

# How The Internet Works:

## IP Addresses, URLs, Domain Names – What Do They Mean? Part 1

Source: High Technology Crime Investigation Association (htcia.org) Newsletter, March 2002, Vol. 3 issue 1, page 3

*Contributed by Mike Menz*

*IT Security Investigator, Hewlett Packard Company*

*Member of the Northern California Chapter*

*Co-written by Assistant District Attorney Stephen Treglia*

*Chief, Technology Crime Unit, Nassau County, New York District Attorney's Office*

*Member of the Northeast Chapter*

An issue that arises from time to time is the reading of a “URL” (short for “Uniform Resource Locator”) and how a URL is distinguishable from an Internet Protocol address (“IP address”) or a domain name. In this two-part article we’ll define the three terms, breakdown a URL into its component parts, and even throw you a few curve balls. (It is spring training time, after all.)

Don’t sweat it. This is intended to be a very basic-level introduction, so the curves are only of the Little League variety. This edition’s article will introduce you to the meaning of an IP address and explain how it differs from a URL.

## **IP Addresses (Internet Protocol Addresses):**

Every computer, when connected to the Internet, is assigned an IP address. When people typically refer to an IP address they mean a series of numbers broken into four groups with no group exceeding the number 255. Each of the four groups of numbers is separated by a period. For example, this is how an IP address might appear:

211.5.172.46

Each group of up to three digits is called an “octet”. How can a three-digit number not exceeding 255 be called an octet? You need to recall some basic algebra.

Each of the 4 octets in an IP address can be any number between 0 to 255, meaning there can be a total of 256 possibilities in each octet. So where did they come up with the term “octet” (which means a “group of eight”)? Well, if you take the number 2 and times it by itself 8 times ( $2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2$ ), you get the number 256.

How many IP addresses are there? Again, it’s simple high school algebra. Take 256 and multiply it by itself 4 times ( $256 \times 256 \times 256 \times 256$ ) and you have over 4.2 billion IP addresses.

Sounds like a lot of addresses right? Well, not really. Each time a computer connects to the Internet there are several IP addresses that must get used in the process. (You’ll just have to take our word on this one. We don’t want to get too technical in this article.)

Discussions occurring in techie circles have been warning about the fact that we will run out of available IP addresses for people to connect to the Internet in less than a decade. Hence, a new IP address structure is already in the works (called “IPv6” as opposed to “IPv4” which is what techies call the present system). Again, we won’t go into any more detail about this other than to mention it to you since you’ll probably be hearing a lot more about it in the years ahead.

Earlier in this article we mentioned how each computer connects to the Internet and is assigned an IP address. Don't fall into a very common trap here. An IP address is not like a telephone number. Each telephone number is assigned to a definite telephone connection and instrument (at least until you forget to pay your bill). Just because you sign on to the Internet and are assigned an IP address, your computer does not permanently keep that IP address.

If you use a telephone modem, or most cable modems, DSL or ADSL lines, you are assigned what is called a "dynamic IP address," which means it changes each time you attach to the Internet. The only computers with "static IP addresses" (meaning they always stay the same) are those typically found in large businesses and universities.

## URLs (Uniform Resource Locator):

Computers don't care that there can be more than 4 billion IP addresses in purely numeric form floating around in cyberspace. We poor humans, on the other hand, don't work as well remembering the difference between 211.5.172.46 and 211.5.127.46.

That's why some clever techie (are there uncles techies?) came up with the idea of translating a numeric IP address into recognizable words and syllables for us homo sapiens to try to remember. Of course, this clever techie would not allow any old translation do (although a computer wouldn't have cared). Instead it has to be a Uniform Resource Locator system (which you may occasionally see called the "Universal Resource Locator" system as it was originally called), meaning there are protocols and rules to follow.

A typical URL might look something like this (we're using a Hewlett-Packard URL here, we hope no one there is offended by our borrowing it):

<http://www.hp.com.us/public/info.html>

As probably most of you know, only a certain portion of the URL shown above comprises the numerical IP address. The "www.hp.com.us" is the human-friendly translation of the purely numerical corresponding IP address.

So you may be wondering how does anyone know the corresponding IP address for each and every URL. Well, somewhere on the Internet (actually, it's located in several places) a really clever techie has posted the translation of each IP address that has been converted into a URL. (And you thought you worked in a place with low job recognition; how would you like that job?)

But you know no clever techie would let it rest there. It is not enough that computers on the Internet can know what the component parts of a URL mean, this techie insisted that each component part of a URL actually has to make sense to humans as well.

Before tackling that task of breaking down each part of the complete URL and explaining what each component means, however, we'll take a rest until the next newsletter, at which time we'll complete this story.

## How The Internet Works:

### IP Addresses, URLs, Domain Names – What Do They Mean? Part 2

Source: High Technology Crime Investigation Association (htcia.org) Newsletter, June 2002, Vol. 3 issue 2, page 2

In our last newsletter we demonstrated the meaning and significance of an Internet Protocol address (“IP address”). We also explained how an IP address forms the basis for a “Uniform Resource Locator” (“URL”), which is the most well-known method of locating a website.

In this conclusion to this two-part series, we will begin by breaking a URL down into its component parts and analyzing the significance of each part. We will also introduce you to the concept behind the term “domain name”.

## URLs:

If you recall, we ended the last newsletter by showing how the purely numeric IP address translates into a URL. We even borrowed, with the kind indulgence of Hewlett-Packard, one of their URLs for demonstration purposes.

<http://www.hp.com.us/public/info.html>

Now let’s take you, piece-by-piece, through each section of this sample URL and show you how there really is a method to this madness. Here’s a short-hand explanation of what each part of the sample URL actually represents:

Access protocol://third-level domain.second-level domain.first-level domain.country code /  
directory/filename

Here’s our explanation for each component part of the URL:

Access protocol (in our example: http)

You probably are familiar with the access protocol known as “http,” which stands for “Hypertext Transfer Protocol”. This is the protocol (meaning the set of technical rules and procedures) that is used to transfer data you requested over the World Wide Web.

Only certain formats of data can be transferred over the Web. These formats include “Hypertext Markup Language” (“HTML”), “Extensible Markup Language” (“XML”), “Java applets” and certain image formats such as “Joint Photographic Experts Group” (“JPEG”) or “Motion Picture Experts Group” (“MPEG”).

As most of you know the Internet consists of much more than just the Web. For example, some of the more familiar “other protocols” are “File Transfer Protocol” (“FTP”), “gopher”, or “Wide Area Information System” (“WAIS”). Some of the less common access protocols are: “news”, “telnet” and “mailto”.

Hence, you will not always see the letters “http” at the start of every URL.

One or two of you might be wondering: “Why do I occasionally see a colon sign and number (like ‘:80’) following the access part of the URL?” The “:80” is an instruction to the computer attempting to access another computer connected to the Internet what specific port of the computer to use when accessing the information

stored on it. (What's a port? No, not an expensive sherry. Think of a spaceship docking into a "port" on a space station and you've probably got the necessary visual imagery of what a computer is doing when accessing information from another computer on the Internet.)

Why don't you see a colon and port number more frequently in a URL? Because all computers on the Internet know when the URL doesn't mention a specific port, "80" is the default port to be used to gain access.

### First-, Second- and Third- level domain names (www.hp.com)

The part of the URL you've probably heard the most about is called the "domain name." That's because a domain name has become part of the commercial aspect of the Internet or (if you're the less cynical type) the more "user friendly" part of a URL. This is the part of the URL you immediately think of when you access a familiar site on the Web.

A domain name, technically, consists of three levels and the sequence of these three levels should properly be considered as you go from right to left. In the domain name "www.hp.com", "com" marks the first-level domain, "hp" is the second-level and "www" makes up the third-level.

Why is it called a "domain" name? The explanation is really not that far-fetched. Think of a domain in its most common terms, that is, a realm. The ".com" domain comprises the "commercial" realm of domain names, the ".gov" comprises the "government" realm of domain names and so on.

Until recently, there were only seven first-level domain names. Besides .com and .gov there were: ".edu" (for educational organizations), ".int" (for international organizations), ".mil" (for military organizations), ".net" (for networking organizations) and ".org" (for miscellaneous organizations). The categories of first-level domain names have just been expanded to include seven more: ".aero" (for the "air-transport industry"), ".biz" (for businesses), ".coop" (for cooperatives), ".info" (unrestricted), ".museum" (for museums), ".name" (for individuals), and ".pro" (for accountants, lawyers and doctors).

In more technical terms, think of the various levels of domain names as you would think of directories stored in a computer's memory. The "first-level" domain is similar to the "parent directory". It represents the highest level in a directory tree from which all subdirectories flow.

The "second-level domain" (the "hp" in the example provided in this article) is the next level of subdirectories below the first-level domain. So, "hp" is a subdirectory under the ".com" first-level domain, sharing the .com first-level domain with all the other ".coms" in the world.

The first and second-level domains in a URL really comprise the numerical "IP address", so those are the only URL components necessary to gain access to a computer on the Web. The purpose behind the presence of the third-level domain is a bit too technical for this basic-level article, but it should be noted that it is often superfluous, meaning many Web sites can be accessed without even typing in the third-level domain, e.g., "hp.com" is the same as "www.hp.com".

A domain name must be registered with one of the registrar organizations accredited by the Internet Corporation for Assigned Names and Numbers ("ICANN"). Obviously, there's a registration fee involved which can last from one to several years, depending on the registrar organization.

## Country codes (us)

Okay, we've thrown you plenty of Little League curves in this article, but here's a legitimate major-league knuckleball. This one can make you cross-eyed if you're not careful.

Why do some URLs have a country code and some don't? Technically, all URLs have a country code, but you don't always have to type one in to get where you want to go.

If you're accessing the Internet from the United States, you never have to type in the letters "us" to get to an American Web site. The "us" country code is understood. The same practice exists in other countries. If you're in England you don't have to type "uk" to get to an English Web site.

What can get you a little cross-eyed at times is when you type in "www.hp.com" in different countries, you are actually being taken to different Web sites. If you're in the United States, you're taken to the American site; if in England you're taken to that site. If you're in America and want to get to the English site you have to add the ".uk" or ".ca" to get the Canadian site. People outside the United States have to add ".us" to get to the American site.

## Directory/filename (public/info.html)

Once you've arrived at the computer(s) where the Web site is housed, what you are really attempting to access is something commonly called a "Web page", which essentially is nothing more than a file stored on a computer. Often a URL will provide you with the specific file name you are attempting to access rather than make you search around the main (or "home") page of the Web site.

In the URL example we previously gave you ("http://www.hp.com.us/public/info.html") you are attempting to look at the file named "info.html" which is stored in the "public" directory of the computer(s) which store data under the domain name of "www.hp.com".

## A few last curves:

Why do some URLs contain some very strange symbols? Many of them go back to the earliest days of mainframe computers which used the UNIX programming language. Certain symbols represented certain commands, words or letters and have been carried over to the Internet era.

For example, the URLs "http://www.micro-share.com" and "http://www.micro%2share.com" both refer to the same site. The %2 stands for a hyphen symbol, but we'll save that explanation for a more advanced article in a subsequent newsletter.