Client Side Site Impact Review Toolkit

Main Title:	Client Side Site Impact Toolkit		
Date:	17/06/2024	**Release Maturity:**	Draft/~~Live/Final~~
Author:	Allen Woods		
Owner:	Allen Woods		
Client:			
Version:	1.50		
CADMID	Concept		
Line of Development	Doctrine		
Organisation			
Release Classification	U		
Document ID/Number			

Introduction

This document is derived from a more complex client side impact review template which can be viewed here (if the reader has downloaded it). The template giving a somewhat wordy view of the reasoning behind it and the various sections that make it up. This document is a summary, based on the full template "Annex A". The aim to provide readers with a simple introduction to how to go about assessing client side impact of code delivery into end user devices based on the authors experience of providing quick and simple illustrations of the risk not checking the nature of code drops into visitor devices nowadays.

Over time, the method of approach to reviewing has matured in that before writing a full template review, a "quick and Dirty" review would be executed which would prompt more detailed review of a site if the author was asked to do so. The first part of this document focuses on "quick and dirty", with the remainder of the tools used to review sites being provided in later pages,

Method of Execution?

The method of execution is quite simple.. If you want to take an under the hood look at a web site, any site, simply cut and paste a site URL into memory, open this document, then use the "Left click here" links in the document to activate the nominated tool, past the URL into the main text box asking for a URL, then click on any associated button. It really is that simple.

The tools are, for the moment, free to use. Searching and assessing a site using each takes seconds few to execute.

Why do it?

Several reasons.

For legals, client confidentiality and security matters.

When considering hiring privacy and cybersec companies? As a means of assessing heir technical competence. Basically, if they drop "stuff" into devices that perhaps they should not then that should raise concerns about their own competence.

For any other site? Commercial risk and its minimisation

Quick and Dirty

Ser	Tool	Purpose	Link Address.
1	Built With	A tool that gives a sophisticated description of the components and tools a web site has been built with.	Left click here
2	IP address geolocation	Used to identify the physical location of a server or other fixed device	Left click here
3	Fou Analytics Page X-Ray	This component will present you with a tree map of the connections, across the world, that a site visit generates.	Left click here
4	Whois is a catalogue of web site registrants.	It will tell visitors who owns a registered domain. It should be noted that because of discussions related to privacy matters, names of individuals are not visible on any returned entry. It should be noted that in the event a search fails, users may be directed to other registrant catalogues.	Left click here
5	Site Security Response Headers	When a site loads, several electronic conversations ensue. Each with their own security risks server side. The URL contains detailed notes on the nature of the risk involved as well as providing a comprehensive, but simple to understand, risk profile of a site	Left click here
6	DNS Viz	The Domain Name System (DNS) is the means by which the location of a web site on a host server, is catalogued in respect of location. However, it generates its own security risks which are often overlooked. With the tool indicated here, the lowest level of concern, your domain name, is the section of the search results to look at.	Left click here

Caveats

There are other tools you could use. The ones listed here are just a few. The point? Use them, learn and experiment.

Bear in mind that while the tools are simple to operate, interpretation may take a bit of practice.

The Rest

The following sections provide short lists of tools that can be used if the need arises. They are grouped by category/subject area. The author would use some or all of the tools listed in the following pages if preparing a full blown brief using the more complex template.

Physical Location

The fact of the matter is, if your site is in a country other than your own, your site is bound by several forms of legislation. Note that there is no relationships between a Top Level Domain (TLD) and where a site may be physically hosted.

Ser	Tool	Purpose	Link Address.
1	Built With	A tool that gives a sophisticated description of the components and tools a web site has been built with.	Left click here
2	IP address geolocation	Used to identify the physical location of a server or other fixed device	Left click here
3	The "Way Back Machine"	One of several web archive platforms that can be used to review a domain and its site history	Left click here
4	Traceroute	Use this tool to track the geophysical route from your device to the host machine the domain you are reviewing is housed. While this on line tool is useful, be aware that there is a similar tool available on windows machines with simple command line arguments set out below that may be preferrable to use: c:\>tracert *Usage: tracert [-d] [-h maximum_hops] [-j host-list] [-w timeout]* *[-R] [-S srcaddr] [-4] [-6] target_name* *Options:* *-d Do not resolve addresses to hostnames.* *-h maximum_hops Maximum number of hops to search for target.* *-j host-list Loose source route along host-list (IPv4-only).* *-w timeout Wait timeout milliseconds for each reply.* *-R Trace round-trip path (IPv6-only).* *-S srcaddr Source address to use (IPv6-only).* *-4 Force using IPv4.* *-6 Force using IPv6.*	Left click here
5	Whois is a catalogue of web site registrants.	It will tell visitors who owns a registered domain. It should be noted that because of discussions related to privacy matters, names of individuals are not visible on any returned entry. It should be noted that in the event a search fails, users may be directed to other registrant catalogues.	Left click here

DNS

Ser	Tool	Purpose	Link Address.
1	DNS Reverse Look up	Find multiple sites at same IP if there	Left click here
2	DNS Viz	A tool to detect and determine DNS security status from TLD down.	Left click here

3	Reverse DNS Look up (comprehensive)	Two tools that can be used to check "reverse DNS". Typically, multiple sites are hosted on single machines nowadays with the end effect being that they share domain level security risks of many kinds. Some of the shared host lists are very, very long. It is YOUR RESPONSIBILITY to secure your domain level DNS set up	Left click here
4	Reverse DNS Look Up (Simple)	See above	Left click here
5	Site Response Header check	Use to prove the viability of site http response headers	Left click here
6	Site security certificate checker	SSL certs are an indicator that the site is using encrypted comms between visitor and host device. They run out or expire. Nor do they provide complete server side protection (they are not meant to.	Left click here

Code Drops

Its not the cookies, it's the code wot does it.

Ser	Tool	Purpose	Link Address.
1	Built With	A tool that gives a sophisticated description of the components and tools a web site has been built with.	Left click here
2	Domain IO Scan	On page load, this tool provides a means to check the nature of connectivity in some detail	Left click here
3	Fou Analytics Page X-Ray	Used to identify, in a tree/node schematic the nature of links between a home domain and any CDNs.	Left click here
4	The "Way Back Machine"	One of several web archive platforms that can be used to review a domain and its site history	Left click here

Security Related

A couple of links on common site security issues and their monitoring. Bear in mind that security along is a massively complex subject area and organisations like the UK NCSC, OWASP and issues like cross-origin resource sharing come into play. The links below are merely an introduction. If a site does not have an SSL certificate, has a poor security response header score and poor domain level DNS security, then using it represents a risk to end users, if those building web sites or offering data protection advice fail the two links below, then that is indicative of systemic failure on their part.

Ser	Tool	Purpose	Link Address.
1	Site Response Header check	Use to prove the viability of site http response headers	Left click here
2	Site security certificate checker	SSL certs are an indicator that the site is using encrypted comms between visitor and host device. Certificates run out or expire but the site remains accessible, but with warnings. Nor do they provide complete server side protection (they are not meant to.	Left click here

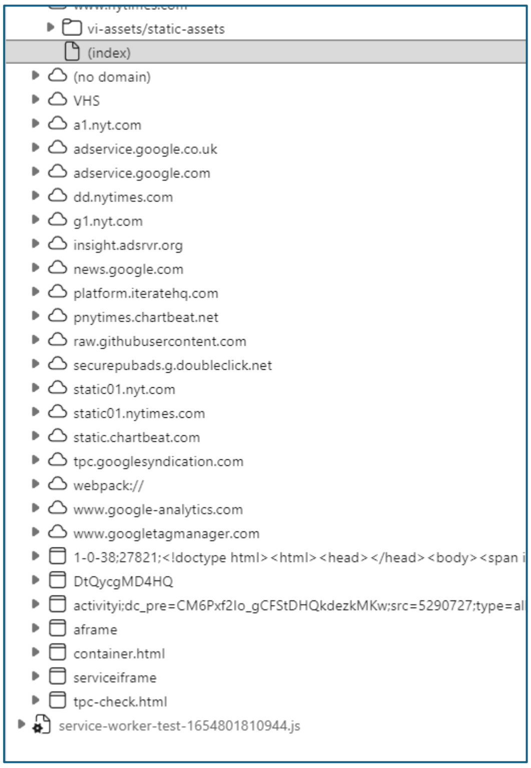
Official

Whilst arguably readers may think to themselves the kind of thing set out here is “over the top”, then the two links below may help to disavow that view. They point to European Data Protection Board (EDPB) downloads, free to use, that will report on many of the matters listed in this document.

Nuff said.

Ser	Tool	Purpose	Link Address.
1	EDPB Web Site Evidence Collector	It took them a while, by the EU Data Protection Board has developed a sophisticated site review toolkit that anyone can download and use. It is highly recommended that people do. Especially “professionals”.	Left click here
2	EDPB Web Site Review Tool		Left click here

Operating System Components



www.nytimes.com
vi-assets/static-assets
(index)
(no domain)
VHS
a1.nyt.com
adservice.google.co.uk
adservice.google.com
dd.nytimes.com
g1.nyt.com
insight.adsrvr.org
news.google.com
platform.iteratehq.com
pnytimes.chartbeat.net
raw.githubusercontent.com
securepubads.g.doubleclick.net
static01.nyt.com
static01.nytimes.com
static.chartbeat.com
tpc.googlesyndication.com
webpack://
www.google-analytics.com
www.googletagmanager.com
1-0-38;27821;<!doctype html><html><head></head><body><span i
DtQycgMD4HQ
activityi;dc_pre=CM6Pxf2Io_gCFStDHQkdezkMKw;src=5290727;type=al
aframe
container.html
serviceiframe
tpc-check.html
service-worker-test-1654801810944.js

Figure 1 - A typical browser developer tool "sources" CDN manifest

Learn to use your browser “debugger” capabilities. There is a lot that can be done with them with practice, but initially, locate and find the “sources tab” which will present a lit something like that illustrated above.

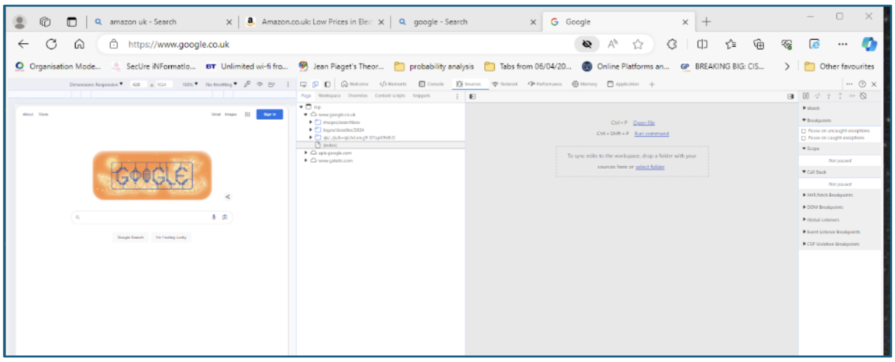


Figure 2 - Googles own CDN manifest

As to why execute this kind of check? Figure 2 illustrates the CDN manifest of Google. Be like Google. Do not use code from outside your domain which is inevitably outside of your control……

Ser	Tool	Purpose	Link
1	Netstat	Windows TCP detector, run as command line software	Left click here
2	NLS Lookup	Name Server Lookup	Left click here List of DNS Record Types Left click here

The phrase "open source" comes into play here. Wile "open source is a "good thing", it does not mean free of responsibility or contractual liability (all opensource tends to be bound by licensing of one form or another) it is the case that calling any software component from an external source, that is placed in your site visitors browser, opens a private conversation between the source code host and your visitor over which site owners have not control. With that surrendering of control being a massive, massive risk to your end users. Google does not do it. Nor should you.

Why Bother – A Simple Overview – Opticians a Market Sector Story

This annex was written as a result of some of the reaction to the use of the term "bumping into" people using the internet as a form of "stalking". In the author's view, the majors are not "stalking" rather the "bumping into" is a matter of trying to rationalise randomness such that patterns of behaviour, of similar kinds of people can be identified and exploited for commercial gain. What follows is an example of the art of the possible, given enough data, using the profession of "optician" as a review example.

If readers use the Google search engine (bearing in mind the data gathering starts there) to search for "How many companies use Google Analytics" the search results will contain many, many listings, the site analysis tool "builtwith" contains the analysis presented here there are other search list entries that return similar figures.

In the UK company register, at any one time, under the SIC code 47782 used to classify opticians for profession ID purposes, there are SIRO 19,500 opticians, the number varies as market conditions change for a variety of reasons. A casual examination of the web sites of every single one within fifty miles of where the author lives, uses one or another of the "free" and "simple" Google components, with their web sites, usually, being constructed using a site builder like Wordpress by developers for them.

For want of an argument, the author assumes some 80% of UK opticians use one or another of the free google components, that gives an approximate number of 15,600 opticians that may be taking advantage of Google's offers.

Given that the way the components are written into web sites, that is to say on every page and given that web sites consist of multiple pages (easy to detect by reading a sitemap.xml if one is installed), then let's assume for ease of calculation, for each site, there are 100 pages. That would suggest as a ballpark figure, that suggests 1,560,000 positionings of such components across the optician profession. For each site, each page load, will provide more detail of each visit, of each optician customer. The behaviour patterns Google seeks to identify build like that.

If a search is made to estimate the number of people wearing spectacles or contact lenses in the UK, the various estimates returned hover around the 60% to 70% of a population of more than 67 million people who will probably visit an optician each year.

Opticians do not just sell spectacles. If you sit and look, they sell all manner of things. It would be a commercial opportunity t be able to find out what those "all manner of things" might be…..

The trick is, for those who provide enough free components that are deployed in enough web site, is that at some point in their lives, each spectacle wearer will be obliged to visit an optician. Typically, though not always, as a precursor, those needing an optician will visit the web site oof the company concerned, book appointments, take advantage of the "best offer" thingies and so on. The purpose of components like Google Analytics being to provide the means to identify them, not necessarily as individuals, but to provide the means to, say, detect market trading volumes and so on.

Many of them will be Google device owners of a phone or chrome book. Many of them will visit other sites that have also deployed Google analytics. Part of the deliberate function of Google Analytics is to profile a site visit and during such visits, place cookies. The purpose of cookies being not to track per se, but to act as a place marker.

Over time, the millions of site visits will be cyclical in that, the recommended check-up period for an eyesight test is 2 years in the UK.

The behaviour patterns over time, become more complete and more sophisticated such that accurate inference becomes more than possible. The key business advantage for Google in giving away millions of copies of Google Analytics being that while site operators get charts and graphs (eye candy), it is specific to their business. Google gets market wide intelligence it can then sell on.

Google, as a search capability has been in business since 1998. By now, given the size of their data centres, it is likely that a significant number of visits to opticians now form predictable patterns of behaviour. And that, is key commercial business intelligence, in just one sector of business that can be sold through auction under the general application of a system of "real time" or "header" bidding.

"Bidding" being the key word, those with the deepest pockets win auctions. The risk of using things like Google Analytics, without understanding it, is, as Google themselves are reported to have discussed, an existential one for small to mediums who, it would seem, do the vast majority of data gathering for them.

A single "First person singular" who may or may not buy a pair of spectacles every two years is a business intelligence irrelevance. Many such "first persons"…. Ah now, that is where the market intelligence really lies.