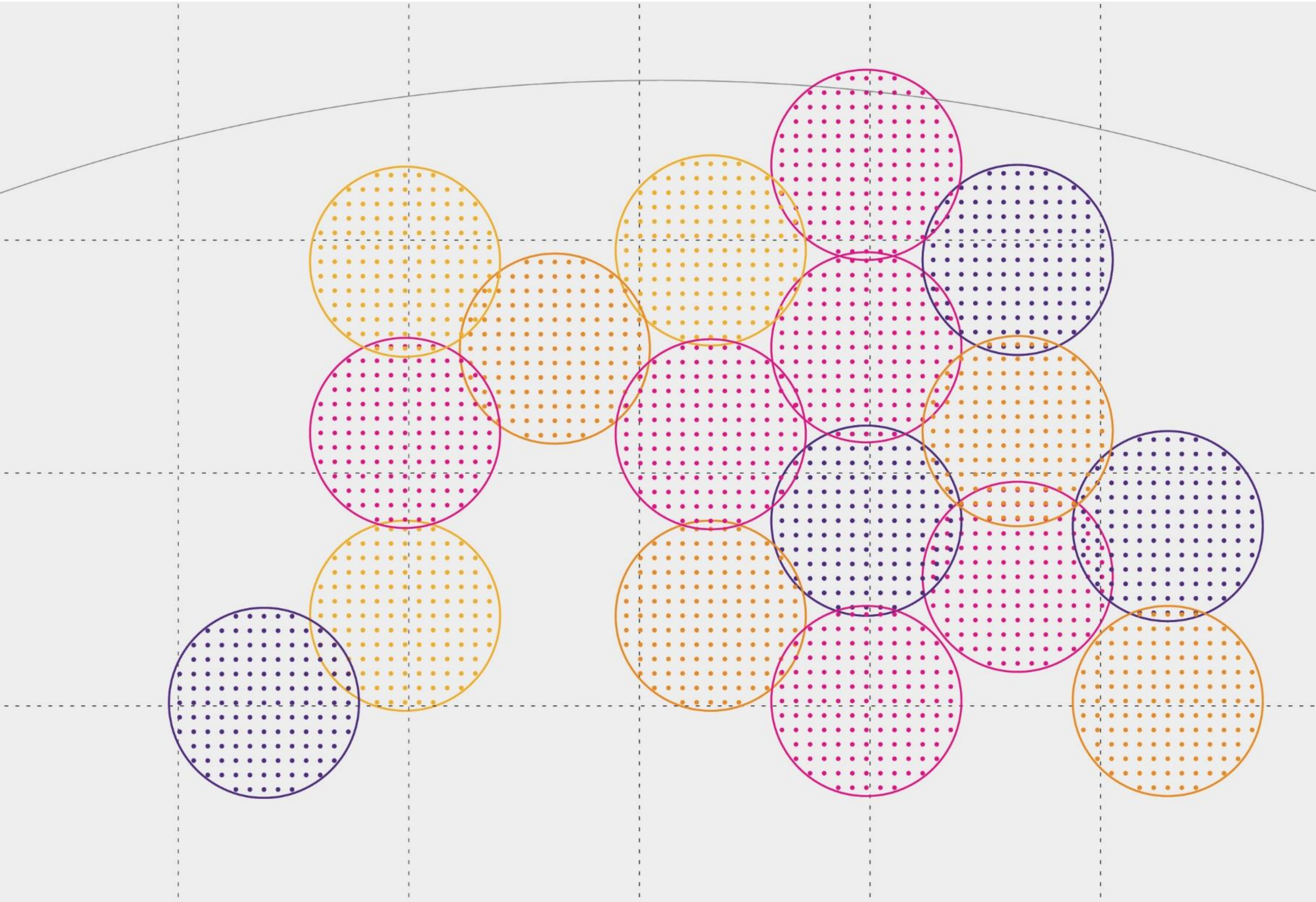
# A Busy Person's Guide to How the Internet Works (and is paid for)

12 July 2022

Mark McFadden, Aude Schoentgen, Karim Bensassi-Nour



## About Plum

Plum offers strategy, policy and regulatory advice on telecoms, spectrum, online and audio-visual media issues. We draw on economics and engineering, our knowledge of the sector and our clients' understanding and perspective to shape and respond to convergence.

## About this study

This is a report for Google that addresses the topic: How the Internet Works (and is paid for) – a 2020s refresh (to inform policymaking).

**This is the first of three parts of the study: a concise introduction to the way the Internet works**.

Plum Consulting
10 Fitzroy Square
London
W1T 5HP

**T** +44 20 7047 1919
**E** info@plumconsulting.co.uk

# Contents

# 1 Introduction

The Internet is a collection of people, nodes, networks and content. Often it is described as a cloud made up of connections between devices, but this is misleading because there are rarely direct connections over the Internet. In fact, the Internet is completely decentralised with many nodes and combinations of direct and indirect connections between them.

In the early days of the Internet, people used computers to construct and send requests to other computers that serviced those requests. This request and response model, over a collection of decentralised networks served the Internet very well in its early decades. Today's Internet is far more complex. We use the term "nodes" because the Internet connects far more than computers. Today's Internet connects things, services, mobile devices, cars and much more. Nodes are simply devices on networks that can send or receive information.

In its simplest form, each node has an address[1] that can be used as a locator to which other nodes can send requests or information. The address allows the Internet to deliver the message to the correct destination even though the message usually passes through many other intervening nodes.

There are special nodes in the network that connect diverse networks to each other. These "routers" have the job of receiving messages and forwarding them on toward their intended destination. The routers use the address on the message as a tool to decide how to send the message onward, and closer, to its final destination.

[1] An Internet Protocol address, or IP address for short.

# 2 How Do You Get Access to the Internet?

Consumers and enterprises both gain access to the Internet through Internet Service Providers (ISPs). ISPs sell access to the Internet as a service. For a consumer, this often means one of two possibilities: first, a fixed Internet access service in the home; and second, Internet access for a mobile device. In the first case the ISP often provides a router or home hub that allows devices in the home to connect. Messages from nodes inside the home are directed to the home router and then out to the Internet through the ISPs own network. Enterprises and businesses usually have connections with more capacity. Still, the basic picture is often the same: the enterprise has a router on the "edge" of its internal network and nodes inside the network send messages through that router and out to the ISP's network connections.

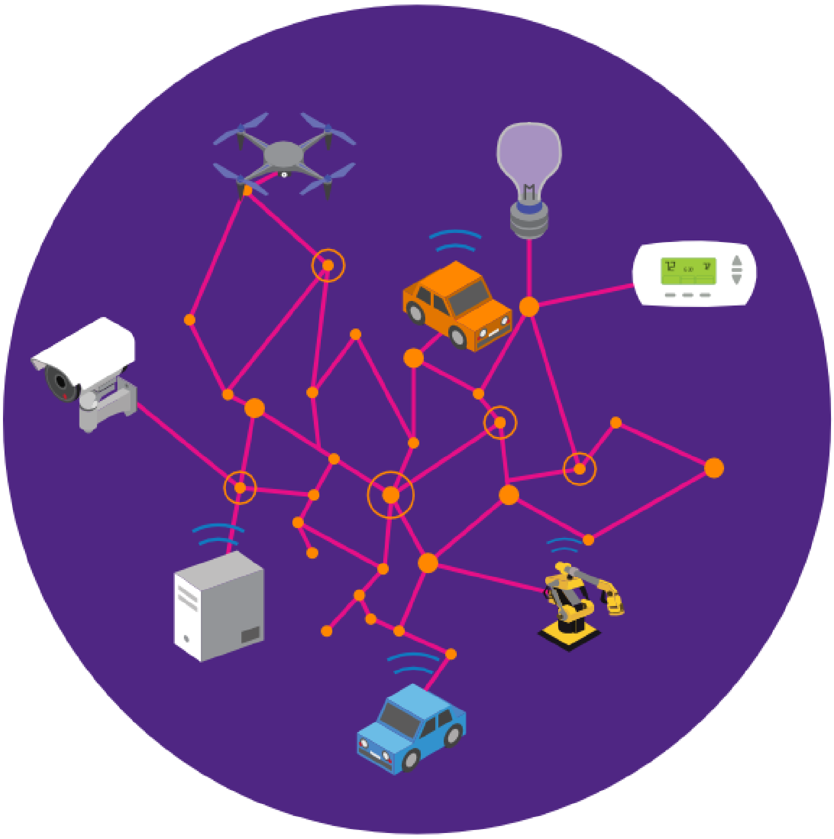
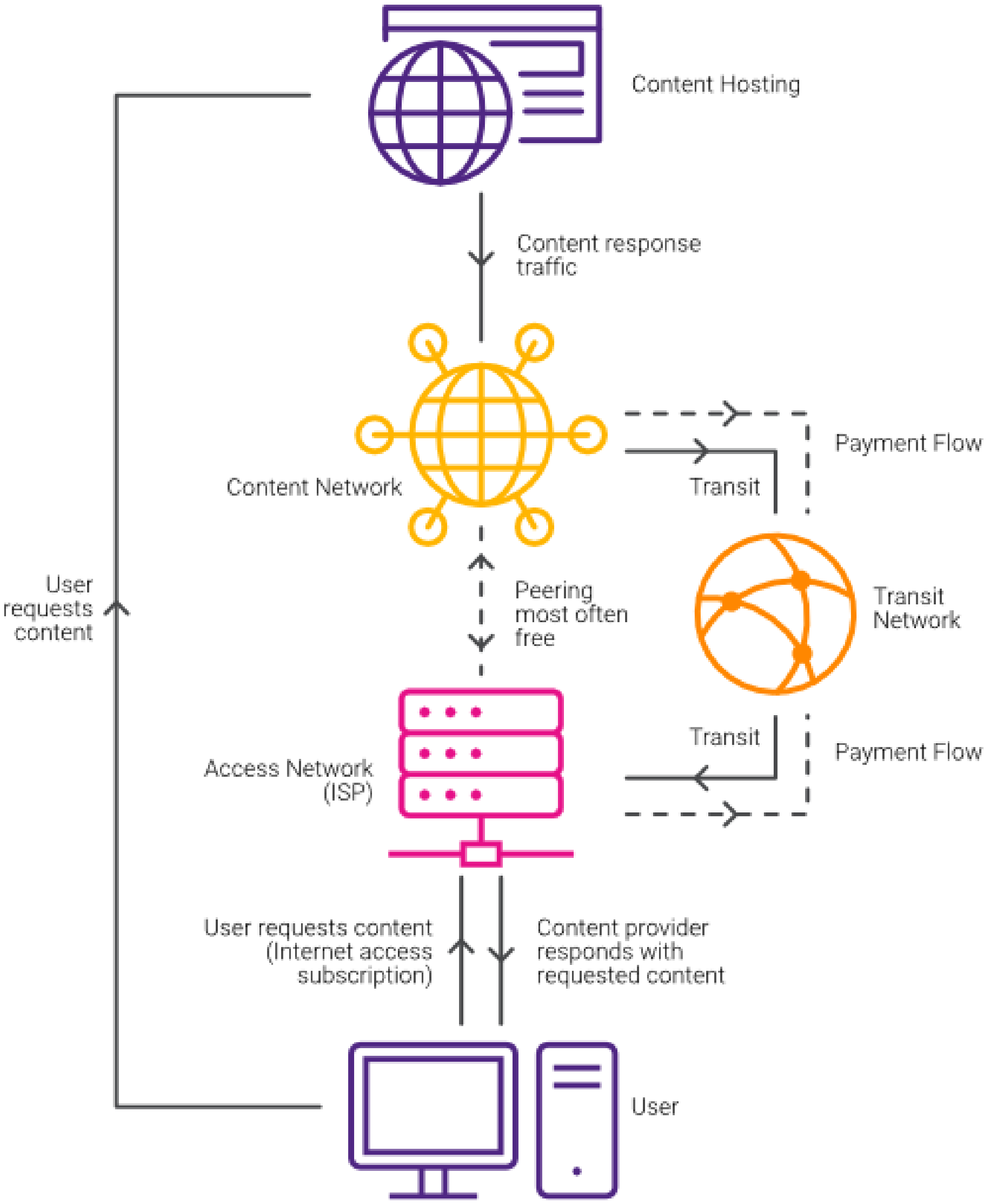


**Figure 2.1: Connected devices of the Internet**

**Figure 2.2: Information Flows on the Internet**




Source: Plum

One of the most significant ways that the Internet has changed is how people connect to their ISPs. Many years ago, most connections used the copper wires that also provided traditional voice communications. Today, the

Internet uses a variety of access technologies that allow people and businesses to connect, including cable services, cellular and mobile Internet, fibre optic, fixed wireless and even access to the Internet via satellites.

ISPs get paid by end-users (either consumers or enterprises) for Internet access provision. End-users are generally charged on an unmetered basis through "flat rates" or on a usage basis. The flat rate for Internet access includes the Internet connection itself and the transfer of an unlimited volume of data. In the early days of the Internet, before the advent of broadband, most customers' monthly Internet access spending had two components: an Internet service subscription fee, and a metered fee related to the amount of time, or amount of traffic, a customer used while connected to the Internet.

In terms of end-user prices for Internet access, there is great variation across different regions of the world. These disparities can be explained by the differences in regulatory regimes, levels of competition, infrastructure development and consumer bases.

# 3 What Does Information Look Like on the Internet?

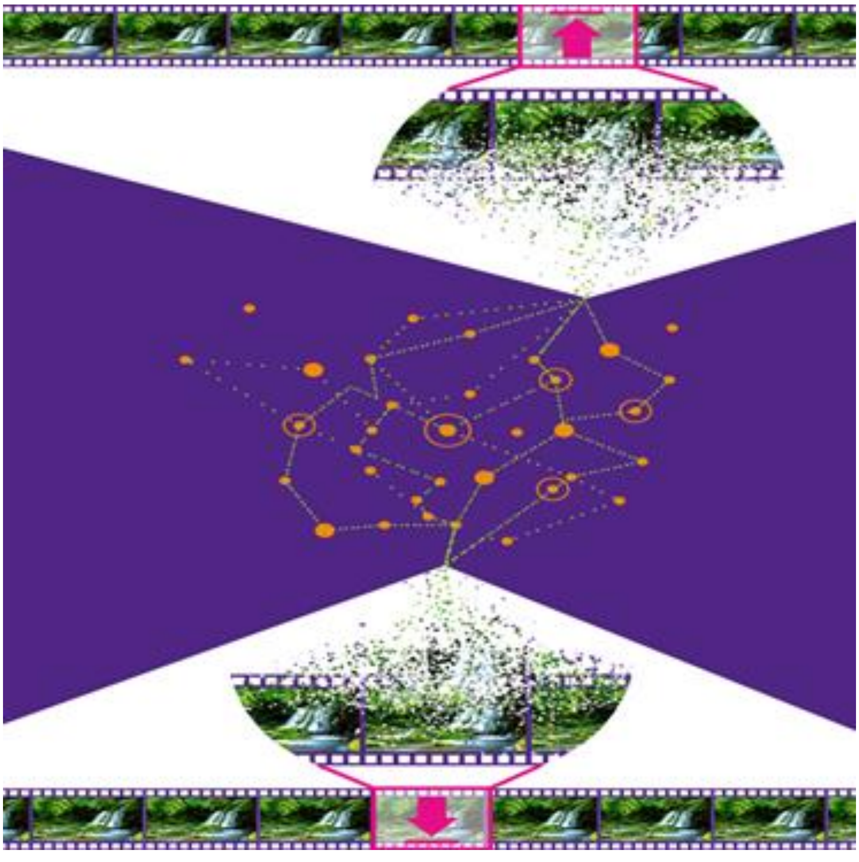
The basic unit of information on the Internet is called a "packet." Every time a node sends messages on the Internet, it constructs a series of packets that function as envelopes containing the information being exchanged. The packets usually have two parts: a "header" that contains information about the packet, and the "payload" which contains the content of the message.

Packets are simply made up of ones and zeroes, known as "bits." All messages traverse the Internet in this binary format. The nodes in the network are responsible for encoding information (email, pictures, video, etc.) from its original format into bits that can be assembled into packets.

One of the most important parts of the header is the addressing information it contains. Each packet's header has the address of the node the message is being sent to, as well as the address the message is being sent from. Routers in the Internet use these addresses to decide how to forward the packet on and closer to its intended destination.

The payload of the message contains the content sent between nodes. For example, a picture would be made up of many packets – each to be sent between the sending and receiving node. The sending node would break the image into multiple packets and send all of them, through the routers toward the receiving node. The destination would then take the packets and reassemble them into the original picture. Most things sent on the Internet are too big to fit in a single packet, so this disassembly and assembly process takes place for almost every service or application on the Internet.

**Figure 3.1: : Sending a Movie Across the Internet**



It's crucial that both the sender and receiver use the same approach for disassembly and assembly of information. To ensure that the sender and receiver can understand each other, a set of standard, interoperable "protocols" have been developed so that all nodes will have a common set of rules for exchanging messages and information. These protocols are simply rules for nodes to use to talk to each other. Very famous protocols include the Hypertext Transport Protocol (HTTP), which allows the exchange of messages for the World Wide Web and the Simple Mail Transport Protocol (SMTP), which allows for the transport of messages for electronic mail.

Applications as complex as video conferencing and as simple as Internet chat all have this defining characteristic in common: the use of standardised protocols to exchange information contained in packets over the networks connected to the Internet.

# 4 How is the Internet Coordinated and Governed?

The protocols that make the Internet possible are developed by a diverse group of international, technical organisations – sometimes called "standards bodies" or "standards development organisations" (SDOs), including:

- The Internet Engineering Task Force (IETF) is the primary SDO for establishing Internet communication and transport protocols.
- The Institute of Electrical and Electronics Engineers (IEEE) is the key SDO for describing the rules for sending bits over wires or wireless connections.
- The International Telecommunications Union's Standardization Sector (ITU-T) is responsible for developing and publishing telecommunications protocols.
- The International Organization for Standardization (ISO) defines standards for application in a very diverse group of sectors, including technology, business, government, and society.
- The World Wide Web Consortium (W3C) creates the rules for exchanging information on the World Wide Web.
- The Internet Corporation for Assigned Names and Numbers (ICANN) is responsible for ensuring the stable and secure operation of the Internet's unique identifier systems, by coordinating the management of the technical elements of the Domain Name and numbering System (DNS) to ensure universal resolvability so that all users of the Internet can find all valid addresses.
- Regional Internet Registries (RIRs) such as RIPE and APnic coordinate the Internet's unique numbering and addressing spaces. The RIRs work through a process of consensus for their regional community.

# 5 How Does Information Move Around the Internet?

The Internet is not one, big, unified network, it is sometimes called a "network of networks." The networks that make up the building blocks of the Internet are called Autonomous Systems (AS), and there are about 100,000 such ASes on the Internet today. Each AS is called "autonomous" because each is a network administered independently. When these networks interconnect, they constitute the public Internet as we know it. Every Internet user is always a part of one of these Autonomous Systems.

To move information around the Internet there needs to be "routes" through the map of interconnected networks. The protocol that makes the interconnections and map possible is called the Border Gateway Protocol (BGP). BGP is the protocol that allows computers to map a route for packets from source to destination using the shortest and cheapest path possible.

Each AS uses BGP to build its own map for traversing the Internet. However, almost no BGP server has a complete global map of all the possible paths for packets on the Internet.

When a node sends a packet onto the Internet, it usually sends the packet to the first router it knows about. This is often the primary router of the ISP that the node is connected to. That router reads the packet and looks at the header to decide if the packet is destined for a node that is connected to the same network as the router. If the packet is destined for a node on the ISP's own network, the ISP can deliver it on its own. If the address indicates that the packet must be sent onward, it uses BGP to decide which new router to send it to. Every router on the path does this until the packets arrive at their destination.

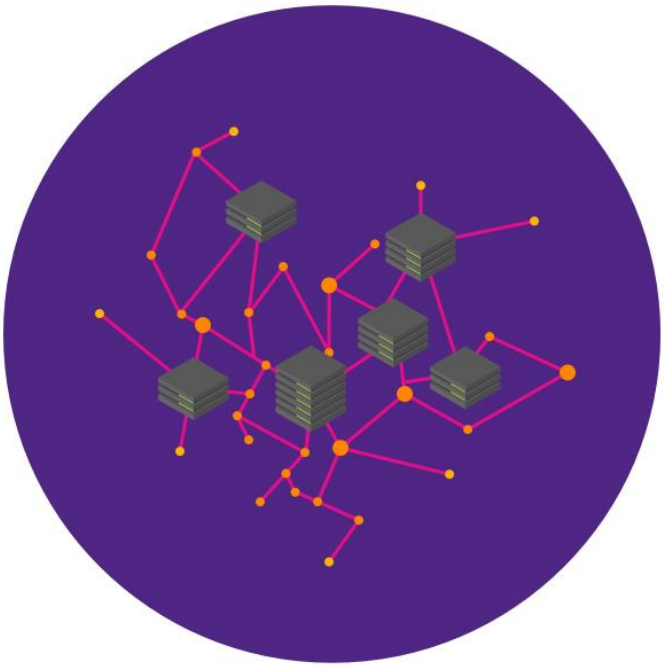


**Figure 5.1: Routers in the Internet**

When BGP servers of different ASes talk to each other, they become neighbours. As neighbours, they exchange maps of the routes they know about and want to share. Shorter paths for the packets are preferred because that results in less time for the packet to get to its intended destination.

Once the ASes have a map of available routes with their neighbours, they can exchange data. In some cases, the ASes will agree to exchange traffic between the ASes for free – an arrangement called "peering."

Often the size of the neighbours will be different. The word "transit" is used to describe connectivity to any destination on the Internet.

Any organisation responsible for part of a packet's journey from source to destination has a stake in peering and transit. This includes access providers of any size. ISPs who provide access – using any access technology – need a way to route the packet toward its destination. In addition, the access provider also provides a path to its own customers. That way, packets headed to an ISP customer's endpoint have a path for delivery. This novel distributed way the Internet works contrasts fundamentally with the way traditional telephony networks worked historically, linking A to B in a linear, dedicated straight line, rather than through several diverse nodes as the Internet does.

# 6 An Example of How Content is Delivered

Imagine an instructional video produced in South Korea being viewed by an Internet user in South Africa. How would the video be delivered?

1. First, the video content would be created and produced by a user or company in South Korea.

2. Once the video was prepared, it would be transmitted to the Internet via that user's local ISP.

3. The ISP would use its transit and peering resources to send the video to the local router of the video service.

4. Then, the video service would use its own backbone and private networks to forward the video to the video service's data centre for processing.

5. Once processed, the video is then transmitted to the video service's local data centres around the world – once again using its own backbone, subsea and private networks to distribute the newly processed video to local servers run by the video service.

6. At this point, the video is now ready to be accessed locally by users around the world.

7. When the user in South Africa requests the video, the request is sent by that user's ISP through the ISP's network to the local video server.

8. The local video server then responds to the request with a local copy of the video and sends it back to the requesting user via the user's ISP. (see Figure 2.2).

# 7 How is Peering and Transit Paid For?

Physical interconnection at a router is only one side of the interconnection process. There needs to be an agreement between the different actors so they can send their data to each other.

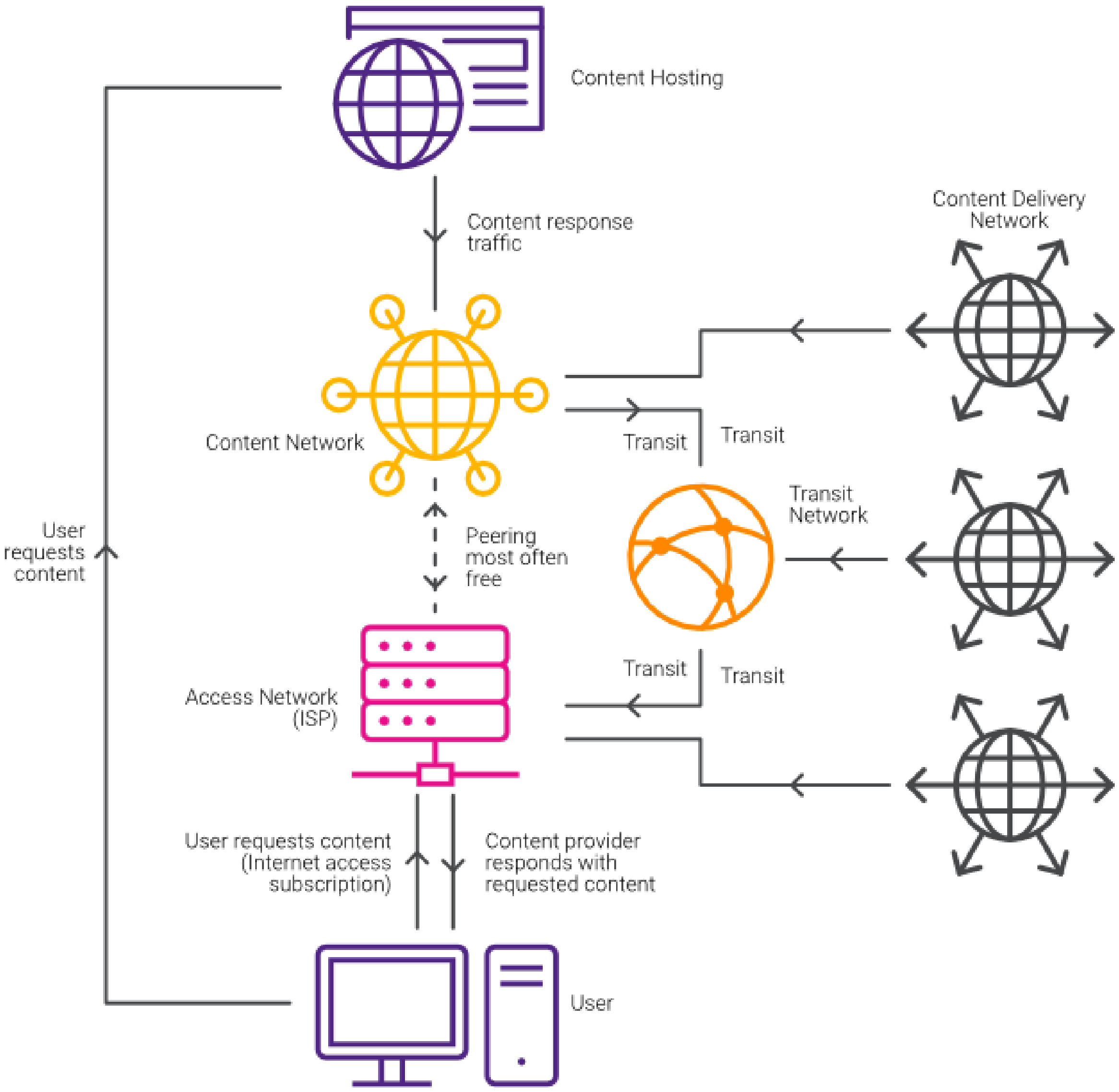
Transit arrangements provide access to the entire Internet, while IP Peering arrangements facilitate the direct mutual exchange of traffic between directly connected players and each party's downstream customers. The two types of interconnections can be both complementary and substitutable arrangements depending on the network configuration chosen by an operator.

A key difference between transit and peering is that in the vast majority of cases, no money is exchanged between networks that agree to peer with one another. The benefit to an operator of an AS to peer is that the arrangement can reduce dependency on upstream transit providers – reducing that dependency has the effect of reducing cost for that operator.

Peering enables the two networks to exchange data and benefit equally. The peering arrangement also has the ASes advertise only their internal customer routes. Although this is a zero-cost arrangement, there are usually costs for both parties associated with co-location, the routes to get to the co-located facility, and the required infrastructure connections. Strategies for peering (and interconnection in general) vary from one network operator to another. Those may be explained in a reference document, which is public most of the time and known as the "peering policy", although there is rarely any requirement to stick to this policy in the network's dealings with others. Cloudflare, for example, has an "open peering" policy and participates at nearly 150 Internet exchanges.

In practice, peering agreements are often not covered by a written contract and are established by informal agreements between the two parties. According to Packet Clearing House (PCH), in 2016 about 99.9 per cent of the peering agreements were done informally using a handshake. While there are cases where one party pays for peering in the contemporary Internet, this is relatively rare.

**Figure 7.1: Information and Intermediaries on the Internet**




Source: Plum

# 8 What are Internet eXchange Points and Content Delivery Networks?

Internet eXchange Points (IXPs) are co-located connections of many ASes for the purpose of peering. The IXP is usually located in a data centre which provides the infrastructure to join multiple networks together at a single location. IXPs make it easier for networks to connect to many different ASes over a single connection.

Content Delivery Networks (CDNs) have become a major part of Internet infrastructure by contributing to a more regionalised Internet interconnect that relies less on relatively expensive international transit. The increasing volume of content going through the global Internet in the past few years has required fresh solutions to be developed to avoid the following issues:

- One technical: the latency of content delivered may increase, resulting in a lower quality of service to the user; and,
- One economic: the overall cost of content delivery would increase.

In this context, CDN services have emerged to move the content “closer” to the final user for improved delivery. A CDN is a network of servers (called edge servers) that has been deployed in different geographical locations that cooperate in providing content/data to their users by handling traffic loads and reducing the time of content delivery, through a process of replication. The CDN determines the shortest possible route between the user and the CDN (located at the edge of the Internet) based on indicators like proximity, speed, or latency.

# 9 What is Content on the Internet?

The content provision environment involves a variety of applications, services and players, but essentially online content refers to everything an Internet user can look for online: this could be cat videos on YouTube, ideas on Pinterest, news articles on the Washington Post or John Lennon's biography on Wikipedia. Content provision services vary by type, size, reach and business model, but they all share one common feature: they all seek to capture and keep the consumer's attention, provide needed services and distribute information.

What follows is a simplified classification based on the consumer's intention when navigating online. It is assumed that a consumer can either be looking for entertainment, for information, for a service, or for socialising.

**Figure 9.1: Different categories of online content**

| Entertainment-based content | Information-based content | Socialising-based content | Service-oriented content |
|---|---|---|---|
| • Video streaming (Netflix, Disney+)<br>• Online gaming<br>• Musique streaming (Spotify, Pandora) | • Websites (Vodafone's website)<br>• Online newspapers<br>• Knowledge sharing platforms (Wikipedia)<br>• E-learning platforms (Coursera, Udemy) | • Social media (FB, Instagram, Linkedin)<br>• User-generated content<br>• Dating platforms (Tinder) | • E-commerce platforms (Amazon, Alibaba)<br>• Apartment rental (Airbnb) |

Source: Plum

Early Internet content was primarily static. A device or application requested content from a server, the server prepared and returned a response, and the device or application presented the results. While some of this static content remains on the Internet, much of the content landscape has changed.

One major change is the rise of mobile applications, where an end-user takes advantage of a mobile device to gain access to services and information. In 2010, there were only 210,000 apps in the Apple App Store. Today the number is more than 4.5 million. The Internet acts as the connectivity foundation for the vast majority of those apps, but the tools and protocols in use are quite different from the ones used by electronic mail or other older services.

If mobility has changed how the Internet works, so have hyper-scale platforms. Platforms such as Facebook and Twitter also use the Internet as the fundamental connectivity tool, but the content of these services is dynamically generated in real-time, often by end-users, and in a state of constant revision.

# 10 Who Pays for Content?

Business models vary greatly from one content provider to another, but they mainly depend on how they allocate the space between content and advertising. There are three different business models that dominate content economics today.

- *Subscription-only business models.* Content Providers (CPs) that operate on a subscription-only business model such as Netflix and Disney+ get paid by end-customers through a flat subscription fee. They usually require the customer to sign-up for monthly, quarterly, or annual automatic payment plans and can generate recurrent revenue from an engaged customer base. These CPs may face the potential for higher customer churn or loss than free services. In some cases, CPs can offer access to exclusive content for paying members, while non-paying members are only given limited access: This is referred to as a "Freemium model."

- *Subscription-based and advertising-based models.* Other CPs, such as Spotify, get paid by both consumers and advertisers. They may choose to use an advertising network or sell ads privately through their own negotiation. Non-paying consumers have access to the proposed content but are targeted with advertisements, while paying customers can usually benefit from an ad-free service.

- *Advertising-funded only models.* Other content providers attract consumers by offering their services for free such as YouTube or Facebook. Although there is no monetary payment involved in this case, there is still an exchange: In exchange for the service (watching videos, communicating with friends etc.), consumers provide their attention and personal data. The data collected enables the content providers to attract advertisers who are willing to pay to reach their customers based on their preferences and their online activities. This model works best when the volume of users is very large or very specialised.

# 11 The Internet and Regulation

Unlike traditional telecommunications, the Internet is entirely decentralized: connecting networks together across the globe without concern for jurisdictional boundaries. A lack of centralized control means that explicit regulatory control by an individual government can sometimes be circumvented. If a website is removed in one country, it can be back in operation nearly instantly in another. If a document is removed from a website, it can be reposted on many more almost trivially.

The result is that many public policy and regulatory issues on the Internet often cannot be handled by traditional territorial, and national institutions. The regulatory and public policy model for the Internet is usually not top-down but characterised by transnational cooperation between many stakeholders, including protocol engineers, ISPs, network operators, users, governments, and international organisations. National policies play an important part in shaping the Internet, but the Internet has also evolved new institutions and governance arrangements that respond to the unique needs of the network.

Internet governance happens at the global, regional, national, and local levels. As a basic rule, infrastructure and the technical layer of the Internet have a global approach to governance. Protocols, cables, and routers are maintained collaboratively by multi-stakeholder organisations that support the Internet as a cross-border and international technical structure. In contrast, applications and content are more susceptible to national or local governance mechanisms. As an example, regulating content that is allowed to be published or viewed online means that Internet users are subject to their countries' laws and regulations when going online.

There is no equivalent to the International Telecommunications Union or national regulatory authorities as there are in traditional telecommunications, in part due to the global nature of the Internet, and the multiplicity of private networks that comprise it.

# 12 Advantages of the Internet's Model

The simplicity of the Internet Protocol means that it can run on top of almost any physical network. This gives the Internet the advantage of tremendous flexibility. The Internet runs over a wide variety of physical carriers including simple copper pairs, coaxial TV networks, cellular mobile systems, satellite systems and a wide variety of wireless networks. The advantage of the Internet model is that there can be improvements in the physical capacity of the underlying network without the need to change anything from the Internet Protocol to the application. This flexibility results in the Internet Protocol dominating as the preferred method to transport most kinds of traffic, including traditional voice, video, and other data.

Simplicity alone is not the only advantage: the Internet has scaled faster and more dramatically than any other network technology in history. Not only has the number of people and devices grown exponentially in the last twenty years, but the amount of traffic has increased dramatically. Once again, scaling the Internet is related to the independence of the layers that make up protocols on the Internet. Both the networks and the applications that use them, can be changed or replaced without impacting the other.

If we use the common characteristic of the Internet as a 'network of networks" the growth rate would be in evidence by counting the number of distinct networks connected together. In 1986 there were 80 Autonomous Systems (AS) attached to the Internet, by 2000 the number had increased to about 10,000 and today the number is nearly 100,000[2]. This expansion combines growth in the number of users with an expanded geographic reach. In addition, to the number of networks, the Internet has also scaled up in capacity in the global backbone. Transoceanic cable capacity is primarily driven by growth in content and Internet applications and content providers are investing in the supply of that part of the backbone – those companies now account for about two-thirds of the total capacity of subsea cables.[3]

The Internet also has an advantage in adapting to new needs and requirements. Before the mid-1990's, information publication and sharing was difficult. The emergence of a new technology, the World Wide Web, meant that nearly anyone could publish on the Internet and that publishing application was available to anyone around the globe. The Web emerged as a dominant technology – so much so that many users think the World Wide Web is the Internet. The Internet has the advantage of becoming the primary tool for distributing services. Traditional services, such as health care, travel reservations, schooling, watching movies or banking have moved from being associated with a specific location to being available anywhere. Support for diverse applications directly results from the permissionless innovation design principle.

Recently, not only has the Internet's scaling advantage been obvious, but so is its resilience in the face of unexpected changes or events. The pandemic has been the most dramatic unexpected event: traffic rates in some regions of the world increased by 50 per cent over the previous, pre-pandemic year. Working from home, homeschooling and home entertainment all combined to create a sudden explosion of traffic. Developers of applications worked to change their applications to work in the face of increased volume. Operators of networks also increased network capacity in the face of the increase in traffic. Studies have shown that the deployment of increased capacity allowed end-users to maintain download speeds without implementing service guarantees or class of service rules. One reason for this resilience is that the Internet is decentralised, and the networks that make up the Internet are autonomous and able to respond to local changes in demand independently of other parts of the network. Another is that the Internet's topology contains no single, central point of control. There are multiple routes to most content on the Internet, and if one of those routes fails, other routes can provide needed content without other changes to the network

[2] IPv6 CIDR Report, May 2022. Available at: https://www.cidr-report.org/as2.0/

[3] Aqua Comms, High Tide: the state of subsea cables. 2016. Available at: https://www.datacenterdynamics.com/en/opinions/high-tide-the-state-of-subsea-cables/