Hur autonoma system och Border Gateway Protocol möjliggör internet

Kevin Koljonen 2201813
Kandidatavhandling i datateknik
Handledare: Jan Westerholm
Fakulteten för naturvetenskap och teknologi
Åbo Akademi
2023

Abstrakt

Denna text beskriver autonoma system och Border Gateway Protocl (BGP) som är en av de viktigaste protokollen som används inom interdomän kommunikation. Texten behandlar hur autonoma system är uppbyggda, deras attribut samt hurdana olika autonoma system det finns. Historien bakom och evolutionen av BGP kommer att behandlar samt hur protokollet fungerar, kommunicerar med andra routrar och hur BGP ställer in rutter till andra autonoma system. Dokumentet behandlar mest om interdomänsk kommunikation, det vill säga kommunikation mellan olika autonoma system och deras anhållande till varandra, men drar också en del från intradomänsk kommunikation som tar plats inom autonoma system för att få en bättre bild på hur datan från olika system på andra sidan världen, i olika autonoma system kommunicerar och skickar data genom rutterna som blivit etablerade i världen.

BGP använder sig av specifika meddelanden för att etablera anslutningar och utbyta ruttinformation mellan BGP peers, var ruttinformationen innehåller flera attribut som beskriver olika rutter. Autonoma system är då objekten som BGP bestämmer den snabbaste rutten mellan. Autonoma systemen har olika typer och relationer, och de används olikt beroende på dessa faktorer.

Med hjälp av autonoma system och BGP fungerar internet på en mycket större nivå än om de inte skulle existera. BGP fungerar som en ryggrad för dagens internet, var BGP möjliggör en snabb, responsiv internet tack vare dess dirigering. Autonoma system delar upp internet till mindre, mer lätthanterliga delar och hjälper också information att sprida beroende på dess relation till andra autonoma system. Det är inte en överskattning att säga att autonoma system samt BGP möjliggör dagens internet.

Ordlista

Autonoma system En grupp av maskiner inom ett nätverk var alla maskiner känner till alla andra maskiner. 2

BGP peer En BGP speaker som etablerat en anslutning med en annan BGP speaker. 2

BGP session Ett tillstånd var en BGP speaker skickar meddelanden med en annan BGP speaker. 2

BGP speaker En router som kör BGP. 2

Border Gateway Protocol nätverksprotokollet som används mellan autonoma system. 2

Dirigerinstabell En tabell som specifierar vilka rutter data skall ta så att det kommer fram till dens destination. 2

IPv4/6 Står för Internet Protocol version 4/6. 2

Maskin Kan vara en dator, router eller switch. En elektronisk apparat som kan uppkopplas till internet. 2

Innehåll

Kapitel 1

Introduktion

I dagens värld är internet ett redskap som används dagligen vilket betyder att nätverk finns allstans. Dessa nätverk är uppdelade till mindre och mindre nätverk tills man kommer till den individuella nivån var vi använder nätet. Nätverk är uppdelade på detta vis för att spara på IPv4 adresser som inte är oändliga och är utdelade till varje maskin. Uppdelningen av nätverk låter en utomstående maskin åtkomma flera maskiner genom en IPv4 adress som har ett prefix/suffix i adressen. Dessa uppdelade nätverk kallas autonoma system(AS) och varje AS har åtminstone en router som är ansluten till en annan router i en annan AS. För att maskiner i olika AS skall hitta varandra så används både intradomäna ruttningsprotokoll och interdomäna ruttningsprotokoll. Både de intradomäna och interdomäna ruttningsprotokollen använder deras egna ruttningstabeller. Medan det finns flera intradomäna ruttningsprotokoller finns det endast en för interdomän ruttningsprotokoll. Border Gateway Protocol är protokollet som används för interdomän ruttning och detta protokoll fungerar med att jämföra rutter, nätverkspolicyer eller regeluppsättningar. BGP är en av de viktigaste protokollerna för att internet idag fungerar eftersom att det är det mest använda protokollet för interdomän kommunikation.

Syftet med denna avhandling är att leta reda på hur olika nätverk är kopplade tillsammans samt hur en maskin från ett nätverk åtkommer en annan maskin inom ett annat nätverk. Detta ämne är viktigt eftersom den mesta kommunikation och lagring av information använder sig av internet. Därför är det viktigt att veta hur kommunikationen inom internet fungerar. BGP är den viktigaste överföringsprotokollet och på ett sätt ryggraden som håller internet förbunden.

1.1 Grunderna av nätverk

Denna sektion kommer handla om hur maskiner blir tilldelade så kallade IP adresser som sedan används för att bilda anslutningar till andra maskiner över internätet. Speciellt med fokus på intradomänsk ruttning. Av OSI modellen så kommer texten fokusera på transport och nätverkslagret.

1.1.1 Internet Protocol och dirigering

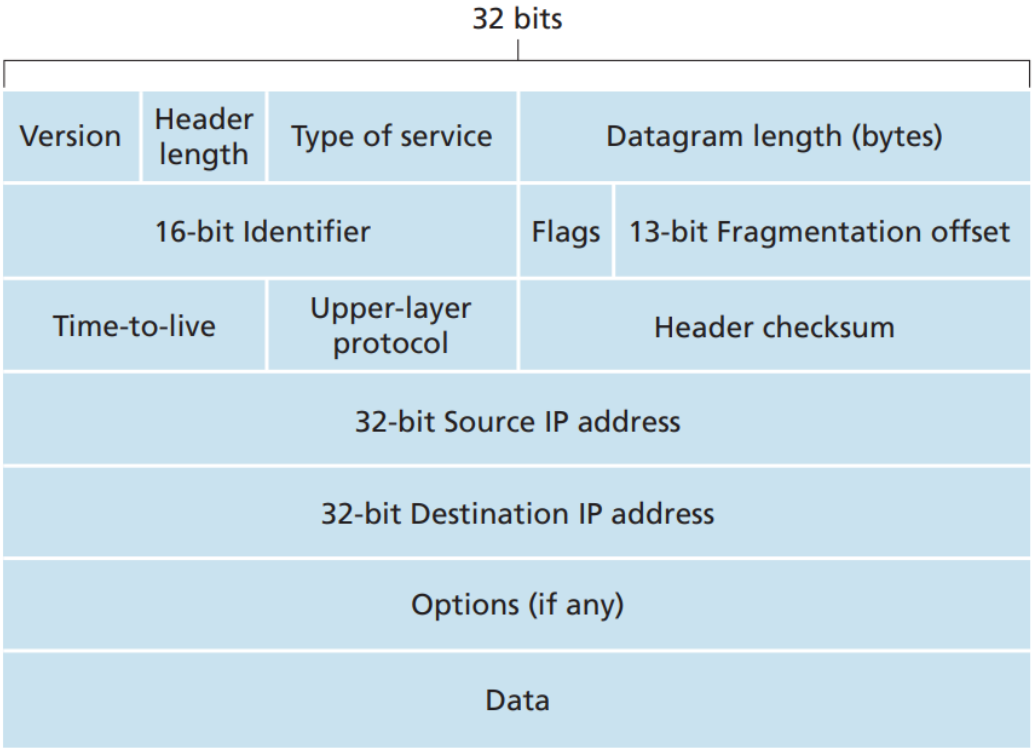
Internet protokollet (IP) är protokollet som möjliggör internet. IP är ansvarig för adressering av värdgränssnitt, inkapslande av data in i datagram och dirigering av datagrammen från en källvärdgränssnitt till en destinationsvärdgränssnitt. IPv4 är det dominerande protokollet av internet och dess efterträdare, IPv6, har varit tillgänglig sedan 1998 men spridningen och i bruk tagande av IPv6 har varit långsam. [1, p.4]

Varje maskin som är uppkopplat till internet har en IP adress som möjliggör att andra maskiner kan hitta den maskin som används. Ihop med maskinernas dirigeringstabeller så skickas paketerna genom en rutt som går genom flera maskiner, routrar och/eller switchar för att komma fram till deras destination.

Eftersom det kan finnas flera maskiner på samma nätverk och uppkopplat till samma router, så kan en router använda sig av en nätmask. Nätmasker kan maskera"en del av IP adressen och på detta vis generalisera hela nätverkets IP adress. Detta gör också dirigeringen till subnät lättare. CIDR (Classless Inter-Domain Routing) så är ett sätt att notera maskningen. CIDR använder sig av /0-24 notation som berättar hur många bitar av IP adressen skall ta i beaktande för att komma till rätt nätverk. t.ex. en IP adress som är 192.168.96.1/24 har subnäts masken 255.255.255.0 vilket säger att IP adresser som börjar med 192.168.96 befinner sig i detta subnät. CIDR notationen berättar också tydligare hur många IP adresser subnätet har blivit allokerat för. Då masken är /22 så finns det 1024 IP adresser men om masken är /24 finns det 256 IP adresser som kan användas i subnätet. Detta kan räknas genom formeln 2^{32-x} var x är nätmasken med CIDR notationen, det vill säga, om nätmasken är /24 så finns det $2^{32-24} = 2^8$ adresser.

Nedan så är figur 1.1 från Kurose och Ross [2] för formatet för datagrammet som IPv4 använder. Varje paket som skickas genom nätverkslagret med IP så kommer ha likadana datagram före själva datan som skickas. Eftersom datagrammernas storlek är relatvit små så måste datan som skall skickas delas upp i flera datagram och sedan skickar man de datagrammen över ett nätverk till mottagaren. Datagrammerna innehåller information om

källan och destinationen av paketet samt instruktioner på hur datan skall sättas tillbaka ihop då datan blir fragmenterad.



Figur 1.1: IPv4 datagram format [2, Figure 4.12]

De allmännaste överföringsprotokollerna för internet är Transmission Control Protocol (TCP) och User Datagram Protocol (UDP). Dessa protokollen säger hur 2 system interagerar med varandra för att kommunicera data mellan varandra. Den ovanstående figuren beskriver ett datagram inom UDP. UDP är en uppkopplingslös överföringsprotokoll eftersom den inte utför handskakningmeddelande utan endast överför datan till destinationen. UDP adderar endast ett meddelande från applikationen som datan skickas från, bifogar käll- och destinations port nummer fält för multiplexering/demultiplexering, adderar två små fält till och sedan skickar det datan vidare till nätverkslagret. [2, p.199] TCP till skillnad från UDP så är mer pålitlig än UDP eftersom TCP använder sig av handskakning som etablerar en anslutning mellan två maskiner. TCP kontrollerar också om datan har kommit fram och om den kommit fram intakt och så att inte endast vissa paket kommit fram till mottagaren. Eftersom TCP har mer information som den sänder samt tre-vägs handskakningen som TCP använder sig av så är TCP långsammare och mer krävande än UDP. [2, p.200]

Dirigeringen av ett paket så sköts av dirigeringstabeller som varje maskin har. Dirige-

ringstabellerna är en lista av rutter till specifika IP adresser. De flesta dirigeringstabeller har en förvald adress som den mesta trafiken till utomstående maskiner kommergå igenom, men dirigeringstabellen kan också använda sig av specifika IP adresser i samband med nätmasker för att skicka data till maskiner som är i ett näraliggande nätverk utan att behöva använda sig av internetleverantörens dirigeringstabeller och på de sättet kan paketen komma fram snabbare. Dirigeringstabellen kan ha både statiska och dynamiska rutter som den använder, dynamiska rutter uppdateras när en snabbare rutt hittas eller en ny maskin blir tillagd till nätverket och statiska rutter använder rutter som har blivit definierade förr. [3, p.59]

Dirigeringstabellerna har flera algoritmer för att räkna ut snabbaste rutten till paketernas destination. Algoritmerna som dirigerinstabellerna använder kan vara Bellman–Ford algoritmen eller Dijkstras algoritm. Båda av dessa algoritmerna hittar den snabbaste rutten mellan källan och destinationen, eller allmänt, olika noder. Den generella skillnaden mellan de två algoritmerna är att Bellman-Ford algoritmen hittar kortaste rutten till en destination per gång medan Djikstras algoritm räknar ut kortaste rutterna till alla destinationer. Effektiviteten av de två algoritmerna beror av hur nätverket är uppbyggt var generellt så är Bellman-Ford algoritmen snabbare eftersom den inte behöver karta ut alla rutter till alla destinationerna men om nätverket är fullt ansluten var alla noder känner till varandra så kommer Djikstras algoritm vara snabbare eftersom algoritmen kan räkna med alla rutterna på en gång istället för att ta en gång i gången. [3, p.43]

IP, dirigeringstabeller samt TCP och UDP är de almännaste verktygen som används inom autonoma system, som kommer förklaras i nästa kapitel. IP och dirigeringstabeller används också mellan autonoma system för att komma till sändarens destination men överföringsprotokollen har blivit ersatt med Border Gateway Protocol. BGP är det ända överföringsprotokollet som sköter om inter AS kommunikation.

Kapitel 2

Autonoma system

Autonoma system (AS) är en samling av datorer eller routrar som alla känner till varandra, dvs. varje maskin känner till varje andra maskin inom AS:en. Alla maskiner inom en AS kan åtkommas genom en eller flera IP prefix och är hanterad av en eller flera nätoperatörer vilka har en, väl definierad dirigeringspolicy. Inom en AS kan det grena ut sig från maskinerna som känner varandra till mindre nätverk som t..ex. en person kan använda. Allt som förklarades i introduktionen händer inuti autonoma system förutom BGP som tar hand om inter-AS kommunikation. [4, kap.3]

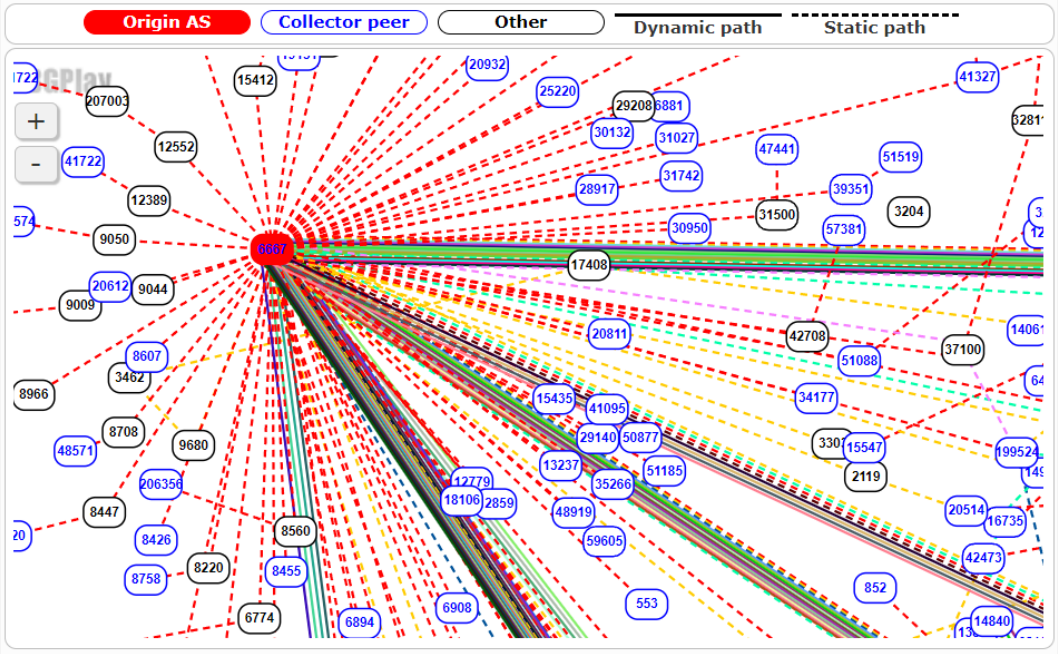
Varje offentliga AS har en äutonomous system number"(ASN) som blir tillagd till AS:n då den skapas. På lika sätt till portar på en dator så är ASN tilldelade autonoma system, dvs. det finns vissa ASN siffror som är reserverade och vissa siffror som är öppna åt publiken. I bilden 2.1 från ipinfo.io [5] så syns de största AS:ar i Finland och deras allokerade ASN samt hur många IP adresser varje AS har allokerat åt sig. AS siffrorna är tilldelade som block av Internet Assigned Number Authority (IANA) till regional Internet registries (RIR). Den lämpliga RIR tilldelar då ASN till enheter inom deras område från blocket som IANA har tilldelat RIR, så t.ex. den RIR som funktionerar i Finland blir tilldelat ett block av IANA varav RIR då tilldelar ASN till applikanter inom Finland. För Europa är det RIPE Network Coordination Centre (RIPENCC) som fungerar som den regionala Internet registret. [6]

Autonoma system använder sig då av routrar som kör med BGP för att bilda rutter mellan AS. AS:arnas dirigeringspolicy är definierade enligt utbyte av dirigeringsinformation

ASN	NAME	NUM IPS
AS719	Elisa Oyj	3,831,040
AS1759	Telia Finland Oyj	3,682,816
AS16086	DNA Oyj	1,818,624
AS1741	CSC - Tieteen tietotekniikan keskus Oy	609,792
AS208722	Global DC Oy	199,936
AS29422	Telia Inmics-Nebula Oy	157,440
AS41701	Capgemini Finland Oy	132,864

Figur 2.1: Lista på de största AS i Finland [5]

mellan AS:er. [4] I bilden 2.2 från RIPENCC [7] så syns autonoma systemet med ASN 'AS6667' och flera av AS:ens definierade ruttar till andra autonoma system. AS6667 är Finlands EUNet stamnät som leverantörer som Elisa Oyj och liknande använder sig av i Finland. Man kan själv kolla upp hur en AS bildar ruttar mellan andra AS:ar med verktyg som BGPlay som hittas vid: `https://stat.ripe.net/widget/bgplay`.



Figur 2.2: Figur på rutter som AS6667 har format [7]

2.1 Olika typer

Det finns olika typer av autonoma system varav flera baserar sig på hur många anslutningar de har till andra AS:ar.

enhemmade (single-homed) Stubba (stub) autonoma system beskriver AS:ar som är ansluten till endast en annan AS, så om anslutningen blir bryten på något sätt kan inte AS:en kommunicera över internet utan endast inom autonoma systemet. enhemmade stubba system behöver sällan skilida ASn eftersom de är ofta anslutna till t.ex. en internetleverantörs AS som har en ASn och kan dirigera sig till den stubba AS:en. Ett flerhemt (multihomed) stubb system så har flera anslutningar till andra autonoma system till skilja från en stubb. Flerhemmade system har då inte samma sårbarheter som en enhemmad stubb eftersom om en av anslutningarna bryts, så kan AS:en ännu kommunicera med internet genom en annan anslutning.

Transit AS:ar är sådana som erbjuder möjligheten att dirigera data mellan olika AS:ar, vilket en stubb AS inte kan göra. Till exempel om ASx kan skicka data till ASy genom ASz, så är ASz ett transit system. Transita autonoma system kan ha finansiella konsekvenser, var t.ex. en internetleverantör kan köpa transittjänster från en annan internetleverantör. [8]

Internetutbytespunkter (Internet Exchange Point) (IXP) är fysisk infrastruktur var internetleverantörer eller innehållsleveransnätverk (Content Delivery Network, CDN) utbyter trafik mellan deras AS:ar. IXP:s möjliggör kortare ruttar till andra nätverk som är anslutna till samma IXP, vilket sänker latens och tur- och returtid vilken som konsekvens kan potentiellt minska på kostnaderna. En IXP är uppbyggt av en eller flera ethernet switchar som är sammankopplade över en eller flera fysiska byggnader. I huvudsak är IXP:s primära uppgift att koppla ihop flera nätverks routers rent och effektivt. Till skillnad från transit AS:ar så är IXP en mer optimiserad system som presenterar kortare vägar till andra AS som ligger fysiskt nära. t.ex. om man inte skulle använda sig av IXP så kan det hända att datan som skickas från Helsingfors till Åbo går via Rovaniemi och tar på detta sätt en mycket längre rutt till ändpunkten istället för att gå via en IXP som har anslutningar mellan start och slutpunkterna. Detta fenomen kallas för 'tromboneffekten'. [9]

2.2 Anslutningar mellan autonoma system

Likadant som hur det finns olika typer av autonoma system så finns det olika anslutningar mellan AS:ar också.

Peering är då ett par av autonoma system bildar ett ömsesidigt avtal med varandra för att utbyta trafik med varandra utan kostnad. Antagandet mellan AS:arna är att båda har ett intresse av att ansluta till den andras kunder, vilket liknar hur postsystem inte tar betalt då post dirigeras från ett land till ett annat. Peering avtal är inte transitiva så att om AS1 pirar med AS2 och AS2 med AS3 så är inte AS2 nödvändigtvis skyldig att transportera trafik till AS3, men detta är ett affärsmässigt beslut istället för ett tekniskt beslut. [8]

Ett transit förhållande, som en transit AS, är då en AS går med på att fungera som en router, var den bär på trafik mellan AS:ar som den är sammankopplade med. Exempel på AS:ar med transit förhållande är internetleverantörernas AS:ar som har transit förhållande med deras kunder då internetleverantören säljer tillgång till internet. En transit AS använder sig typiskt av mätare som mäter trafiken från olika anslutningar och sedan ta en transitavgift. [8]

2.2.1 Internetnivåer

Som förhållandena säger så finns det anslutningar som säljs och andra som är kostnadsfria. Internetnivåerna baserar sig på hurdana förhållanden internetleverantörerna har med andra autonoma system.

Nivå 1 internetleverantörer är de som inte behöver betala för andra nätverk för transit. Istället för att köpa transit mellan AS:ar så bildar de peers med alla andra nivå 1 nätverk, av vilka det fanns cirka 14 över hela världen. Given en valfri IP-adress kommer en nivå 1 internetleverantör kunna ansluta sig direkt till en toppnivå internetleverantör som kan dirigera trafiken till den adressen. t.ex. Förenta staterna har 8 sammankopplade regioner vilket bildar en förvalsfri zon var nivå 1 internetleverantörer ansluter deras nätverk i peering avtal.

Nivå 2 internetleverantörer är sådana som behöver köpa transit förhållanden för att ansluta sig till åtminstone någon del av internet. Eftersom nivå 2 internetleverantörer måste betala transitkostnader så försöker de bilda så många direkta peering förhållanden som de kan med andra nivå 1 och nivå 2 internetleverantörer så att de kan utbyta trafik utan kostnader med dem. Kostnader kan ändå finnas som en peeringavgift vilket bestäms affärsmässigt. t.ex. det är vanligt för kabel och telefonbolag att pira sig med innehållsleverantörer såsom Google eller Amazon. [8]

Kapitel 3

Border Gateway Protocol (BGP)

Detta kapitel berättar om Border Gateway Protocol (BGP) som är designat att utbyta dirigerings- och nåbarhetsinformation mellan autonoma system och dess utveckling sedan protokollets uppfinnelse i 1989. BGP gör dirigeringsbeslut baserat på rutter, nätverkspolycier eller regler vilka är konfigurerade av en nätverksadministratör. [10]

BGP är en 'Exterior Gateway Protocol' (EGP) vilket beskrivs som ett protokoll som hanterar information om rutter och trafik mellan autonoma system. Utvecklat i stor del av Yakov Rekhter och Kirk Lougheed. BGP är ett protokoll som använder sig av TCP för att förmedla meddelanden mellan gränsroutrar (Gateway Routers). Information om nåbarhet innebär information om autonoma system var nåbarhetsinformationen traverserar. BGP-4 har mekanismer som stöder användningen av CIDR IP prefix och aggregering av rutter inklusive AS rutter.

Information om rutter som är utbytta med BGP stöder ändast destinationsbaserat vidarebefodring under antagandet att routern vidarebefodrar ett paket endast med basen av IP-adressen som hittas i headern som konsekvent bestämmer hurudana regler man kan inställa med BGP i routern. [11] BGP använder sig av både externa och interna anslutningar när de bildar BGP peers. En BGP peer är en BGP speaker som har etablerat en peer relation med en annan BGP speaker. Av BGP peers finns det två olika versioner: IBGP peers och EBGP peers. IBGP peers hänvisar till BGP peers inom ett autonomt system och med EBGP peers menar man BGP peers som existerar i ett annat autonomt system. [12]

3.1 Utvecklingen av BGP

Uppfinnelsen av BGP började i 1989 då Yakov Rekhter och Kirk Lougheed ritade ut protokollet på baksidan av servetter, därför känns BGP också med namnet 'three-napkin protocol' eller 'Two-Napkin Protocol' beroende på vem man frågar. Det första menade användningsfallet med BGP var som ett snabbt fix till internetets problem vid tiden och skulle bli ersatt med något bättre så snabbt ett bättre alternativ uppfanns. Detta hände aldrig utan BGP används ännu idag fast man har utvecklat och itererat på BGP för att fixa fel som fanns, klargöra oklarheter och uppdatera protokollet till vanliga praxis inom branschen. [13]

Första beskrivningen av BGP kom ut 1989 i RFC 1105 [10] och kallas för BGP version 1. BGP-1 fick sin första stora revision i 1990 vilket kallas för BGP version 2. 1991 blir BGP pånytt uppdaterad med 'BGP algorithm Analysis' och BGP version 3. I 1994 släpps BGP version 4 ut som tas i bruk samma år och används ännu idag med några viktiga uppdateringar som kom ut i 1995 samt en revision i 2006. [14]

Skillnaderna mellan BGP-1 och BGP-2 finns i formaterna av meddelanden som utbyts mellan gränsroutrar samt har mycket av specifikationerna blivit utvecklade och lättare förklarade. Skillnaderna från BGP-1 och BGP-2 till BGP-3 är större än från 1 till 2. BGP-3 använder sig av nya fält i meddelandernas format samt har hierarkiska anslutningar förblivit och så har BGP:s finita tillståndsmaskin blivit utvecklad och ändrad med mera specifika händelser. Utvecklingar inom BGP i BGP-4 så är den största utvecklingen förmågan att hantera CIDR prefix vilket eliminerar helt nätverks 'klasser' inom BGP. BGP-4 introducerar också mekanismer vilka tillåter aggregering av rutter, inklusive AS rutter. [10], [11], [15], [16]

3.2 Hur BGP-4 fungerar

Eftersom BGP-4 är den version som är ännu i användning idag samt den största EGP i använding så analyserar jag hur BGP-4 fungerar enligt RFC 4271 vilket är skrivet av Rekhter, Hares och Li [11]. Detta dokument innehåller specifikationerna för BGP som ännu används idag. När det skrivs BGP så refererar jag till BGP-4 specifikt.

BGP använder TCP som sitt transport-protokoll vilket eliminerar behovet av att implementera explicita fragmenterings, återsändnings, bekräftnings och sekvenserings funktioner. BGP lyssnar på TCP port 179 och felmeddelningsmekanismen i BGP antar att TCP

versionen stöder en 'graciös' stängning, vilket innebär att all data kommer bli skickat före anslutningen stängs. [11]

3.2.1 BGP meddelanden

För att andra BGP speakers (routrar som kör BGP) ska kunna hitta varandra så annonserar en BGP speaker alla de rutter som den har en peer anslutning med. Informationen meddelas till de sammankopplade BGP speakers genom UPDATE meddelanden vilket är ett sorts meddelande mellan BGP speakers för att uppdatera ruttinformation medsammans.

Det finns 4 stycken meddelanden som BGP speakers kan skicka mellan varandra för att kommunicera och alla dessa meddelanden är också headers vilka beskriver kompatibilitet, meddelandets längd och vilket typ av meddelande skickas. Meddelanden i ordning av typ koderna i headern är OPEN, UPDATE, NOTIFICATION och KEEPALIVE. Det finns också en till typ av meddelande som heter ROUTE-REFRESH vilket anhåller om att BGP speakern blir omanonnserad. De olika typerna av meddelanden har då olika format som adderas till efter headern. [11], [17]

Den första typen av meddelande kallas för OPEN och detta skickas efter en TCP anslutning har blivit etablerad. OPEN meddelandet skickas till en router som sedan svarar tillbaka med ett KEEPALIVE meddelande om OPEN meddelandet var acceptabelt. Detta öppnar en BGP anslutning mellan två BGP speakers. OPEN meddelandet innehåller information om BGP version, AS nummer, 'hold timer' vilket bestämmer hur länge routern skall vänta på KEEPALIVE meddelandet, BGP indentifierare som är definierad med IP-adressen som BGP speakern har vilket är definierat vid uppstart och är samma för varje lokalt gränssnitt och BGP peer. Det finns också valfria parametrar i OPEN meddelandet som beskriver vilka förmågenheter BGP speakern stöder. [11], [18]

UPDATE meddelandet används för att skicka dirigeringsinformation mellan BGP peers. Informationen i UPDATE meddelandet kan användas för att rita upp en graf som beskriver relationerna av de olika autonoma systemen. UPDATE meddelandet används också för annonsering av möjliga rutter som har gemensamma ruttattributer till en peer och också dra tillbaka flera rutter som inte är längre möjliga. UPDATE meddelandet innehåller fält om tillbakadragna rutter och deras längd, totala rutter som BGP speakern har och dess längd och information om nätverkslagrets nåbarhet ('Network Layer Reachability Information', NLRI). Om ett UPDATE meddelande endast annonserar rutter som är tillbakadragna från service så inkluderas det inte NLRI info i meddelandet. [11]

NOTIFICATION meddelandet skickas då ett feltillstånd uppstår och består av en felkod, felunderkod och data som diagnoserar orsaken för NOTIFICATION meddelandet. Det finns 6 felkoder som kan uppstå som är 'Message Header Error', 'OPEN Message Error', 'UPDATE Message Error', 'Hold Timer Expired', 'Finite State Machine Error' och 'Cease'. De första 3 felkoderna har felunderkoder vilket då mera specifierar vad som gått fel. 'Hold Timer Expired' betyder att tiden för att skicka ett KEEPALIVE meddelande har gått ut, 'Finite State Machine Error' innebär att t.ex. en oförväntad händelse hände och 'Cease' felkoden skickas om en BGP peer har valt att stänga sin anslutning till någon anna BGP speaker. [11]

KEEPALIVE meddelandet skickas mellan BGP peers med specificerade tidintervall för att hålla anslutningen öppen. Intervallet för att KEEPALIVE meddelandet ska omskickas bestäms av hur lång 'Hold timer' de förra meddelanden har, ett rimligt sätt att bestämma maximala tidsintervallet är att ta en tredjedel av hold timern. KEEPALIVE meddelandet består av endast meddelandets header. [11]

3.2.2 Attribut av rutter

Inom UPDATE meddelandet specificeras attributerna av rutterna som etableras av BGP och dessa ruttattributer kan delas upp i 4 olika kategorier. Välkända obligatoriska ('Well-known mandatory'), välkända diskretionära ('Well-known discretionary'), valfria transitiva ('Optional transitive') och valfria icke-transitiva ('Optional non-transitive').

Alla implementationer av BGP måste känna igen alla välkända attribut och de attributen som är välkända och obligatoriska måste finnas i varje UPDATE meddelande som använder sig av NLRI fältet. När en BGP peer har uppdaterat någon av deras välkända attribut måste den skicka dessa attribut vidare till dess peers i alla uppdateringar som den skickar ut. [11]

I addition till de välkända attributerna som en BGP speaker måste veta kan det också inkluderas en eller flera valfria attributer men det är inte förväntat att alla implementationer av BGP kommer stöda alla valfria attributerna. Hanteringen av ett okänt valfritt attribut bestäms av inställningen av den transitiva biten i attributflaggans innehåll, vilket finns inuti UPDATE meddelandets ruttattribut fält. Rutter med okända transitiva valfria attribut borde accepteras och skickas vidare till andra BGP speakers. Okända icke-transitiva valfria attribut måste tyst ignoreras och inte skickas vidare till andra BGP peers. Nya valfria transitiva attribut må bifogas till en rutt av skaparen av rutten eller vilken som helst annan

BGP speaker inkluderad i rutten.

3.2.3 Typer av attributer

Attribut som används i UPDATE meddelanden är dessa: ORIGIN, AS_PATH, NEXT_HOP, MULTI_EXIT_DISC, LOCAL_PREF, ATOMIC_AGGREGATE och AGGREGATOR. Utav dessa så är ORIGIN, AS_PATH, NEXT_HOP är välkända obligatoriska attribut, LOCAL_PREF och ATOMIC_AGGREGATE är också välkända attribut men är inte obligatoriska att ha med i meddelandet. Resten av attributen är valfria medan MULTI_EXIT_DISC är icke-transitiv och AGGREGATOR är transitiv.

ORIGIN är ett attribut genererad av speakern var rutten hade sitt ursprung och ska inte modifieras av någon annan speaker.

AS_PATH attributet är ett attribut vilket identifierar de autonoma system av vilka UPDATE meddelandet har gått igenom. När UPDATE meddelandet skickas igenom en AS så adderar AS:en sitt eget ASn till UPDATE meddelandet, men detta händer inte om meddelandet har sin destination inom AS:en.

NEXT_HOP attributet definierar IP-adressen av routern som borde användas som nästa hoppet på väg till destinationerna listade i UPDATE meddelandet. NEXT_HOP attributet modifieras på olika sätt beroende på hur många hopp meddelandet har kvar till sin destination, generellt om slutdestinationen är inom AS:en kommer NEXT_HOP attributet inte modifieras medan och hoppet är externt så modifieras attributet med data från själva attributet och ruttninginformation som routern har.

MULTI_EXIT_DISC attributet är menat att bli använt för externa länkar för att urskilja mellan en eller flera utgångspunkter till samma grann-AS. Detta attribut skall inte propageras vidare till andra AS:ar.

LOCAL_PREF måste vara med i meddelanden som en BGP speaker skickar till interna peers. Attributet anger ursprungsrouterns preferenser angånde externa rutter. Detta attribut måste vara tomt när UPDATE meddelandet skickas till externa peers med undantag om speakern är med i en 'BGP confederation' vilket gör så att flera AS kan samlas ihop till en stor AS.

ATOMIC_AGGREGATE inkkluderas då en BGP speaker aggregerar flera rutter för att

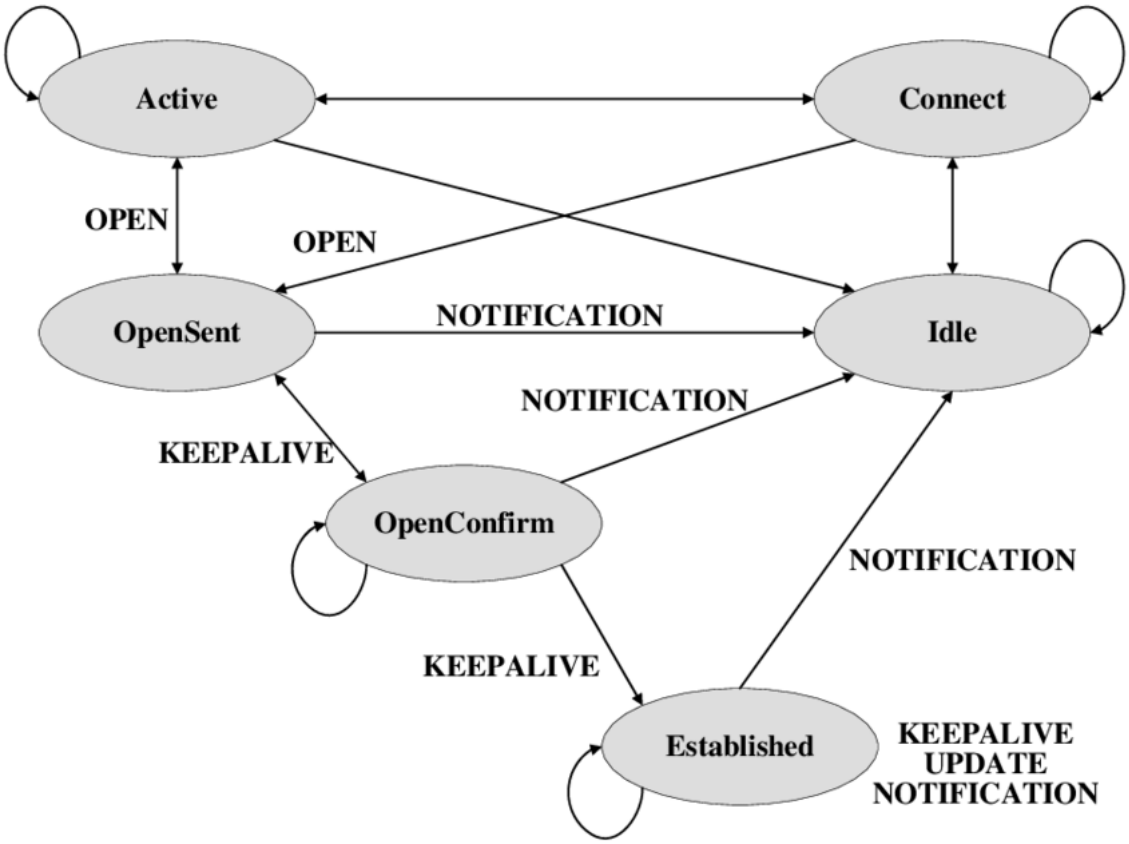
annonsera dem till en specifik peer, attributet innehåller information om de AS:ar som rutterna har använt sig av då den aggrerade rutten formades. Attributet ska också inte bli borttaget då rutten propageras vidare till andra speakers.

AGGREGATOR attributet adderas valfritt till då en AS aggregerar rutter för att identifiera vilken AS har utfört aggregationen. AGGREGATOR attributet innehåller både ASn och IP-adress.

3.2.4 BGP:s finita tillståndsmaskin

För att hålla BGP konsistent över hela nätet fungerar handlingarna som protokollet gör som en tillståndsmaskin. Tillståndsmaskinen är livsviktig för BGP för att kunna avgöra nästa handling.

I figur 3.1 taget från 'DECENTRALIZED MODULAR ROUTER ARCHITECTURES' skrivet av Hidell [19] så visar hur BGP:s tillståndsmaskin ser ut och också förenklar hur de olika tillstånden leder till varandra.



Figur 3.1: Figur över BGP:s finita tillståndsmaskin [19]

Border Gateway Protocol startar i ett 'IDLE' tillstånd var protokollet väntar på 'ManualStart' eller 'AutomaticStart' händelsen. När detta sker så börjar BGP med att initialisera alla resurser som behövs för att ansluta till en peer, lägger 'ConnectRetryCounter' till noll och startar den med det initiala värdet, startar en TCP anslutning med den andra BGP peeren, lyssnar efter anslutningar som kan ha startats av andra BGP peers och till sist byter tillståndet till 'CONNECT'.

'CONNECT' tillståndet innebär att BGP startar och väntar på att TCP anslutningen till BGP peeren ska bli färdig. Tillståndet hålls vid 'CONNECT' ända tills TCP:s trevägshandskakning har utförts färdigt. Det är antaget att båda sidorna av anslutningen kommer försöka starta en BGP session med varandra och då kommer peeren med högre router ID att ta hand om BGP sessionen.

Border Gateway Protocol hoppar till 'ACTIVE' tillståndet då den första trevägshandskakningen misslyckades och det skickas en ny trevägshandskakning till den andra BGP peeren. I 'ACTIVE' tillståndet kommer BGP försöka att skicka ett nytt OPEN meddelande för att öppna en BGP session och om detta försök också misslyckas faller BGP tillbaka till 'CONNECT' tillståndet. Generellt så kommer BGP endast vara i detta tillstånd om den första anslutningen misslyckades.

I tillståndet 'OPEN SENT' så har ett OPEN meddelande blivit skickat av båda routrarna. Då en router har skickat samt mottagit ett OPEN meddelande flyttar sig BGP till tillståndet 'OPEN CONFIRM' var ett KEEPALIVE meddelande är skickat av båda routrarna.

När båda routrarna har mottagit KEEPALIVE meddelandet så flyttar sig BGP till tillståndet 'ESTABLISHED' var BGP kan börja utbyta information om rutter. 'ESTABLISHED' tillståndet är ända tillståndet som tillåter utbytet av dirigeringsinformation, d.v.s om BGP ligger i vilket som helst annat tillstånd räknas BGP sessionen som icke-funktionell. [20]

Kapitel 4

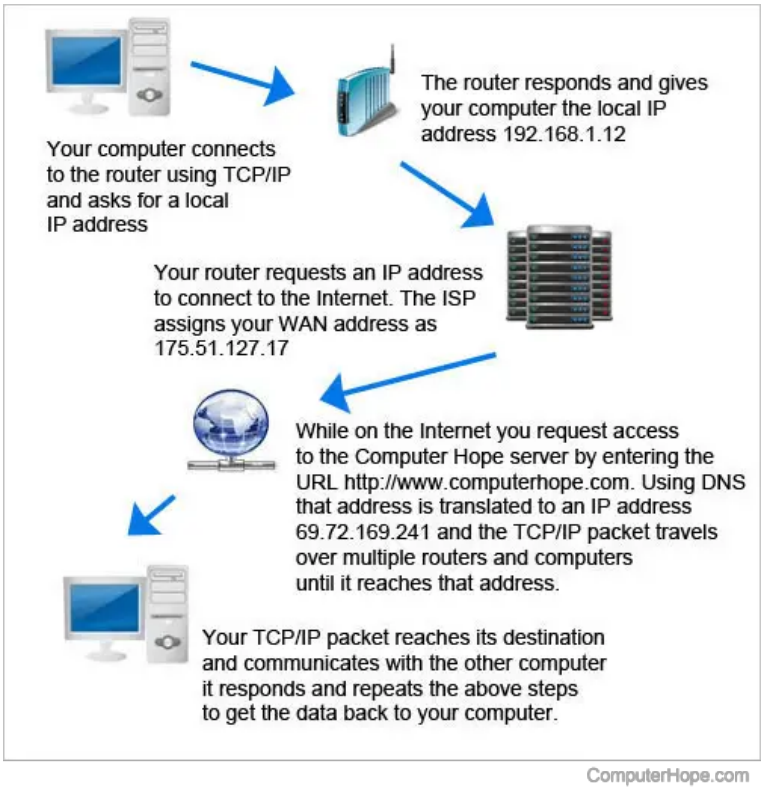
Interdomän ruttning

Detta kapitel förklarar mer konkret vad som händer när en dator kommunicerar med internet eller någon annan över internet var det ingår inter-AS överföring. Hur anlutningen mellan internetleverantören och resten av internet trafikerar och dirigerar paketerna genom olika autonoma system för att komma fram till dess slutdestination.

När en router, i ett hem eller liknande, blir ansluten till internet genom en internetleverantör så blir routern tilldelad en IP-adress av internetleverantören, samt så tilldelar routern lokala IP-adresser till de maskiner som är anslutna till routern. Från routern skickas paketerna enligt dirigeringstabellen i routern via internetleverantörens routrar vilka har deras egna dirigeringstabeller och för paketerna vidare. Om slutdestinationen av vart man vill ansluta är inom internetleverantörens AS så kommer paketerna inte lämna internetleverantörens AS utan kommer antagligen gå via en BGP speaker inuti AS:en som i sin tur kommer börja med en intern BGP session för att skicka paketerna till dess slutdestination.

I figuren 4.1 av ComputerHope [21] så visas hur en dator ansluter till internet, men vad figuren inte visar är hur autonoma system och BGP påverkar dirigeringen och visar inte klart om paketerna måste hoppa till eller genom andra AS:ar för att komma till dess slutdestination. I figuren 4.1 förklaras hur en dator blir tillagd en lokal IP-adress samt hur internetleverantören ger hemroutern en IP-adress. Dessa IP-adresser existerar då inom internetleverantörens AS såsom alla andra som är leverantörens kunder. Webbsidor använder URL (Uniform Resource Locator) för att åtkommas på internet och dessa är transformationer av IP-adresser med hjälp av DNS (Domain Name System) omvandlar

URL länken tillbaka till en IP-adress så att trafik kan dirigeras dit. Till sist kommer paketerna fram till slutdestinations datorn vilket då repeterar processen med källdatorn som slutdestination för att utbyta information. [21]

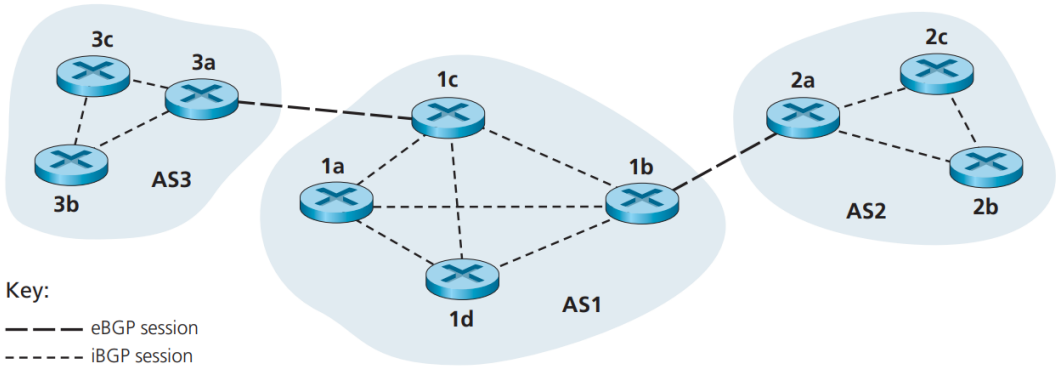


Figur 4.1: Förklaring av hur en dator ansluter till webbsidan computerhope.com [21]

Den flesta datan som skickas görs med hjälp av TCP/IP för att etablera anslutningar mellan routrar och datorer. Beroende på hur långt borta slutdestinationen kan vara, måste datan som skickas gå via BGP speakers och hoppa till andra AS:ar genom endera peer AS:ar eller transit AS:ar. Varje BGP speaker har då deras egna dirigeringstabell och regler på hur de ska dirigera trafik beroende på vilken AS de är i och på personen som definierat dessa regler. BGP speakers fungerar som gränsroutrar för AS:arna. Metoden för överföringen av data hålls samma förutom att rutten som väljs är bestämd av BGP när det gäller interdomän kommunikation. Det vill säga att TCP/IP tar hand om överföringen av data medan BGP tar hand om dirigeringen av datan.

När en BGP speaker har hittat och etablerat en anslutning med en annan BGP speaker så kan data överföras mellan dem. I figur 4.2 av Kurose och Ross [2] demonstreras hur interdomän kommunikation kan se ut som. De tjockare linjerna är rutter som BGP har valt och de tunnare beskriver hur BGP drigierar trafik inuti en AS. BGP har då interna BGP sessioner med alla routrar inuti en AS samt externa BGP sessioner med andra BGP

peers i andra AS:ar. Dessa rutter hjälper då BGP för att bestämma den bästa rutten för inkommande data. t.ex. om en dator ansluten till routern 3b och vill skicka ett meddelande till en dator som ligger ansluten till routern 2b dirigeras paketen först till 3b varav med hjälp av dirigeringstabellen skickar paketen vidare till 3a som fungerar som gränsroutern, därifrån skcikas paketen vidare till 1c - 1b - 2a - 2b och därifrån till datorn som är slutdestinationen. I figuren 4.2 är routrarna 3a och 1c samt 1b och 2a BGP peers.



Figur 4.2: Figur över hur inter-AS kommunikation kan se ut som [2, Figure 4.40, s.391]

Kapitel 5

Sammanfattning

Dagens internet är format av tusentals autonoma system med deras egna regler på hur informationen ska dirigeras igenom dem och med hjälp av BGP och CIDR kan man allokera mera IPv4-adresser än förr. Tack vare BGP kan man använda internet på ett snabbt och smidigt sätt. BGP sessioner hålls mellan BGP speakers för att etablera anslutningar och utbyta dirigeringsinformation för att bättre kunna dirigera inkommande trafik till dess slutdestination. Autonoma system existerar för att kunna organisera internet och spara på IPv4-adresser som håller på ta slut.

Början av BGP som en snabb fix för 1990-tals problem över internet utritat på servetter har utvecklats till en av de viktigaste protokollen för internet. Protokollet som tar hand om den största delen av trafiken över internet.

Utan autonoma system, BGP och CIDR skulle man inte kunna använda internet som på modernt vis. Tack vare insatser av Y. Rekhter och K. Lougheed samt många andra ingenjörer kan vi njuta och använda oss av internet på modernt vis.

Källförteckning

[1] OECD, ”The Economics of Transition to Internet Protocol version 6 (IPv6),” nr 244, 2014. DOI: https://doi.org/https://doi.org/10.1787/5jxt46d07bhc-en. URL: https://www.oecd-ilibrary.org/content/paper/5jxt46d07bhc-en.

[2] J. F. Kurose och K. W. Ross, *Computer Networking: A Top-Down Approach (6th Edition)*, 6th. Pearson, 2012, ISBN: 0132856204.

[3] F. P. Kelly, ”Network routing,” *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*, årg. 337, nr 1647, s. 343–367, 1991.

[4] J. A. Hawkinson och T. J. Bates, *Guidelines for creation, selection, and registration of an Autonomous System (AS)*, RFC 1930, mars 1996. DOI: 10.17487/RFC1930. URL: https://www.rfc-editor.org/info/rfc1930.

[5] ipinfo.io, *Finland ASN Summary*. URL: https://ipinfo.io/countries/fi.

[6] RIPENCC, *What We Do*. URL: https://www.ripe.net/about-us/what-we-do.

[7] RIPENCC, *BGPlay*. URL: https://stat.ripe.net/widget/bgplay#w.resource=6667.

[8] P. Krzyzanowski, *Understanding Autonomous Systems - Routing and Peering*, mars 2016. URL: https://people.cs.rutgers.edu/~pxk/352/notes/autonomous_systems.html.

[9] Cloudflare, *What is an Internet exchange point? | How do IXPs work?* URL: https://www.cloudflare.com/en-gb/learning/cdn/glossary/internet-exchange-point-ixp/.

[10] Y. R. K. Lougheed, *Border Gateway Protocol (BGP)*, RFC 1105, juni 1989. DOI: 10.17487/RFC1105. URL: https://www.rfc-editor.org/info/rfc1105.

[11] Y. Rekhter, S. Hares och T. Li, *A Border Gateway Protocol 4 (BGP-4)*, RFC 4271, jan. 2006. DOI: 10.17487/RFC4271. URL: https://www.rfc-editor.org/info/rfc4271.

[12] H. P. E. Development, *BGP speaker and BGP peer*. URL: https://techhub.hpe.com/eginfolib/networking/docs/switches/5710/5200-4992_13-ip-rtng_cg/content/517702401.htm.

[13] C. Timberg, *The long life of a quick 'fix'*. URL: https://www.washingtonpost.com/sf/business/2015/05/31/net-of-insecurity-part-2/.

[14] Datapath.io, *The History of Border Gateway Protocol*. URL: https://medium.com/@datapath_io/the-history-of-border-gateway-protocol-a212b7ee6208.

[15] *Border Gateway Protocol (BGP)*, RFC 1163, juni 1990. DOI: 10.17487/RFC1163. URL: https://www.rfc-editor.org/info/rfc1163.

[16] *Border Gateway Protocol 3 (BGP-3)*, RFC 1267, okt. 1991. DOI: 10.17487/RFC1267. URL: https://www.rfc-editor.org/info/rfc1267.

[17] E. Chen, *Route Refresh Capability for BGP-4*, RFC 2918, sept. 2000. DOI: 10.17487/RFC2918. URL: https://www.rfc-editor.org/info/rfc2918.

[18] J. Scudder och R. Chandra, *Capabilities Advertisement with BGP-4*, RFC 3392, nov. 2002. DOI: 10.17487/RFC3392. URL: https://www.rfc-editor.org/info/rfc3392.

[19] M. Hidell, "DECENTRALIZED MODULAR ROUTER ARCHITECTURES," mars 2023. URL: https://www.researchgate.net/publication/242367521_DECENTRALIZED_MODULAR_ROUTER_ARCHITECTURES.

[20] InetDaemon, *BGP Finite State Model*. URL: https://www.inetdaemon.com/tutorials/internet/ip/routing/bgp/operation/finite_state_model.shtml.

[21] ComputerHope, *How do computers connect over the Internet?* Juni 2020. URL: https://www.computerhope.com/issues/ch001358.htm.