Deep analysis of a modern web site

Patrick Lambert

[http://dendory.net]

November 28, 2015

Abstract

This paper studies in details the process of loading a single popular web site, along with the vast amount of HTTP requests resulting from this single action, to sites all across the Internet. We will look at some of the code being loaded, what it accomplishes, and the impact it has on reliability and performance.

1   Introduction

Web pages used to be very static documents containing text and images that were typically loaded from the same server where the HTML page lived on. This has obviously evolved over time, where large sites quickly became overwhelmed with the amount of visitors that hammered against that one server. Many different techniques were introduced, including storing information in databases, load balancing, all the way to fully distributed Content Delivery Networks (CDNs).

Now, it's rare to see a popular site that doesn't have some kind of distributed delivery system. Many large corporations such as Akamai[1] exist for the sole purpose of providing content from all around the world, so that no single server ever gets overloaded, while people from one region of the world get content from a nearby server, and not one on the other side of the planet.

This works well when a single content delivery network serves the entirety of a site's content, but lately even this isn't true anymore. Instead, modern web sites use content not only from their own databases and servers, but also from others through APIs and web calls. Now, it's not rare to connect to a single site, and in the background, get content from dozens of different locations. This is of course all fully transparent, handled by our browser behind the scenes as we type in a URL in the location bar, but what exactly is that content doing, and how does it impact the web as a whole, both from a reliability and performance standpoint?

In the following sections, we will take a single web site, `http://www.techcrunch.com`, and dig down into those multitudes of connections, the calls that our browser makes across the web. We will explore what the content and snippets of code that come from these various servers do, and how they affect the experience of viewing that site.

1.1   Audience

This document was writing for a wide audience. We will go down into some networking concepts along with source code like HTML, CSS and JavaScript, and as such some experience with those notions can be useful, but not mandatory. The goal of this experiment is to explore how the modern web works, and show what goes on behind the scenes when you type in that URL. As such anyone interested in the technology will be able to see how a modern web page works, and how many dependencies there truly are on a typical large site.

1.2 Disclaimer

The site selected for this paper is the Techcrunch web site. The reason is simply because it is one of the most popular news sites out there, and seems like a good representation of how a modern site is designed. This study is not affiliated with Techcrunch in any way, and any conclusion from this experiment should not be taken as a statement about the performance or reliability of this particular site. Instead, the goal is to focus on a typical popular web site, as a general measure.

2 List of connections

Before we can start digging down into each specific call, we need to understand what happens when you type in the URL in your browser and press *ENTER*, or click on a link to that same site. The very first call goes to the main Techcrunch server, or more precisely, one of their servers. If you do a DNS lookup of their host name, you will see that they actually use DNS round-robin, a technique to split out where users will end up when they type in that URL:

```
> nslookup www.techcrunch.com
Server:  localhost
Address:  127.0.0.1

Non-authoritative answer:
Name:  techcrunch.com
Addresses:  192.0.82.250
 192.0.83.250
 76.74.255.117
 76.74.255.123
 66.155.9.244
 66.155.11.244
Aliases:  www.techcrunch.com
```

So in this case, the call will be at the first IP address, or *192.0.82.250*. On the other end, the server answers the request with the HTML code of the main page. The full page has around 2,000 lines of code, but we'll focus on a few of them in the coming sections. For now, let's just look at the total amount of connections that our browser makes during a full load. Note that you can see the list for yourself by using a network scanner or a web development tool such as the *Network Inspector* in Firefox or the *DevTools* in Chrome.

Here is a table with a list of connections that occurred when I loaded the page. As the site evolves, this list will likely change over time:

Host name	File name	Type
techcrunch.com	/	HTML
r-login.wordpress.com	/remote-login.php	JavaScript
s1.wp.com	/_static/	CSS
s1.wp.com	/_static/	JavaScript
o.aolcdn.com	/os_merge/	JavaScript
s0.wp.com	/wp-content/themes/vip/techcrunch-2013/assets/images/logo.svg	SVG
tctechcrunch2011.files.wordpress.com	/2014/10/wearables-blueprint.png	PNG
tctechcrunch2011.files.wordpress.com	/2014/10/built-in-brooklyn-logo.jpg	JPG
tctechcrunch2011.files.wordpress.com	/2014/10/chrome-stars.jpg	JPG
tctechcrunch2011.files.wordpress.com	/2014/10/2.png	PNG
s0.wp.com	/wp-content/themes/vip/techcrunch-2013/assets/images/210x210.png	PNG
s0.wp.com	/wp-content/themes/vip/techcrunch-2013/assets/images/300x169.png	PNG
s0.wp.com	/wp-content/themes/vip/techcrunch-2013/assets/images/1x1.png	PNG
s0.wp.com	/wp-content/themes/vip/techcrunch-2013/assets/images/logo-crunchbase-flat.png	PNG
s0.wp.com	/wp-content/themes/vip/techcrunch-2013/assets/images/crunch-network.jpg	JPG
0.gravatar.com	/js/gprofiles.js	JavaScript
platform.twitter.com	/widgets.js	JavaScript
s0.wp.com	/_static/	JavaScript
b.scorecardresearch.com	/b	HTML
o.sa.aol.com	/b/ss/aoltechcrunch,aolsvc/1/H.25.4/s66782871651621	GIF
techcrunch.com	/wp-content/themes/vip/techcrunch-2013/assets/fonts/b38504ed-66a1-4884-999a-07cca7997408-3.woff	Font
techcrunch.com	/wp-content/themes/vip/techcrunch-2013/assets/fonts/adb262fa-fc57-47bc-b34e-33dff61a8c3c-3.woff	Font
techcrunch.com	/wp-content/themes/vip/techcrunch-2013/assets/fonts/0ade293c-ee37-42ec-990f-13b6e9fb854c-3.woff	Font
b.grvcdn.com	/moth-min.js	JavaScript
pixel.wp.com	/g.gif	GIF

pixel.wp.com	/g.gif	GIF
0.gravatar.com	/css/hovercard.css	CSS
0.gravatar.com	/css/services.css	CSS
zor.livefyre.com	/wjs/v1.0/javascripts /CommentCount.js	JavaScript
connect.facebook.net	/en_US/all.js	JavaScript
b.scorecardresearch.com	/beacon.js	JavaScript
s.aolcdn.com	/os/aol/unb.min.js	JavaScript
www.google-analytics.com	/r/__utm.gif	GIF
static.ak.facebook.com	/connect/xd_arbiter/ehazDpFPEnK.js	JavaScript
tctechcrunch2011.files .wordpress.com	/2014/10/slack-large.png	PNG
tctechcrunch2011.files .wordpress.com	/2014/07/sequence-02-still005.jpg	JPG
tctechcrunch2011.files .wordpress.com	/2014/10/india-traffic.jpg	JPG
tctechcrunch2011.files .wordpress.com	/2014/03/htc-one-m8-4.jpg	JPG
pthumbnails.5min.com	/10369555/518477742_c_300_170.jpg	JPG
pthumbnails.5min.com	/10369293/518464642_43_300_170.jpg	JPG
pthumbnails.5min.com	/10369503/518475113_9_300_170.jpg	JPG
graph.facebook.com	/	JSON
s0.wp.com	/wp-content/themes/vip/techcrunch-2013/assets/images/210x210.png	PNG
s0.wp.com	/wp-content/themes/vip/techcrunch-2013/assets/images/300x169.png	PNG
s0.wp.com	/wp-content/themes/vip/techcrunch-2013/assets/images/1x1.png	PNG
tctechcrunch2011.files .wordpress.com	/2014/10/sony-logo.jpg	JPG
amch.questionmarket.com	/adsc/d1042020/12/1108335/ad_radar.php	GIF
www.google-analytics.com	ga.js	JavaScript
ads.tw.adsonar.com	/adserving/getAdsAPI.jsp	JSON
a02.korrelate.net	/a/e/d2a.ads	GIF
js.adsonar.com	/js/aslJSON.js	JavaScript
apx.moatads.com	/pixel.gif	GIF
www.google-analytics.com	/__utm.gif	GIF
s.aolcdn.com	/os/SponsoredListings//AdGallery/ 27899558/125x125/4930254.png	PNG
s.aolcdn.com	/os/SponsoredListings//AdGallery/ 27899134/125x125/4927093.jpeg	JPG
s.aolcdn.com	/os/SponsoredListings//AdGallery/ 27861361/125x125/4926600.jpg	JPG
a02.korrelate.net	/1x1.gif	GIF
ads.adsonar.com	/adserving/getAdsAPI.jsp	HTML
cdn.at.atwola.com	/_media/uac/guid.html	HTML

As you can see, there is a large number of calls made by our browser in order to display the home page of that site. 59 calls in total. The number of different servers is also quite impressive, as is the amount of file types. In order to go through all that in an orderly fashion, and not just in order of page loaded, we will go through each type of content and see what exactly gets loaded, and what is happening in the core of our browser's engine.

3 HTML and images

Perhaps the easiest types of content to understand are the actual web and image documents. As you can see in that table, the very first network call is to the *techcrunch.com* site, requesting the / file name. This is the root of the site, and there's no obvious way to find out what exactly produces that content. It could be a simple *index.html* file, but more often than not, it's actually a script gathering the content from templates and databases.

Images are loaded from many different sites as well. The primary ones are in the *wp.com* and *wordpress.com* domains. The reason for this is that Techcrunch actually runs on a WordPress site, more precisely, their VIP service[2], a content delivery network aimed at large companies. This particular choice of CDN is specific to Techcrunch, and is one option out of many. As described earlier, the main benefit of relying on such a service is that WordPress can handle the scaling, security and optimization of the infrastructure, while the site designers can focus on the site itself.

WordPress doesn't seem to talk much about their network, but their FAQ[3] does say this:

> *WordPress.com uses multiple data centers for content creation and serving, with additional data centers used for DNS and other functions. Data centers are run active-active and WordPress.com is built for N+1 redundancy, allowing for one data center to fail at a time.*

As for the images themselves, the main page of any site typically has lots of different images, so it's no wonder why there are so many calls going to get JPG, PNG, GIF and even a SVG image. Those are all file types that your browser can display and make up the bulk of the network traffic. Some of them compose the main theme of the site,

while others are specific to the posts that happen to be on the main page at the time.

The theme images are typically included in whatever template their designers created, while the image posts are included in each individual post, typically stored in a database somewhere. Since that all happens in the back end, this isn't something our browser can tell us.

One detail we can see when looking at the HTML source however is that the site uses a WordPress plugin called *Batcache* which implements a system called *Memcached*[4]. This is a way to cache files into memory so that the back end system doesn't have to reload content every time a user connects. This is a typical process implemented by large web sites.

4   Style sheets

The very next thing any web page needs in order to display properly, past the HTML code and images, is at least one style sheet. Here, the main style sheet is a call to this CSS file:

```
1 <link rel='stylesheet' id='all-css-0' href='http://s1.wp.com/_static/??-
2 eJx9kOFuwjAMhF9owWOgij/TniWkJnFx4ihxqXh7UhAwxFb5z110n30KTNk4SYpJIY4m8+
3 gpVaiRGH+5yIBOX93K1foBf2NMR6wwoGbrjubq3uJ7Fv8ExHvsZVRzEGaZYKLe4+
4 INJwXbe8xW50TEniwyxhZbwmLu7tQsQyv43k1D21PhRBkUXXBlTC6Yr8/
5 1BqqeGVeROg0iSKLUuPoQS9c9imFxVknSizEHtlSW0ILzjzXpoaV+2f/
6 ahy14lr3lOfATv9fbNpvdruuGCx28uEw=' type='text/css' media='all' />
```

This is a massive 1,794 lines file containing a number of CSS definitions telling your browser how the page should be displayed. It has everything from background colors, browser specific codes, typography, layout, icon information and responsive design for mobile devices.

The comments in this file specify that this is part of the *Smiley* theme, a WordPress theme that Techcrunch is likely based on. This CSS also goes along with a JavaScript file declared just a bit later on the main page:

```
1 <script type='text/javascript' src='http://s1.wp.com/_static/??-
2 eJyNkF1OxDAMhC+ENy2LhHhAnKVN3dQhiUPsbLScnoBY8VdpebLl+
3 Twj27QMlpNiUuPFRJ4pIFTBMrk+AOorH7zcmM5RsqEuKO+gf6lYzp/
4 lKgCRXJkUD5HSBf6Wmlk0okiP3FF/RlE6EbarmEfNk32GgkKvf1znwA5yqI6SmN47X
5 LgqrBwCN9Nocah7R0lbefZo9behbtgPMCfKRtFuttRkN7gdxqOZRFA/tgPN8KXu/
6 eJfNo7Zhcsrn+LjeDcOx4fxfhj8G6UIq0I='></script>
```

It's hard to say what that blob of data passed along to both the CSS and JavaScript files are, but it's clearly Base-64 encoded text, probably encrypted. Either way, these two files provide most of the information that your browser needs to process in order to display the site.

5 Fonts

Fonts used to be a pretty simple deal. The browser would have one default font, loaded from your local system, and every page would be displayed using that font. Slowly, designers started to include other font families through CSS calls. But what if you want to include fonts that people typically don't have installed? If you call for a font that isn't available on the browser's machine, then the browser isn't going to be able to display it.

This is where web fonts come in. A style sheet is able to tell the browser to load a font file directly from the web. There are three popular font types: True Type, Open Type, and a newer format called Web Open Font Format (WOFF). This latest type is getting popular because it includes compression, which makes the file size smaller. In our case, we see calls to three font files:

- `b38504ed-66a1-4884-999a-07cca7997408-3.woff`

- `adb262fa-fc57-47bc-b34e-33dff61a8c3c-3.woff`

- `0ade293c-ee37-42ec-990f-13b6e9fb854c-3.woff`

These are part of the template Techcrunch designed in 2013 for their site, but one good source of fonts you can use in your own projects is the Google Fonts[5] site which has hundreds of fonts available freely for designers.

6 Profiles

We've already seen that the theme they use includes a large JavaScript file, but as with most modern sites, this one also includes JavaScript from a myriad of different sources, along with what's included in the main page as well. One such code snippet is used for profiles and logged in users:

```
1  <script type='text/javascript'
2  src='//0.gravatar.com/js/gprofiles.js?ver=201444x'></script>
3  <script type='text/javascript'>
4  /* <![CDATA[ */
5  var WPGroHo = {"my_hash":""};
6  /* ]]> */
7  </script>
8  <script type='text/javascript' src='http://s2.wp.com/wp-content/mu-plugins/
9  gravatar-hovercards/wpgroho.js?m=1380573781g'></script>
10
11 <script>
12 //initialize and attach hovercards to all gravatars
13 jQuery( document ).ready( function( $ ) {
14         if ( typeof Gravatar.init !== "function" ) {
15                 return;
16         }
17
18         Gravatar.profile_cb = function( hash, id ) {
19                 WPGroHo.syncProfileData( hash, id );
20         };
21         Gravatar.my_hash = WPGroHo.my_hash;
22         Gravatar.init( 'body', '#wp-admin-bar-my-account' );
23 });
24 </script>
```

As you can see, this code calls functions from Gravatar[6], an open profile system that many web sites use. Their site describes the service as such:

> *Gravatar is a free service for site owners, developers, and users. It is automatically included in every WordPress.com account and is run and supported by Automattic.*

But Gravatar isn't the only third party plugin that Techcrunch uses. It also integrates with Livefyre[7] which counts comments, and AOL, their parent company, to display a navigation bar of other AOL sites.

7 Analytics

Any popular web site owner will want to know how popular their site is. More than that, they typically want detailed statistics on what individual posts are most popular, where their traffic comes from, and so on. The most popular analytics site is Google Analytics[8], and sure enough, this line of code shows that Techcrunch uses them as well:

```
1  var TC_Google_Analytics_Config = {"account":"UA-991406-1",
2  "domain":"techcrunch.com"};
```

They also include a WordPress analytics plugin as well. These stats are typically closely guarded by any public site, but there are a number of companies out there doing studies on top web sites. For example, Alexa[9] shows us that Techcrunch is ranked 301 worldwide and 149 in the United States.

8 Ads

Ads are a necessary evil online. We all dislike having to go through five minutes of ads on TV, similarly the web also relies on ads to pay the bills. Several of those network calls we've seen are used to load ads. Some load JavaScript which your browser parses and executes, others are HTML snippets, and others are simple images displayed on the screen with a link to the advertiser.

In fact, there are two main types of ads network calls. The first type is the normal ads, for example the call to the *atwola.com* site at `http://cdn.at.atwola.com/_media/uac/guid.html` which is a full network connection your browser has to make in order to retrieve the following page:

```
1   <html>
2   <body>
3   <script type='text/javascript'>
4   try {
5   var dt=new Date(),t=0,d=document,l;
6   dt.setFullYear(dt.getFullYear()+1);
7   d.cookie='ads3PTest=yes; path=/; expires='+dt.toGMTString();
8   if (d.cookie.indexOf('ads3PTest=')!=-1){
9     t=1;
10    d.cookie='ads3PTest=; path=/; expires=Thu, 01 Jan 1970 00:00:01 GMT;';
11  }
12  if (!t) {
13    l=localStorage.getItem('adsGUID');
14    if (!l){
15    var l='xxx-xx-4'.replace(/[xy]/g, function(c){
16  var r=Math.random()*16|0,v=c=='x'?r:(r&0x3|0x8);
17  return v.toString(16);
18    });
19    localStorage.setItem('adsGUID',l);
20    }
21    var x=("guid="+l).toString();
```

```
22    window.top.postMessage(x, "*");
23  }
24  }
25  catch (e) {}
26  </script>
27  </body>
28  </html>
```

As you can see, this code sets a tracking cookie in your browser. This is how advertisers can track you across web sites, since any site which includes a call to this particular page will get the tracking cookie updated in your browser. Since it's an HTML page including JavaScript, and not just a JavaScript page, this helps the advertiser bypass third-party cookie restrictions that your browser may be using to protect you.

The second type of calls that can be included in the ads section are things like this: `http://a02.korrelate.net/1x1.gif` As the file name indicates, this is an image of one pixel by one pixel. Your browser will display this basically invisible image, and the server it resides on will know that your browser loaded it. This is yet another way to track your actions across web sites, without even needing to run JavaScript or set cookies. Even Google Analytics uses this trick with its `__utm.gif` file.

9   Social media

Of course, what popular web site doesn't include social media plugins these days? Not to be outdone by others, Techcrunch includes both Twitter and Facebook integration. It used to be that social networks would provide simple buttons that linked to their sites. Now, integration is far, far deeper.

9.1   Twitter

Twitter has a whole developer center[10] with libraries that web sites can use to integrate Twitter into the workflow for their users. The process includes two steps. First, you include Twitter's own JavaScript code by calling their library:

```
1  <script>window.twttr = (function (d, s, id) {
2    var t, js, fjs = d.getElementsByTagName(s)[0];
3    if (d.getElementById(id)) return;
4    js = d.createElement(s); js.id = id;
```

```
5    js.src= "https://platform.twitter.com/widgets.js";
6    fjs.parentNode.insertBefore(js, fjs);
7    return window.twttr || (t = { _e: [], ready: function (f) { t._e.push(f) } });
8  }(document, "script", "twitter-wjs"));</script>
```

Then, you can add elements from Twitter's site by using the proper classes to your HTML elements, such as sharing with `twitter-share-button` or following with `twitter-follow-button`. Of course, the JavaScript code loaded from Twitter, or from any of those external APIs, does a lot more than just provide CSS tags. In fact if you look at the page displayed by your browser, and the source of the very first HTML loaded, they are drastically different.

Twitter, for example, includes *iframes* to display Twitter specific content. Here is one such snippet of HTML code, in this case displaying a Tweet button, fully generated on the fly as the page gets loaded:

```
1  <iframe style="width: 109px; height: 20px;" data-twttr-rendered="true"
2  title="Twitter Tweet Button" class="twitter-share-button
3  twitter-tweet-button twitter-share-button twitter-count-horizontal"
4  src="http://platform.twitter.com/widgets/tweet_button.
5  d58098f8a7f0ff5a206e7f15442a6b30.en.html#_=1414778911738&amp;
6  count=horizontal&amp;counturl=http%3A%2F%2Ftechcrunch.com%2F2014%2F10
7  %2F30%2Fequity-crowdfunding-service-seedrs-acquires-junction-investments-
8  plots-us-expansion%2F&amp;id=twitter-widget-1&amp;lang=en&amp;
9  original_referer=http%3A%2F%2Ftechcrunch.com%2F2014%2F10%2F30%2Fequity-
10  crowdfunding-service-seedrs-acquires-junction-investments-plots-us-
11  expansion%2F&amp;size=m&amp;text=Equity%20Crowdfunding%20Service%20Seedrs
12  %20Acquires%20Junction%20Investments%2C%20Plots%20US%C2%A0Expansion&amp;
13  url=http%3A%2F%2Ftcrn.ch%2F1tGCsjo&amp;via=techcrunch" allowtransparency=
14  "true" scrolling="no" id="twitter-widget-1" frameborder="0"></iframe>
```

9.2 Facebook

The integration with Facebook is even deeper, but the process is very similar. The JavaScript is loaded from `connect.facebook.net` and then each widget or element is inserted with JavaScript calls and CSS classes. Facebook also uses the Open Graph protocol[11] to get extra information from the site using META tags:

```
1  <meta property="fb:app_id" content="187288694643718" />
2  <meta property="fb:admins" content="1076790301,543710097,500024101,771265067,
3  1661021707,1550970059,663677613,10219991,1178144075,726995222,506404657,4700188"
4  />
5  <meta property="og:site_name" content="TechCrunch" />
6  <meta property="og:title" content="TechCrunch" />
```

```
7  <meta property="og:description" content="TechCrunch is a leading technology
8  media property, dedicated to obsessively profiling startups, reviewing new
9  Internet products, and breaking tech news." />
10 <meta property="og:image" content="http://s0.wp.com/wp-content/themes/vip/
11 techcrunch-2013/assets/images/logo-large.png?m=1391183173g" />
12 <meta property="og:url" content="http://techcrunch.com" />
13 <meta property="og:type" content="website" />
```

Again, Facebook also generates *iframes* in order to display code on the page as it gets loaded. Techcrunch not only uses its social sharing widgets, but also comments, so there is more code being loaded. The commenting section, for example, comes from hundreds of lines of code.

Of course there are a lot of benefits for sites to include social plugins. The first is that many people are logged into their social portals, so information can be imported automatically. When you post a comment on a site that uses Facebook comments, your Facebook name and avatar are automatically added. Also, you can share that comment automatically on your feed.

10   Conclusion

As we've seen, the process of loading a single web site can be quite an affair. Fortunately, this all happens behind the scenes at a very fast pace. With our modern devices, going through everything we've covered typically takes between one and four seconds.

Of course, each site does things slightly differently, but this particular site is probably average of what happens when your browser requests the home page of a top 500 site on the Internet. The more integration, social widgets, ads, and other elements, the more network calls your browser will need to do in order to load the page. But what does this all mean as far as where the web is headed?

If a web site relies on dozens of other servers in order to be displayed properly, there is always a risk that one of those servers will fail. Fortunately, your browser is smart enough to ignore these failures and do its best in order to display the web page. For example in this case, one of those network calls actually failed. At the time of this experiment, the site `scorecardresearch.com` returned a 403 error. So that content was ignored, and the browser moved on.

But a lot of this extra content is fairly important. A static site relies

only on one thing, the web server that serves the page. But a modern site also relies on an extensive back end including database systems, caching, load balancers, and so on. Then, they typically rely on CDNs to deliver key assets, social plugins, analytics scripts, and more. One example of things going wrong is Malvertising[12], the process of hacking into an ads network, and serving malware through ads on hundreds or thousands of sites loading ads from that network.

   Of course it would be foolish to try and reverse this trend. As CDNs become more common, they also become faster and more secure, ensuring that these events are less likely. But having a single point of failure is never good design. So these things must be kept in mind whenever developers create web sites, especially popular ones.

11   References

References

[1] Akamai Technologies: *Content Delivery Network*
    http://www.akamai.com/

[2] WordPress: *VIP Service*
    http://vip.wordpress.com/

[3] WordPress: *Infrastructure FAQs*
    http://vip.wordpress.com/documentation/infrastructure-faqs/

[4] Memcached: *Free & open source, high-performance, distributed memory object caching system*
    http://memcached.org/

[5] Google Fonts: *Hundreds of free, open-source fonts optimized for the web*
    http://www.google.com/fonts

[6] Gravatar: *A Globally Recognized Avatar*
    https://en.gravatar.com/

[7] Livefyre: *Engage Consumers Through Real-Time Conversation, Content, Social Curation and Ads*
    http://web.livefyre.com/about/

[8] Google Analytics: *Find the audience you want*
    http://www.google.ca/analytics/

[9] Alexa: *How popular is techcrunch.com?*
http://www.alexa.com/siteinfo/techcrunch.com

[10] Twitter: *Twitter for Websites*
https://dev.twitter.com/web/overview

[11] OG: *The Open Graph Protocol*
http://ogp.me/

[12] Proofpoint: *Malware in Ad Networks Infects Visitors and Jeopardizes Brands*
http://www.proofpoint.com/threatinsight/posts/
malware-in-ad-networks-infects-visitors-and-jeopardizes-brands.
php