How the Internet Works: Headers and Data; Packets and Routing

1. Transmitting Information through the Ages

How can you get information from one place to another? How, for example, can you get an image from one place to another? Two hundred years ago, you would have to draw, paint, or print the image on paper or canvas, then carry, on foot, horseback, boat, or wagon, this analog version of view.

One hundred years later, technology had advanced. By the early part of the 20th century, fax machines existed. With fax, you could draw the image and send it over wires. What new scientific developments made that change possible? Here are a few milestones:

<table>
<thead>
<tr>
<th>YEAR</th>
<th>WHAT</th>
<th>WHO</th>
<th>WHERE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1800</td>
<td>Invents electric battery</td>
<td>Allesandro Volta</td>
<td>Italy</td>
</tr>
<tr>
<td>1820</td>
<td>Discovers electric current creates magnetism</td>
<td>Hans Ørsted</td>
<td>Denmark</td>
</tr>
<tr>
<td>1837</td>
<td>Invents electric telegraph</td>
<td>Samuel Morse</td>
<td>USA</td>
</tr>
</tbody>
</table>

Before the battery, electricity was available from lightning and by shuffling across rugs on dry days, but that static electricity could not provide a steady, even flow. Volta’s battery made it possible to create a steady electrical current. Twenty years later, Ørsted discovered that a steady current of electricity produced a magnetic effect. Later, Morse connected a battery, a switch, and a coil so that closing the switch at one end of the wire would create magnetism that would attract a piece of metal at the other end of the wire. By closing the switch for short and long durations, he was able to transmit bits a long distance. This system, the electric telegraph, made long-distance, all-weather, private communication possible. This technology led to the Internet.

2. Headers and Data

Say you have a telegraph, and you want to send an image of the letter T to a friend back home. Using 1-bit color, you can represent each pixel in the image as a 1 or 0. Here is the image and the binary representation:

```
<table>
<thead>
<tr>
<th>BINARY</th>
<th>IMAGE HEADER</th>
<th>IMAGE DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000001</td>
<td>111010010</td>
<td>111010010</td>
</tr>
</tbody>
</table>
```

The 9 bits of data represent the 9 pixels, but you need to send more than just 9 bits to the friend back home. How will that person know what to do with the 9 bits? You need to specify the dimensions of the image and the color depth (i.e. the number of bits per pixel). Therefore you have to send a header and then the data. The header explains what to do with the data.

Got that? The header explains what to do with the data. Simply sending a bunch of data bits to someone is not enough. You also have to include some bits that describe tell the receiver what you want done with the data bits.

The Internet, as well as most digital forms of information, relies heavily on this idea of headers and data. In fact, the idea of headers to explain what to do with bits is the coding system for the Internet. In the big picture of problem → procedure → code, The procedure is how to get data from place to place, and the code is the header. Now we move from the era of the telegraph to the modern day.
3. The Internet - A System for Sending Bits Long Distances

Information can be represented as lists of numbers. Numbers can be represented as lists of 1’s and 0’s. When you send or receive data over the Internet, you send or receive sequences of 1’s and 0’s. How are 1’s and 0’s sent over a wire, and how does information get from one computer to another? Here is a concrete example:

Examine this picture carefully, we shall discuss its contents in detail. In the upper right corner is a picture of you sitting at your desk IMing person Y, at the same time, viewing an image in some web page. The IM text is arriving simultaneously with the image appearing in the browser.

The text you are receiving from person Y (who lives downstairs in your building) is being typed on a keyboard, and the image you are viewing in a web page, though, is stored as magnetism on a disk on a machine at the lower middle of the picture.

How does this all work? In particular,

a. How do you receive instant messages and an image at the same time? and
b. How does the sequence of bits get to you from these remote machines? and
c. While we’re at it, how to all the other connections work?

Today, we examine and answer those questions.

4. Connecting to Nearby Machines: Local Area Networks

First, how does data get from one computer to a nearby computer? One technique for sending data between nearby computers is a technology called ethernet. When you plug a network cable from your computer into a jack in the wall in your dorm, you are usually using Ethernet. The principle of ethernet is the party line, developed in earlier days of the telephone network.

All computers on the ethernet cable share one wire, so anything one says is heard by all other computers on the wire. Just as all the telephones in your house share one wire, when you pick up a telephone, you can speak to someone else in the house who picks up a telephone.

Therefore, when the computer being used by person X wants to send a message to the computer used by person Y, that computer sends a message over the wire that consists of two parts.

HEADER: address of computer used by Y
DATA: data to send

More precisely, the entire message is a sequence of 1’s and 0’s expressed as pulses of electricity or light. So a message might look like:
Here, spaces are included for readability only, the actual bits sent over the wire do not have spaces. This message has 48 bits of machine address, 4 bits of message length, and 16 bits of ASCII data. The message is "HI".

**Computer Addresses: Each Network Card has an ID**

Every computer network card has a built-in ID number. This number is called the MAC address (for Media Access Control address). The circuitry of the computer has that address wired into it. The card is capable of recognizing messages containing its own MAC address. A MAC address is to a network card what a Social Security number is to humans -- it never changes and is unique for each person.

**Headers and Data: Envelopes containing Envelopes**

Notice how this string of 1’s and 0’s consists of two parts. The first 48 bits are the address of the recipient. The next 20 bits are the message. But that message has two parts: the length and then the text. So there are really two headers. The first header is the header for the entire message. This header specifies which machine should receive the message. The data for that message consists of two parts: the length and the data. In this message, the length describes the data, and the rest is the text of the data.

This simplified model shows how ethernet works. A sender picks up its connection to the party line, states the address of the machine to send to, then states the message. All machines on the wire hear the entire message, but only the recipient bothers to listen to anything more than the address part of the packet.

**Sharing a Wire: Collision Detection and Retries**

What happens if two computers try to send messages at the same time? Like two people at a dinner party beginning to speak at the same time, both stop, wait a short amount of time, then try again. The one who tries again first gets to finish the statement. The other person can get a chance later.

**Summary: Local Area Networks**

A Local Area Network (aka a LAN) connects several nearby computers in such a way that any computer on the wire can speak directly to any other computer on the wire. Other computers on the wire may listen, but the important idea is that any computer on the local network can send data directly to any other one. They communicate by sending a string of 1’s and 0’s that consists of a header part (includes the recipient’s address) and then the actual data.

Think of a letter in the postal system. The letter has an envelope that contains the address, and the letter has contents that contains the message. Of course the letter can have an introduction and then its contents.

5. **But How Does the Internet Work?**

So far, we have seen how a LAN works: one computer sends a message directly to another computer. Each machine has a unique MAC address, so any machine on the wire can send a message to any other machine by preceding the data with the address of the destination machine.

We return now to the big picture at the start of this document. We can understand how people X and Y can use computers to send messages back and forth.

*But how can person X view an image file stored on another network?*

The answer consists of two simple ideas

**FACT ONE:** Routers

**FACT TWO:** Universal Addressing

**Routers**

A router is a computer that is connected to two or more networks. On the diagram, draw a line from the machine marked R1 to the LAN in the middle. In physical terms, one puts a second network card into that machine and has two wires running from the computer -- one wire runs to a jack for the middle LAN and the other wire runs to the LAN on the right.
Now, if the web server at the bottom of the diagram wants to send an image file to person X, the web server can send the data directly to R1 and R1 can then send the data directly to person X's computer.

This two-trip process is similar to taking a subway ride from a station on one line to a station on a different line. A subway or bus line is like a LAN -- you can get from any station on the line to any other station in one trip. But if you want to travel from a station on one line to a station on a different line, you need to make one trip to a common station, then take a second trip to the destination. Of course, in some cases, you need to change lines more than once. A trip from your dorm to your home may involve taking a bus to a train, a train to a different train, a bus to a plane, a plane to a hub, another plane to your home city, then a train to your neighborhood, then a bus to your house. In each case, you are on a single LAN, but the entire trip involves changing from one LAN to another LAN.

**How can person Y receive or send mail via gmail.com?**

**Add one more connection to make these three LANs into a single WAN**

**Add one more connection so there are two paths from gmail to person Y**

**A WAN is a Wide Area Network**

A network composed of several connected LANs is called a WAN -- a Wide Area Network. The world-wide WAN connecting thousands and thousands of LANs is called the Internet.

**Internet Addressing**

Now we come to a new problem: addressing. On an ethernet-based LAN, each machine has a MAC address and any other machine on the LAN can send a message by preceding the data with that MAC address. But, what about the web server that has to route data through R1 to get to person X? That web server cannot put address 00:10:A4:E7:E0:02 in the header for the image file.

**Why not?**

**What address does the web server put in the message?**

**What is the problem with that?**

With several networks, each with several machines, we need a way to specify an address for each machine. The answer is to assign an ID number to each network and then an ID number to each machine on each network.

**Assign an ID number (in range 1-127) to each network shown in the picture.**

**Assign an ID number to each computer on each network.**

An internet address (aka IP address) is a pair of numbers of the form:

```
network_number.machine_number
```

An example of an IP address is 10.0.0.1 which is network 10, computer 0.0.1. And IP address 12.0.20.3 is network 12, computer 0.20.3. †

**Given your numbering system, what is the IP address of X’s computer?**

**What IP address is assigned to the web server computer?**

**What IP address is the gmail computer?**

**Internet Routing**

Now we can understand how the webserver gets the image file to X’s computer. The webserver creates a packet that looks like:

```
00:14:4F:4F:D2:13 netID.comPID length img-info img-data
```

When the web server broadcasts the packet on its LAN, machine R1 receives it. That machine looks reads the next

† The IP addressing system is a little bit complicated. Some networks get longer numbers but make up for it with shorter computer numbers. For example 130.64.1.83 is network number 130.64 with computer number 1.83. Some network numbers use three parts leaving only one part for the computer number, so 220.110.95.94 is network 220.110.95 computer 94. The rules are that network numbers 1-127 use one part of the four-part address leaving three parts for computer ID, numbers 128-191 use two parts for the network and two parts for the computer, and numbers 192-255 use three parts leaving one part for the computer ID.
part of the header to see the ultimate destination of the packet. It sees that the destination is on the rightmost LAN
so it repackages the packet so the packet now looks like:

00:14:A4:E7:E0:02 netID.compID length img-info img-data

and broadcasts this message on the party line that is the rightmost network. The computer X is using recognizes its
MAC address and reads the packet.

That computer then inspects the IP address of the destination. That computer sees the packet is addressed to it, so it
then reads the image file data and displays the picture.

[Explain how a packet gets from gmail.com to person Y.]
[Explain how a packet gets from Y’s computer to X’s computer]

**Summary of IP Routing**

The solution to the problem of getting information from a computer on one LAN to a computer on another LAN is:

a. Assign ID numbers to every network
b. Assign ID numbers to every computer on each network
c. Every computer knows its own network and computer number
d. Every computer knows the MAC addresses of computers on its LAN
e. When a computer wants to send a message to another computer
   1. It examines the IP address of the destination
   2. If the destination is on the same LAN, it sends the packet directly
   3. Otherwise it sends the packet to a router on its LAN
   4. The router repeats the process

Therefore, like a traveler going from bus line to train line to train line to plane to train to bus to train, a packet of
data travels from LAN to LAN by changing at routers. But unlike a traveler, the packet does not set out with a route
in mind. Instead, the packet arrives at each router, and the router decides where to send the packet next.

**6. Problem: Numbers are Hard to Remember; Solution: DNS**

But knowing the network ID and computer ID of the gmail server is not so easy. That’s a lot of numbers to remem-
ber. It’s easier to remember a word like gmail.com. So the Internet has created a system called DNS (domain name
system) that is a directory assistance service. Just as you can dial 411 and ask for the number for John Smith, your
computer calls a DNS server and asks for the address of www.gmail.com. The DNS server sends back an IP
address. Your computer can now address the packet. You never have to know the IP address of gmail, just its name.

**7. Problems: Big Files and Lost Packets; Solution: TCP**

IP is amazing technology. Every computer on the Internet has a unique IP address consisting of a network number
and a computer number. DNS attaches readable names to numbers. Each router forwards packets from one LAN to
another. Each packet contains the IP address of its sender and its destination. On each leg of the journey, the packet
gets a different ethernet header to get the packet to the next router, but the content of the packet stays the same.

But there are two problems. First, IP routers can get really busy sometimes. In that case, they may just throw away
packets it is too busy to process. How can you be sure your email message or image file gets to the destination?

Second, in the 1750 fax system, we saw round-robin packet switching is an effective way for several people to trans-
mits data over one line. We have not explained how round-robin scheduling works in the Internet.

The solution is as follows:

a. The sender splits a big file (over 1500 bytes) into ‘pages’
b. Each page is numbered
c. Each is marked “return receipt requested”
d. Each page is put in its own IP packet and sent off
e. The receiver sends back an acknowledgment for each packet
f. The receiver reassembles the pages into a complete document
This procedure is called the *transmission control protocol* and is known by the acronym TCP. This combination of splitting a large file into smaller pages and requesting confirmation of arrival solves the problem of sharing one wire among many users and solves the problem of lost packets.

Therefore, a packet containing one `page` in a long file has the structure of:

```
00:14:4F:4F:D2:13  netID.compID length page#  RRRQ  img-info  img-data
```

The first part is the ethernet address, used to get the packet to the next stop on the route. The next part is the IP address of the destination, used to figure out where to go next. The page number is used to reassemble the pages upon arrival, the return receipt request tells the receiver to send back confirmation, and finally the img-info tells the browser the dimensions of the image, the color depth, and other useful information for displaying the bits stored in the image data.

8. Summary: TCP/IP and Ethernet

*Ethernet* is a system for transferring packets directly from one computer to another on a LAN. There are other systems used for LANs. We shall not discuss those here. You can read about how WiFi works, another popular system for local networks.

*Internet Protocol* (IP) is a system for addressing machines in terms of the network and machine ID and a system for routing packets from LAN to LAN.

*Transmission Control Protocol* (TCP) is the system of breaking large files into smaller packets, using round-robin scheduling to share a wire, and using return receipts to make sure lost packets are retransmitted.

9. Summary Diagrams

This collection of images summarizes some of what we have looked at and fill in some details we did not discuss. The diagrams on the next pages show how packets can travel from one network to another.

An image file has a heading that contains the dimensions and depth of the image followed by the pixels as a sequence of bits. If the image file is more than about 1500 bytes, it is split into several packets. The first packet in the collection might look like the following.

The first part gets the packet to the router on the local network. The next part is read by the routers to tell them to what network and machine on the Internet the packet should be sent.

The TCP part tells the receiving machine which part of the file this packet contains and the mailbox number (port) to which the file should be delivered. The TCP part also asks for a confirmation that the destination machine received the packet.

```
Each Section is a sequence of Bits
```

- EP HDR
- IP HDR
- TCP HDR
- DNS
- BITMAP

Complete diagram of TCP and IP headers:
Computer Networks

- A local area network (LAN) is a collection of computers connected by a common wire. Each LAN is usually contained within a single building or floor.

- A wide area network (WAN) is a collection of LANs where:
  - a group of LANs are connected by routers, which are computers that perform special network services, and
  - there is a path from every LAN in the WAN to every other LAN, possibly by passing through intermediary LANs.

- The Internet is a WAN.

![Diagram of computer networks]

Ethernet Protocol

All computers monitor the LAN at all times. They all detect that a packet is being put onto the LAN and they all listen. After the hardware address of the recipient, D, is broadcast, only B and D continue listening. The other computers wait for a new packet to be broadcast.

<table>
<thead>
<tr>
<th>Ethernet Header</th>
<th>Ethernet Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>D's hardware address</td>
<td>B's hardware address</td>
</tr>
<tr>
<td>Size</td>
<td>Checksum</td>
</tr>
</tbody>
</table>

Then, B and D each read the entire packet.
Internet Protocol (IP)

*This is a WAN composed of two LANs and a router.
*Computer A wishes to send data to computer C.

<table>
<thead>
<tr>
<th>Ethernet Header</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware address of B</td>
</tr>
<tr>
<td>Hardware address of A</td>
</tr>
<tr>
<td>Size of ethernet packet</td>
</tr>
<tr>
<td>Checksum</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ethernet Packet Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP Header</td>
</tr>
<tr>
<td>IP address of C</td>
</tr>
<tr>
<td>IP address of A</td>
</tr>
<tr>
<td>IP version (4)</td>
</tr>
<tr>
<td>Size of packet</td>
</tr>
<tr>
<td>Checksum</td>
</tr>
</tbody>
</table>

| IP Data for C |

Every computer on the Internet has its own unique IP address.

A prepares an ethernet packet whose data is an IP packet, and sends it to router B.
Then, B receives the ethernet packet and strips out the IP packet.
B sees that the recipient's IP address must be passed onto LAN 2.
B creates a second ethernet packet from B to C whose data is the original IP packet.

Finally, C receives the second ethernet packet, strips out the IP packet, sees that C itself is the recipient of the IP packet, and processes its data.

Transmission Control Protocol (TCP)

- Computer A wants to send data to computer B.
- Sender & receiver first establish a connection, or "exchange.
- The data is broken up into some number of pieces.
- Each piece is sent as a separate TCP packet.
- As B receives each packet, it sends an acknowledgement to A if the data passes the checksum test, and requests a resend if not.
- If A does not get an acknowledgement from B for a given packet, it resends the packet.
- B assembles the data and makes it available to its own processes as soon as it is available.