

The Internet

Tim Capes

November 7, 2011

What is the Internet?

The internet is a global system consisting of millions of interconnected networks. These individual networks are anything from local (a Home Network) to global (some academic, business and government networks).

How does the internet work?

The internet is built out of protocols each of which governs the behavior of different tasks. The most important of these is the internet protocol suite, which is the basis of how to find where information is located. Some well known protocols are:

How does the internet work?

The internet is built out of protocols each of which governs the behavior of different tasks. The most important of these is the internet protocol suite, which is the basis of how to find where information is located. Some well known protocols are:

- ▶ VoIP: Voice over internet protocol

How does the internet work?

The internet is built out of protocols each of which governs the behavior of different tasks. The most important of these is the internet protocol suite, which is the basis of how to find where information is located. Some well known protocols are:

- ▶ VoIP: Voice over internet protocol
- ▶ HTTP: Hyper-Text Transfer Protocol

How does the internet work?

The internet is built out of protocols each of which governs the behavior of different tasks. The most important of these is the internet protocol suite, which is the basis of how to find where information is located. Some well known protocols are:

- ▶ VoIP: Voice over internet protocol
- ▶ HTTP: Hyper-Text Transfer Protocol
- ▶ FTP: File Transfer Protocol

How does the internet work?

The internet is built out of protocols each of which governs the behavior of different tasks. The most important of these is the internet protocol suite, which is the basis of how to find where information is located. Some well known protocols are:

- ▶ VoIP: Voice over internet protocol
- ▶ HTTP: Hyper-Text Transfer Protocol
- ▶ FTP: File Transfer Protocol
- ▶ POP3: Post-Office Protocol version 3 (e-mail)

How does the internet work?

The internet is built out of protocols each of which governs the behavior of different tasks. The most important of these is the internet protocol suite, which is the basis of how to find where information is located. Some well known protocols are:

- ▶ VoIP: Voice over internet protocol
- ▶ HTTP: Hyper-Text Transfer Protocol
- ▶ FTP: File Transfer Protocol
- ▶ POP3: Post-Office Protocol version 3 (e-mail)
- ▶ SSH: Secure Shell (remote access)

The internet suite

The fundamental problem of a big network is figuring out where everything you want is and how to get at that material. For the internet this is accomplished by the internet protocol suite (IPS or TCP/IP). This protocol suite is made up of two protocols:

The internet suite

The fundamental problem of a big network is figuring out where everything you want is and how to get at that material. For the internet this is accomplished by the internet protocol suite (IPS or TCP/IP). This protocol suite is made up of two protocols:

- ▶ TCP: Transmission Control Protocol

The internet suite

The fundamental problem of a big network is figuring out where everything you want is and how to get at that material. For the internet this is accomplished by the internet protocol suite (IPS or TCP/IP). This protocol suite is made up of two protocols:

- ▶ TCP: Transmission Control Protocol
- ▶ IP: Internet Protocol

Internet Protocol

The problem of where everything is handled by the Internet Protocol. This protocol is a system of addressing the entire internet. There are several versions of this protocol.

IPv4 is the current version which is in use. Addresses consist of four numbers between 0 and 255 separated by dots. For example 127.0.0.1 is a special address reserved for local communications. For single computer networks it will represent the local computer regardless of what computer it is on. For non-reserved addresses it is typically the case that the following convention applies:

IPv4 is the current version which is in use. Addresses consist of four numbers between 0 and 255 separated by dots. For example 127.0.0.1 is a special address reserved for local communications. For single computer networks it will represent the local computer regardless of what computer it is on. For non-reserved addresses it is typically the case that the following convention applies:

- ▶ The first three digits are a country code (digits before the first dot). This gives information about what country the network address is from.

IPv4 is the current version which is in use. Addresses consist of four numbers between 0 and 255 separated by dots. For example 127.0.0.1 is a special address reserved for local communications. For single computer networks it will represent the local computer regardless of what computer it is on. For non-reserved addresses it is typically the case that the following convention applies:

- ▶ The first three digits are a country code (digits before the first dot). This gives information about what country the network address is from.
- ▶ The next digits are typically region digits.

IPv4 is the current version which is in use. Addresses consist of four numbers between 0 and 255 separated by dots. For example 127.0.0.1 is a special address reserved for local communications. For single computer networks it will represent the local computer regardless of what computer it is on. For non-reserved addresses it is typically the case that the following convention applies:

- ▶ The first three digits are a country code (digits before the first dot). This gives information about what country the network address is from.
- ▶ The next digits are typically region digits.
- ▶ The next three digits indicate information about the area network.

IPv4 is the current version which is in use. Addresses consist of four numbers between 0 and 255 separated by dots. For example 127.0.0.1 is a special address reserved for local communications. For single computer networks it will represent the local computer regardless of what computer it is on. For non-reserved addresses it is typically the case that the following convention applies:

- ▶ The first three digits are a country code (digits before the first dot). This gives information about what country the network address is from.
- ▶ The next digits are typically region digits.
- ▶ The next three digits indicate information about the area network.
- ▶ The last three digits are often local information (refer to a specific local network on the network). Note that historically this typically meant a specific computer, but in most cases these days it's the hub of local network (on which several computers may use the same IP to connect to the internet)

Internet Protocol v6

There is a problem where version 4 of the protocol is actually running out of addresses. There are only 4,294,967,296 addresses. This seems like a lot at first, but millions of them are reserved for specific tasks, and even without those millions this number is smaller than the population of the earth.

Furthermore, each website server needs an address. IPv6 uses far more digits in an address, which allows for 2^{128} possible addresses. In base 10 that's approximately $3.4 * 10^{38}$ or 340 undecillion. The way IPv6 works is that it uses 64 bits for the MAC address of the hardware, and has 64 bits in reserve or padding to create new addresses for future needs.

Transmission Control Protocol

This protocol allows for the transmission of larger chunks of data. IP handles specific requests (typically small text instructions), and breaking up all data into chunks that small is a task that would be very difficult to implement non-automatically in every single program using the internet. Instead larger requests are handled by the TCP. TCP will do this breaking into chunks and use IP to send all the tiny pieces. It is a protocol that is flexible and can handle lost packets, which makes it useful for most applications (if a protocol handles lost packets poorly there are potential serious problems with missing data).

How it works for the web

- ▶ A hyperlink is a method of accessing data remotely: a web address represents the IP address of the website server.

How it works for the web

- ▶ A hyperlink is a method of accessing data remotely: a web address represents the IP address of the website server.
- ▶ The slashes in the address indicate directory changes on that machine.

How it works for the web

- ▶ A hyperlink is a method of accessing data remotely: a web address represents the IP address of the website server.
- ▶ The slashes in the address indicate directory changes on that machine.
- ▶ For example, en.wikipedia.org corresponds to an IP address for the wikipedia server.

The Web vs The Internet

Lots of technical distinctions are often glossed over or ignored in the media and general public. The internet and the world wide web (www) are two very different things.

The Web vs The Internet

Lots of technical distinctions are often glossed over or ignored in the media and general public. The internet and the world wide web (www) are two very different things. The internet is a worldwide network as described before.

The Web vs The Internet

Lots of technical distinctions are often glossed over or ignored in the media and general public. The internet and the world wide web (www) are two very different things. The internet is a worldwide network as described before. www is a system of addressing that was developed for the use of HTML webpages.

The Web vs The Internet

Lots of technical distinctions are often glossed over or ignored in the media and general public. The internet and the world wide web (www) are two very different things. The internet is a worldwide network as described before. www is a system of addressing that was developed for the use of HTML webpages. Basically, the distinction is that The Web is just the part you browse with a web-browser and the internet is the entire network and everything else done on it.

The World Wide Web

First proposed in 1989, this was developed in 1991 by Tim Berners Lee at CERN.

The World Wide Web

First proposed in 1989, this was developed in 1991 by Tim Berners Lee at CERN. The first standards were developed in 1994 as a free and open technology.

The World Wide Web

First proposed in 1989, this was developed in 1991 by Tim Berners Lee at CERN. The first standards were developed in 1994 as a free and open technology. This is important as it meant that one could adopt standards for the web on any site without having to pay royalties. If one needed to pay royalties to be able to use an addressing system for the internet, the commercialization of the internet that followed this technology would have been impossible.

Content for the Web

Webpages are written in a markup-language called hyper-text markup language (HTML).

Content for the Web

Webpages are written in a markup-language called hyper-text markup language (HTML). We will cover this in depth later.

Connections to UI

User interface is everywhere, and here there is no exception.

1. The underlying language is text-based.

Connections to UI

User interface is everywhere, and here there is no exception.

1. The underlying language is text-based.
2. There are graphical editors that enable one to do this visually and generate the text based on the appearance of the page.

Connections to UI

User interface is everywhere, and here there is no exception.

1. The underlying language is text-based.
2. There are graphical editors that enable one to do this visually and generate the text based on the appearance of the page.
3. The graphical editors are easier to use, but often generate code that is much worse at handling different browser types, screen resolutions or other format changes.

Philosophy of HTML

The basic idea of HTML is to be a flexible language which specifies the pertinent details and emphasis of the document (the webpage) but allows the users (and/or browsers) custom settings to be applied.

Philosophy of HTML

The basic idea of HTML is to be a flexible language which specifies the pertinent details and emphasis of the document (the webpage) but allows the users (and/or browsers) custom settings to be applied. So for example if you have a font preference you can see text in that font without the web-page designer imposing their font on you. If you have a resolution setting the webpage will be shown in that setting (although some webpages suggest using specific settings).

CSS

The problem with this philosophy is that it is rather hard to allow for sleek design. Designers spend far too much time micro-managing settings, sizes, locations and other factors to try and present a nice UI. Cascading Style Sheets (CSS) are a solution to this problem. Without getting into much detail, you can think of them as a way to classify each object on a page, and associate default settings to object classes in a nice way. Often saves thousands of lines of code on sleek webpages.

HTML5 and Mobile Development

Another problem is that it's hard to design something that scales incredibly well. HTML5 is designed to be able to write pages that will rescale well to mobile browsers, touchpads, full laptops, or even desktop screens.

HTTPS

Getting away from HTML and into transferring information.
HyperText Transfer Protocol (HTTP) allows for the efficient transfer of information.

HTTPS

Getting away from HTML and into transferring information. HyperText Transfer Protocol (HTTP) allows for the efficient transfer of information. However, it does not do so securely. What if the information being exchanged is a password? Other sensitive data?

HTTPS

Getting away from HTML and into transferring information. HyperText Transfer Protocol (HTTP) allows for the efficient transfer of information. However, it does not do so securely. What if the information being exchanged is a password? Other sensitive data? Need a secure connection. The S in HTTPS adds encryption and stands for Secure.

RSA

RSA is a remarkable security algorithm named for its inventors: Rivest, Shamir, Aldeman.

1. The key needs to be established publically.. anyone can listen to you pass this information.

RSA

RSA is a remarkable security algorithm named for its inventors: Rivest, Shamir, Aldeman.

1. The key needs to be established publically.. anyone can listen to you pass this information.
2. The basis of the security technique is:

RSA

RSA is a remarkable security algorithm named for its inventors: Rivest, Shamir, Aldeman.

1. The key needs to be established publically.. anyone can listen to you pass this information.
2. The basis of the security technique is:
 - 2.1 It is easy to multiply prime numbers to make a large number.

RSA

RSA is a remarkable security algorithm named for its inventors:
Rivest, Shamir, Aldeman.

1. The key needs to be established publically.. anyone can listen to you pass this information.
2. The basis of the security technique is:
 - 2.1 It is easy to multiply prime numbers to make a large number.
 - 2.2 It is hard to guess which primes make up that number.

RSA

RSA is a remarkable security algorithm named for its inventors:
Rivest, Shamir, Aldeman.

1. The key needs to be established publically.. anyone can listen to you pass this information.
2. The basis of the security technique is:
 - 2.1 It is easy to multiply prime numbers to make a large number.
 - 2.2 It is hard to guess which primes make up that number.
3. Pass the large number publically, and use the set of primes to decrypt.

RSA

RSA is a remarkable security algorithm named for its inventors: Rivest, Shamir, Aldeman.

1. The key needs to be established publically.. anyone can listen to you pass this information.
2. The basis of the security technique is:
 - 2.1 It is easy to multiply prime numbers to make a large number.
 - 2.2 It is hard to guess which primes make up that number.
3. Pass the large number publically, and use the set of primes to decrypt.

This is a simplification. The algorithm itself involves some rather advanced mathematics