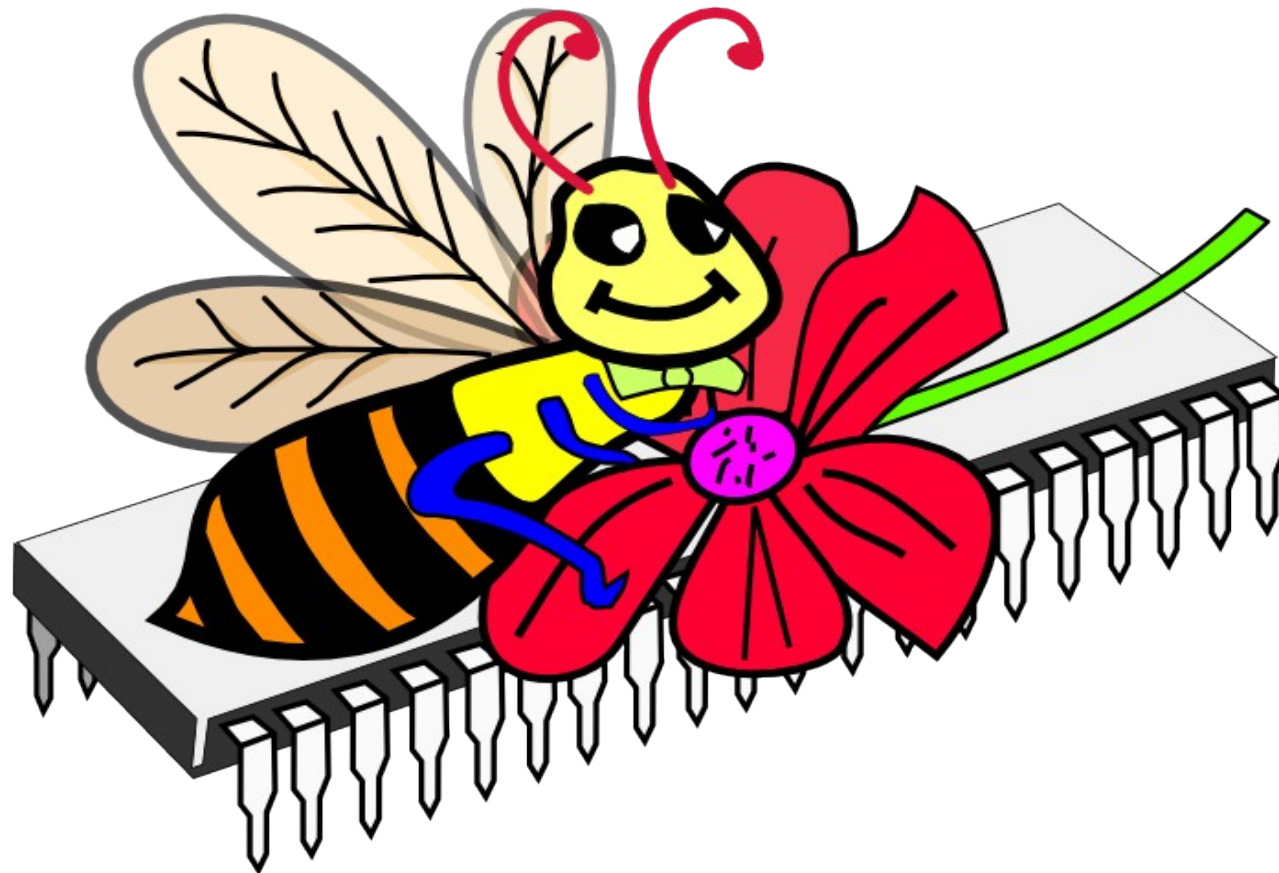


BCUG Internet Workshop 2010-03-05

Web Browsing - Under The Hood

By Gilbert Heaton < gilbert@heaton.net >

<http://www.bcug.com/>



Disclaimers

Information prepared for casual workshop on a quick effort basis and intended to supplement the presentation. Use at your own risk. Also some wording has been crunched down to fit the page/slides better.

BCUG is not a part of the college it happens to meet at.

Copyright 2010 by Gilbert Heaton under [Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported](http://creativecommons.org/licenses/by-nc-sa/3.0/)



Contents

Table of Contents

1 Contents.....	2
2 Preview.....	2
3 Overview.....	3
4 What Is Html.....	4
5 What Is Http.....	5
What Is Https.....	7
TLS Tips.....	9
2 Cookies: What They Are.....	10
3 Cookie Security.....	12
4 Tracking Cookies.....	13

1 Preview

http://ars.userfriendly.org/cartoons/?id=20070413	UF: Unlimited Yahoo E-mail
http://ars.userfriendly.org/cartoons/?id=20040815	UF: This LAN Was Made For You And Me
http://ars.userfriendly.org/cartoons/?id=20100131	UF: If the Internet was an amusement ride (this tall)
http://blogs.oreilly.com/wateringhole/2010/01/strip226.html	WH: Feedback loop

Microsoft®, Windows®, and Outlook® are registered trademarks of Microsoft Corporation.

Mac and Mac OS are trademarks of Apple Inc., registered in the U.S. and other countries.

Pine and Pico are trademarks of the University Of Washington.

2 Introduction

Surfing the web is a lot like driving a car. You don't need to know how the car works, just how to point and steer it. Yet knowing more about cars opens up more ways to drive them, protect them, and help give information to mechanics to fix them.

Computers, and “the web”, are very much the same. The more you know the better off you are. This workshop presents some more details about web browsing. It doesn't matter if you are using Internet Explorer, Fire Fox, Chrome, Opera, or one of many other browsers. While browsers provide better suspension, easier controls, better crash tests, and all sorts of optional features, when you get down to it they all travel the same electronic highway.

“You don't how to know how the computer works,
just how to work the computer”

3 What Is Html

Publishers getting ready to print something “mark up” the original manuscripts with codes to tell the typesetters (people actually preparing the galleys to print) what to do. Where to put italic type. Where paragraph breaks are. Why styles of type (font, sizes, spacing between lines) It was made as a simpler version of SGML, a mark up language used by publishers.

Hypertext Markup Language is the “mark up language” for web pages. HTML pages are strait text files, like Notepad puts out, but with HTML markup in them enclosed in <angle bracket quotes>. Use *View Source* on your browser to see.

```
<HTML>
<HEAD><TITLE>Title For Browser Titlebar</TITLE></HEAD>
<BODY><H1>Top Level Heading</H1>
<P>First Paragraph</P><P>Second Paragraph</P><P>Third
Paragraph.      Note      that any type of spaces, line endings,
tabs, or spaces, are treated as a simple space. Many spaces
in a row are squeezed into a single space.</P><P>The markup is
vital to making everything look right.</P></BODY></HTML>
```

The presence of “links” you can click on makes the “hypertext” part of web pages. Hyperlinks were a huge part in the making of the web.

4 What Is Http

Hypertext Transport Protocol not only brings web pages from the web server to your computer, it allows your web browser to tell web servers what you want.

It all starts with the web browser asking a web server for something:

Request Page

GET /SomePage.html?form=foo&field=bar HTTP/1.1

Page On Server

Host: www.healton.net

Host name

User-Agent: Mozilla/5.0 (X11; U; Linux i686;

self so sever can tune its reply to the browser.

en-US; rv:1.8.1.19) Gecko/20081216

Fedora/2.0.0.19-1.fc8 Firefox/2.0.0.19

Accept: text/xml,application/xml,

features the web browser supports.

application/xhtml+xml, text/html

Any Form Fields

text/plain;q=0.8, image/*, */* (GET method, POST method puts file

Accept-Language: en-us,en;q=0.5

language(s) users prefer replies in

Accept-Encoding: gzip,deflate

Accept-Charset: ISO-8859-1,utf-8;q=0.7,*;q=0.7

Other technical things

Keep-Alive: 300

Connection: keep-alive

Cookie: cookie-name=cookie-value

any prior cookies it received from THIS domain

Web servers reply with a heading describing the reply which is followed with the main reply.

HTTP/1.1 200 OK Reply status: 200 is normal. 404 not found

Date: Fri, 05 Mar 2010 04:38:53 GMT

Server: Apache/1.3.26 (Unix) FrontPage/5.0.2.2510

Last-Modified: Sat, 20 Feb 2010 02:38:25 GMT

Etag: "11e0f45-aea-4b7f4b21"

Accept-Ranges: bytes

Content-Length: 2794

Connection: close

Content-Type: text/html

Set-Cookie: cookie-name=cookie-value

Set-Cookie: another-cookie-name=another-cookie-value

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">

<html><head><title>Brookdale Computer Users Group Home Page</title>

<base target="_self">

<link rel="stylesheet" type="text/css" href="stylesheets/general.css">

</head>

<body>

Time stamp (often not on smart pages)

Main reply (web page, image, sound, video)

Setting cookies it wants browser to return

Start of web page

Server id, and often options in the server (FrontPage, PHP, perl, ...)

Blank line ends header

5 What Is Https

HTTP is limited to just the stuff needed to move web pages around networks. HTTPS is the TLS version of HTTP.

OK, what is [TLS](#)?

Underneath HTTP are lower “network layers” that do the actual dirty work of assuring your information gets moved about correctly. [TCP/IP](#) are the next layers down (Transmission Control Protocol, and Internet Protocol).

The difficulty is TCP/IP works in “clear text”. Anyone looking at the “packets” TCP/IP uses can read them clear as day. Not good for passwords, bank accounts, personal data, and many other things.

	Data unit	Layer	Function	Samples
Host layers	Data	7. Application	Network process to application	FTP, HTTP, SMTP, ...
		6. Presentation	Data representation and encryption	MIME, XDR
		5. Session	Interhost communications	NetBIOS, Named Pipes
	Segment	4. Transport	End-to-end connections and reliability.	TCP, UDP, SSL, TLS.
Media layers	Packet	3. Network	Path determination and logical addrs.	Bits on net: IP, Apple Talk
	Frame	2. Data Link	Physical addressing	Bit sending: ARP, Ethernet, PPP
	Bit	1. Physical	Media, signal and binary transmission.	Pins, voltages, RS232, WiFi, ...

To assure privacy a “Transport Layer Security” extension to the TCP layer was created (Layer 4) that encodes the data in packets using fancy programming and special “certificates”. (TLS is Open Source and supersedes secure socket layer” or SSL)

- Each web site has a certificate that is “digitally signed” by companies meeting security requirements and known to web browsers (Signing Authorities).
 - Public key: browser gets public key from server just by asking. The public key allows browsers to decode information from the server and encode

information being sent to the server.

- Private key: only the server knows this. This key is used to encode information sent to, and received from browsers.
- TLS is tuned for each session. The client and server agree on a secret “session key” to encrypt their information. This session key can not be used with any other session.
- The private key involves to large prime numbers. The public key just one of the primes.
- Applications using TLS (Web browsers, E-mail, many others) do not need to know the fierce math and deep rules of TLS... just a *relative* handful of special things to make TLS work (still not trivial).
- Browsers show a “lock” icon when https is being used. Broken lock if not.
- [Extended Validation SSL Certificates](#) require a more intensive check of the requesting entity by the certificate authority before being issued.
 - These show a green background on the SSL notification.
 - Left click green area for additional details.



6 TLS Tips

- You sometimes get a pop-up when browsers go to a new secure server for the first time. Other programs like E-mail and SSH are more likely to do this.
- You get a warning message if the certificate has expired, is not properly signed, or has other troubles.
 - Includes if signing authority is not known to the web browser.
 - You can add new authority certificates to your browser.
- If you have been there before an unexpected warning might indicate malware deleted your keys and is trying to spoof you.
- You should get a message if the certificate changes;
 - Having clients remember certificates greatly helps security.
 - This can indicate hackers between you and the server is trying to spoof you.
 - Or it may indicate the server needed a new certificate.
 - “Snake Oil” certificates are the default sample that comes with Apache.

7 Cookies: What They Are

Cookies are a small, crunchy byte-sized nuggets of information servers send to browsers. Browsers are expected to return these cookies to servers to provide a sweet burst of information.

- Original Intent:
 - #1: helps server issue unique identifiers to individual browsers.
 - If OS understands individual user logins, cookies provide unique identifiers to individual users.
 - Popular way to keep private sessions private.
 - Helps servers track what users are doing during complex or secure sessions. Helps secure logins, track shopping cart activity, etc.
 - Cookies can self-destruct at the end of a session (session cookies).
 - Cookies can survive for longer periods of time (persistent cookies).
 - Persistent cookies allow users to quickly return to what they were doing

(shopping carts, looking at maps, passwords) when they come back later.

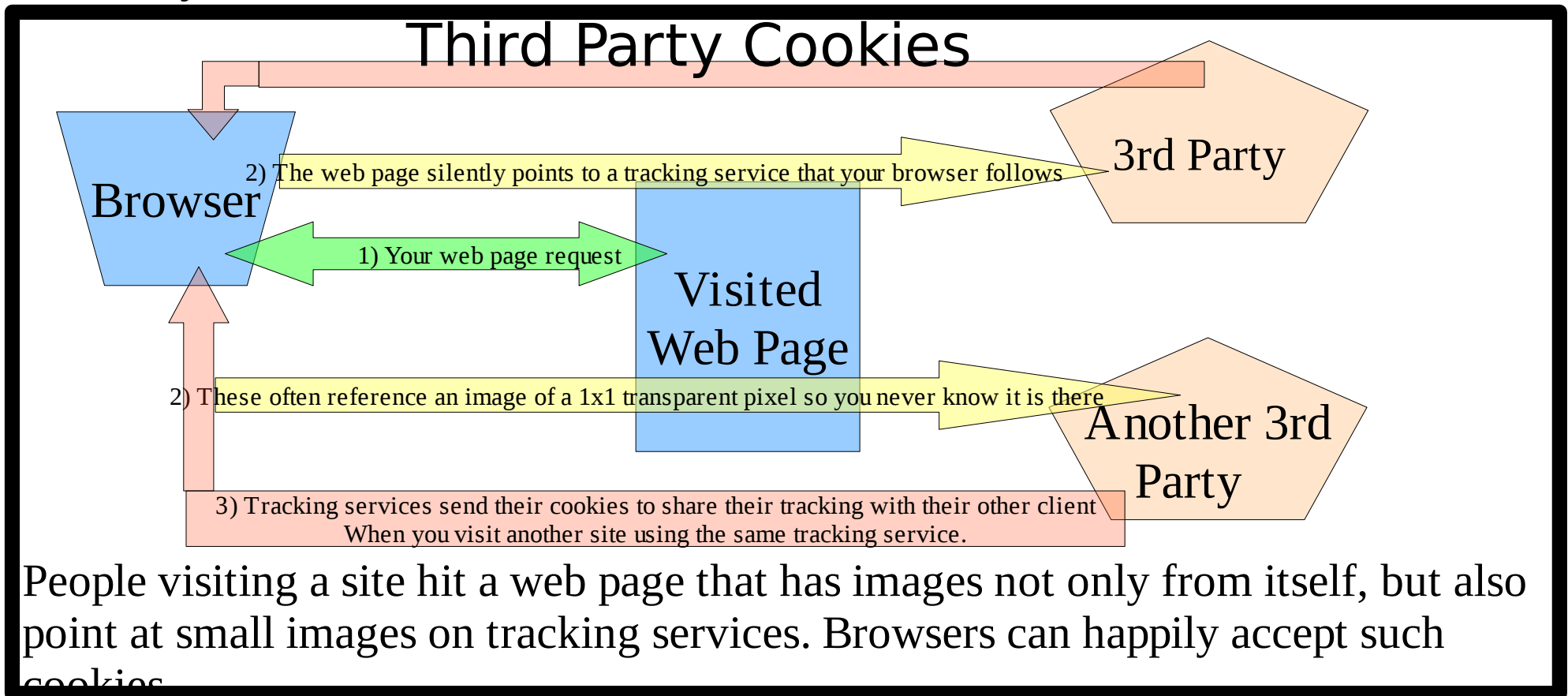
- Cookies are text send in headers of replies send to browsers. They are not programs. It is the programs reading the cookies that give cookies their powers.
- Set-Cookie: NAME=VALUE; expires=DATE; path=PATH; domain=DOMAIN NAME; secure; HttpOnly
 - NAME: what the server calls cookie by. VALUE associated value(s).
 - Expires: when to die. Session cookies have no date. Browsers can expire cookies before this. Cookies lasting years are all too often ugly cookies.
 - Path: only return cookies to web pages starting with this name (/hello/ for /hello, /hello/world.html, ...).
 - Domain: only return cookies to domains starting with this name (must be part of domain cookie came from; bcug.com includes www.bcug.com, secure.bcug.com, ...). Defaults to full domain.
 - Secure: only https protocol gets these cookies. *Vital for proper protection!*
 - *HttpOnly: cookie hidden from JavaScript, etc., programs in browsers.*

8 Cookie Security

- Browsers are sensitive to domain names: only returns cookies to the sending domain (information could be gleaned from server logs).
- Cookies should have reasonably limited lifetimes.
- Browsers can often warn about incoming cookies or restrict cookies.
- Cookies can be stolen, hijacked, and otherwise corrupted. Usually by malware on your computer *or* by lack of thought on web programmers.
- Lack of **secure** on secure pages, sends critical cookies in the clear.
- Lack of **HttpOnly** on cookies browser programs do not need give malware access to vital information.
- Every user on a Windows computers tends to have the ability to read cookies of any other user once the cookie jar is found. Great way for parents to tell where their kids have been. Not so hot for secure information.

9 Tracking Cookies

- All the good and bad you expect of tremendous power.
- Allows completely different servers, across completely different companies, to track your movements across different web sites.



- Tracking sites include: doubleclick.com, *advertising*, ...

The Good	The Bad
Can not harm the user's computer and they power a some of the webs advanced functions.	Read on...
If you are searching for something across different web sites, the opening page on that site can focus on what you are interested in to save you time.	Both the store and the tracking service can remember what you were doing.
For web store owners, the easier it is to find something the more likely it is you will buy.	Some privacy statements are worse than others.
Away from stores, a lot of tracking services can provide anonymous user profiles to tune advertisements to what you are interested in.	How anonymous this is in the age of data mining is an open question. Depends on how serious the tracking service is at keeping their secrets from their clients.
Web browsers can be set to give users more control over cookies.	Loose some power of web... browse longer. If configured to prompt for every cookie you will spend a lot of time deciding to accept or not. Prompt on 3 rd party cookies was not that bad, once upon a time, but heavier use of 3 rd party cookies is making for more and more prompts.

For house-wide blocking, routers can be configured to block the more annoying sites.

10 Form Fields

- Web pages with fill-in-the-blank, or select from the list, check boxes, etc., are called “forms”.
- When the “form” is submitted a GET or POST request is sent by the browser to get the results.
 - `<form action="/cgi-bin/submit-receiver.pl" method="GET">...</FORM>`
 - GET requests include all form fields after a question mark (?) in the URL.
 - Fields are in the name=value style.
 - Many special character in the name or value are “escaped” using %xx notation, where “xx” is the two-digit hex value for the character (&, %, <, >, +, space, though space can also be replaced with “+”).
 - The & delimiter within a URL should use **&** escapes:
``
 - These URLs clutter up web server logs and present security difficulties as they show any secure field values in clear text.

- POST request just have simple URL.
- Additional POST fields provided after POST header and are not logged.
Basically these fields are treated as a file upload.
- Web servers providing forms must have all fields in the HTML, properly filled out.

11 Web Server Tricks

- You can use telnet, connecting to port 80, and just provide the first two lines shown in the previous server reply (GET and host name).

```
telnet www.bcug.com 80
HEAD / HTTP/1.1
Host: www.bcug.com
(blank line)
```

- A HEAD request just gets the server heading, without the HTML reply.

<http://yro.slashdot.org/story/10/02/23/2236254/Criminals-Hide-Payment-Card-Skimmers-In-Gas-Pumps>

Your Rights Online: Criminals Hide Payment-Card Skimmers In Gas Pumps

on Tuesday February 23, @07:36PM

Posted by kdawson on Tuesday February 23, @07:36PM from the swipe-and-get-swiped dept.

ttugfoigel writes

"Wave of recent bank-card skimming incidents demonstrate how sophisticated the scam has become. Criminals [hid bank card-skimming devices inside gas pumps](#) — in at least one case, even completely replacing the front panel of a pump — in a recent wave of attacks that demonstrate a more sophisticated, insidious method of stealing money from unsuspecting victims filling up their gas tanks. Some 180 gas stations in Utah, from Salt Lake City to Provo, were reportedly found with these skimming devices sitting inside the gas pumps. The scam was first discovered when a California bank's fraud department discovered that multiple bank card victims reporting problems had all used the same gas pump at a 7-Eleven store in Utah."

[Read More...](#)   [238](#) comments