BITAG

Broadband Internet Technical Advisory Group

Interconnection and Traffic Exchange on the Internet

A BROADBAND INTERNET TECHNICAL ADVISORY GROUP
TECHNICAL WORKING GROUP REPORT

A Uniform Agreement Report

Issued:
November 2014

Copyright / Legal Notice

This document and the information contained herein is provided on an "AS IS" basis and BITAG AND THE CONTRIBUTORS TO THIS REPORT MAKE NO (AND HEREBY EXPRESSLY DISCLAIM ANY) WARRANTIES (EXPRESS, IMPLIED OR OTHERWISE), INCLUDING IMPLIED WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT, FITNESS FOR A PARTICULAR PURPOSE, OR TITLE, RELATED TO THIS REPORT, AND THE ENTIRE RISK OF RELYING UPON THIS REPORT OR IMPLEMENTING OR USING THE TECHNOLOGY DESCRIBED IN THIS REPORT IS ASSUMED BY THE USER OR IMPLEMENTER.

The information contained in this Report was made available from contributions from various sources, including members of Broadband Internet Technical Advisory Group, Inc.'s Technical Working Group and others. Broadband Internet Technical Advisory Group, Inc. takes no position regarding the validity or scope of any intellectual property rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this Report or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights.

About the BITAG

The Broadband Internet Technical Advisory Group (BITAG) is a non-profit, multi-stakeholder organization focused on bringing together engineers and technologists in a Technical Working Group (TWG) to develop consensus on broadband network management practices and other related technical issues that can affect users' Internet experience, including the impact to and from applications, content and devices that utilize the Internet.

The BITAG's mission includes: (a) educating policymakers on such technical issues; (b) addressing specific technical matters in an effort to minimize related policy disputes; and (c) serving as a sounding board for new ideas and network management practices. Specific TWG functions also may include: (i) identifying "best practices" by broadband providers and other entities; (ii) interpreting and applying "safe harbor" practices; (iii) otherwise providing technical guidance to industry and to the public; and/or (iv) issuing advisory opinions on the technical issues germane to the TWG's mission that may underlie disputes concerning broadband network management practices.

The BITAG Technical Working Group and its individual Committees make decisions through a consensus process, with the corresponding levels of agreement represented on the cover of each report. Each TWG Representative works towards achieving consensus around recommendations their respective organizations support, although even at the highest level of agreement, BITAG consensus does not require that all TWG member organizations agree with each and every sentence of a document. The Chair of each TWG Committee determines if consensus has been reached. In the case there is disagreement within a Committee as to whether there is consensus, BITAG has a voting process with which various levels of agreement may be more formally achieved and indicated. For more information please see the BITAG Technical Working Group Manual, available on the BITAG website at www.bitag.org.

BITAG TWG reports focus primarily on technical issues, especially those with the potential to be construed as anti-competitive, discriminatory, or otherwise motivated by non-technical factors. While the reports may touch on a broad range of questions associated with a particular network management practice, the reports are not intended to address or analyze in a comprehensive fashion the economic, legal, regulatory or public policy issues that the practice may raise.

BITAG welcomes public comment. Please feel free to submit comments in writing via email at comments@bitag.org.

Executive Summary

The term "interconnection" refers to the various means by which network providers attach to and move traffic between one another, and is a collection of business practices and technical mechanisms that allow individually managed networks to connect together for this purpose. There is no central authority that manages Internet interconnection – the overall system arises because of the many bilateral and multilateral decisions that various actors make to interconnect. In contrast to the telephone system, where interconnection is performed in a highly regulated environment, interconnection in the Internet (in most parts of the world) remains a private-sector matter.

The topic of Internet interconnection is receiving increased attention as the Internet ecosystem continues to evolve. Networks of all types interconnect among one another, including those of Internet access providers, content providers, academic institutions, and commercial enterprises. Internet connectivity is achieved by passing pieces of data, called packets, from a connected device through networking equipment, known as routers, operated by one or more network providers until those packets are delivered to the desired destination. The mechanisms that implement interconnection thus serve both technical and business purposes, and discussion concerning the technology of interconnection must, of necessity, refer to business issues to some extent – as many of the mechanisms can only be understood in that context.

Network interconnection in the United States has evolved significantly since the early days of the Internet, and today is a complex amalgam of models incorporating new connectivity options, delivery options, traffic management requirements and business practices. It is important to note the difference between the two dominant forms of interconnection, which are: (1) *transit* – where access to every publicly reachable destination on the Internet is provided for a fee; and (2) *peering* – where customer traffic is exchanged between two networks and the access provided is only to each other's network and customers. Further, when two networks peer there can also be both "settlement free" (without requiring payment) and paid arrangements.

Network operators are motivated to peer for a variety of reasons that may include both business and technical motivations. Each network operator stipulates the technical and operational criteria used to evaluate what networks they will interconnect with, and many of these requirements are made publicly available online. Connecting networks does not come without costs, and a decision to interconnect requires careful consideration of the benefits compared to the costs incurred to connect at each location.

The two common options for interconnection are either through a private bilateral arrangement between two networks using a dedicated physical connection (called a "direct interconnection"), or a multilateral arrangement where all networks connect into a public Internet Exchange switch. An Internet Exchange is a service that uses a switch infrastructure (sometimes referred to as a switch fabric or peering exchange) to provide connectivity between multiple networks. Interconnection methods are constantly evolving,

and one of the more important developments in interconnection is the use of content delivery networks (CDNs). CDNs provide a more efficient means of distributing content by placing content and applications on servers distributed closer to, and sometimes within, the destination network – essentially bringing data (e.g., popular content) closer to the requestor instead of delivering the data across a potentially significant portion of the Internet. The introduction of CDNs and IXs has contributed to the "flattening" of the historic hierarchical model of Internet interconnection.

Internet traffic has grown rapidly since the Internet's inception, and this has often been driven by the growth of popular applications. Managing the exchange of Internet traffic between networks is accomplished primarily through the use of an inter-network routing protocol called the Boarder Gateway Protocol (BGP). BGP offers network administrators the ability to implement routing policy, or in other words how traffic flows through a network. BGP's design offers limited support for inbound (traffic destined into one's network) traffic control.

Peering connections are increasingly the primary interconnection paths between networks, supplanting the model of hierarchical interconnection via a small group of long-distance network providers. In most cases, two parties seeking to interconnect are able to come to terms. In some cases after an agreement is reached, however, traffic volumes or other factors may change, which in rare cases may lead to "de-peering" events if the changes are significant enough. More commonly, such changes lead to a renegotiation of the manner or type of interconnection agreement between the two parties. Although peering disputes over traffic imbalances, and other reasons, are not new, peering disputes in the U.S. have been increasingly publicized in recent years.

In some cases traffic can flow contrary to the intentions of network operators, either in error or due to malicious activity. There are a number of important security considerations when connecting two networks. There are numerous types of attacks, as well as various motivations that may drive attackers. There are also a number of potential mitigations, as well as efforts to make routing more secure through new routing protocol extensions, notably RPKI and BGPSec.

This report provides a reference on the subject of Internet interconnection, and presents a detailed review on how networks connect, the development and changes in connection models, motivations for connection, how networks manage traffic between each other and some of the challenges that arise as networks evolve.

Table of Contents

1 Issue Overview

The topic of Internet interconnection is receiving increased attention as the Internet ecosystem continues to evolve. Growing end user interest in higher bandwidth Internet applications such as streaming video and similar cloud-based applications is altering Internet traffic growth rates and the methods used to deliver that traffic to the end user. This report provides a reference on the subject of network interconnection for the Internet. It presents a detailed review on how networks connect, the development and changes in connection models, motivations for connecting, how networks manage traffic between each other and some of the challenges that arise as networks evolve.

The Internet is a global network of networks that enables any connected device, identified by a unique IP address, to communicate to any other connected device, anywhere in the world. The connectivity is achieved by passing pieces of data, called packets, from a connected device through networking equipment, known as routers, operated by one or more network providers until those packets are delivered to the desired destination. Routers implement software called a routing protocol that maintains up-to-date network information in order for this to happen. In routers, individual IP addresses representing devices are grouped together into prefixes.[1] A routing protocol enables routers the ability to share these prefixes among one another for the purposes of creating a network map that directs packets towards their destination.

The term "interconnection" refers to the various means by which network providers attach and move traffic between one another. There is no central authority that manages Internet interconnection – the overall system arises because of the many bilateral and multilateral decisions made by various actors that interconnect. In contrast to the telephone system, where interconnection is performed in a highly regulated environment, interconnection in the Internet (in most parts of the world) remains a private-sector matter. This is due in part to both the decentralized architecture of the Internet, as well as the historically unregulated nature of the Internet (as there is no one coordinating body or governance model).

Networks of all types interconnect among one another, including those of access providers, content providers, academic institutions, and commercial enterprises. Some of the interconnections that make up the Internet are among customers and providers, and some

[1] For example, for the (32-bit) IPv4 addresses from 255.255.255.1 to 255.255.255.255, there are 255 individual IP addresses corresponding to the prefix 255.255.255/24 (the /24 at the end means that the first 24 bits, or 3 octets, of the IP address are fixed and the AS operator can allocate the remaining 8 bits, or 1 octet, as they see fit).

are between firms that view each other as competitors. The mechanisms that have arisen to implement interconnection thus serve both technical and business purposes, and discussion concerning the technology of interconnection must, of necessity, refer to business issues to some extent – as many of the mechanisms can only be understood in that context.

Network interconnection in the United States has evolved significantly since the early days of the Internet (see the "History of Interconnection" box before Section 2 below). The initial form of the Internet was a single, wide area or backbone network (the Internet equivalent of a long-distance provider) that was operated by the U.S. government. Smaller, regional networks connected to this network forming a simple hierarchical structure (see Figure 1). Traffic from one part of the Internet to another was handed off to this backbone network, which carried it to the destination network. For many years, the technical requirements on the routing protocol providing this function were simple, and there was no need to deal with business issues.

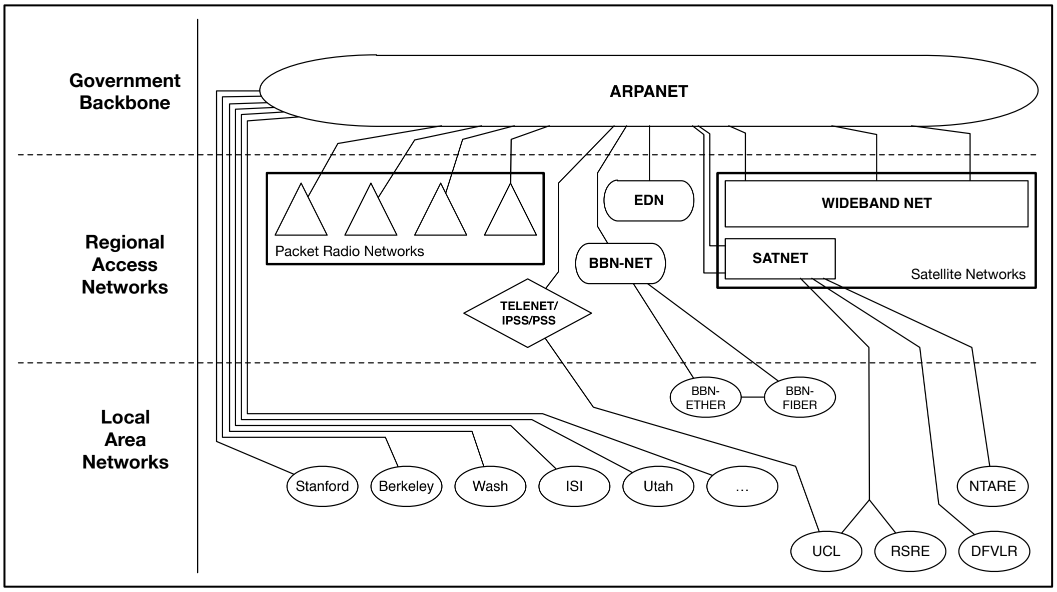

Figure 1: The very early Internet, circa 1983. (adapted from an "Internet Topographic Map, 1983", Computer History Museum, available at: http://www.computerhistory.org/internet_history/internet_history_80s.html.)

The backbone eventually transitioned from a single government-operated backbone to a federated backbone model comprised of multiple commercial network operators (see Figure 2), and as a result the routing protocol responsible for maintaining the map of destination networks required improvement. A new routing protocol, called Border Gateway Protocol (BGP), was created, allowing for commercial provision of backbone connectivity by multiple parties. BGP allowed network operators to manage this more complex and competitive space, and to express at least a limited set of business constraints on routing. For example, using BGP a network operator could specify how it preferred traffic flow into and out of its network. At the time of this report, BGP remains the standard routing protocol used to exchange information about IP routing between networks.

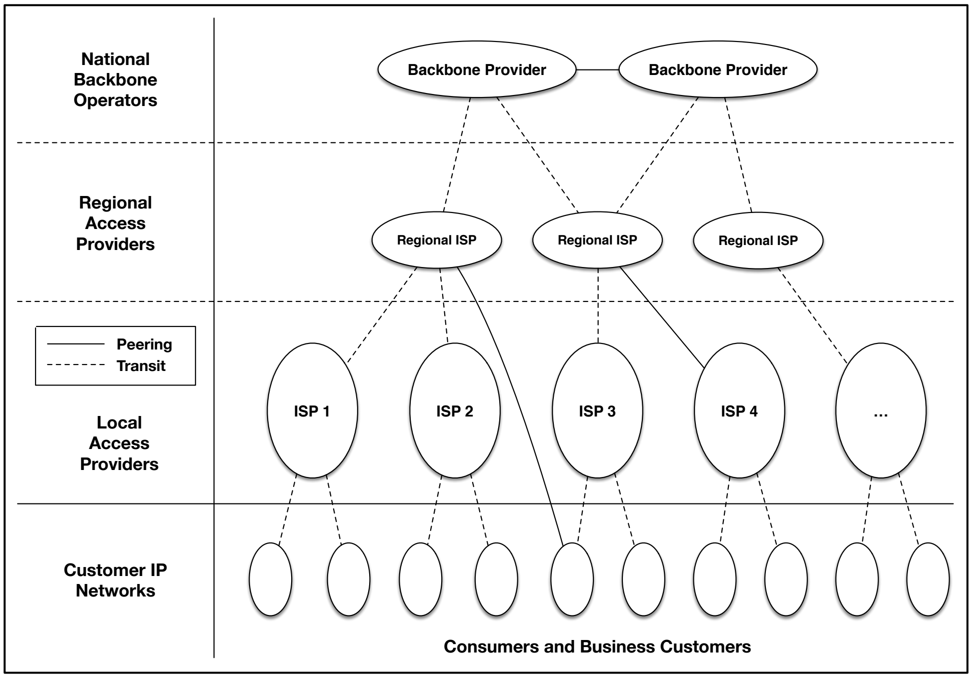

Figure 2: The hierarchical state of Internet interconnection in the 1990s.

The original commercial providers forming the backbone, sometimes called "Tier 1" networks, sold "transit" service as a product – they delivered their customers' traffic to the rest of the Internet for a fee. These networks also interconnected with each other, typically without any reimbursement charges, a practice known as Settlement-Free Interconnection (SFI). As the regional networks grew and became more diversified (e.g., serving enterprise customers or selling residential broadband access), the drive for cost reduction and improved performance, among other reasons, led to an increase in direct interconnections among networks of all sizes.

The transition to a more commercial backbone model also called for the creation of a small number of network access points (NAPs). These access points provided an infrastructure for locating network equipment and facilitating interconnection among networks. Later, the move to more direct connection among network providers drove the creation of many more locations where interconnection could be engineered. These became known as Internet Exchanges (IX) and will be discussed in more detail later in the report.

The next transition occurred in the late 1990's, as user interest grew in online content and high-performance applications. To meet the demand for higher traffic volume and better performance, content delivery networks (CDNs) were created to provide a more efficient means of distributing content. A CDN places content and applications on servers and distributes those servers closer to, and sometimes within, the destination network, essentially bringing data (e.g., popular content) to the regional networks instead of sending the data on-demand from a centralized data center. Like IXes before them, the CDN model is an evolutionary example of innovative technology designed to lower long-haul costs of traffic distribution and improve the consumer experience. As the emergence of the CDN model illustrates, changes in methods of interconnection are driven by changes in business practices, application requirements for data transmission, and by changes in user behavior and traffic patterns. Traffic flows on the Internet have evolved for other reasons as well, with fewer networks now accounting for a greater proportion of both the sources and destinations of traffic.

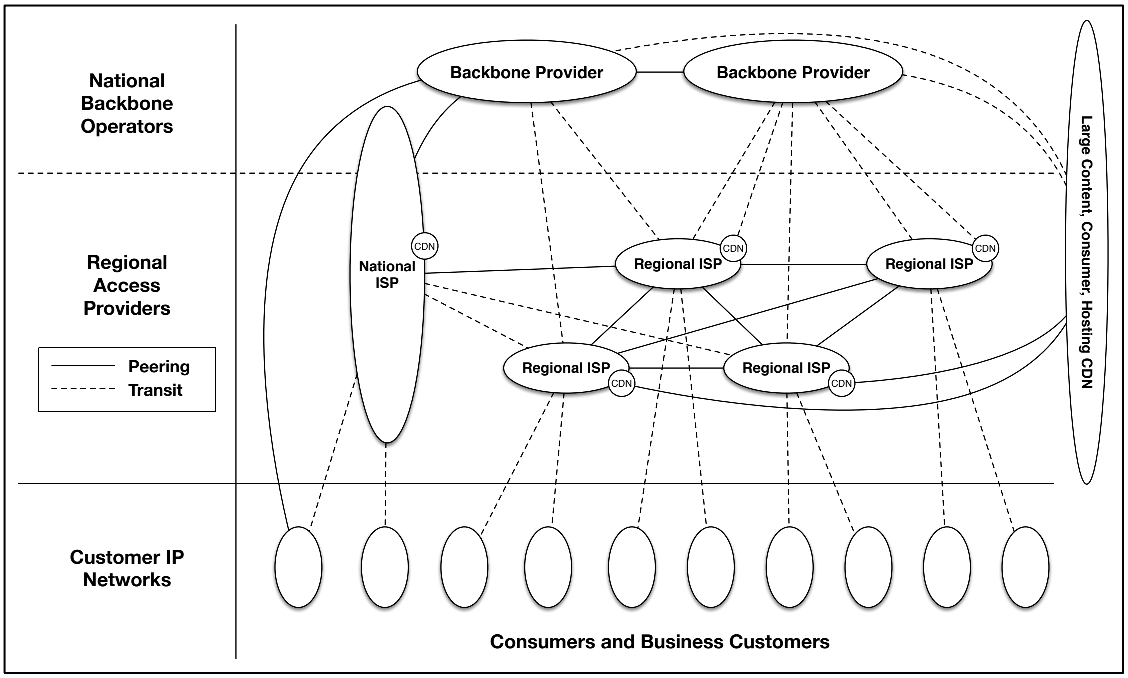

Figure 3: The flatter, highly interconnected Internet of today.

Interconnection today (see Figure 3) has evolved into a complex amalgam of models incorporating new connectivity options, delivery options, traffic management requirements and business practices. This report addresses several aspects important to the understanding of the current state of Internet interconnection. Section 2 of the report provides an overview of the different interconnection models as well as the means to implement the physical connection itself, or in other words what a pair of networks must do to engineer a path (an actual circuit) between them. Section 2 also discusses the issues that arise in the implementation of that physical path, including models, motivations, and costs of peering as well as the role of Internet Exchanges and CDNs. Another aspect of interconnection is the management of the routing protocols. Creating a physical connection does not automatically mean that all, or even any, traffic will flow over it. The term “routing” refers to the control mechanisms in the Internet that regulate which traffic flows over which paths. Routing uses the physical connectivity of the Internet together with administrator configuration to build a connectivity map, which the Internet traffic then follows. Section 3 discusses the technical mechanisms for implementing routing decisions (called "traffic engineering"), and how such mechanisms are used both to achieve efficient routing, as well as to implement routing decisions reflective of the underlying business considerations. Finally, this report discusses some critical issues that arise as a result of interconnecting two or more independently operated networks, such as security and measuring network performance.

The goal of this report is to provide a consensus-based technical overview of Internet interconnection, and to provide an educational reference document on the technical details of Internet network interconnection and related traffic management. The economic, business or policy aspects of interconnection will only be mentioned to the extent necessary to describe the technical elements. No value judgments on particular economic models or agreements, or peering policies are intended. This report is also not intended as an advisory report on measurement methodology or technical practices, although those topics may be addressed in subsequent BITAG reports.

1.1 A brief glossary of terms

Interconnection can be complex and a common vocabulary is important. This section provides a list of commonly used terms related to interconnection and their definitions, in order to provide standard terminology throughout the report. A number of these terms are used interchangeably by members of the Internet community, contributing to some degree of confusion when discussing the topic of Interconnection.

- *Autonomous System* (or AS) – An "Autonomous System" (AS) is a network, operated by a single organization, that presents a single coherent interior routing plan and a consistent picture of what networks are reachable through it. In BGP routing, it is identified by an AS Number (or ASN), which is allocated through a Regional Internet Registry (RIR). A network operator may utilize one or more ASNs in its network.
- *Internet Service Provider* (or ISP) – A network operator (consisting of one or more ASes) that offers Internet access to other customers. This is in contrast to the many ASes on the Internet that are edge networks serving other purposes, for example enterprises, educational institutions, or governments.
- *Transit* – A form of interconnection in which one network purchases network connectivity service, offering access to every publicly reachable destination on the Internet, from an ISP. In many cases, what is offered is "full" transit, meaning access to every publicly reachable destination on the Internet.
- *Transit Provider* – An ISP that provides transit to customers as a product.
- *Backbone Network* – The collection of high-speed transmission links and supporting network infrastructure that connects a network's geographically distributed points of presence and/or regional networks.
- *Peering* – A form of interconnection in which two networks agree to exchange their customer traffic among one another. In contrast to transit, where the offered service

is access to “all the Internet,” in peering the access is only to the other network and its customers.

  - *Settlement-Free Peering* (also *Settlement-Free Interconnection* (SFI)) – A form of peering in which two parties agree to exchange traffic for no direct monetary exchange.
  - *Paid peering* – A form of peering in which two parties agree to exchange traffic between their networks with some form of monetary compensation involved.

- *Colocation facility* – A physical location (e.g., a building) where multiple network operators install their networking equipment for the purpose of interconnection. Also called a Colocation Center.
- *Private Interconnect* – A “private” form of bilateral interconnection in which direct physical connections are used to interconnect two networks, without the need of an Internet Exchange. If used for peering, this is known as “private peering.”
- *Internet Exchange* (IX) – A “public” form of interconnection that allows for efficient connectivity among any ASes capable of physically reaching the colocation facility in which the Internet Exchange switching equipment is located. An Internet Exchange, in technical terms, is accomplished using a piece of computer networking equipment called *switch* (or collection of switches) – also called a *switch fabric*. Each AS connects from their network to this equipment, which in turn provides the ability to connect to every other AS connected to the switch fabric. Also called an Internet Exchange Point (IXP). Peering interconnections that are set up over Internet Exchanges are referred to as “public peering.”
- *Network Access Point* (NAP) – An early term (no longer current) for the facilities that interconnected the early Internet. These were a combination of colocation facility and Internet Exchange.
- *Content Delivery Network* (CDN, also sometimes called a Content Distribution Network) – A CDN places and distributes content as close as practical to the end users requesting that content. A CDN contains an organized set of servers in multiple locations that collectively provide caching or other services. Depending on the CDN business model, its servers may be distributed into colocation facilities, directly into ISP networks or housed in its own facilities.
- *Point of Presence* (PoP) – An access point to a communication provider’s network. It is a physical location that houses network equipment with interfaces that are administered as network connection points. A PoP may serve strictly as a location offering customers access to the network. Alternatively, it could be a room or cage within a colocation facility for the purpose of interconnection, or it could serve both purposes.
- *De-peering* – The act of two peered ASes discontinuing a peering relationship.

Text box: History of Interconnection

In the early 1970's ARPA (the Advanced Research Projects Agency of the U.S. Department of Defense) developed a number of network technologies in addition to the original ARPAnet, including a packet satellite network and a spread spectrum packet radio network. However, the data transmission protocols of the ARPAnet did not allow for the interconnection of these networks. The initial development of the Internet protocols (IP and TCP) in the 1970's had the goal of allowing end-to-end connectivity across interconnected heterogeneous network technologies.

After almost a decade of research and evolution, the specifications of TCP and IP were published in 1981, and in 1983 became the mandatory protocols for communication across the ARPAnet as well as the larger set of interconnected networks. At this time, researchers began experimenting with rudimentary routing protocols. An Exterior Gateway Protocol (EGP) was developed to control routing among these various networks, while each network had a internal routing scheme specific to the technology – for example the satellite network used broadcast. However, Internet routers (or gateways, as they were then called) began to be used to build individual networks, not just interconnect networks, so routing protocols for use inside a network were also developed as part of the Internet effort. Like today's routing protocols, most historic routing protocols attempted to calculate the best path between two points in a network, using various algorithms and metrics. The Exterior Gateway Protocol (EGP) was unusual in that it simply advertised reachability to a set of remote prefixes. EGP was limited, however, as it assumed there was a single backbone network, which created issues as the backbone became more complex.

During the first half of the 1980's, several experimental networks were deployed in the United States, including CSNET, USAN, and NSFNET. Building on the success of the 56 kbps NSFNET, several regional consortia came into existence. As the early network became more complex, the weaknesses of EGP were increasingly evident. In response, the Internet research community developed the Border Gateway Protocol [1]. BGP worked to allow, among other things, routing of data traffic according to business policy through a potentially complex network of arbitrarily connected ASes. Initially, BGP selected routes based on the shortest path measured as a sequence of ASes, but over time other attributes were added to further optimize routing, for example labeling prefixes (called "communities").

The backbone of the Internet was provided by NSF during the 1980's and the early 1990's, but the NSF backbone was decommissioned and replaced by commercial wide area networks in the mid 1990's. As a part of this transition, NSF sought bids for Network Access Point (NAP) Managers [2], who would establish "a highspeed network or switch to which a number of networks can be connected via routers for the purpose of traffic exchange and interoperation." The solicitation specified three priority locations in California, Chicago and New York City, as well as a number of other desirable locations. The four institutions that received the NSF's NAP Manager awards were: Pacific Bell in San Francisco, CA; Ameritech (Telecordia) in Chicago, IL; MFS Datanet in Washington, D.C., and Sprint in New York City, NY.

In 1998, Akamai introduced an important means of optimization called a Content Distribution Network (CDN) (see Section 2.7). A CDN is a network of servers that deliver cached content to users who request that content. The goal of a CDN is to place the cached content as close as possible to the user. CDNs and other types of caches can substantially change the pattern of traffic exchanged, taking some loads off long-haul and transit links and increasing traffic delivered on local networks.

Today, interconnection uses both dedicated ("private") interconnection and shared ("public") Internet Exchange at multiple connection points between any two carriers, and is generally at rates of tens to hundreds of gigabits per second (Gbps). A number of companies operate CDNs to facilitate the delivery of various kinds of content. Internet exchanges are common worldwide, regardless of a region's level of economic development.

2 Internet Interconnection Overview

Internet interconnection is a collection of business practices and technical mechanisms that allow individually managed networks (ASes) the ability to connect together for the purpose of exchanging Internet traffic. The relationship between two networks when they interconnect takes two dominant forms (Section 2.1, below): peering and transit. Peering may form as the result of agreement among networks after considering the motivations for the networks involved (Section 2.2). Network operators especially the larger ones, may expect a set of technical conditions or criteria for peering (Section 2.3) to be met before agreeing to connect to another network. Interconnection occurs at a colocation facility or Internet exchange (Section 2.5) through connections made between network equipment located in networks' Point of Presence (Section 2.4). A decision to interconnect requires careful consideration of the benefits discussed in the motivations section, compared to the costs incurred to connect at each interconnection location (Section 2.6). Interconnection methods are constantly evolving, and one of the most important developments in interconnection is the use of CDNs and caches (Section 2.7). The introduction of CDNs into the Internet landscape has been one of the factors in the "flattening" of the historic hierarchical model of Internet interconnection. Further evolution of interconnection has come about as a result of shifts in Internet traffic patterns (Section 2.8), including the rise in streaming video services. The business relationships developed between operators through interconnection have typically been amicable, but, on occasion, they have resulted in conflict between various parties (Section 2.9).

2.1 Peering and Transit

When networks interconnect, they make joint decisions about how to exchange traffic based on their respective business relationships, such as transit provider, peer or customer. If an AS provides "transit" service for a customer AS, it can carry traffic between that customer's network and all other Internet endpoints. To implement this relationship, a transit provider will advertise the routes for a customer's network (i.e., IP prefixes) to the rest of the Internet using an Inter-AS routing protocol call the Border Gateway Protocol (BGP), described in Sec. 3.1. The transit provider will also advertise all of the Internet routes (typically as a single default route representing all Internet destinations) to the customer. Transit relationships may be "full" (the customer receives routes for all Internet destinations from its transit provider), or "partial" (the customer receives routes for some subset of all Internet endpoints). Transit is always a service offered for a fee.

If an AS "peers" with another AS, the two ASes agree to exchange traffic only between their own endpoints and the endpoints in their customers' networks. This agreement can be formal or informal. Peering agreements have historically been informal "hand shake" agreements[2] but appear to increasingly involve contractual relationships [3]. Where a peering agreement is formalized, it will usually include confidentiality and non-disclosure terms – most operators do not make their interconnection arrangements public.

[2] For example, Woodcock and Adhikari find that only 0.5% of a large sample of peering agreements had been formalized in written agreements [4]. While this sample contained over 100,000 peering arrangements, it should be noted, however, that this was a convenience sample, and as such may not be generalizable to the larger population of peering agreements.

These arrangements are implemented with routing policies that control which routes are advertised to a neighboring AS. In a transit relationship, a transit provider will typically advertise all of its routes to a customer, whereas in a peering relationship a peer will only advertise its customer routes to another peer.

Peering relationships may be settlement-free or paid, involving either monetary or other types of value exchange. These are essentially barter transactions where both sides negotiate until they perceive equal value in the relationship. The routes that are exchanged in a paid peering relationship are the same as in a settlement-free peering relationship.

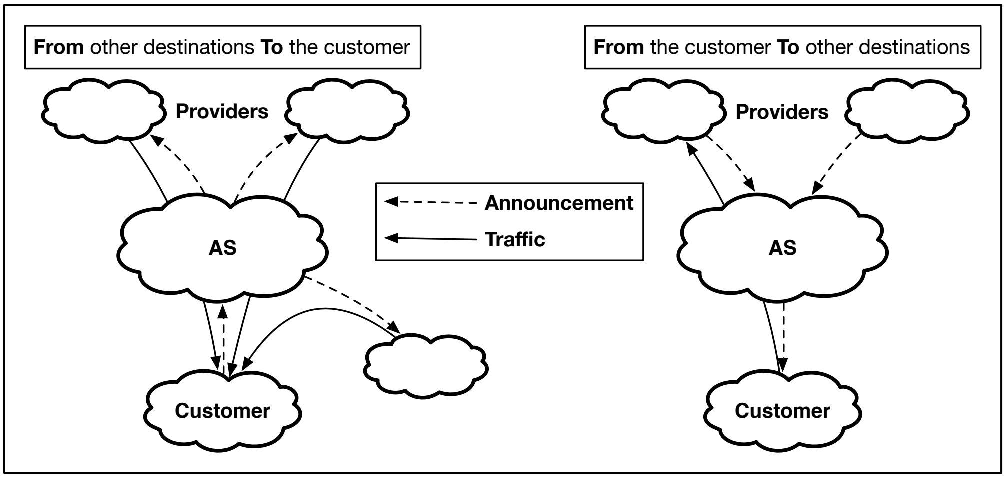
From other destinations To the customer
Providers
AS
Customer
Announcement
Traffic
From the customer To other destinations
Providers
AS
Customer

Figure 4: Business relationships also influence routing announcements, dotted lines depict route announcements and solid lines traffic flows.

Figure 4 shows how these business relationships further translate to decisions about which routes to advertise to neighboring ASes. When an AS learns a route from one of its customers, it typically advertises that route to all of its connected networks, since the customer pays the AS for transit connectivity to other Internet destinations. On the other hand, when the AS learns a route from a provider, the AS only advertises that route to its customers, but not to its providers or peers; doing so would result in the AS having to provide transit between its providers or peers.

Figure 5 (below) demonstrates how business relationships translate to preferences between different routes to a destination. AS A connects to a prefix in Destination AS's network via three intermediate ASes, Peer, Customer and Provider. In this figure, Destination AS is in a transit relationship with all three of these networks. AS A has to pay to use its Provider link and gets paid for traffic on the Customer link, regardless of the direction that traffic flows on these links. The peering link may not require payment to use ("settlement-free") (or it may involve a payment ("settlement"), which is not depicted in Figure 5). In terms of preferring routes to different destinations, AS A will generally prefer the route to the destination via its customer (which it gets paid to use) over the route to the destination through either its peer (which it does not pay to use) or its provider (which it pays to use).

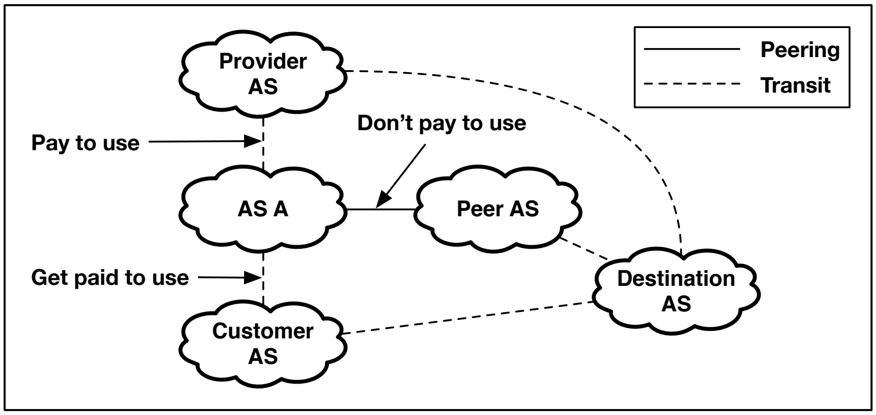

Figure 5: A simple example of how business relationships can influence traffic routing.

2.1.1 Basic Technical Requirements to Peer

In order to participate in a peering relationship on the Internet with any party, some basic technical requirements need to be fulfilled. A network looking to peer must have:

- A public AS number assigned by a Regional Internet Registry (RIR) [5]. Without this, the network will not have a unique "identity" on the Internet for the purposes of routing traffic.
- At least one block of public IP addresses (independent of any upstream provider) assigned by an RIR. These addresses are what the network "announces" or "advertises" to other networks it interconnects with.
- A network edge router capable of running the BGP protocol, and the technical capability to configure and manage BGP interconnections.

These basic requirements to peer may also apply to a network purchasing transit service, for example from multiple ISPs. This advanced network configuration is often used by larger enterprises for redundancy and traffic management purposes.

2.2 Motivations for Peering

Network operators may be motivated to peer for a variety reasons, both business and technical. For some complex transit relationships, similar motivations may apply. The motivations for peering networks can include:

- ***Network effects*** – Put simply, a network (or a network of networks) is not particularly useful if no one is connected to it or those who are connected to it cannot reach each other. A related concept is that a network becomes more valuable to the connected participants as more and more connect to the network.

- ***Increased redundancy* –** Network operators strive for as redundant and as resilient a network connectivity map (called a "topology") as economically possible. If one network interconnects to another in a single location this poses some risk, as a single point of interconnection has a higher likelihood of failure compared to two or more points. Increasing the number of interconnection points between networks results in a more resilient and reliable network for both parties.
- ***Increased routing control* –** The ability of a network operator to influence path selection between endpoints in the network can lead to lower latency, decreased packet loss and increased quality of experience for network users. Each new peering connection increases network operator's control over the traffic routed between them, given that their traffic no longer has to traverse intermediary networks.
- ***Reduced latency* –** Users' quality of experience generally improves with decreased latency, especially for applications that communicate time-sensitive data, such as interactive voice or video. Relying upon intermediaries for traffic exchange between two networks means that latency characteristics of the connection cannot be as tightly controlled, which can lead to degraded quality of experience for network users. Introducing direct interconnection between the networks provides for well-known and clearly measurable paths between each network's users. With these clear pathways under the control of the network operators who have an interest in optimizing the quality of experience, latency can be reduced.
- ***Reduced congestion* –** Congestion along the path between the source and destination through intermediaries can degrade the quality of experience for network users. Direct interconnections between the networks allows for both networks to ensure sufficient bandwidth is available between network users for consistent performance.
- ***Improved traffic management and predictability of traffic* –** When networks interconnect, they have direct control in managing the flow of traffic between them. Because of this, both networks may reduce the need to maintain spare capacity on their transit and other connections to the Internet, as the traffic that is directly peered is now less likely to appear on other routes.
- ***Reduced costs* –** Peering can reduce the costs of routing traffic between certain endpoints. For example, if an AS exchanges a significant amount of customer traffic with another AS, rather than sending that traffic up through its transit provider and into the other network through their transit connection, the two networks could peer at a common facility and each will eliminate the transit cost for that traffic exchange. In the case of paid peering, the new cost may be less expensive for the AS that is paying than its previous transit cost.

2.3 Peering Policies

Each network operator, as a result of the decentralized design of the Internet, establishes a set of criteria to determine the networks with which they will interconnect or peer. These peering policies can reflect how the technical and often contractual arrangements that network operators negotiate with each other to interconnect are influenced by the business objectives and policies of each of the parties. Peering policies are frequently used to achieve consistency in judging requests for interconnection, and typically represent an attempt to set reasonable and transparent boundaries – thereby working to avoid future disagreements if traffic volumes or other factors shift. PeeringDB, a resource where many network operators post their peering policies and locations to make them easily accessible to other network operators [6, 7], categorizes peering policies as open, selective, or restrictive [8]:

- **Open –** An open peering policy is where the party will peer with any other party.
- **Selective –** A selective peering policy is where the party will peer with any other party that meets the criteria articulated in the peering policy, such as minimum traffic exchanged, number of peering points, staffed network operations center (NOC), etc. (see below).
- **Restrictive –** A restrictive peering policy is where the party does not generally peer with other parties, or in other words where peering is the exception and not the norm.

Network operators will often include in their peering policy a set of technical requirements and operational requirements. The specific requirements to interconnect two networks are determined by each respective network. Assuming a network meets the basic technical requirements to participate in peering on the Internet (see above), peering policy requirements may also include:

- **Network capacity:** ASes may impose minimal requirements on the size of a potential peer's network capacity. ASes sometimes also require that the potential peer AS operate a fully redundant backbone network. In addition to imposing capacity requirements at peering links, an AS might also impose requirements on links between major hubs.
- **Geographic scope:** ASes may impose requirements that state the potential peer must have a backbone presence in an expansive and diverse set of geographies. This requirement is often expressed in terms of metropolitan statistical areas (MSAs), time zones, or census bureau divisions. In addition to having a backbone network in these locations, the peering requirements often stipulate that the AS peer in some fraction of these locations where the potential peer AS operates its backbone network, and that each interconnection point have a minimum throughput of a certain traffic rate.
- **Routing:** A potential peer is generally required to operate an IP network between the interconnection points and use the Border Gateway Protocol (BGP) to exchange

routes at the interconnection points where the peering occurs. Typically, the potential peer AS must use the same AS number at all peering points and announce a consistent set of routes at all peering points, thus ensuring that each AS can implement "hot potato" routing – a form of "shortest-path" routing that allows each network to minimize carriage costs (see Section 3.1.4 below) – at any of the peering points where the agreement is in place.

- **Network operations centers (NOC):** The potential peer AS must typically maintain a network operations center, where staff is on duty at all times to resolve any problems that may arise within a reasonable timeframe. Cooperation to resolve security incidents is required in many peering agreements.
- **Network traffic ratios and volumes:** Peering policies may stipulate that the potential peer not exceed an aggregate traffic ratio in a certain direction (e.g., "traffic ratios must not exceed 2:1", meaning that aggregate outbound traffic on the peering links must be no more than twice the volume of aggregate inbound traffic). Peering agreements sometimes also require that the potential peer send a *minimum* volume of traffic over the peering links. Some peers may have preference as to how they exchange traffic, depending on traffic volume – typically small volumes use Internet Exchanges, large volumes use private interconnections.
- **Filtering:** Peering policies may also require the potential peer AS to filter route announcements from its customers by prefix, to ensure that incorrect route announcements do not "leak" across the peering link and that no transit or third-party routes are announced.

As mentioned, many large networks make their peering policies available publicly online. Additionally, groups of ISPs typically meet together in person at operators' group forums (e.g., the North American Network Operators Group or the Global Peering Forum) to identify mutually beneficial peering relationships [9,10]. These events often host an informal meeting where ASes can advertise the capabilities of their networks to other ASes who might be interested in peering.

In most cases, two parties seeking to interconnect are able to come to terms, either on a paid transit, paid peering, or settlement-free interconnection (SFI) basis. In some cases after a peering agreement is reached, however, traffic volumes or other factors may change, which in rare cases and after some time may lead to "de-peering" events if the changes are significant enough (see Section 2.9). More commonly, such changes lead to a renegotiation of the manner or type of interconnection agreement between the two parties.

2.4 Interconnection Facilities

For networks to interconnect, they have to physically connect their networking equipment with each other. This requires the networks to meet in a common location, in facilities capable of supporting the equipment required for interconnection. These colocation facilities lease their customers secure space to locate and operate equipment, often in the form of lockable rack cabinets or cages [11]. In addition to floor space, a colocation facility provides power, cooling, structured cabling to support network connection, real-time physical security monitoring, and fire protection. Colocation facilities are provided by commercial enterprises to house multiple network providers in the same or adjacent buildings.

A Point of Presence (PoP) is the location where one network can connect with other networks. In order for two (or more) networks to interconnect, their networks need PoPs in close proximity to one another for local interconnection. Another option is the acquisition of a circuit providing the extra network reach needed to connect with other PoPs. Network operators often maintain PoPs located in multiple colocation facilities across the country and/or globe.[3]

2.5 Private Interconnect and Internet Exchanges

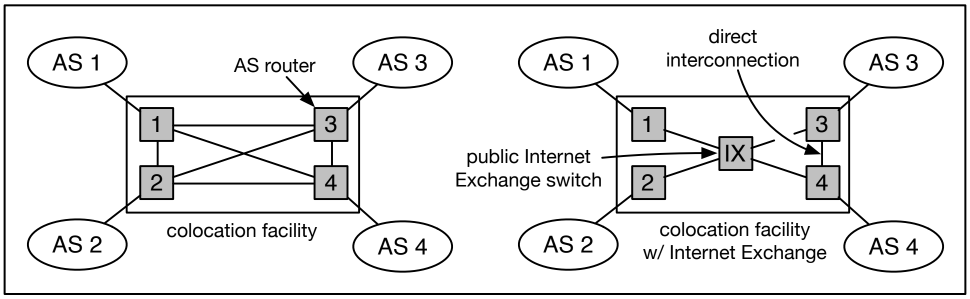

Figure 6: At left, a simple colocation facility where each AS directly interconnects. At right, a colocation facility that also offers Internet Exchange through a public switch (or "switching fabric").

The two common options for interconnection are either through a private bilateral arrangement between two networks using a dedicated physical connection (called a "direct interconnection"), or a multilateral arrangement where all networks connect into a public Internet Exchange switch ("public interconnection"). Interconnecting two networks requires establishing both (1) physical connectivity and (2) network connectivity. Figure 6 illustrates the differences between the types of connections at a colocation facility where

[3] In the United States the most widely used colocation facilities are located in New York City, NY; Ashburn, VA; Atlanta, GA; Miami, FL; Chicago, IL; Dallas/FortWorth, TX; Los Angeles, CA; San Jose, CA; and Seattle, WA. These colocation facilities developed around the original four NAPs of the National Research and Education Network (NREN) that attracted the most interest based on their proximity to major networks or technology hubs [12]. In addition to these colocation facilities there are a number of regional colocation facilities in the United States [13].

ASes can privately interconnect and a colocation facility that provides a public Internet Exchange switch.

A private interconnect is accomplished using one or more dedicated physical (see the left images in Figure 6 and 7) cables to interconnect two networks in close proximity, and it is used exclusively for the exchange of traffic between those two networks. The physical media that connects the two networks at that site is often referred to as a “cross-connect” and can be copper or fiber. The data speed of a cross-connect can range from Mbps up to many Gbps [14]. It is common in larger networks for the traffic volume between networks at a single location to exceed the capabilities of any single router interface. This limit is typically overcome through the combination of multiple physical connections into a single logical one. The downside of this arrangement is the additional operational overhead of managing multiple (sometimes dozens) physical connections, as well as added complexity in troubleshooting faults in the connections.

Alternatively, multiple networks can interconnect through a public Internet Exchange. An Internet Exchange is a service that uses a switch infrastructure (sometimes referred to as a switch fabric or peering exchange) to provide connectivity between multiple networks. The multi-party switching fabric often comprises one or more Layer 2 Ethernet switches (see the right images of Figure 6 and 7) and provides an alternative to the dedicated cross-connect required for private interconnection. For some networks the ability to reach multiple networks using only a single connection into the IX is advantageous, resulting in lower operating expenses. Conversely, for very large networks with large data flows with a few networks, it may be more advantageous to use private interconnects.

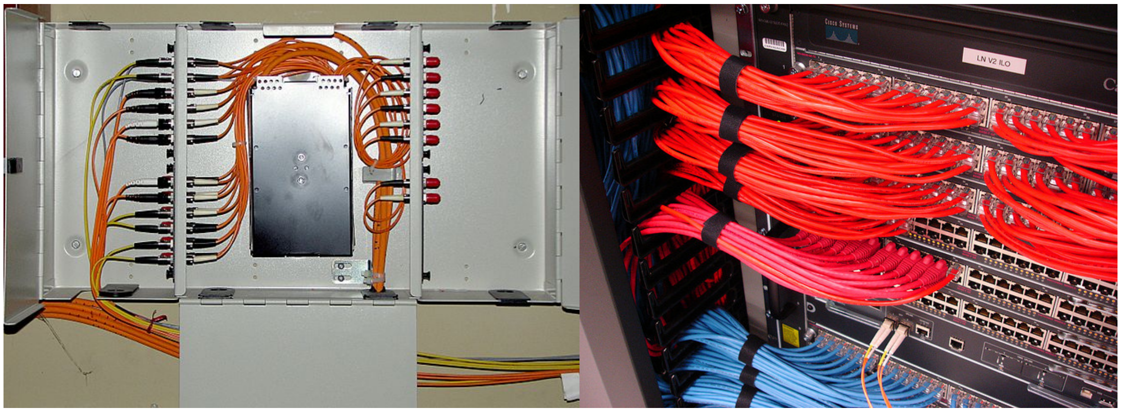

Figure 7: At left, an optical distribution frame with network cross-connect cables. At right, a switch. (Left: (cc) 2007, Eigens Werk: http://de.wikipedia.org/wiki/Datei:Optical-fiber-distribution-frame-0a.jpg; Right, (cc) 2008, Adrian Black: https://flic.kr/p/67N1ZV).

Internet Exchange

Physical connectivity to an Internet Exchange does not automatically entitle access to every other network on the exchange, or even mean that any traffic will flow over that connection at all. After establishing physical connectivity, the network operator must establish network connectivity with other network(s) present on the exchange. Some Internet Exchanges have mandatory multilateral peering agreements as part of the service and therefore require no additional approval for connectivity with the other networks at that exchange, while other Internet Exchanges will require each network to establish a separate peering agreement with each other network.

Internet Exchanges (IXes) can be operated as for-profit businesses, as cooperative non-profits, or even as voluntary arrangements between operators with no formal organizational structure – though this normally occurs only in the "startup" phase of an exchange. The for-profit IX model is dominant in the United States, while the non-profit cooperative model is dominant in other parts of the world. For-profit IXes are commercial operations, often run by the colocation company that houses them and offered as an additional value-added service to incent companies to colocate their equipment there. Network operators pay a fee to connect to the IX, with pricing determined by the colocation provider operating the exchange.

Conversely, exchanges outside the United States are typically operated as non-profit cooperatives, owned and operated for the benefit of the member ISPs, and funded by the membership and connection fees paid by the connected ISPs [15]. Most European IXes are separate organizations from the colocation facilities that host them, and they typically only provide public peering services.[4] Fees paid to the organization cover the peering exchange infrastructure costs and the costs of running the IX itself. Many European IXes have a broader "community" remit to undertake activities for the benefit of their members, for example engaging with policymakers and regulators on Internet policy and governance issues.

The recently-formed non-profit "Open-IX" organization is an attempt to facilitate peering and interconnection, in particular in the United States, and was formed by a cross-industry group of content providers, telecommunications providers, and datacenter network operators [17]. It develops standards, specifications and certifications for datacenters, "meet me rooms" and IXes. There is some movement in the United States in this direction and an increasing number of US Internet Exchanges are setting up in the co-operative model. Certain European IX organizations have also started to expand into North America.

[4] One notable exception being LINX, who provides a "private interconnect" service to its members as an alternative to purchasing private interconnect cross-connects from certain colocation vendors [16].

Some IXes offer the use of a "route server" as a value added service to ease the process of peering with other participants over an Internet Exchange [18,19,20]. These platforms are typically run by the Internet Exchange administrators, and are connected to the dedicated peering fabric. Route servers redistribute BGP routes between all members who connect to it. Route servers lower the maintenance overhead for multi-lateral peering, which can significantly increase the rate of multilateral peering for medium to small ASes [20].

2.6 Costs of Peering

Connecting two networks in a peering relationship does not come without costs. As mentioned, both networks must have PoPs at a common colocation facility or purchase connectivity to the colocation facility in order to be able to interconnect. This may mean that one or both networks must create a PoP that may require laying or leasing fiber-optic cable, purchasing network infrastructure and installing that equipment in a colocation facility.

Peering costs for each location often include things such as (this is not an exhaustive or exclusive list):

- Networking equipment for interconnecting the networks, which in turn can include additional router/switch ports, additional router/switch capacity (chassis, CPUs, etc.), and equipment racks;
- The leasing costs for space and power at the colocation site for the network equipment;
- Interconnection fees charged by the colocation site or IX, which can be structured as private peering – a one-off and/or monthly cross-connect fee charged by the colocation site, or in the public peering model – a one-off and/or monthly fee at the IX for the additional ports on the IX's switch;
- Network connectivity (transit, leased circuits, and/or fiber) capacity from the PoP to the rest of the network operator's network for the additional peered traffic;
- Operational fees, typically paid to the colocation provider in locations where a network operator may not employ local support staff but requires support from on-site technicians; and
- Engineering labor to design and deploy the network for the new interconnect.

2.7 Evolution in Interconnection - Caches and CDNs

When particular content (text, document, picture, video, etc.) is in high demand on the Internet, it can cause a bottleneck, flooding the originator of the content with a large number of requests to view that content, and frustrating end-users who may find that content unreachable or slow to download. Caches and content delivery networks (CDNs) were developed in order to better locate and distribute popular content closer to the end users requesting that content.

CDNs are designed to mitigate the demands that popular content can place on servers and networks. For example, video of a major breaking news item or sporting event, or release of a popular piece of software, can generate demand that could overwhelm limited, dedicated server and network capacity if the content was repeatedly sourced from a single server. Instead, CDNs locate copies of popular content close to consumer endpoints, reducing delivery costs, increasing the ability of a network to scale traffic, and improving performance for end-users.

Similarly, a cache is a temporary data store that holds frequently accessed data, and addresses the same issues but locally. For example, many web browsers include a cache to store frequently accessed objects such as images, without having to repeatedly fetch those images from the network. Caches can also be based in the network and can provide the same functionality and benefits for many users – for example a cache in an educational institution network (a "LAN cache") may serve all users on that network. This will save bandwidth on the school's Internet link, as the same object can be retrieved by different users while only needing to be retrieved over the network once; it may also improve performance if the Internet link is congested, slow, or lossy.

The CDN is an evolution of the simple cache that can provide optimizations across multiple networks or the whole Internet (for the remainder of this section, we use the term CDN to refer to both a CDN and a simple cache). As explained in Section 1, a CDN consists of an organized set of servers that collectively provide caching or other services in multiple locations (see Figure 3). These CDN servers may run on a contiguous underlying network, or may be a "virtual" network relying on one or more underlying ASes to link the different servers together. As a result, a content provider can simply maintain a small number of "origin" servers that update (or "seed") a large CDN. The CDN can dynamically adjust how many servers the content is stored on and how much bandwidth is available.

2.7.1 CDN Models

Typical services provided by a CDN may include caching, as described above, but can also include application acceleration, content optimization, and Denial of Service (DoS) protection [21,22]. Content optimization, for example, may involve services like down-sampling of images or content compression to reduce bandwidth usage, which is particularly common on cellular networks. DoS protection is a natural function of CDNs; a denial-of-service attack involves flooding the content originator with more traffic than it can handle, however, a CDN can serve that content relatively seamlessly from a distributed CDN-based resource.

There are a variety of places in the network where CDNs can be placed and there are a number of CDN models. Network operators, commercial CDN operators, or content providers can run CDNs. Network operators may run their own CDN on their own network, for their own content (e.g. IPTV CDNs for cable companies) and for any third party who wishes to purchase the service for reaching users on that network. Commercial CDN operators, such as Akamai or Limelight, sell their CDN capacity to content providers such as the BBC and CNN. Content providers themselves can establish CDNs; for example, Google's "Global Cache" and Netflix's "Open Connect," each use their CDNs only for their own content [23,24].

CDNs may use different deployment models and locate servers in various places in the network, depending on their technical and operational requirements. Some may use multiple models. CDN platforms may be deployed:

1. Within an operator's network or networks, distributed as close to the user as possible. This model relies on the operator's facilities for operation including rack space, power, cooling, and connectivity.
2. Within the CDN operator's own facilities and supported by its own backbone network, and interconnected with peers using either public or private peering, generally at the same locations where other peering takes place.
3. Within colocation facilities and connected to peers using either public or private peering. This differs from (2) in that the various CDN deployments are not themselves interconnected, they are "islands" of servers, relying on other providers (typically wholesale transit providers) to provide back-end connectivity to fill the CDNs with content and for management.

2.8 Evolution of Internet Traffic Patterns and Impact on Connection Models

Internet traffic has grown rapidly since the Internet's inception, and this has often been driven by the growth of popular applications. In the early 1990's, the World Wide Web was very popular. Later in the decade early voice applications drove traffic growth. In the first decade of this century peer-to-peer applications, cloud storage, and social networks were important traffic drivers.

Recent evolution in Internet traffic patterns can be attributed to numerous factors including the advent of new consumer devices (smartphones, tablets) with intelligent features and functionalities. The introduction of mobile-to-mobile (M2M) applications in vertical markets such as energy (smart meters), security (video surveillance), fitness (wearables), healthcare, and transportation industries, is driving growth in connections. The increasing popularity of streaming music services (Spotify) and streaming video games as a spectator sport (Twitch.TV) are also traffic drivers [25,26].

Video has emerged as a key driver shaping current consumer bandwidth consumption. Nearly half of the peak downstream traffic on fixed access networks in North America in the first half of 2014 was attributed to streaming video from Netflix and YouTube [27]. For fixed access, the mean monthly downstream usage grew from 18.6 GB in the beginning of 2011 to 43.8 GB in the beginning of 2014 [27,28]. Peak hour Internet traffic has grown more quickly than average Internet traffic and, in 2013, peak hour traffic increased 32% as compared to 25% for average traffic growth [27,28,29].

Video resolution impacts bandwidth requirements. As 4K Ultra-High-Definition (UHD) video streaming technology is introduced, higher bit rate streams will drive increased bandwidth requirements, although new compression algorithms may help ease this growth factor as they are introduced into the market.

Residential broadband traffic is characterized by asymmetric patterns due to the nature of consumer data consumption. Since the creation of the World Wide Web residential users have been more interested in receiving data, whether it be Internet radio, Internet video, or playing online games, rather than sending data back towards the Internet. The current trend in traffic volume shows a greater percentage increase in the inbound direction towards the broadband user vs. outbound away from the broadband user [27]. As seen in the peak upstream & downstream trends reported in North America in the first half of 2014, the mean monthly upstream traffic (7.6 GB) is significantly lower than the downstream traffic (43.8 GB) [27]. It is expected that residential Internet traffic will remain asymmetric for the foreseeable future.

These changes in Internet traffic patterns and the underlying technological advances and market forces have coincided with a dramatic change in the Internet connectivity model. Research showed that by 2009 half of all the Internet traffic content originated in less than 150 large content and content distribution companies [30]. And that by May of 2014 this number had dropped by a factor of 5, with just 30 companies including Netflix and Google contributing on average more than one half of all the Internet traffic in the United States during prime time hours [31,32].

One example cited in this research was that in 2007 Google used transit providers for more than 70% of their traffic to consumers and that by 2010 more than 80% of all Google traffic was directly peered between Google infrastructure and access networks. This reflects the internalization of network costs by some content providers – rather than simply buying transit, content providers have invested in network infrastructure to enable them to exchange traffic via peering rather than transit. This trend was not limited to Google, but observed across the entire Internet ecosystem [33].

The same research has also shown an increase in the direct interconnections between access networks (e.g., Earthlink, Comcast) and content or edge providers (e.g., Hulu, Google, Facebook)[31]. Among network providers, peering connections are increasingly the primary interconnection paths between networks, supplanting the model of hierarchical interconnection via a small group of tier-1 providers [34,35] (see Figures 1-3 and associated discussion earlier in this report).

The growth of IXes and CDNs, two initially independent infrastructure sectors, has contributed to this shift over the last decade and a half. As discussed previously, IXes facilitate interconnection among networks within a region, allowing traffic to flow along less expensive and lower-latency routes. This has resulted in a symbiotic relationship between Internet Exchanges, Content Delivery Networks, and ISPs - caching content closer to users, optimizing bandwidth efficiency and performance. There are over 400 IXes in the world, with diverse architectures, membership, and business/governance models [6]. The diversity in connectivity that IXes enable has not yet been rigorously quantified, but recent studies show evidence of a significantly large number of peering connections at IXes [34,36].

All of these factors have resulted in "flattening" the Internet's hierarchical structure as described in Section 1 and shown in Figure 3.

2.9 Peering Disputes and Outcomes

Although peering disputes over traffic imbalances, or other reasons, are not new [37,38,39], a number of peering disputes in the U.S. have been increasingly publicized in recent years [40,41,42,43,44,45,46,47,48].

As discussed previously, a peering relationship represents an agreement between two networks to exchange traffic in some agreed upon manner. If one or both of the networks determine that the peering relationship no longer meets the agreed upon terms or is no longer mutually beneficial, several things can happen.

Typically the parties will engage and attempt to negotiate new terms agreeable to both parties – this may result in changes in the technical characteristics of the interconnection point (e.g., capacity or location, number of routes exchanged), or the commercial terms of the interconnection. Most disputes of this nature are resolved through this process.

In some cases, the parties do not agree on new terms. When negotiation fails or has not completed, a party may take one of three possible actions, from a technical perspective:

Option 1: Suspend Network Capacity Augmentation – In this scenario, one network may decide to no longer augment capacity between the peering connections. The result is full network connectivity is maintained throughout the period in dispute, but congestion may result if traffic volumes exchanged over the interconnects grow to the current capacity limits. Through traffic engineering, it may be possible to manage traffic levels on the interconnects, and other connections, such that congestion does not occur.

Option 2: Augment Network Capacity – In this scenario, both networks may provisionally agree to augment capacity between the peering connections. The result is full network connectivity is maintained throughout the period in dispute, and congestion may be avoided even as traffic volumes continue to grow.

Option 3: Disconnection (De-Peering) – In this scenario the peering connection(s) between the networks will be disconnected (either logically or physically) by one or both networks. Disconnection can cause two different results for the networks involved and their respective customers:

Result 1: Partial Loss of Internet Reachability – This occurs after a peering disconnection when customers of each network no longer have a path to communicate with customers of the other network. This will only occur if the de-peering is between "Tier 1" networks who both have no transit connectivity – in such a case, any customers each network has who are "single-homed" only to either transit provider will not be able to reach single-homed customers on the other network.

Result 2: No Loss of Internet Reachability – This occurs after a peering disconnection when at least one of the networks has a transit connection offering a backup path to the other network and its customers. In this scenario, customers of each network are still able to communicate, but there may be a performance impact in the form of greater latency, and any other benefits of peering (Section 2.2) will also be lost.

3 Internet Traffic Engineering

Any given network may connect to other Internet destinations via multiple neighboring networks, to balance traffic load on its links and to achieve redundancy against failures in other networks. Connecting to multiple networks, or "multihoming," allows the network to control the links on which given traffic enters and leaves the network; this practice is typically called "traffic engineering." Traffic engineering can be either inbound (i.e., affecting how traffic from other Internet endpoints reach the network in question) or outbound (i.e., affecting how traffic leaves the network en route to other Internet destinations).

Managing the exchange of Internet traffic between networks is accomplished primarily through the use of BGP, an inter-AS routing protocol. BGP's design offers limited support for the control of inbound traffic (destined into one's network). This limited control allows the sender of the traffic the ability to exert greater control over the path the traffic takes to get to the receiver. In the past, when networks of equal size peered with one another this was rarely an issue as both networks exchanged traffic in roughly equivalent proportions. The advent of newer specialized networks that produce very high volumes of asymmetric traffic (such as CDNs) can make traffic engineering for the receiver (e.g. a broadband access provider) more challenging. Effective traffic engineering of high volume asymmetric traffic benefits both networks involved to work together to create a joint capacity planning and traffic engineering plan. The remainder of this section provides a technical overview of BGP and the tools it providers for traffic engineering. This section then provides the logic behind hot and cold potato routing, and how CDNs perform load-balancing and traffic engineering.

In addition to performing traffic engineering by tuning routing configurations (see Sec. 3.1.3), CDNs may have additional "content routing" mechanisms at their disposal. Specifically, given a request from a client for content, the CDN can use DNS to route the client to the “best” cache node that contains the required piece of content; often, the best node is the one that is geographically closest, but it may also be one that is more lightly loaded. A CDN can thus optimize delivery of content to its clients using a combination of content routing (i.e., directing the client to the best cache node) and traffic engineering (i.e., optimizing the underlying network path between the client and the cache to which it is directed). CDNs may perform content routing using DNS, by resolving the client’s DNS request for a domain name to an IP address corresponding to the best cache node, or may use other more granular mechanisms (see Section 3.1.5).

Additionally, although many ASes perform multihoming and traffic engineering via direct interconnection with multiple neighboring networks, some ASes may use other networks to perform these services on their behalf. For example, Internap and Equinix offer connectivity services that perform inbound and outbound traffic engineering on behalf of customers [49,50].

3.1 Routing Protocols

Routing protocols can be classified into two categories: Interior Gateway Protocols (IGP), which address traffic routing within an AS; and Inter-AS Protocols, which address traffic routing between networks.

An Interior Gateway Protocol (IGP) such as IS-IS or OSPF focuses on routing traffic within an AS [51,52,53]. These protocols distribute prefixes between routers within a specific network, and use algorithms for identifying the best path within the network or to the best exit point if the destination is another network.

An Inter-AS protocol provides the mechanism for managing connectivity between networks [54]. Inter-AS protocols rely on information communicated by other networks to determine the best path between networks. Trusting the information they receive from other networks allows an inter-AS protocol to not require a complete and accurate view of the internal structure of each network.

3.1.1 BGP Routing Overview

BGP is the Internet inter-AS routing protocol that allows for the exchange of traffic between ASes [54]. In BGP, routes are the combination of prefixes representing network endpoints and path attributes that provide additional information about the path to the prefix. Routes are exchanged between ASes using BGP peering sessions, which are long-term connections established between routers located in two interconnected ASes. BGP provides an algorithm for selecting the best path to reach the destination network. BGP offers AS administrators the ability to implement routing policy, or in other words how traffic flows through a network, by influencing the path selection. BGP also implements a loop avoidance mechanism to help ensure that Internet traffic does not get stuck in a circular pattern between two or more ASes. This section will focus on the BGP attributes used in path selection and that help implement routing policy, and give a general overview of how those attributes are used to manage traffic exchange between networks.

3.1.2 BGP Path Selection

BGP Path selection is the method used for selecting the best path for sending traffic to a given prefix. BGP analyzes each new or replacement route it receives over a BGP peering session and compares that route to any similar route stored in a local database. One by one the paths for each prefix are compared using their BGP Path attributes, detailed later in this section.

First, the "length" of an IP prefix advertised over each BGP session is analyzed. If a router receives multiple prefixes and some are more "specific" (smaller in size) than others, BGP will prefer the more specific route for that smaller prefix. For example, a router may receive 10.0.0.0/23 (512 IP addresses) over route 1, and 10.0.1.0/24 (256 IP addresses) over route 2 – which is also part of the other larger prefix. The router will prefer route 2 for 10.0.1.0/24, and route 1 for 10.0.0.0/24 (note this is only half of the original prefix advertised over route 1).

Assuming prefix length is the same, the path selection process compares the path attributes of like prefixes in the following order; 1) LOCAL_PREF 2) AS_PATH 3) ORIGIN 4) MED 5) NEXT_HOP and 6) Router ID. Matching prefixes are compared attribute by attribute in a serial fashion through the above list, the preferred path is then selected and all other paths are eliminated from consideration. If more than one path remains, those paths are passed on to the next attribute where a similar process takes place, until only one path remains. When the path selection process reaches the NEXT_HOP attribute, all destinations with longer metrics are eliminated leaving the shortest path remaining as the path selection. The last attribute, Router ID may serve as a tie-breaker if all other attributes are equal.

3.1.3 BGP Traffic Engineering

If a network has routes via different neighboring networks to a given destination IP prefix, it can perform outbound traffic engineering by adjusting the "local preference" attribute for each BGP route. However, controlling how traffic reaches the network from other destinations ("inbound traffic engineering") is far more difficult and indirect because it involves controlling the routes that *other* ASes choose. To discourage traffic from arriving via a given neighboring AS, a network can change or "tune" the path length by adding an additional route to the AS path (referred to as "AS path prepending," see Section 3.1.3.2 below). Unfortunately, changing the path length is not always effective as neighboring networks' routing policies may still ignore it. Another mechanism for controlling how neighboring ASes set their routing policy is to use the BGP COMMUNITIES attribute to "tag" a route as a backup route, which effectively asks a neighboring AS to assign a lower local preference to that particular route. In many cases these COMMUNITIES are only offered to paid customers such as those in a transit agreement.

3.1.3.1 Managing Traffic using LOCAL_PREF

When a network has several connections to other networks the BGP LOCAL_PREF attribute can be utilized to help manage traffic towards or away from one or more of these connections. As previously mentioned, LOCAL_PREF favors a higher value over a lower value. One example is moving traffic away from a congested link as it leaves the network. In order to shift traffic off a congested link, a network operator may lower the LOCAL_PREF assigned to an entire link or for a subset of prefixes reachable over the link. Some or all of the outbound traffic on the congested link will then redistribute across links with higher LOCAL_PREF values, alleviating the congestion. It is important to note that this works for outbound traffic only. If inbound traffic is the cause of congestion on the interconnection link, the network administrator will often contact the operations department of the other network and request they shift traffic using the same procedure.

3.1.3.2 Managing Traffic using AS_PATH length

Depending on how networks are interconnected, inbound traffic may favor one network over the other. The primary method for changing or tuning traffic as it enters a network is to advertise routes that look farther away than other routes through a practice called “AS path prepending.” AS path prepending through the use of the AS_PATH attribute makes that particular route advertisement look “longer,” as if it has traversed multiple networks. The BGP path selection algorithm favors routes with shorter paths assuming the LOCAL_PREF attributes for both are the same. This technique can often influence how neighboring ASes select routes. To balance the traffic the network operator can add duplicates of their own AS onto the existing AS path before announcing it to a neighbor. If the administrator prepends their AS path on the network with higher utilization, traffic may shift to connections with shorter AS paths with the hopes of more evenly distributing the traffic across both networks. If the neighboring network already prefers the route with a higher local preference value, AS path prepending will likely have no effect.

3.1.3.3 Managing Traffic using BGP COMMUNITIES

The BGP COMMUNITIES attribute is used as a way to implement routing policy across an administrator’s network and to allow customers that are running BGP with that network to have some control over their routing decisions. Communities are often used for customers of a network as well. The administrator would publish a list of communities to all of their respective customers that implicitly gives instructions to the network on how to handle the traffic. For instance a customer could use a community to prepend its own AS two times for certain routes, to discourage use of those routes. The customer-requested prepending would have the effect of shifting their traffic to another network upstream of their ISP to help balance incoming traffic as mentioned above.

3.1.3.4 Overview of BGP Path Attributes

BGP path attributes in addition to prefix and prefix length information are exchanged in the form of BGP routes in messages exchanged between routers running BGP. The primary purpose of these attributes is to provide network reachability information in order to determine the exit point for Internet traffic sent between networks. Some of the BGP path attributes are mandatory and will always be included with a BGP route, while others are optional. Mandatory path attributes provide critical information about a BGP route such as the route origin, the AS path detailing all ASes that need to be traversed, and the network address of the next router known to be one step closer to the destination. Other path attributes are used to affect path selection when multiple paths exist, to append additional information to a BGP route, or to signal when a BGP route represents multiple related routes that are able to be combined for efficiency purposes. Commonly used BGP path attributes include:

- *ORIGIN* – This attribute contains information on the source from which a route was received. The potential values are IGP, EGP and Incomplete. In Path Selection an IGP is favored over an EGP and EGP is preferred over an Incomplete value.
- *AS_PATH* – This attribute contains an ordered list of every AS needed to reach the destination prefix. It is a list of networks that must be traversed to reach the network announcing the final destination. Routes with shorter AS_PATHs are preferred over longer ones.
- *NEXT_HOP* – This attribute contains the IP address of the next router to be used to reach the prefix contained in a BGP route. There may be more routers in the path beyond the one represented by this attribute, but this is the next router that must be reached in order to determine the further subsequent router in the path, if any.
- *LOCAL_PREF* – This attribute allows the administrator of a network to configure a preference used for selecting one path over another when multiple paths are avaiable. In path selection, routes with higher LOCAL_PREF are preferred over routes with lower values.
- *MULTI_EXIT_DISC* (MED) – This attribute allows the administrator of a network a mechanism for suggesting the preferred entry point for traffic inbound from another peer when multiple connections are available. It should be noted that the final decision for which connection to use is in the hands of the AS sending the traffic and this attribute may be ignored by the sender.
- *COMMUNITIES* – This attribute allows an administrator of a network to pass additional information along to BGP peers. This information is useful as a way to simplify the implementation of complex routing policies.

3.1.4 Routing Exit Design – Hot vs. Cold Potato Routing

A well-known concept of BGP routing is “nearest exit” or “Hot Potato” routing. Paul Baran from RAND coined this term while developing network interconnection theories for the early ARPANET [55]. “Hot potato” routing is typically explained as “if you are passed a too-hot potato, you try and give it away to someone else as soon as possible”.

Networks are often interconnected in multiple geographically diverse locations, with the same destinations advertised in multiple locations, meaning traffic can take multiple routes over very different distances to reach the same destination (for example, from Los Angeles to Seattle through San Francisco instead of Houston). BGP examines all of the routes received, and by default chooses the shortest path to the destination AS, i.e., the connection that passes the traffic off of the network as soon as possible. This has the primary benefit of reducing the cost on the network sending traffic, by minimizing transit cost payments for the traffic by instead using a peering connection for that traffic. As an example, suppose two networks A and B have interconnections in Atlanta and San Jose. Traffic originating from Network A in Atlanta and destined for Network B in San Jose, would leave Network A locally in Atlanta and Network B would be burdened with carrying the traffic to San Jose. Traffic in the reverse direction would leave Network B in San Jose, with Network A having the burden of carrying the traffic to Atlanta.

It should be noted that "hot potato" routing behavior, and the associated economics of the sender of traffic having less of a burden than those receiving, is often cited in peering disputes when one network is receiving significantly more traffic than it is sending. It is the primary reason for balanced traffic ratio requirements in some peering contracts.

Some networks choose to use "cold potato" routing. This is the opposite of "hot potato" routing, with the network looking to keep traffic on their network as long as possible, and the exchange of traffic as close to its destination as possible. This enables the source network to manage the quality of delivery of traffic on as long of a portion of the network path as possible.

Content providers are the most frequent users of cold potato routing, to optimize their users experience by carrying the traffic on their network as close to the user as possible. The content provider bears the cost of the additional traffic load on their network, and the peered network benefits from reduced costs compared to if the content provider had used "hot potato" routing.

3.1.5 CDN Traffic Flow & Management

As mentioned in Section 2.7, CDNs and caches direct user requests to the optimal serving platform for each user. The simplest way to direct traffic is to operate as an "in-line" cache that intercepts all traffic to and from a particular set of users, also as discussed in Section 2.7, above. This becomes problematic when scaled to large numbers of users on an ISP or network operator, as the platform can become overloaded, representing a potential single point of failure, and costs can escalate as the platform needs to be scaled to manage all traffic flowing through it (rather than just the cacheable traffic). Cellular networks often use in-line caches, especially with a "content optimization" feature to reduce bandwidth usage on the radio network [56].

Some CDNs employ the user's DNS resolver to determine the nearest CDN platform to serve that user. The majority of residential broadband subscribers accept the DNS server's information provided by their ISP, as this information is configured automatically as part of the broadband Internet service. ISPs deploy DNS servers regionally in close proximity to the user in order to reduce latency for DNS lookups and thereby improve the customer experience. CDNs can associate users to the provider's regional DNS server in order to determine the general location of the user. The CDN provider uses this information, and other information from their traffic routing algorithms, to direct the user to an appropriate CDN node. Traffic steering via DNS redirection has limitations, as it depends on the ISP deploying numerous regional DNS servers in order for the CDN operator to have enough granularity to steer user traffic accurately, and it depends on customers using their ISP's DNS rather than a third-party DNS.

Alternatively, for video streaming and some other applications, users may be directed to the appropriate CDN platform by direct interaction between the client application and the CDN, using the actual IP address of the user, without having to rely on DNS. This generally results in more accurate and granular management of CDN traffic flows.

CDNs can also use optimization algorithms in order to minimize costs or best utilize certain capacities at certain times of day or in accordance with other rules established in the CDN's service logic. For example, a user in Miami may get a close response from a CDN server in Miami that is directly connected to their ISP over a no-cost link. Another user in Chicago may get different answers at different times of day; at peak times they may be served by a Denver server over a low-cost link and at off-peak they may be served by a New York server via a higher cost link. In some cases these choices may not necessarily appear to be the "best" from either a cost or performance standpoint; there are often many other technical, non-technical, and business criteria and objectives involved. In addition, content customers of a CDN in some cases will use multiple CDNs, and similar logic may be used to split traffic between multiple CDNs to optimize costs or other criteria.

3.1.6 Traffic Engineering Implications of CDNs

For network operators, deploying CDNs and caches can have a variety of technical implications. When a CDN is connected to an operator as a peer at an IX (options 2 and 3 in Sec. 2.7.1 above), the interconnection decision and technical complexity is similar to a regular network peer at an IX.

When CDNs and caches are deployed inside an operator's network or networks, deeper into the operator's AS than the IX, and closer to users when compared to the location of colocation facilities, additional benefits may be provided to both the network operator and content provider, in the form of reduced long-haul traffic costs and improved user experience.

However, such CDN deployments may result in additional challenges for a network operator, in terms of locating suitable space and power for the CDN or cache deployment, and management and security issues – for example the CDN equipment may have maintenance and support requirements, and may complicate security risks. Technical considerations also play a part, given that most CDNs and caches operate at the IP layer and can only be deployed where the IP protocol is present. For example, on cellular networks in particular, there are often only a few IP "gateways" (perhaps one per region) between the mobile network and the Internet network (that can translate into IP protocol), restricting the ability for CDNs to be deployed very close to users.

The trend of consolidation of content origination and increase in video traffic is causing a significant change in the dynamics of the Internet ecosystem, and in the balance of sent and received traffic on network interconnections. Especially with CDNs that have multiple copies of content in different locations, traffic originators typically control the source of traffic. Within the limitations of BGP path selection (Section 3.1.2, controlled by the receiver of traffic), the sender therefore may control the path of data sent to another network, so they can either increase or alleviate loading and congestion on different points of interconnection.

ASes receiving traffic have fewer options in terms of controlling traffic that enters their network, for example by BGP path prepending or other BGP decision criteria (see Section 3.1.3). CDNs and other originators of large traffic volumes also play a growing role in traffic management today, performing complex optimizations, measuring download performance and adjusting traffic flows in near-real-time, including from which nodes they source content [57,58,59]. To be sure, transit providers also play a role with the sale of capacity to their customers and peers. The capacity at each interconnection point is negotiated between peers. The resulting choices for routing to networks may have implications for network stability and performance [60,61,62,63].

As with peering, agreements between CDN providers and network operators are usually covered by non-disclosure agreements, so determining the extent of deployments or the benefit of CDNs for network operators is difficult.

Given the growth of video and high-bandwidth content discussed earlier in this report, it is clear that CDNs will continue to develop as a way to help the Internet scale efficiently and cost-effectively.

3.1.7 IP-based load-balancing

To avoid overwhelming a single source of content, a service may deploy more than one copy of the resource in a single colocation facility. However, there are still concerns with this model of balancing load as a single colocation facility can see power, cooling, and/or network failures. For many services, such a low level of redundancy is not acceptable; so multiple physical locations are necessary.

When more than one instance of the service is available at a single location, it is possible to load-balance with either multiple IP addresses for the service, or with devices which will spread the requests for service across a set of available instances of the service. In the latter case of multiple physical locations, there could be different IP addresses on each geographically disparate version of the service, and thus users are routed to different copies depending on which version is closest to them.

4 Security Concerns around Interconnection and Traffic Exchange

Interconnection involves engineering physical and logical connections between networks, but in some cases traffic can flow contrary to the intentions of network operators, either in error or due to malicious activity. There are a number of important security considerations when connecting two networks. In this section, we first discuss types of attacks involving interconnection as well as motivations that may drive attackers. We then discuss potential mitigations for these attacks and efforts to make routing more secure through new routing protocol extensions, notably RPKI and BGPSec.

4.1 Attacks on Routing

Routing protocols like BGP are susceptible to incorrect, malformed, or malicious information being included in the routing information exchanged. This can lead to attacks, such as the following [64]:

- **IP prefix hijacking** occurs when an AS advertises one or more IP prefixes that it does not actually own. This can be either intentional or accidental (i.e., as a result of misconfiguration) with the result being that an AS announces routes to incorrect IP prefixes via a routing protocol like BGP, called "BGP route hijacking." An attacker can implement a route hijack attack either by injecting announcements on a point-to-point BGP session between peering routers, or by connecting a rogue AS to an upstream ISP that is not filtering routes (see below).
- **AS path falsification** occurs when an AS announces a BGP route advertisement with an AS path that does not correspond to the sequence of ASes that advertised the route. When mounting this attack, an AS might shorten the AS path to attract traffic for the purposes of disruption or surveillance. This attack is sometimes called a "traffic attraction" attack.

These attacks can be mounted either through a compromised AS (as occurred with the Beltelecom [AS 6697] and Siminn [AS 6677] attacks in 2013) or via rogue ASes which advertise routes through less well-managed upstream ASes (as occurred with Atrivo/Intercage through 2011 and continues to persist with the Russian Business Network). An attacker can use BGP route hijacking to disrupt (or "blackhole") traffic for an IP prefix, to gain network IP address space from which to launch attacks (e.g., spam, other attacks), or to become a "man in the middle" AS for the traffic.

Manipulation of BGP route advertisements has also been featured in several high-profile censorship events. In an attempt to censor YouTube, Pakistan inadvertently hijacked YouTube's IP prefixes and began erroneously routing traffic to these prefixes from networks all around the world [65]. Other countries, such as Egypt, have used BGP to implement widespread censorship within the country [66].

There are some valid reasons for BGP route "hijacks." Many companies (Arbor Networks, Radware, Verisign, and even some large carrier networks) implement opt-in services that hijack the IP address of another network (with permission) to redirect traffic through middleboxes that perform additional services (e.g., traffic scrubbing) [67,68,69].

4.2 Defenses

Certain operational enhancements, such as the BGP "TTL Security Hack" and extensions to TCP to perform per-packet HMACs on point-to-point BGP peering sessions have reduced the risk of point-to-point hijacks [70,71]. Nevertheless, these enhancements protect only the BGP messages exchanged between two directly connected routers. They do *not* mitigate attacks from rogue ASes.

ISPs may also filter BGP routes, allowing BGP advertisements from their downstream networks only for prefixes that those networks (or their customers) own. Such filtering is more feasible for "stub" ASes at the edge of the network, where it is easy to determine (e.g., by consulting an Internet Route Registry (IRR)) whether the AS owns the IP prefix that it is trying to advertise. In the "core" of the network, it can be difficult to determine whether the AS that advertises an IP prefix can legitimately reach that prefix.

The Resource Public Key Infrastructure (RPKI) is an approach to build a formally verifiable database of IP addresses and AS numbers [72]. RPKI has (1) a public key infrastructure (PKI), with the necessary certificate objects; (2) digitally signed routing objects; and (3) a distributed repository for the objects.

RPKI is based on resource certificates that define extensions to a standard certificate format (X.509) to represent IP addresses and AS identifiers [73]. Authorizations for route origination, called Route Origination Authorizations (ROAs), contain information about which autonomous system (AS) is authorized to originate certain IP prefixes, as well as the maximum prefix length that the AS is authorized to advertise [74].

In RPKI, a ROA conveys only the simple authority to originate a route announcement. It does not provide any additional authorizations concerning routing policy. For example, the ROA contains no information about whether the AS can delegate the right to advertise the IP prefix to another entity. It also operates only at the level of IP prefixes, and cannot authorize information at finer levels of granularity (e.g., per application, based on time of day, etc.). Existing authentication frameworks such as RPKI make it difficult to express authentication logic that supports the type of delegation that is required for legitimate uses of route hijacking, such as DoS mitigation [75].

In addition to these technical limitations, RPKI does not prevent all attacks on BGP (e.g., it does not authenticate the AS path in BGP route announcements), thereby leaving BGP vulnerable to certain attacks such as path shortening attacks, whereby an AS can shorten an AS path to attract traffic (e.g., as part of a man-in-the-middle attack).

The Internet Engineering Task Force (IETF) is designing the BGP security protocol (BGPSEC), which extends RPKI by securing the AS path of BGP route advertisements [76]. Using RPKI, BGPSEC ensures that the autonomous system announcing the prefix is the legitimate owner of the prefix (*origin authentication*). BGPSEC extends RPKI by verifying that the AS path associated with the announcement has not been tampered with (*path authentication*)[77]. Hurdles for BGPSEC deployment include the computational overhead associated with signing and verifying BGP routing messages in real-time, and its ineffectiveness in partial deployment.

5 Measuring Interconnection

Measuring interconnection can benefit network operators, content and application providers, researchers, policymakers, users, and the general public. Such measurements can help people understand the state and health of interconnections; measurements may also shed more light on issues such as how specific traffic is routed and how traffic shifts over time across various networks. This subject is complex, and much work remains. Given the complexity of this topic, exploring it in-depth is out of scope for this document.

6 Future Developments in Interconnection

Interconnection policies and mechanisms continue to evolve. Advances in computing power may ultimately make some protocol improvements more feasible. For example, protocols to secure the Internet's routing infrastructure that require checking cryptographic signatures on routes may become computationally more efficient. Advances in computing power may also make it easier to monitor traffic and interconnection, although increased transmission speeds may continue to present challenges.

The research and operations communities are experimenting with what would be significant changes to interconnection as well. Recent developments in software defined networking (SDN) and network functions virtualization (NFV) point to new possibilities for more sophisticated and fine-grained interconnection arrangements than are possible with BGP today. BGP allows operators only limited control over traffic: it operates only at a prefix-level granularity, its mechanisms are indirect, and it is difficult to influence routing decisions beyond the immediate next-hop AS. SDN facilitates more fine-grained, direct control over forwarding decisions, and it also allows ASes to influence routing decisions along points of an end-to-end path, even if they are not directly connected at those points [77,78,79]. These capabilities will make it possible to control traffic in ways that are difficult or impossible today. For example, SDN allows an AS to allow only certain application traffic to arrive on a particular interconnection point. By virtue of NFV's ability to allow for virtual network functions to be easily created, modified, deployed and flexibly applied (e.g., for applications like Network-as-a-Service, NaaS) – NFV could make it easier to modify traffic flows at interconnection points and may also involve re-routing traffic to support network services. Such a capability introduces both challenges and opportunities, as interconnection agreements may need to include provisions for how traffic can (or cannot) be modified “in flight” by another AS. Another direction for future improvement would be development of algorithms that allow for increased cooperation between content providers and ISPs to achieve overall more efficient content delivery [80,81].

As these and other developments create new ways for ASes to both interconnect and control traffic at interconnection points, it will be important to continually revisit their implications for new interconnection mechanisms and policies.

7 References

[1] Lougheed, K., and Y. Rekhter, "A Border Gateway Protocol (BGP)," RFC 1105, June 1989, http://tools.ietf.org/html/rfc1105.

[2] National Science Foundation, "NSF 93-52 – Network Access Point Manager, Routing Arbiter, Regional Network Providers, and Very High Speed Backbone Network Services Provider for NSFNET and the NREN(SM) Program," Program Solicitation, https://w2.eff.org/Infrastructure/Govt_docs/nsf_nren.rfp.

[3] Schwartz, M. and E. Mershon, "Paid Internet Peering on the Rise, Disputes Possible," Communications Daily, vol. 33, No. 126 (July 1, 2013), http://www.cs.columbia.edu/~misra/news/CD070113.pdf.

[4] Woodcock, B. and V. Adhikan, "Survey of Characteristics of Internet Carrier Interconnection Agreements," 2011, https://www.pch.net/resources/papers/peering-survey/PCH-Peering-Survey-2011.pdf.

[5] ARIN, Regional Internet Registry, https://www.arin.net/knowledge/rirs.html (last visited Oct. 31, 2014).

[6] PeeringDB, https://www.peeringdb.com (last visited Oct. 31, 2014).

[7] Hurricane Electric, PeeringDB and Why Everyone Should Use It, 2011, http://www.internetsociety.org/sites/default/files/PeeringDB%20and%20why%20everyone%20should%20use%20it.pdf.

[8] Dr. Peering, Peering Policy, http://drpeering.net/white-papers/Peering-Policies/Peering-Policy.html (last visited Oct. 31, 2014).

[9] North American Network Operators Group, https://www.nanog.org/ (last visited Oct. 31, 2014).

[10] Global Peering Forum, www.peeringforum.com (last visited Oct. 31, 2014).

[11] Equinix, IBX Data Centers and Colocation Services, http://www.equinix.com/services/data-centers-colocation/ (last visited Oct. 31, 2014).

[12] Wikipedia contributors, "Network access point," *Wikipedia, The Free Encyclopedia,* http://en.wikipedia.org/w/index.php?title=Network_access_point&oldid=609507980 (last visited Oct. 31, 2014).

[13] Telegeography, Internet Exchange Map, http://internetexchangemap.com (last visited Oct. 31, 2014).

[14] Internet Exchange Point, http://en.wikipedia.org/w/index.php?title=Internet_exchange_point&oldid=628044019 (last visited Oct. 31, 2014).

[15] Euro Internet Exchange Association, "What is an IXP?", https://www.euro-ix.net/what-is-an-ixp (last visited Oct. 31, 2014).

[16] LINX, Private Peering Services, LINX Private Interconnect, http://www.linx.net/service/privatepeering.html (last visited Oct. 31, 2014).

[17] Open-IX, http://www.open-ix.org/ (last visited Oct. 31, 2014).

[18] Seattle Internet Exchange, Route Servers, https://www.seattleix.net/rs.html (last visited Oct. 31, 2014).

[19] Amsterdam Internet Exchange, Specifications and Descriptions, AMS-IX Route Servers, https://ams-ix.net/technical/specifications-descriptions/ams-ix-route-servers (last visited Oct. 31, 2014).

[20] P. Richter, G. Smaragdakis, Anja Feldmann, Nikolaos Chatzis, J. Boettger, W. Willinger, "Peering at Peerings: On the Role of IXP Route Servers," 2014, http://net.t-labs.tu-berlin.de/~prichter/imc238-richterA.pdf.

[21] CloudFlare, An Overview of CloudFlare, https://www.cloudflare.com/overview (last visited Oct. 31, 2014).

[22] Akamai, Solutions, http://www.akamai.com/html/solutions/index.html (last visited Oct. 31, 2014).

[23] Google, Peering and Content Delivery, Google Caching Overview, https://peering.google.com/about/ggc.html (last visited Oct. 31, 2014).

[24] Netflix, Netflix Open Connect, https://openconnect.itp.netflix.com/ (last visited Oct. 31, 2014).

[25] Cisco, "The Zettabyte Era – Trends and Analysis," May 29, 2013, http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI_Hyperconnectivity_WP.html (last visited Oct. 31, 2014).

[26] Kleiner Perkins Caufield & Byers, Mary Meeker/Liang Wu, "Internet Trends D11 Conference," May 29, 2013, http://www.kpcb.com/insights/2013-internet-trends (last visited Oct. 31, 2014).

[27] Sandvine, Global Internet Phenomena Report 1H2014, https://www.sandvine.com/downloads/general/global-internet-phenomena/2014/1h-2014-global-internet-phenomena-report.pdf.

[28] Sandvine, Global Internet Phenomena Report, Spring 2011, http://www.wired.com/images_blogs/business/2011/05/SandvineGlobalInternetSpringReport2011.pdf.

[29] Cisco, "Cisco Visual Networking Index: Forecast and Methodology, 2013-2018," June 10, 2014, http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.pdf.

[30] Craig Labovitz, Scott-Iekel Johnson, Danny McPherson, Jon Oberheide, and Farnam Jahanina. "Internet Inter-Domain Traffic". *In Proc. of the ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, 2010, http://doi.acm.org/10.1145/1851275.1851194 (last visited Oct. 31, 2014).

[31] Craig Labovitz, Hearing on Competition in the Video and Broadband Market: the Proposed Merger of Comcast and Time Warner Cable, Statement, May 8, 2013, http://judiciary.house.gov/_cache/files/d76b57df-eece-4ed2-ade3-e3c6b3a91fbe/labovitz-testimony.pdf.

[32] Craig Labovitz, Massive Ongoing Changes in Content Distribution, Content Delivery Summit Spring 2013, http://conferences.infotoday.com/documents/172/2013CDNSummit-B102A.pdf

[33] Analsys Mason, "Content and Applications Providers are Major Investors in the Networks that Make Up the Internet," Oct. 13, 2014, http://www.analysysmason.com/About-Us/News/Newsletter/Internet-infrastructure-investment-Oct2014/ (last visited Oct. 31, 2014).

[34] B. Ager, N. Chatzis, A. Feldmann, N. Sarrar, S. Uhlig, and W. Willinger, "Anatomy of a Large European IX," in *Proc. of ACM SIGCOMM*, 2012.

[35] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian, "Internet Inter-domain Traffic," in *Proc. of ACM SIGCOMM*, 2010.

[36] V. Giotsas, S. Zhou, M. Luckie, and K. Claffy, "Inferring Multilateral Peering," in *Proc. of ACM CoNEXT*, 2013.

[37] "France Telecom severs all network links to competitor Cogent," 2005. http://morse.colorado.edu/~epperson/courses/routing-protocols/handouts/cogent-ft.html.

[38] S. Cowley, “ISP spat blacks out Net connections,” Oct. 6, 2005, InfoWorld, http://www.infoworld.com/t/networking/isp-spat-blacks-out-net-connections-492 (last visited Oct. 31, 2014).

[39] M. Ricknas, “Sprint-Cogent Dispute Putes Small Rip in Fabric of Internet,” Oct. 31, 2008, PCWorld, http://www.pcworld.com/article/153123/sprint_cogent_dispute.html (last visited Oct. 31, 2014).

[40] Merit Network, “Some Truth About Comcast - WikiLeaks Style,” Email List Archives, 2010, http://www.merit.edu/mail.archives/nanog/msg15911.html (last visited Oct. 31, 2014).

[41] J. Engebretson, “Level 3/Comcast dispute revives eyeball vs. content debate,” Nov. 30, 2010, Telecompetitor, http://www.telecompetitor.com/level-3comcast-dispute-revives-eyeball-vs-content-debate/(last visited Oct. 31, 2014).

[42] J. Engebretson, “Behind the Level 3-Comcast peering settlement,” July 17, 2013, Telecompetitor, http://www.telecompetitor.com/behind-the-level-3-comcast-peering-settlement/ (last visited Oct. 31, 2014).

[43] D. Young, “Unbalanced Peering, and the Real Story Behind the Verizon/Cogent Dispute,” Verizon Public Policy Blog, June 2013, http://publicpolicy.verizon.com/blog/entry/unbalanced-peering-and-the-real-story-behind-the-verizon-cogent-dispute (last visited Oct. 31, 2014).

[44] J. Brodkin, “Why YouTube buffers: The secret deals that make-and break-online video,” July 28, 2013, ArsTechnica, http://arstechnica.com/information-technology/2013/07/why-youtube-buffers-the-secret-deals-that-make-and-break-online-video/ (last visited Oct. 31, 2014).

[45] S. Buckley, “Cogent and Orange France fight over interconnection issues,” Aug. 21, 2011, FierceTelecom, http://www.fiercetelecom.com/story/cogent-and-orange-france-fight-over-interconnection-issues/2011-08-31 (last visited Oct. 31, 2014).

[46] Andrews, J., and S. Higgenhotham, “YouTube sucks on French ISP Free, and French regulators want to know why,” Jan. 2, 2013, GigaOm, http://gigaom.com/2013/01/02/youtube-sucks-on-french-isp-free-french-regulators-want-to-know-why/ (last visited Oct. 31, 2014).

[47] S. Buckley, “France Telecom and Google entangled in peering fight,” 2013. http://www.fiercetelecom.com/story/france-telecom-and-google-entangled-peering-fight/2013-01-07.

[48] J. Brodkin, “Time Warner, net neutrality foes cry foul over Netflix Super HD demands,” Jan. 17, 2013, ArsTechnica,

http://arstechnica.com/business/2013/01/timewarner-net-neutrality-foes-cry-foul-netflix-requirements-for-super-hd/ (last visited Oct. 31, 2014).

[49] Internap, Flow Control Platform, http://www.internap.com/network-services/ip-services/flow-control-platform/ (last visited Oct. 31, 2014).

[50] Equinix, http://www.equinix.com/ (last visited Oct. 31, 2014).

[51] D. Oran, "OSI IS-IS Intra-domain Routing Protocol," RFC 1142, Feb. 1990, http://tools.ietf.org/html/rfc1142.

[52] J. Moy, "OSPF Version 2," RFC 2328, Apr. 1998, http://tools.ietf.org/html/rfc2328.

[53] Coltun, R., D. Ferguson, J. Moy, and A. Lindem, "OSPF for IPv6," RFC 5340, July 2008, tools.ietf.org/html/rfc5340.

[54] Rhekter, Y., and T. Li, "A Border Gateway Protocol 4 (BGP-4), RFC 1771, Mar. 1995, https://tools.ietf.org/html/rfc1771.

[55] "Digital Simulation of Hot-Potato Routing in a Broadband Distributed Communications Network", Paul Baran and Sharla P. Boehm.

[56] Claudia Bacco, EMEA: CDN in the RAN with Saguna and Akamai, RCRWireless News, Sept. 2, 2014, http://www.rcrwireless.com/20140902/carriers/emea-cdn-ran-saguna-akamai-tag11

[57] V. K. Adhikari, Y. Chen, S. Jain, and Z.-L. Zhang, "Reverse-Engineering the YouTube Video Delivery Cloud," in *Proc. of the Workshop on Hot Topics in Media Delivery (HotMD)*, Jul 2011.

[58] V. K. Adhikari, Y. Chen, S. Jain, and Z.-L. Zhang, "Where Do You 'Tube'? Uncovering YouTube Server Selection Strategy," in *Proc. of IEEE ICCCN*, Aug 2011.

[59] Y. Chen, S. Jain, V. K. Adhikari, , and Z.-L. Zhang, "Characterizing Roles of Front-end Servers in End-to-End Performance of Dynamic Content Distribution," in *Proc. of the USENIX/ACM SIGCOMM Internet Measurement Conference (IMC)*, Nov 2011.

[60] J. Jiang, V. Sekar, and H. Zhang, "Improving Fairness, Efficiency, and Stability in HTTP-based Adaptive Video Streaming with FESTIVE," in *ACM CoNEXT 2012*, 2012.

[61] T.-Y. Huang, N. Handigol, B. Heller, N. McKeown, and R. Johari, "Confused, Timid, and Unstable: Picking a Video Streaming Rate is Hard," in *Proc. of the 2012 ACM conference on Internet measurement conference*, IMC '12, (New York, NY, USA), pp. 225-238, ACM, 2012.

[62] V. K. Adhikari, Y. Guo, F. Hao, M. Varvello, V. Hilt, M. Steiner, and Z.-L. Zhang, "Unreeling netflix: Understanding and improving multi-CDN movie delivery," in *Proc. of IEEE INFOCOM*, pp. 1620-1628, 2012.

[63] V. K. Adhikari, Y. Guo, F. Hao, V. Hilt, and Z.-L. Zhang, "A tale of three CDNs: An active measurement study of Hulu and its CDNs," in *Proc. of IEEE INFOCOM Workshop*, 2012.

[64] U.S. Department of Commerce, National Institute of Standards and Technology, "Border Gateway Protocol Security," Recommendations, Special Publication 800-54, July 2007, http://csrc.nist.gov/publications/nistpubs/800-54/SP800-54.pdf.

[65] RIPE, "YouTube Hijacking: A RIPE NCC RIS Case Study," Mar. 2008, http://www.ripe.net/internet-coordination/news/industry-developments/youtube-hijacking-a-ripe-ncc-ris-case-study (last visited Oct 31, 2014).

[66] A. Dainotti, C. Squarcella, E. Aben, K. Claffy, M. Chiesa, M. Russo, and A. Pescape, "Analysis of Country-wide Internet Outages Caused by Censorship," 2011, http://www.caida.org/publications/papers/2011/outages_censorship/outages_censorship.pdf

[67] Arbor, Arbor Cloud, http://www.arbornetworks.com/products/arbor-cloud (last visited Oct 31, 2014).

[68] Radware, DDoSPedia, Scrubbing Center, http://security.radware.com/knowledge-center/DDoSPedia/scrubbing-center/ (last visited Oct 31, 2014).

[69] Verisign, Features, https://www.verisigninc.com/en_US/website-availability/ddos-protection/ddos-product-features/index.xhtml (last visited Oct 31, 2014).

[70] Gill, V., J. Heasley, and D. Meyer, "Internet Draft, The BGP TTL Security Hack (BTSH)," May 2003, http://tools.ietf.org/html/draft-gill-btsh-02.

[71] Ford, A., C. Raiciu, M. Handley, and O. Bonaventure, "TCP Extensions for Multipath Operation with Multiple Addresses," RFC 6824, Jan. 2013, https://tools.ietf.org/html/rfc6824.

[72] Weiler, S., and D. Blacka, "Clarifications and Implementation Notes for DNS Security (DNSSEC)," RFC 6840, Feb. 2013, https://tools.ietf.org/html/rfc6840.

[73] Lynn, C., S. Kent, and K. Seo, "x.509 Extensions for IP Addresses and AS Identifiers," RFC 3779, June 2004, https://tools.ietf.org/html/rfc3779.

[74] Lepinksi, M., S. Kent, and D. Kong, "A Profile for Route Origin Authorizations (ROAs),"

RFC 6482, Feb. 2012, https://tools.ietf.org/html/rfc6482.

[75] Bush, R., and R. Austein, “The Resource Public Key Infrastructure (RPKI) to Router Protocol,” RFC 6810, Jan. 2013, https://tools.ietf.org/html/rfc6810.

[76] Lepinski, M., and S. Turner, “An Overview of BGPSEC,” Internet Draft, July 2014, https://tools.ietf.org/html/draft-ietf-sidr-bgpsec-overview-05.

[77] Project CARDIGAN: An SDN-Controlled Internet Exchange, North American Network Operators Group Meeting (NANOG 57), Feb. 4-6, 2013, https://www.nanog.org/meetings/nanog57/presentations/Wednesday/wed.lightning3.whyte.sdn.controlled.exchange.fabric.pdf.

[78] Gupta, A., L. Vanbever, M. Shahbaz, S. Donovan, B. Schlinker, N. Feamster, J. Rexford, S. Shenker, R. Clark , and E. Katz-Bassett, “SDX: A Software Defined Internet Exchange,” ACM SIGCOMM, Aug. 2014, http://gtnoise.net/papers/2014/gupta-sigcomm2014.pdf

[79] Bailey, J., D. Pemberton, A. Linton, C. Pelsser, and R. Bush, “Enforcing RPKI-Based Routing Policy on the Data Plane at an Internet Exchange,” *ACM SIGCOMM Workshop on Hot Topics in Networking (HotSDN),* Aug. 2014, http://dl.acm.org/authorize?N71315.

[80] Jiang, W., R. Zhang-Shen, J. Rexford, and M. Chiang, “Cooperative content distribution and traffic engineering in an ISP network,” *In Proc. ACM SIGMETRICS*, June 2009.

[81] Valancius, V., B. Ravi, N. Feamster, and A. Snoeren, “Quantifying the benefits of joint content and network routing,” *In Proc. of the ACM SIGMETRICS, International Conference on Measurement and Modeling of Computer Systems*, 2013, pgs. 243-254.

8 Glossary of terms

All definitions of terms are solely for the purposes of this report, and many are adapted from publications of the Internet Engineering Task Force (www.ietf.org). Readers should be aware that a number of terms have alternate definitions, particularly when used in different or non-networking contexts.

Autonomous System (or “AS”)	An “Autonomous System” (AS) is a network, operated by a single organization, that presents a single coherent interior routing plan and a consistent picture of what networks are reachable through it. In BGP routing, it is identified by an AS Number (or ASN), which is allocated through a Regional Internet Registry (RIR). A network operator may utilize one or more ASNs in its network.

Backbone Network	The collection of high speed transmission links and supporting network infrastructure that connects a network’s geographically distributed points of presence and/or regional networks.
Colocation facility	A physical location (e.g., a building) where multiple network operators install their networking equipment for the purpose of interconnection. Also called a Colocation Center.
Content Delivery Network (CDN, also sometimes called a Content Distribution Network)	A CDN places and distributes content as close as practical to the end users requesting that content. A CDN contains an organized set of servers in multiple locations that collectively provide caching or other services. Depending on the CDN business model, its servers may be distributed into colocation facilities, directly into ISP networks or in housed in its own facilities.
De-peering	The act of two peered ASes discontinuing a peering relationship.
Internet Exchange (IX)	A “public” form of interconnection that allows for efficient connectivity among any ASes capable of physically reaching the colocation facility in which the Internet Exchange switching equipment is located. An Internet Exchange, in technical terms, is accomplished using a piece of computer networking equipment called *switch* (or collection of switches) – also called a *switch fabric*. Each AS connects from their network to this equipment, which in turn provides the ability to connect to every other AS connected to the switch fabric. Also called an Internet Exchange Point (IXP). Peering interconnections that are set up over Internet Exchanges are referred to as “public peering.”
Internet Service Provider (or ISP)	A network operator (consisting of one or more ASes) that offers Internet access to other customers. This is in contrast to the many ASes on the Internet that are edge networks serving other purposes, for example enterprises, educational institutions, or governments.
Network Access Point (NAP)	An early term (no longer current) for the facilities that interconnected the early Internet. These were a combination

	of colocation facility and Internet Exchange.
Peering	A form of interconnection in which two networks agree to exchange their customer traffic among one another. In contrast to transit, where the offered service is access to "all the Internet," in peering the access is only to the other network and its customers. ○ *Settlement-Free Peering* (also *Settlement-Free Interconnection* (SFI)) – A form of peering in which two parties agree to exchange traffic for no direct monetary exchange. ○ *Paid peering* – A form of peering in which two parties agree to exchange traffic between their networks with some form of monetary compensation involved.
Point of Presence (PoP)	An access point to a communication provider's network. It is a physical location that houses network equipment with interfaces that are administered as network connection points. A PoP may serve strictly as a location offering customers access to the network. Alternatively, it could be a room or cage within a colocation facility for the purpose of interconnection, or it could serve both purposes.
Private Interconnect	A "private" form of bilateral interconnection in which direct physical connections are used to interconnect two networks, without the need of an Internet Exchange. If used for peering, this is known as "private peering."
Transit	A form of interconnection in which one network purchases network connectivity service, offering access to every publicly reachable destination on the Internet, from an ISP. In many cases, what is offered is "full" transit, meaning access to every publicly reachable destination on the Internet
Transit Provider	An ISP that provides transit to customers as a product.

9 Document Contributors and Reviewers

- Dean Bahizad, *Charter Communications*
- Fred Baker, *Cisco*
- Richard Bennett, *American Enterprise Institute*
- Michael Blanche, *Google*
- William Check, *National Cable and Telecommunications Association (NCTA)*
- Lily Chen, *Verizon*
- kc claffy, *University of California, San Diego; Center for Applied Internet Data Analysis (CAIDA)*
- David Clark, *Massachusetts Institute of Technology (MIT)*
- David Farber, *Carnegie-Mellon University*
- Michael Fargano, *CenturyLink*
- Nick Feamster, *Georgia Institute of Technology*
- Jeff Finkelstein, *Cox Communications*
- Jeffrey Good, *Disney*
- Joseph Lorenzo Hall, *Center for Democracy & Technology (CDT)*
- Dale Hatfield, *University of Colorado*
- Arianne Hinds, *CableLabs*
- Kevin Kleinsmith, *Union Wireless*
- Ken Ko, *ADTRAN*
- Gary Langille, *EchoStar*
- Matt Larsen, *Vistabeam*
- Jason Livingood, *Comcast*
- Kevin McElearney, *Comcast*
- Chris Morrow, *Google*
- Brad Noyes, *CenturyLink*
- Jon Peha, *Carnegie-Mellon University*
- Barbara Stark, *AT&T*
- Matthew Tooley, *National Cable and Telecommunications Association (NCTA)*
- Scott Weber, *Charter Communications*
- Jason Weil, *Time Warner Cable*
- Greg White, *CableLabs*