How to Stream Video over a Network or the Internet

How Video Travels Across the Internet

As noted in earlier chapters and as any end user would be quick to point out, viewing streaming video over the Internet is hardly a seamless experience. Streaming video suffers from hiccups, delays, drop-outs, skips, and connection loss. In this section, we explain how the Internet moves data and how this affects video playback.

It's sometimes hard to understand why the Internet has trouble moving audio and video when radio, television, and telephones do it fairly well and have existed for almost 100 years. So first let's look at the mechanisms of these traditional media.

Radio

router
a network device
that groups computers
in the network
and establishes an area
inside it

Radio works simply because a single tower broadcasts the same signal to many receivers. Everyone listens to the same thing at the same time. All the stations are available at any time; you simply have to tune into a different frequency signal. The main barriers to radio transmission are distance; physical barriers such as hills, buildings, and tunnels that block the signal; and interference between two strong signals near each other on the dial. In terms of communication, radio is a one-way broadcast transmission.

Television

Television works much like radio, except that television broadcasting is organized into national networks. The same program is delivered to television receivers around the country by broadcasting the originating signal to branch offices, which broadcasts it out from towers, out through cable companies, or to people with satellite dishes. In any case, the same signal is sent to everyone at the same time - a one-way broadcast. All the channels are available at any time; there is no noticeable delay caused by changing channels. The main barriers to television reception are bent or frayed cables, badly aimed antennas or dishes, physical barriers as in radio, and interference of stations with each other.

Telephone

Telephone calls use many of the same wires used by the Internet. The telephone central office maintains devices called switches (automated versions of the classic telephone switchboard) that are used to connect the call to the next location. Telephone calls create a two-way circuit all the way from caller to receiver. The message "All circuits are busy" - usually heard only during disasters or radio call-in concert ticket giveaways - means the switch does not have any more slots in which to carry this call. The main barriers to telephone transmission are found at the beginning of the call— if there are not enough circuits to place the call. While a call is in progress, the entire route between the caller and recipient is reserved for their use only, even if there is silence and no one is talking. Telephones use what is called a circuit-switched connection.

Internet Basics

The path from a website to a web browser is different than these other systems. Conceptually, it is similar to the telephone conversation: It's a two-way conversation in which the browser asks for a document and the server sends it. Unlike the telephone call, however, there is no reserved circuit. Data, in the form of requests and responses, are organized into chunks called packets and sent between the requesting web browser and the web server.

In between the requester and the server is a series of routers. These machines route traffic between different smaller networks. Each time a packet crosses the boundary from one ISP to another, or from one kind of network to another, it goes through a router. The packets "hop" from router to router like a bucket brigade. This type of data transmission is called packet switching, instead of circuit switching. Internet packet switching has some attributes that make it reliable and unreliable at the same time.

The Internet is an extremely heterogeneous network, consisting of several different kinds of networks and ways of connecting networks to the Internet, as described in the next few sections.

The Internet Backbone

The Internet backbone (as much as a large, shapeless and ever-shifting cloud of networks can have a backbone!) consists of long-haul connections that carry large volumes of Internet traffic (packets) across and between continents.

routing
 the process of selecting paths in a network along which data can be sent between computers (through the router)

bandwidth
the amount of data that can be transferred through a specific path in the network (usually expressed in kb/s [kilo bits per second])

Public Exchange Points

Public exchange points exist at various points on continents and are the major nerve centers where many regional private networks, Internet providers, corporations, schools, and government divisions—large and small—converge to exchange traffic destined for other points on the Internet. You can compare these centers to major public airports, where international and domestic flights arrive 24 hours a day and trade passengers from different airlines.

Peering

The process of connecting a network to the Internet at one of these exchange points is called peering, and connecting to the backbone this way makes one a Tier-1 Internet provider. ISPs that rent their connection from a Tier-1 provider are called Tier-2 providers, and so on. The policies, prices, and agreements that cover how data is treated on these connections are as numerous as there are companies involved. This is the first source of variability for our packet switching.

Private Peering

Peering is simply two networks connecting to each other with routers. Public peering occurs at large exchange points, but any two networks that find a lot of traffic flowing between them can choose to create a direct private link between the networks (called private peering). This reduces the cost of access through a public exchange point or other provider for all the bandwidth that travels between these two networks. It also decreases the number of intermediate connections between the networks. For instance, when several schools in the same organization link together, their inter-campus network traffic does not have to go out to the Internet at large, and is often more reliable as a result. In this scenario, though, each school has its own connection to the Internet. What if one of the school's Internet connections went down? Would it be fair to send its traffic through the private peering connection and use another school's Internet connection? The way these kinds of questions are answered and the internal policies in this regard are another contributing factor to the variability of Internet packet switching.

Internet Complexity

As everything "goes digital," the distinction between cable TV wires, telephone wires, radio waves, and satellite transmission blurs. However, there are many ways to send data over these media. Internet data transmission can be complicated, leading to a variety of undesirable transmission characteristics.

Packet Loss

Circuit switching on the Internet is described as "best-effort," meaning that one of the routers along the way can lose a packet before it reaches its destination. In this case, the sender or receiver must somehow note that the packet was lost (perhaps by receiving the next packet and noting that it is out of context) and re-requesting the lost packet. This mechanism is fairly reliable in that two machines will usually (and eventually) figure out what went wrong and resend the missing packets. Packet loss causes audio and video to pause if the packets are eventually resent, and it causes video to pause, drop out, and skip if the packets are not resent at all. In our analogy of a public exchange point being a major airport, if it's a "foggy day" at that exchange, the part of the Internet that goes through that exchange can be slowed down (called a brownout) by the data that can't "take off."

Different Routes

Not all packets in a file follow the same route to the destination computer. This is not unlike the airline's hub and spoke system: One packet might go "direct" from San Francisco to Washington; others might "transfer" in Atlanta or Chicago to get to Washington. Contributing to this issue is private peering and the variable rules and costs associated with all the choices to be made. Alternate routes can be excellent when one path between two machines goes down and a packet can use another path. It can also cause strange effects, such as when a packet is sent down a slow route, is assumed lost, is resent - and then later reappears as a duplicate packet! Audio can stutter and skip if duplicate packets are not detected and discarded. Also, some paths travel far out of the way, hopping through many more routers than necessary and causing large delays. The more "hops" or routers between two machines, the higher the chance of unexpected delays.

Delay (Latency)

Because of the many different routers a packet has to go through to get from sender to receiver and because there are no reserved circuits on the Internet like there are for telephones, the delay of any given packet can be high or low, or change unexpectedly. This can be caused by a variety of factors such as: A router is too busy and can't keep up with traffic. A particular link between sender and receiver becomes saturated. A link goes down, causing traffic to be rerouted to a different link. One or more routers in between can't think fast enough. A firewall looks at all the packets for viruses. Delay is added due to the use of older technology, such as modems. Other downloads on a pipe cause it to delay. Packets are lost, resulting in resends, and other packets get bunched up behind them. These factors make predicting how long it will take to get packets back from a server difficult. Because of varying latency, video can take a

peering
the arrangement
of traffic exchange
between the Internet
service providers (ISPs).

broadcast
the action of sending data
by one computer
in the network to
all the others
that are available
inside the network

long time to start playing; fast-forward and rewind features can be slow and clunky; and video can pause, stutter, skip, and stop altogether. Bandwidth Variation Another factor on the Internet is the variability of bandwidth. With broadcast media, such as radio or television, as well as telephones, the bandwidth is always the same - just enough to carry the channel or the conversation. There is no wasted bandwidth; the size of the channel is just enough to carry the data. It was designed to be that way. Because the Internet is designed to allow different computers of different speeds and different channel sizes communicate, it is possible to have bottlenecks, not just due to traffic that the size of the channel varies from sender to receiver. The Internet link for a major website's hosting provider might be excellent. The links between the host's ISP and its branch in a particular city might be high-capacity. However, the Internet link provided by a small ISP to the end user might be very small due to oversubscription. If that Internet provider has incurred a good deal of customer growth without upgrading its own connection to the Internet backbone, the potentially high-bandwidth connection from the website host is lowered to the slowest intermediate link in the chain. In other words, the bandwidth between a website and a client is no faster than its slowest link.

NOTE

Fundamentally, the Internet is far better suited for sending web pages than real-time media because web pages are far smaller and far less sensitive to delays. There is not much difference between a one- and two-second delay in getting a web page, but a one-second pause in real-time video is unacceptable. The brute-force approach of keeping the bit rate of the video far below the maximum bandwidth of the Internet connection can be effective in getting Internet video to perform predictably.