prefiksów z procesów BGP. Problem z widocznością może być spowodowany niewłaściwą pracą urządzenia sieciowego znajdującego się fizycznie w odległym zakątku globu. Komunikacja ze administratorem takiego urządzenia może być utrudniona z wielu przyczyn, chociażby z powodu bariery językowej lub trudności z odnalezieniem właściwych danych (wpisy w bazach whois RIR są często nieaktualne).

Drugim zagadnieniem jest wpływ ilości informacji routingowych na zużycie zasobów routerów importujących prefiksy. Aby routery mogły przetwarzać takie ilości informacji potrzebne są coraz szybsze procesory oraz większe ilości pamięci operacyjnej. Zdarza się, że jak w przypadku opisanych incydentów urządzenia sieciowe „zawieszały” się po wyczerpaniu się pamięci operacyjnej przeznaczonej dla tablic BGP.

### **11.2. Możliwe rozwiązania**

W celu ograniczenia liczby informacji routingowych można posłużyć się kilkoma metodami.

Pierwszą z nich jest zmiana sposobu przyznawania nowych klas adresowych. W chwili obecnej RIR przyznają dwa rodzaje klas adresowych. Klasa PA (Provider Assigned) jest powiązana z konkretnym LIRem (LIR - Local Internet Registrar). Przyznanie klientowi LIRa puli adresów z takiej klasy oznacza przekazanie klientowi podsieci w ramach klasy przyznanej już wcześniej LIRowi.

Drugi rodzaj klas adresowych (rozgłaszanych jako prefiksy) to PI - Provider Independent. Przyznawane są one organizacjom niekoniecznie posiadającym status systemu autonomicznego.

Uniemożliwienie przyznawania klas PI organizacjom nie posiadającym nu-

meru ASN może drastycznie zmniejszyć ilość rozgłaszanych prefiksów. Takie organizacje będą musiały opierać się na podsieciach wydzielonych z klas adresowych operatora LIR. Wyeliminuje to konieczność rozgłaszania klas adresowych o długich maskach (najczęściej /24), ponieważ organizacje będą używały klas adresowych wydzielonych z klasy adresowej operatora, rozgłaszanej jako prefiks o krótkiej masce (np. poniżej 20bitów).

Zaletą tego rozwiązania może być wspomniane już wcześniej znaczne ograniczenie liczby rozgłaszanych prefiksów w BGP. Wadą jest wyeliminowanie z użycia klas PI, czyli zmuszenie organizacji, które nie posiadają przyznaných klas PI oraz numerów ASN do zmian adresacji w obrębie własnej sieci, przy każdej zmianie operatora ISP.

Druga metoda opiera się na weryfikacji i agregacji przyznaných klas adresowych. Wiele sieci posiada przyznaných kilka klas adresowych niemożliwych do rozgłaszania jako jeden prefiks. Przykładem może być sieć Telekomunikacji Polskiej S.A., której sukcesywnie, wraz z rozwojem sieci, przyznawane były nowe prefiksy. W chwili obecnej numery AS TPNET rozgłaszają następujące klasy adresowe.

80.48.0.0/13  
83.0.0.0/11  
194.204.128.0/18  
195.116.0.0/16  
195.117.0.0/16  
195.205.0.0/16  
212.160.0.0/16  
212.244.0.0/16  
213.25.0.0/16  
213.76.0.0/16  
213.77.0.0/16  
217.96.0.0/16  
217.97.0.0/16  
217.98.0.0/16  
217.99.0.0/16  
195.35.80.0/24  
195.149.232.0/21

Tak dużą liczbę klas można zamienić na jedną, lecz o krótszej masce bitowej. W przypadku przeprowadzenia takiej operacji dla wszystkich posiadaczy klas

adresowych przyznanych przez RIR, liczba rozgłaszanych prefiksów zmniejszyłaby się co najmniej kilkukrotnie.

Tak jak poprzednia metoda, tak i ta posiada wadę. Zmiana adresacji wewnętrznej, częstokroć dużych operatorów, to ogromny problem logistyczny, który mógłby skutkować długotrwałymi problemami z usługami IP u klientów operatora jak i użytkowników Internetu korzystających z usług ulokowanych w reorganizowanej sieci.

## 12. Ograniczony zasób - ASN

### 12.1. Zarys problemu

Najważniejsze obecnie problemy związane z ograniczonymi zasobami „numerycznymi” Internetu to wyczerpywanie się adresów IPv4 oraz wyczerpywanie się ASN. Pierwsze wspomniane zagadnienie zostało dostrzeżone o wiele wcześniej i skutkiem badań jest nowy protokół IP w wersji 6, który jest sukcesywnie wprowadzany w Internecie. W chwili obecnej w testowej sieci 6bone jak i produkcyjnych wdrożeniach IPv6 uczestniczy około 500 systemów autonomicznych, rozgłaszających ok 550 prefiksów. Różnica pomiędzy liczbą prefiksów i liczbą systemów autonomicznych IPv6 jest spowodowana faktem, że niektóre AS rozgłaszają pod swoimi ASN jednocześnie prefiksy z klas produkcyjnych 2001::/16 jak i testowych klas projektu 6bone 3ffe::/16.

Klasy adresowe IPv6 przyznawane operatorom są na tyle duże, by nie istniała konieczność przyznawania nowych klas a tym samym rozgłaszania wielokrotnie większej liczby prefiksów, niż przyznanych ASN. Problem ten jest bardzo duży dla IPv4 gdzie ok 25tys aktywnych ASN rozgłasza ponad 150tys. prefiksów. Zastosowane rozwiązanie dla klas IPv6 jest próbą rozwiązania problemu dużej ilości informacji routingowych związanych z prefiksami.

Wyczerpywanie się wolnych numerów ASN stanowi kolejny, ważny problem przed którymi operatorzy internetowi oraz RIRs staną w ciągu kilku/kilkunastu lat.

W chwili obecnej przyznanych jest prawie 35tys numerów AS. Aktualny stan rozdysponowania numerów AS można na bieżąco śledzić na stronach internetowych <http://www.iana.org/assignments/as-numbers>.

ASN jest 16bitową liczbą, numerowaną od 0 do 65535. Niektóre numery AS są zarezerwowane dla specjalnych celów, jak np AS\_TRAN, RIPE AS Beacon itp. Numery od 64512 do 65534 są tzw. prywatnymi numerami AS i nie mogą

być używane w światowym routingu BGP tak jak i zarezerwowany numer 65535.

Dynamika rozdysponowania numerów AS w ostatnich latach wynosi ok 3tys nowych dyspozycji na rok. To oznacza, że przewidywany termin wyczerpania się tego zasobu przy liniowym wzroście przydziałów przypadnie na lata 2012-2016, zaś przy wzroście wykładniczym termin ten może okazać się znacznie bliższy. W związku z bardzo dynamicznym rozwojem usług internetowych oraz zapotrzebowaniem na niezawodność, posiadanie łączy do dwóch niezależnych operatorów internetowych staje się standardem w dużej liczbie przejawów działalności edukacyjnej i komercyjnej w Internecie. To pozwala przypuszczać, że wolne zasoby ASN skończą się znacznie szybciej niż data zakładana przy liniowym wzroście przydziałów numerów AS.

## 12.2. Proponowane działania

### 12.2.1. Weryfikacja przyznanych zasobów

Zjawisko bardzo szybkiego wyczerpywania się wolnych numerów ASN jest mocno związane z brakiem weryfikacji już przyznanych zasobów. Organizacje typu RIR powinny cyklicznie prowadzić badania przyznanej numeracji AS pod kątem:

- istnienia ASN w tablicach BGP
- sposobu rozgłaszania, ze szczególnym uwzględnieniem sieci, posiadających tylko jeden peering BGP.

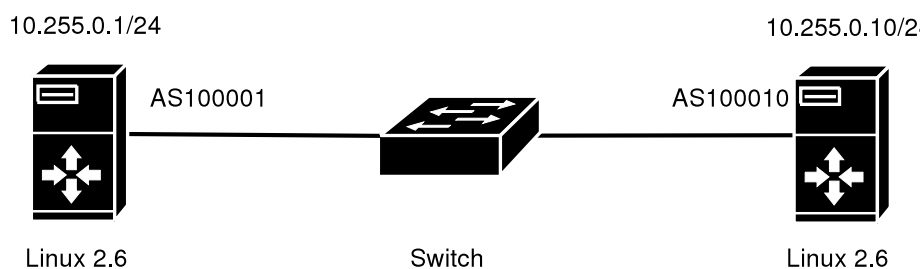
W obu tych przypadkach powinny zostać podjęte działania zmierzające do odzyskania przydziału zasobu ASN, który nie jest właściwie wykorzystywany.

### 12.2.2. 32bit ASN

Rozwiązaniem problemu wyczerpywania się wolnych zasobów w dziedzinie numeracji ASN, może być rozszerzenie zakresu tej wartości. Dlatego też od pewnego czasu na forum IETF, zajmującym się routinguem dynamicznym w Internecie, omawiane są możliwości wprowadzenia do architektury „światowego” BGP, numerów ASN o rozmiarze 32 bitów.

W wyniku tych dyskusji został stworzony draft internetowy opisujący proponowane sposoby rozszerzenia możliwości speakerów BGP o użycie ASN w rozmiarze 32 bitów.

Draft ten jest dostępny na stronach internetowych IETF, pod nazwą *draft-ietf-idr-as4bytes*. Na podstawie zawartych w nim zaleceń, w części praktycznej tej pracy, przystosowano daemon BGP z ogólnodostępnym kodem źródłowym tak, by wspierał 32bitowe numery ASN.



Rysunek 13: Schemat konfiguracji testowej

## 13. Modyfikacja daemona Quagga

Daemon ten został wybrany jako podstawa modyfikacji ze względu na liberalną licencję - GNU GPL v.2 - pozwalającą na modyfikację kodu źródłowego bez ograniczeń, zachowując jedynie obowiązek udostępnienia zmodyfikowanych kodów źródłowych. Kody te zostaną umieszczone na stronach domowych WWW autora niniejszej pracy.

Quagga jako całość jest zbiorem programów, które realizują routing dynamiczny na uniksowych systemach operacyjnych (najpopularniejsze z nich to: Linux, \*BSD, Solaris).

Jest to pochodna projektu GNU Zebra, tworzona przez Kunihiro Ishiguro. Quagga ma na celu „przyciągnięcie” do projektu szerszej społeczności programistów.

Ostatnią wersją, na którą zostały nałożone stworzone rozszerzenia to stabilne wydanie 0.98.

### 13.1. Zakres modyfikacji

Wprowadzone zmiany implementują dużą część zaleceń draftu IETF, ważną ze względu na możliwość przeprowadzenia testów współpracy zmodyfikowanych daemonów BGP pomiędzy sobą. Zaniechane zostały modyfikacje odnoszące się do atrybutów innych niż AS\_PATH.

Zmodyfikowany został interfejs konfiguracyjny, w szczególności możliwość wprowadzania numerów AS o rozmiarze 32 bit. Do testowania stworzonych rozszerzeń użyty został zestaw komponentów przedstawionych na Rysunku 13. Testowa konfiguracja daemona bgpd na routerze o numerze IP 10.255.0.10 zamieszczona jest poniżej:

```
!
debug bgp
```

```
debug bgp events
!  
router bgp 100010  
  bgp router-id 10.255.0.10  
  network 192.168.10.0/24  
  neighbor 10.255.0.1 remote-as 100001  
  neighbor 10.255.0.1 capability 4byte-as  
  neighbor 10.255.0.1 soft-reconfiguration inbound  
!
```

Proces bgpd na routerze B dysponował następującą konfiguracją:

```
!  
debug bgp  
debug bgp events  
!  
router bgp 100001  
  bgp router-id 10.255.0.1  
  network 192.168.1.0/24  
  neighbor 10.255.0.10 remote-as 100010  
  neighbor 10.255.0.10 capability 4byte-as  
  neighbor 10.255.0.10 soft-reconfiguration inbound  
!
```

W przedstawionych powyżej konfiguracjach słowo kluczowe

```
router bgp <nr>
```

tworzy nową instancję routingu BGP. Dalsze parametry to:

```
bgp router-id <IP>
```

określający router-id lokalnego peera BGP.

```
network <prefix>
```

określający rozgłaszany prefix.

```
neighbor <IP> remote-as <AS>
```

określający IP peera BGP oraz jego ASN

```
neighbor <IP> capability 4byte-as
```

nakazujący rozgłaszanie zdolności (capability) obsługi 4bajtowych rozmiarów ASN.

```
neighbor <IP> soft-reconfiguration inbound
```

umożliwiający odtworzenie tablicy BGP bez rozłączania sesji TCP.

## 13.2. Zestawienie sesji BGP

Po uruchomieniu daemonów bgpd na obu systemach sesje zestawiły się. Output polecenia

```
sh ip bgp summary
```

na urządzeniu o numerze IP 10.255.0.10 wyglądał następująco

```
BGP router identifier 10.255.0.10, local AS number 100010
2 BGP AS-PATH entries
0 BGP community entries
Neighbor    V AS      MsgRcvd MsgSent  TblVer  InQ  OutQ  Up/Down  State/PfxRcd
10.255.0.1 4 100001 5         7        0     0    00:00:13 1
```

to samo polecenie na urządzeniu o numerze IP 10.255.0.1 dawało następujący wynik

```
BGP router identifier 10.255.0.1, local AS number 100001
2 BGP AS-PATH entries
0 BGP community entries
Neighbor    V AS      MsgRcvd MsgSent  TblVer  InQ  OutQ  Up/Down  State/PfxRcd
10.255.0.10 4 100010 6         8        0     0    00:04:30 1
```

W obu przypadkach ASN peera BGP jest poprawny, co pozwala sądzić, że ta część rozszerzenia daemona bgpd pracuje poprawnie.

Zrzut sesji BGP wraz z podstawową analizą protokołu BGP dla zapytania BPF `'host 10.255.0.10 and port 179'`, programem `tcpdump` z opcją `-vvv`, daje następujący wynik:

```
14:26:23.188909 IP (tos 0x0, ttl 1, id 58644,
  offset 0, flags [DF], length: 105)
  10.255.0.1.2858 > 10.255.0.10.179: P [tcp sum ok]
  0:53(53) ack 1 win 5840 <nop,nop,timestamp
  260562761 68023584>: BGP, length: 53
Open Message (1), length: 53
  Version 4, my AS 23456, Holdtime 180s, ID 10.255.0.1
Optional parameters, length: 24
  Option Capabilities Advertisement (2), length: 6
  Multiprotocol Extensions, length: 4
  AFI IPv4 (1), SAFI Unicast (1)
  0x0000: 0001 0001
  Option Capabilities Advertisement (2), length: 2
  Route Refresh (Cisco), length: 0
  Option Capabilities Advertisement (2), length: 2
  Route Refresh, length: 0
  Option Capabilities Advertisement (2), length: 6
  Unknown, length: 4
  no decoder for Capability 65
  0x0000: 0001 86a1
```

Analiza ta pokazuje, że w wiadomości OPEN skierowanej do hosta 10.255.0.10 znajdują się następujące dane.

**Wersja:** Wersja 4 protokołu BGP

**AS:** 23456, jest to zarezerwowany przez IANA ASN AS\_TRAN, wypełniający pole AS pakietu HELLO w przypadku wymiany danych pomiędzy peerami obsługującymi 32bit ASN.

**Holdtime:** 180s

**Router-ID:** w tym przypadku 10.255.0.1 (IP peera BGP)

**Opcje:** Ostatnia z opcji posiada kod u numerze 65, jest to zarezerwowany przez IANA kod (zdefiniowany w RFC 3392) służący do obsługi capability: „Support for 4-octet AS number capability”. Rozmiar wartości tej opcji to 4 bajty. Wartość zaś, to przedstawiona w postaci szesnastkowej liczba: 0001 86a1, która po przeliczeniu na postać dziesiętną daje wynik postaci 100001 będący wartością ASN rozgłaszaną przez router o IP 10.255.0.1

### 13.3. Rozgłaszanie prefiksów

Sprawdzona też została poprawność przekazania atrybutu AS\_PATH\_NEW poprzez wydanie polecenia

```
sh ip bgp
```

Dla urządzenia o numerze IP 10.255.0.10

```
BGP table version is 0, local router ID is 10.255.0.10
Status codes: s suppressed, d damped, h history,
* valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete
   Network      Next Hop      Metric LocPrf Weight Path
*> 192.168.1.0  10.255.0.1    0                0    100001 i
*> 192.168.10.0 0.0.0.0       0                32768 i
Total number of prefixes 2
```

Wartość 100001 pola Path dla prefiksu 192.168.1.0 oznacza, że ścieżka została pobrana z atrybutu AS\_PATH\_NEW skojarzonego z tym prefiksem.

### 13.4. Wnioski

Stworzenie oprogramowania zgodnego z wytycznymi draftu IETF dotyczącego 4 bajtowych rozmiarów ASN jest możliwe w rozsądnym czasie, wykorzystując dotychczasowe implementacje protokołu BGP. Chociaż oprogramowanie nie zostało uzupełnione o pełne wsparcie dla 32bit AS, to podstawowe schematy zmian w oprogramowaniu, czyli format przekazywania informacji o lokalnym AS oraz atrybutów zawierających ten parametr zostały zweryfikowane jako możliwe do implementacji na bazie dostępnego oprogramowania.

Kwestia implementacji rozszerzonego protokołu BGP o wsparcie dla 32it ASN to problem nie związany z aspektem technicznym (programistycznym) przedsięwzięcia, ponieważ ten jest dobrze udokumentowany, lecz ze stroną organizacyjną. Strach przed zmianami w sieciach (i związanymi z nimi problemami jak chwilowe braki łączności) będą na pewno czynnikiem hamującym wdrażanie podobnych do opisanego powyżej rozwiązań.

## 14. Podsumowanie analizy aspektu skalowalności protokołu BGP

Skalowalność protokołu BGP nie przysparza tak wielu problemów jak kwestia jego bezpieczeństwa. Pomimo kilku problemów, jak np. panujący obecnie „chaos” informacyjny spowodowany gwałtownym przyrostem liczby uczestników światowego BGP, protokół ten spisuje się bardzo dobrze pod względem skalowalności w nowej rzeczywistości dynamicznie rozwijającego się Internetu. Jest to głównie zasługa architektów protokołu, którzy przewidzieli w nim miejsce na rozszerzenia. Dzięki nim implementacja, dzisiaj już często stosowanych konfederacji AS czy też router-reflectors, jest prosta i doskonale rozwiązuje problemy dzisiejszych operatorów.

Ciągłym problemem jest ograniczona liczba dostępnych ASN. Stworzone rozszerzenie daemona bgpd pokazują, że implementacja 32bitowego ASN nie powinna sprawić trudności producentom sprzętu sieciowego. Tutaj także uwiadamia się elastyczność protokołu w implementowaniu nowinek i reagowaniu na pojawiające się potrzeby (mechanizm capabilities oraz opcjonalnych atrybutów ścieżek).

Do czasu wprowadzenia tego rozwiązania jednostki RIRs powinny położyć nacisk na efektywne wykorzystanie posiadanych numerów AS. Jest wiele AS, które nie są widoczne w globalnym Internecie. Wiele jest także tych, które są przyłączone do jednego ISP. Wzmoczona aktywność RIRs na polu weryfikacji przyznanych zasobów mogłaby dać więcej czas producentom i użytkownikom BGP na „oswojenie” się z nowym schematem adresacji systemów autonomicznych i łagodne przejście w świat 4 bajtowych ASN.

---

# Załączniki

## Słowniczek

**Adjacent** z ang. przyległy - Systemy Autonomiczne (ASs) posiadające zestawioną sesję BGP; będące w peeringu BGP.

**AS** System Autonomiczny (ang. *Autonomous System*) - grupa sieci lub sieć znajdująca się pod zarządem jednej organizacji i posiadająca wspólną politykę routingową.

**ASN** ang. *Autonomous System Number* - Numer Systemu Autonomicznego. Zawiera się w przedziale  $0 < x < 65536$ .

**BGP** ang. *Border Gateway Protocol* - protokół służący do wymiany informacji routingowych w Internecie.

**CIDR** ang. *Classless Internet Domain Routing* - Bezklasowy Routing Międzydomenowy. Forma opisu podsieci odrzucająca podział na klasy A/B/C.

**eBGP** ang. *External BGP* - konfiguracja BGP stosowana pomiędzy różnymi AS.

**FIB** ang. *Forwarding Information Base* - tablica routingu urządzenia sieciowego.

**IANA** ang. *The Internet Assigned Numbers Authority* - organizacja zajmująca się przydzielaniem i opisywaniem zasobów sieci internetowych. IANA w szczególności rozdziela numery IP oraz ASN.

**iBGP** ang. *Internal BGP* - konfiguracja BGP stosowana w obrębie jednego AS.

**IETF** ang. *Internet Engineering Task Force* - grupa organizacji, inżynierów, przedsiębiorstw i administratorów zainteresowanych rozwojem architektury sieci internetowych. Zatwierdza RFC do publikacji i nadaje im numery porządkowe.

**HOP** zazwyczaj router lub inne urządzenie sieciowe przekazujące pakiety IP. Określany też jako "przeskok". Ważną funkcją HOPa jest zmniejszanie

---

wartości parametru TTL pakietów IP, co pozwala na analizę tras routin-  
gowych.

**Peer** Urządzenie sieciowe (najczęściej router) obsługujące sesję BGP z innym  
peerem; BGP neighbour.

**Prefix** Zbór, zawierający parę, składającą się z IP oraz długości maski sieci.

**RFC** ang. *The Request for Comments* - Jest zbiorem technicznych i organi-  
zacyjnych informacji dotyczących Internetu. RFC jest najpierw publiko-  
wany jako tzw. Internet Draft.

**RIR** ang. *Regional Internet Register* - zarządca zasobów internetowych (ASy,  
adresy IP) na poziomie "regionalnym" obejmującym jeden lub kilka kon-  
tynentów. W Europie jest to RIPE.

**Ścieżka** ang. *AS Path* - Lista ASN, przez które prowadzi trasa ze źródłowego  
AS do docelowego AS. Najlepsza trasa to z ang. *best path (best route)*

## Spis rysunków

1.	Sesja eBGP . . . . .	14
2.	Sesja iBGP . . . . .	15
3.	Route-reflector . . . . .	16
4.	Incydent 7007. Faza 1 . . . . .	26
5.	Incydent 7007. Faza 2 . . . . .	26
6.	Incydent 7007. Wybór tras . . . . .	27
7.	128/9 disaster. Dystrybucja prefixów . . . . .	32
8.	Prefix based configuration . . . . .	36
9.	Schemat działania serwera RBL . . . . .	39
10.	Filtry BGP . . . . .	45
11.	Konfederacja AS . . . . .	52
12.	Wzrost liczby rozgłaszanych prefixów . . . . .	54
13.	Schemat konfiguracji testowej . . . . .	60

## Literatura

- [1] RFC 1771 A Border Gateway Protocol 4 (BGP-4)
- [2] RFC 1772 Application of the Border Gateway Protocol in the Internet
- [3] RFC 1773 Experience with the BGP-4 protocol
- [4] RFC 1774 BGP-4 Protocol Analysis
- [5] RFC 1930 Guidelines for creation, selection, and registration of an Autonomous System (AS)
- [6] RFC 1966 BGP Route Reflection: An alternative to full-mesh iBGP
- [7] RFC 1997 BGP Communities Attribute
- [8] RFC 1998 An Application of the BGP Community Attribute in Multi-home Routing
- [9] RFC 2270 Using a Dedicated AS for Sites Homed to a Single Provider
- [10] RFC 2385 Protection of BGP Sessions via the TCP MD5 Signature Option
- [11] RFC 2858 Multiprotocol Extensions for BGP-4 Inter-Domain Routing
- [12] RFC 3065 Autonomous System Confederations for BGP
- [13] RFC 3562 Key Management Considerations for the TCP MD5 Signature Option
- [14] Understanding BGP Misconfiguration - <http://www.cs.washington.edu/homes/ratul/bgp/>
- [15]
- [16] NANOG Archives <http://www.merit.edu/mail.archives/nanog/historical.html>
- [17] BGP support for four-octet AS number space - draft-ietf-idr-as4bytes-09.txt
- [18] An Analysis of BGP Multiple Origin AS (MOAS) Conflicts [http://www.cs.colostate.edu/~massey/pubs/conf/massey\\_imw01.pdf](http://www.cs.colostate.edu/~massey/pubs/conf/massey_imw01.pdf)