

Re: Traffic Log–Legitimate Traffic or Data Mining???

Source: <http://www.derkeiler.com/Newsgroups/alt.computer.security/2004-08/0182.html>

From: Mike (nospam_at_nospam.org)

Date: 08/10/04

Date: Mon, 9 Aug 2004 21:08:16 -0400

09Aug2004

Most web sites embed links to other web sites. This link describes what happens when you load a page.

http://www.surferprotectionprogram.com/Proxy/DOCS/user_manual/user_manual.htm#qstart_whathappens

(Note that you may have to concatenate the url listed above. It starts with "http" and ends with "whathappens".)

The reason your pages are now taking so long is that you have stopped portions of the pages from loading. These portions must timeout before the page finishes loading. For the dilbert example, you need to allow the sites that begin with:

<http://adsremote.scripps.com/html.ng/>

and

<http://adsremote.scripps.com/js.ng/>

If you allow these urls, they will load then tell the browser to go get a page from

<http://adfarm.mediaplex.com/ad/>

This page you want to block.

It helps to have a tool that can show you all of the pages that load when you load www.dilbert.com. The http proxy from the manual above lets you use regular expressions to block sites and to permit sites. It turns out, there are just a handful of patterns required to efficiently block advertisers and fewer patterns that must be allowed. There is no standard, it just happens that everyone uses similar naming conventions. In dilbert, the `/html.ng/` pattern is permitted, but the `/ad/` pattern is blocked. When using the http proxy, even though you permit the `/html.ng/`, the proxy strips several tags out of the outgoing http request – so your privacy is maintained.

Two nice things about using an http proxy:

1. all of your http tools can be directed to use it, not just your browser.
2. once you start blocking advertisers, most of the pop-ups stop

appearing.

"Jeff" <jeff@nospam.net> wrote in message
news:QwtRc.250317\$JR4.100228@attbi_s54...

<deleted>

> *If I try to download the comic from www.dilbert.com*

> *(65.114.4.69), my computer tries to send data packets to*

> *adsremote.scripps.com (204.78.38.15). The list goes on and on and on;*

these

> *are just a few examples.*

>

> *Now that I'm blocking these 'extraneous' data packets from being sent, the*

> *web pages I want to see take 30 seconds to 5 minutes to download, instead*

of

> *the usual couple seconds. But they do download eventually.*

<deleted>

> *And I'm steering clear of sites that want data packets sent to*

> *various alternative IPs when I try to download a webpage, looking for*

> *alternative sites for reading news and other activities.*

>

> *So the key question I have is this: is there a legitimate reason why my*

> *computer should be sending a data packet to adsremote.scripps.com*

> *(204.78.38.15) when I try to read the daily Dilbert comic (65.114.4.69)?*

> *Other than the initial request from my browser to download the .html*

file(s)

> *from a website, why should my browser be sending anything to anywhere*

else?

>

<deleted>

>