Czytaj lub słuchaj artykuł online

Zeskanuj kod QR, aby otworzyć artykuł na XO.PL

Źródło artykułu:

https://xo.pl/it-na-wesolo-net-bgp

SIECI KOMPUTEROWE

BGP, czyli kto decyduje, którędy pojedzie Internet między operatorami?

Witamy na Międzynarodowym Kongresie Zarządców Autostrad! W jednej sali siedzą właściciele dróg z całego świata. Każdy ma własną mapę, własne zasady i niezachwiane przekonanie, że to właśnie jego trasa jest najlepsza. A gdzieś daleko kierowca chce po prostu dojechać do celu. Jakim cudem cały Internet nie stoi w korku? Poznajmy BGP.

Każdy ma własną autostradę. I własne zdanie

Przy ogromnym stole konferencyjnym zasiadło kilkunastu zarządców dróg. Przed każdym leżała tabliczka z numerem.

— Szanowni państwo, rozpoczynam Międzynarodowy Kongres Zarządców Dróg — oznajmił przewodniczący.
— Dlaczego ja mam numer 64501, a nie nazwisko?!
— To numer pańskiego systemu autonomicznego.
— A CO TO JEST SYSTEM AUTONOMICZNY?!
— Sieć zarządzana według własnych zasad.
— Czyli jestem niezależny?!
— W dużym uproszczeniu.
— I MOGĘ ROBIĆ, CO CHCĘ?!
— We własnej sieci, w granicach obowiązujących uzgodnień.
— TO DLACZEGO MAM SĄSIADÓW?!

🤣🤣

Internet nie jest jedną gigantyczną siecią zarządzaną przez jednego kierownika. Składa się z wielu niezależnie zarządzanych sieci. Taką sieć lub grupę sieci o wspólnej polityce **routingu** nazywamy **AS — Autonomous System**, czyli systemem autonomicznym. Każdy AS ma swój numer **ASN**.

Numer 64501 jest tutaj przykładowy i należy do zakresu prywatnych ASN. Prawdziwi operatorzy mogą używać publicznych numerów przydzielanych zgodnie z obowiązującymi zasadami.

BGP — Border Gateway Protocol pomaga systemom autonomicznym informować się nawzajem, do których grup **adresów IP** znają drogę. To nie jest centralny GPS dla całego świata. To raczej sieć porozumień między zarządcami.

Przeze mnie dojedzie pan do Warszawy!

Pierwszy zarządca wstaje.

— Szanowni państwo! Informuję, że przeze mnie można dojechać do Warszawy!
— Skąd pan wie?
— Bo droga do Warszawy należy do mojej sieci!
— A ja mogę dojechać przez niego! — krzyczy sąsiad.
— A ja przez tego drugiego!
— A JA PRZEZ TEGO TRZECIEGO!
— PANOWIE! — przewodniczący uderza w stół. — TO NIE JEST KONKURS NA NAJDŁUŻSZĄ DROGĘ DO WARSZAWY!

🤣

W BGP ogłasza się **prefiksy IP**, czyli zakresy adresów. Ogłoszenie można rozumieć tak: „Umiem dostarczyć pakiety do sieci `203.0.113.0/24`”. To przykładowa adresacja dokumentacyjna, nie prawdziwa trasa do Warszawy.

Sąsiad może przyjąć tę informację i, jeżeli jego polityka na to pozwala, ogłosić ją dalej. Dzięki temu informacje o osiągalności rozchodzą się przez kolejne sieci.

To jednak nie obietnica, że każdy samochód będzie miał zielone światło. Każdy operator może zdecydować, które ogłoszenia zaakceptuje i które przekaże innym.

AS_PATH, czyli lista zarządców po drodze

Na ekranie pojawiły się dwie propozycje dojazdu.

— Moja droga prowadzi przez AS 64502, 64503 i 64504!
— A moja przez 64505 i 64504!
— MOJA JEST LEPSZA!
— Dlaczego?!
— Bo mam mapę w kolorze!

AS_PATH to atrybut ogłoszenia BGP zawierający listę systemów autonomicznych, przez które ogłoszenie przechodziło. W uproszczeniu pokazuje, jakimi sieciami można dotrzeć do ogłaszającego prefiks.

Przydaje się także do wykrywania pętli między systemami autonomicznymi. Gdy AS zobaczy własny numer na otrzymanej liście, standardowo odrzuci taką trasę.

— Panie kierowniku! Nasza trasa wróciła do nas!
— Świetnie! Czyli działa!
— Nie. Ona prowadzi przez nas z powrotem do nas.
— TO MOŻE WYŚLIJMY JĄ JESZCZE RAZ?!
— WŁAŚNIE TEMU MA ZAPOBIEGAĆ AS_PATH!

🤣🤣

Uwaga: liczba numerów AS w AS_PATH nie mówi, ile kilometrów ani ile routerów dzieli nas od celu. To inny rodzaj informacji.

Najkrótsza? Najszybsza? Nie, zgodna z polityką!

Na stole położono dwie mapy. Pierwsza droga przechodzi przez dwa kraje. Druga — przez pięć.

— Bierzemy tę przez dwa! — oznajmił kierownik.
— Ale ta przez pięć ma u nas wyższy priorytet.
— JAK TO WYŻSZY?! TAM JEST PIĘĆ KRAJÓW!
— Zgadza się.
— A TUTAJ DWA!
— Zgadza się.
— TO DLACZEGO MAMY JECHAĆ PRZEZ PIĘĆ?!
— Bo BGP to nie nawigacja szukająca po prostu najkrótszej drogi.
— TO CO TO ZA NAWIGACJA?!
— Biznesowa, panie kierowniku.

🤣🤣

W BGP działa **polityka wyboru tras**. Operatorzy mogą preferować jednego sąsiada, odrzucać część ogłoszeń, zmieniać atrybuty i decydować, którędy popłynie ruch. Wybór nie jest prostym liczeniem kilometrów, opóźnienia lub liczby routerów. Poszczególne implementacje BGP mają własne, szczegółowe procedury wyboru najlepszej trasy.

Jednym z ważnych atrybutów jest **LOCAL_PREF** — lokalna preferencja używana wewnątrz danego AS. Wyższa wartość zwykle oznacza drogę bardziej pożądaną przez tego operatora i może mieć pierwszeństwo przed krótszym AS_PATH.

— Dlaczego jedziemy przez sąsiada A?
— Bo ma wyższy LOCAL_PREF.
— A kto go ustawił?
— Pan.
— KIEDY?!
— Wczoraj.
— I DLACZEGO?!
— Powiedział pan: „Tędy ma jeździć wszystko, bo tak”.
— AHA. TO ZNACZY, ŻE TRASA JEST DOBRA.

🤣

Można więc mieć trasę krótszą na mapie i świadomie wybrać inną. Taka jest natura BGP: **ważne jest nie tylko dokąd, ale też według czyich zasad**.

Sąsiedzi BGP: nie wystarczy wejść i krzyczeć

Spóźniony uczestnik wbiegł na konferencję i zaczął wymachiwać kartkami.

— HALO! PRZEZE MNIE TEŻ MOŻNA DOJECHAĆ DO WARSZAWY!
— A z kim zestawił pan sesję BGP? — spytał przewodniczący.
— Jaką sesję?!
— Z którym sąsiadem wymienia pan informacje?
— Z NIKIM! PRZYSZEDŁEM I MÓWIĘ!
— To nie zebranie wspólnoty mieszkaniowej, proszę pana.

🤣

BGP wymienia informacje między **skonfigurowanymi sąsiadami**, zwanymi peers. Routery zestawiają sesję, zwykle wykorzystując **TCP na porcie 179**. Po jej ustanowieniu wymieniają informacje o trasach. Nie rozsyłają swoich ogłoszeń każdemu urządzeniu w Internecie na oślep.

Jeżeli sesja z sąsiadem przestanie działać, trasy poznane wyłącznie przez nią zwykle przestają być dostępne, a router szuka innych możliwości zgodnie ze swoją polityką.

eBGP i iBGP, czyli rozmowy z sąsiadem i własnym biurem

Na konferencji kierownik dzwoni do własnej centrali.

— Halo! Mam nową trasę od sąsiada!
— To przekaż ją naszym innym oddziałom.
— A CZY TO NIE JEST ZDRADA?!
— Nie. Oni pracują w tej samej firmie.
— A z kim właściwie przed chwilą rozmawiałem?
— Z operatorem zewnętrznym.
— CHOLERA, I OBA POŁĄCZENIA NAZYWAJĄ SIĘ BGP?!

🤣

eBGP to wymiana tras BGP pomiędzy różnymi systemami autonomicznymi — na przykład między dwoma operatorami. **iBGP** to wymiana informacji BGP wewnątrz jednego AS, dzięki której różne routery tego operatora mogą znać potrzebne trasy.

W rozbudowanych sieciach iBGP wymaga odpowiedniego projektu: bez dodatkowych mechanizmów, takich jak **route reflectors**, nie wystarczy założyć, że każdy router bez ograniczeń przekaże każdemu wszystkie informacje. To już jednak temat na osobne posiedzenie komisji.

NEXT_HOP, czyli do którego sąsiada w ogóle skręcić?

Kierownik z dumą prezentuje trasę prowadzącą przez pół kontynentu.

— Wiem, przez jakie kraje pojedziemy!
— Doskonale. A do kogo wysyłamy pierwszy samochód?
— ...
— Do którego sąsiada? W którą stronę?
— PANOWIE, JA TU PLANOWAŁEM CAŁĄ EUROPĘ, A WY MNIE PYTACIE O PIERWSZY ZJAZD?!

🤣

Atrybut **NEXT_HOP** wskazuje adres następnego punktu, do którego należy kierować ruch dla danej trasy BGP. Router musi umieć do niego dotrzeć poprzez swoją tablicę **routingu**. Sama pięknie ogłoszona trasa nie wystarczy, jeśli następny krok jest nieosiągalny.

To rozdziela dwie sprawy: BGP informuje, jakie prefiksy są osiągalne i jaką trasę wybrać, a mechanizmy przekazywania pakietów muszą jeszcze wiedzieć, jak dotrzeć do właściwego następnego routera.

Kto płaci za autostradę? Operatorzy mają umowy

Do sali wchodzi księgowy.

— Przepraszam, ale dlaczego wysyłamy cudze samochody przez naszą płatną autostradę za darmo?!
— Bo kolega poprosił!
— A mamy z nim umowę?
— Nie.
— TO PROSZĘ NATYCHMIAST ZAPOZNAĆ SIĘ Z DZIAŁEM FAKTUR!

🤣

W Internecie istnieją różne relacje biznesowe między sieciami. **Tranzyt** oznacza, że jeden operator zapewnia drugiemu osiągalność dalszych części Internetu. **Peering** oznacza bezpośrednią wymianę ruchu między sieciami według uzgodnionych zasad. Nie musi przy tym chodzić o tę samą cenę ani o jednakowy zakres dostępu.

Dlatego operatorzy filtrują, które prefiksy przyjmują i które ogłaszają dalej. Nie każdy sąsiad ma prawo używać naszej sieci jako bezpłatnej autostrady do wszystkich innych sąsiadów.

— Czyli nie każdemu wolno ogłaszać wszystko?
— Dokładnie.
— NAWET JEŚLI BARDZO ŁADNIE POPROSI?!
— Nawet wtedy.

AS_PATH prepend i MED, czyli uprzejme sugestie dla sąsiadów

Kierownik bardzo chciał, żeby zagraniczni kierowcy rzadziej wybierali jeden z jego wjazdów.

— Czy mogę im powiedzieć, że moja droga jest długa?
— Może pan sztucznie wydłużyć AS_PATH w ogłoszeniu, dopisując swój numer kilka razy.
— A to legalne?!
— To stosowana technika zwana **AS_PATH prepending**.
— A oni na pewno wybiorą inną trasę?
— Niekoniecznie.
— JAK TO NIEKONIECZNIE?!
— To ich polityka. Nie pan wybiera za cały Internet.

🤣🤣

Prepending może sprawić, że dana droga wyda się mniej atrakcyjna ze względu na długość AS_PATH, lecz nie daje gwarancji — inne polityki mogą mieć większe znaczenie.

Istnieje także atrybut **MED (Multi-Exit Discriminator)**, którym można zasugerować sąsiedniemu AS preferowany punkt wejścia, gdy istnieje kilka połączeń. To sugestia, nie rozkaz obowiązujący wszystkich operatorów.

— Panie kierowniku, sąsiad nie posłuchał naszej sugestii!
— TO MU WYŚLIJMY SUGESTIĘ WIĘKSZĄ CZCIONKĄ!
— To tak nie działa.

Ruch w jedną stronę i w drugą, czyli dwie różne decyzje

Na mapie przedstawiciel firmy rysuje strzałkę z Warszawy do Krakowa.

— Tędy wyślemy samochody!
— A którędy wrócą?
— TĘDY SAMO!
— Niekoniecznie. Drogę powrotną mogą wybierać inne sieci według własnych zasad.
— ALE PRZECIEŻ NARYSOWAŁEM JEDNĄ STRZAŁKĘ!
— Jedną, panie kierowniku. W jedną stronę.

BGP nie gwarantuje **symetrii tras**. Ruch wychodzący z naszej sieci w dużej mierze zależy od naszych preferencji i dostępnych tras, natomiast na ruch przychodzący wpływają decyzje sieci po drugiej stronie. Prepending i MED mogą wpływać na wybory innych operatorów, ale nie zapewniają pełnej kontroli nad nimi.

To bardzo ważne przy diagnozowaniu awarii: pakiet i odpowiedź nie muszą jechać tą samą drogą.

Wycofujemy trasę! Most się zawalił!

Do sali konferencyjnej wpada zdyszany urzędnik.

— PANIE KIEROWNIKU! MOST SIĘ ZAWALIŁ!
— JAKI MOST?!
— TEN, PRZEZ KTÓRY WSZYSCY JEŻDŻĄ DO WARSZAWY!
— To postawcie znak „MOSTU NIE MA”!
— Ale kierowcy z zagranicy nie wiedzą!
— TO ICH POINFORMUJCIE!
— WŁAŚNIE PO TO MAMY BGP!

🤣🤣

BGP umożliwia **wycofywanie ogłoszeń tras** (withdrawals). Gdy znika możliwość dotarcia do prefiksu, odpowiedni router może powiadomić sąsiadów. Ci przeliczają swoje trasy i mogą przekazać zmianę dalej. Proces dochodzenia sieci do nowego, stabilnego stanu nazywa się **konwergencją**.

Nie jest ona zawsze natychmiastowa. Liczą się wykrycie awarii, zachowanie sesji, topologia, filtry, dostępne drogi zapasowe i mechanizmy protokołów.

— Ile to potrwa?!
— To zależy.
— OD CZEGO?!
— Od konfiguracji i tego, jak szybko sieci dowiedzą się o zmianie.
— CZY NIE MOŻE PAN PO PROSTU POWIEDZIEĆ „JUŻ”?!
— Mogę. Tylko nie będzie to prawda.

BGP hijacking, czyli samozwańczy zarządca Warszawy

Drzwi sali otwierają się z hukiem. Wchodzi nieznany mężczyzna z mapą pod pachą.

— OD DZIŚ WSZYSTKIE DROGI DO WARSZAWY PROWADZĄ PRZEZE MNIE!
— A kim pan jest?
— Zarządcą Warszawy.
— Ale Warszawa jest w innej sieci!
— JUŻ NIE!
— JAK TO NIE?!
— BO JA TAK OGŁOSIŁEM!

🤣🤣

BGP hijacking polega na tym, że ktoś błędnie lub bez upoważnienia ogłasza osiągalność prefiksu należącego do innej sieci. Może to prowadzić do przechwycenia, przekierowania lub utraty ruchu, jeśli inni operatorzy zaakceptują nieprawidłowe ogłoszenie.

Nie wystarczy jednak krzyknąć „Warszawa jest moja", aby automatycznie przejąć cały światowy Internet. Efekt zależy od filtrów, polityk i **zabezpieczeń** stosowanych przez pozostałych operatorów. Nie wszystkie błędne ogłoszenia są też celowym atakiem — bywają zwykłymi pomyłkami w konfiguracji.

RPKI i ROA, czyli proszę pokazać upoważnienie

Przewodniczący spogląda na samozwańczego zarządcę.

— Dokumenty poproszę.
— Jakie dokumenty?!
— Potwierdzenie, że pański AS może ogłaszać ten prefiks.
— MAM WIZYTÓWKĘ!
— To nie jest ROA.
— A PIECZĄTKĘ?!
— Też nie.
— CHOLERA! CZY WY WSZYSCY MUSICIE WSZYSTKO SPRAWDZAĆ?!

🤣🤣

RPKI (Resource Public Key Infrastructure) i tworzone w nim **ROA (Route Origin Authorization)** pomagają sprawdzić, czy określony ASN jest upoważniony do rozpoczynania ogłoszenia danego prefiksu i o jakiej maksymalnej długości. Operatorzy mogą używać wyniku walidacji do odrzucania części nieprawidłowych tras.

Ważne ograniczenie: **ROA dotyczy przede wszystkim autoryzacji AS źródłowego ogłoszenia, a nie poświadczenia każdego kolejnego odcinka AS_PATH**. RPKI z walidacją pochodzenia znacząco pomaga, lecz nie rozwiązuje wszystkich problemów bezpieczeństwa BGP. Dodatkowo potrzebne są dobre filtry, właściwe polityki i kontrola własnych ogłoszeń.

Co BGP naprawdę robi, a czego nie robi?

Na zakończenie kierownik otrzymuje krótką ściągawkę.

BGP: wymienia informacje o osiągalności prefiksów; pozwala stosować polityki wyboru i propagacji tras; łączy niezależnie zarządzane sieci; pomaga reagować na wycofanie dróg. **BGP nie gwarantuje:** najkrótszej drogi w kilometrach, najmniejszego **pingu**, identycznej trasy powrotnej ani tego, że każde usłyszane ogłoszenie jest prawdziwe.

— Czyli BGP nie wozi sam samochodów?
— Nie. Informuje routery, jakimi drogami można próbować dotrzeć do sieci docelowych.
— TO KTO JE WOZI?!
— Routery przekazujące pakiety.
— I ONE CZYTAJĄ NASZE UCHWAŁY?!
— W pewnym sensie ich tablice **routingu** uwzględniają wyniki naszych decyzji.

Finał: najdziwniejszy system dróg na świecie

Po sześciu godzinach obrad przewodniczący ogłasza sukces.

— Szanowni państwo! Uzgodniliśmy zasady wymiany tras!
— WSPANIALE! Czy teraz każdy pojedzie najkrótszą drogą?!
— Nie.
— Najszybszą?!
— Niekoniecznie.
— Najtańszą?!
— To zależy dla kogo.
— TO JAKĄ?!
— Zgodną z polityką operatorów.

Kierownik długo patrzy na mapę.

— Czyli Internet działa dlatego, że tysiące niezależnych sieci wymieniają informacje o trasach i jakoś się dogadują?
— W dużym uproszczeniu tak.
— I ONE NAPRAWDĘ SIĘ ZE SOBĄ DOGADUJĄ?!
— Przeważnie.
— TO JEST NAJBARDZIEJ NIEPRAWDOPODOBNA RZECZ, JAKĄ DZISIAJ USŁYSZAŁEM!

🤣🤣🤣

Komitet Międzynarodowego Kongresu Zarządców Autostrad informuje, że następne spotkanie odbędzie się po uzgodnieniu trasy dojazdu. Ze względu na rozbieżne polityki BGP termin pozostaje nieznany.