

# Privacy and the Web – Are you doing what it takes?

Amelia Andersdotter & Anders Jensen-Urstad

NightlyBuild 2016, Cologne, September 2

**dataskydd.net**

## *Three paradigms for privacy*

The right to be left alone (1870s)

The right to keep things to yourself. To be secret and unrevealed.

## *Three paradigms for privacy*

### The right to be left alone (1870s)

The right to keep things to yourself. To be secret and unrevealed.

### The right to privacy as control (1970s)

The origin of data protection laws: rules for transparency and accountability even after information is revealed.

## *Three paradigms for privacy*

### The right to be left alone (1870s)

The right to keep things to yourself. To be secret and unrevealed.

### The right to privacy as control (1970s)

The origin of data protection laws: rules for transparency and accountability even after information is revealed.

### The right to identity and identity development (2000s)

The opportunity to develop one's own personality without undue interference.

# How privacy-friendly is your site?

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

Check

<https://webbkoll.dataskydd.net/en>  
(<https://github.com/andersju/webbkoll>)

Swedish municipalities (*in Swedish*):

<https://dataskydd.net/kommuner>  
(<https://github.com/andersju/municipality-privacy>)



Funding: Internetfonden / The Internet Foundation IIS



# Hur privatlivsvänlig är din kommun?

Vi har undersökt webbplatserna för Sveriges 290 kommuner och tagit reda på vilka dataskyddande funktioner de använder — eller *inte* använder — för att hjälpa dig utöva makt över ditt privatliv.

Webbplatserna [betygsattes](#) enligt en skala A-E. Klicka på ett kommunnamn för detaljerad information.

I korthet:

0 **A**

0 **B**

16 **C**

56 **D**

217 **E**

Antal med HTTPS: 14

*Tips: använd [Dataskydd.net:s Webb koll](#) för att testa din egen sajt (eller någon annans)!*

Visa  kommuner

Sök :

Kommun	Betyg	HTTP/HTTPS	Referrers	Kakor totalt	Kakor 1:a	Kakor 3:e	Tredjeparter
<a href="#">Ale</a>	<b>D</b>	HTTP	Ja	8	8	0	10
<a href="#">Alingsås</a>	<b>D</b>	HTTP	Ja	8	8	0	4
<a href="#">Alvesta</a>	<b>E</b>	HTTP	Ja	8	3	5	6
<a href="#">Aneby</a>	<b>D</b>	HTTP	Ja	5	5	0	4

Data leakage to ISPs, schools, work, etc.

Data leakage to adjacent websites.

Data leakage to advertisers, CDNs, font providers, etc.

*“All browsing activity should be considered private and sensitive.”*

— [HTTPS://HTTPS.CIO.GOV/](https://https.cio.gov/)

*“Pervasive monitoring is a technical attack that should be mitigated in the design of IETF protocols, where possible.”*

— INTERNET ENGINEERING TASK FORCE, RFC 7258, “PERVASIVE MONITORING IS AN ATTACK”



*“We’re in a world where if your adversary can see your traffic ... and your traffic is unencrypted, that is an attack vector – not an information leak. This is key: **unencrypted traffic is a vulnerability.**”*

— NICHOLAS WEAVER, “THE GOLDEN AGE OF BULK SURVEILLANCE”, USENIX ENIGMA 2016

www.openbsd.org

norwegian

0h 42m Flight Tracker

-1°C at Oslo


Hardware Platforms  
Security [Crypto](#)  
Events [Papers](#) [Innovations](#)

Getting OpenBSD  
[Buy CDs/Shirts/Posters](#)  
[Download](#)

Getting Source  
[AnonCVS](#)  
[CVSync](#)  
[CVS on Web](#)  
[Daily Changelog](#)

OpenBSD Resources

Free, functional, and secure



# OpenBSD 5.6

Only two remote holes in the default install, in a heck of a long time!

The OpenBSD project produces a **FREE**, multi-platform 4.4BSD-based UNIX-like operating system. Our efforts emphasize portability, standardization, correctness, [proactive security](#) and [integrated cryptography](#). As an example of the effect OpenBSD has, the popular [OpenSSH](#) software comes fr

view-source:http://www.openbsd.org/

```
1 <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
2 <html>
3 <head><link href="http://wifi.norwegian.com/unb/unb.css"
4   rel="stylesheet" type="text/css">
5 <script type="text/javascript" src="http://wifi.norwegian.com
6   /unb/jqr44-1.8.3.js"></script>
7 <script type="text/javascript">var r44_btime=new Date();var
8   r44_smu_time=1455916733.232</script>
9 <script type="text/javascript" src="http://wifi.norwegian.com
10  /unb/unb.js"></script>
11 <title>OpenBSD</title>
```

# GitHub battles “largest DDoS” in site’s history, targeted at anti-censorship tools

HTTP hijacking used to redirect Baidu search engine traffic into a massive DDoS.

by **Sebastian Anthony** - Mar 30, 2015 1:19pm CEST



Share



Tweet



Email



61

GitHub, the largest public code repository in the world, is currently battling against the largest and most gnarly distributed denial of service (DDoS) attack in the site's history. The attack started on Thursday morning (March 26) and has continued unabated since then, evolving several times to circumvent

# Meet “Great Cannon,” the man-in-the-middle weapon China used on GitHub

Powerful weapon could easily be used to inject malware attacks into traffic.

by **Dan Goodin** - Apr 10, 2015 6:32pm CEST



Share



Tweet



Email



61

(Ars Technica)

[Chromium](#) > [Chromium Security](#) >

## Deprecating Powerful Features on Insecure Origins

# Mozilla Security Blog



## Deprecating Non-Secure HTTP

[Chromium](#) > [Chromium Security](#) >

## Marking HTTP As Non-Secure

Geolocation API removed from unsecured origins in Chrome 50



# Let's Encrypt

Data leakage to ISPs, schools, work, etc.

Data leakage to adjacent websites.

Data leakage to advertisers, CDNs, font providers, etc.

**Request URL:** https://maxcdn.bootstrapcdn.com/font-awesome/4.3.0/css/font-awesome.min.css?

**Request method:** GET

**Remote address:** 108.161.188.218:443

**Status code:** ▲ 304 Not Modified

Edit and Re

**Version:** HTTP/1.1

🔍 Filter headers

▶ Response headers (0.443 KB)

▼ Request headers (0.493 KB)

**Host:** "maxcdn.bootstrapcdn.com"

**User-Agent:** "Mozilla/5.0 (X11; Linux x86\_64; rv:49.0) Gecko/20100101 Firefox/49.0"

**Accept:** "text/css,\*/\*;q=0.1"

**Accept-Language:** "en-US,en;q=0.5"

**Accept-Encoding:** "gzip, deflate, br"

**Referer:** "https://afsp.org/find-support/im-having-thoughts-of-suicide/"

**Connection:** "keep-alive"

*“Note: Because the source of a link may be private information or may reveal an otherwise private information source, it is strongly recommended that the user be able to select whether or not the Referer field is sent.”*

— RFC 1945, HYPERTEXT TRANSFER PROTOCOL–HTTP/1.0, 10.13,  
MAY 1996

# *Referrer Policy*, W3C draft

`<meta name="referrer" content="no-referrer">`

*or*

HTTP header:

`Content-Security-Policy: referrer no-referrer`

*(Soon: Referrer-Policy: no-referrer)*



Data leakage to ISPs, schools, work, etc.

Data leakage to adjacent websites.

Data leakage to advertisers, CDNs, font providers, etc.

# Results for **edition.cnn.com**

Input URL: <http://cnn.com/>

Final URL: <http://edition.cnn.com/>



Insecure



Referrers leaked

72

Cookies

194

Third-party requests

67

Third-parties contacted



**Pinboard**

@Pinboard



Following

At this point people without ad and script blockers are like those poor kids born without an immune system

RETWEETS

87

LIKES

117



NEW  
STORY  
STUDIO

6:38 PM - 10 Jun 2016

Google Analytics Solutions

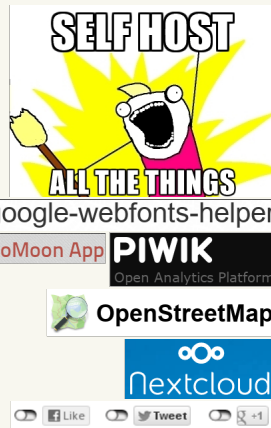
maxCDN

Google Fonts

DISQUS

Google Maps

f Like Share

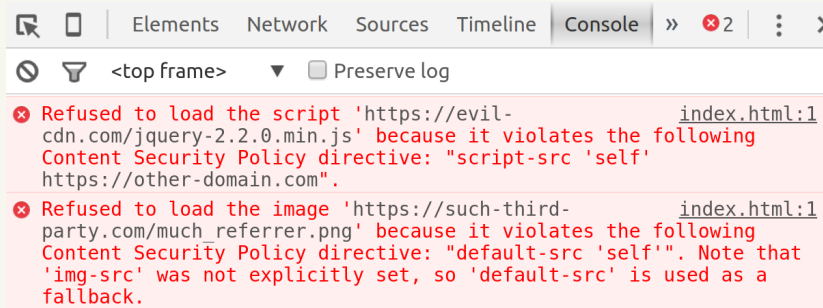


*Alternatives:*

<https://dataskydd.net/nightlybuild2016>

# Content Security Policy

Content-Security-Policy: default-src 'self';  
script-src 'self' https://other-domain.com



Check & build: <https://report-uri.io/home/tools>



# How privacy-friendly is your site?

[Check](#)

This tool helps you check what data-protecting measures a site has taken to help you exercise control over your privacy. [Read more.](#)

*Please note that this service is still under development. Some sites (sometimes) don't work; sometimes results are incorrect. We're working on it! **Also note** that the backend is currently running on only one server with very limited resources, so in case of usage spikes, waiting times can be long. (But you can [run your own instance](#)!) [Feedback](#) is appreciated.*

Test results are stored in our database for a week. We don't show a list of tested URLs. We don't use URLs or test results. We don't log accesses and we don't use cookies.

<https://webbkoll.dataskydd.net/en>



Internetfonden



Funding: Internetfonden / The Internet Foundation IIS

Results for **www.sueddeutsche.de**Input URL: <http://sueddeutsche.de/>Final URL: <http://www.sueddeutsche.de/>

Check again

© 2016-09-02 10:12:27



Insecure



Referrers leaked

34

Cookies

98

Third-party requests

40

Third-parties contacted

## Insecure connection

[www.sueddeutsche.de](http://www.sueddeutsche.de) does **not** use HTTPS by default.

HTTPS encrypts nearly all information sent between a client and a web service. Properly configured, it guarantees three things:

- **Confidentiality.** The visitor's connection is encrypted, obscuring URLs, cookies, and other sensitive metadata.
- **Authenticity.** The visitor is talking to the "real" website, and not to an impersonator or through a "man-in-the-middle".
- **Integrity.** The data sent between the visitor and the website has not been tampered with or modified.

A plain HTTP connection can be easily monitored, modified, and impersonated. Every unencrypted HTTP request reveals information about a user's behavior, and the interception and tracking of unencrypted browsing has become commonplace.

The goal of the Internet community is to establish encryption as the norm, and to phase out unencrypted connections. See [W3C](#), [IETF](#), [IAB](#). Also:

- Browsers support [HTTP/2](#) — which improves page loading speeds — only over encrypted connections.
- Google Chrome (1, 2) and Mozilla Firefox (1) will mark plain HTTP as affirmatively non-secure and make powerful features impossible to use on non-secure sites.
- Google has begun to [favor HTTPS websites in search rankings](#).

To enable HTTPS on a website, a **certificate** for the domain needs to be installed on the web server. To get a certificate that browsers will trust, you need one issued by a trusted certificate authority (otherwise a visitor's browser will show a warning).

[Let's Encrypt](#) is a non-profit certificate authority (sponsored by Mozilla, EFF, Cisco, Facebook and others) providing free domain-validated (

To get a DV certificate, you only need to prove that you control the [domain](#). To get an [Extended Validation \(EV\)](#) certificate, you must pass a more thorough identity verification process.

There is no difference in encryption between DV and EV certificates, but they are typically displayed differently in browsers. EV certificates generally result in the domain owner's name appearing in the browser URL bar that visitors see.

DV certificates are the most common. Let's Encrypt only issues DV certificates.

## Referrers leaked

When you click a link, your browser will typically send the HTTP `referrer` [sic] header to the webserver where the destination webpage is at. The header contains the full URL of the page you came from. This lets sites see where traffic comes from. The header is also sent when external resources (such as images, fonts, JS and CSS) are loaded.

The referrer header is privacy nightmare as it allows websites and services to track you across the web and learn about your browsing habits (and thus possibly private, sensitive information), particularly when combined with cookies.

Let's say you're logged in on Facebook. You visit a page with the URL `http://www.some-hospital.com/some-medical-condition`. On that page, you click a link to their Facebook page. Your browser then sends `Referer: http://www.some-hospital.com/some-medical-condition` to `facebook.com`, along with your Facebook cookies, allowing Facebook to associate your identity with that particular page.

The problem is made worse by the fact that many websites load resources like images and scripts from dozens of third-parties, sending referrer information to all of them, with the typical visitor having no idea that this is happening.

## Third-party services

The site is loading libraries from one or more CDNs.

The site is using Google Analytics. While this is a powerful tool, we think you should respect your users' privacy and not tell Google about them — at least not without your users' consent.

The site loads fonts from Google Fonts. While these are hosted on resource-specific domains and no cookies are sent, Google *could* possibly cross-reference the data (IP and browser fingerprint) with other Google services to identify visitors. Do they? Their own [FAQ](#) is vague: "We do log records of the CSS and the font file requests, and access to this data is on a need-to-know basis and kept secure." What "need-to-know basis" means is not explained.

Thanks to a fairly recent development, [Referrer Policy](#), it's finally possible for websites to tell browsers to not leak referrers. It lets you specify a policy that's applied to all links clicked, as well as all other requests generated by the page (images, JS, etc.).

A few different policies are offered, such as `origin` (strips everything except the `origin`) and `origin-when-cross-origin` (sends full URL with same-origin requests, otherwise stripped). The only one we recommend is `no-referrer`, which kills the referrer header entirely for all requests, no matter the destination.

A referrer policy can easily be set with a `<meta>` element in your HTML. Simply include this inside the `<head>` section:

```
<meta name="referrer" content="no-referrer">
```

While still a work in progress, Referrer Policy is now [supported by all major browsers](#) (except Internet Explorer, although it is supported by Edge, the new browser in Windows 10).

Self-host the files.

Piwik is an excellent alternative. It's free software (PHP & MySQL) and you run it on your own server, meaning you are in control of the data. It offers various privacy settings and, unlike Google Analytics, it can be used without cookies. (While analytics might be considered essential by some websites, another alternative is don't track people just because you can. Visitors do not, in fact, have an implicit obligation to help you optimize things.)

Fonts can easily be self-hosted. The tool [google-webfonts-helper](#) lets you select one or more fonts, generates the proper CSS and prepares a zip file with the fonts. For a command-line alternative, the shell script [google-font-download](#) provides similar functionality.

## First-party cookies

21 first-party cookies.

Domain	Name	Value	Expires on
sueddeutsche.de	__ga	GA1.2.1069439662.147...	2018-09-02 10:12:16Z
sueddeutsche.de	__cGtmAD	1	2016-09-02 10:42:16Z
sueddeutsche.de	__gat_UA-19474199-5	1	2016-09-02 10:22:16Z
sueddeutsche.de	__dc_gtm_UA-19474199-5	1	2016-09-02 10:22:16Z
sueddeutsche.de	__lpl4_u	Kg4dm2qz7	2017-09-02 10:12:16Z
sueddeutsche.de	__gads	ID=85c011f2e6792c88...	2018-09-02 10:12:16Z
sueddeutsche.de	__cGtmS	676982192.1	2016-09-02 10:42:16Z
sueddeutsche.de	__utmtz	6611437.1472811136.1...	2017-03-03 22:12:15Z
sueddeutsche.de	__utmc	6611437	session
sueddeutsche.de	__utmb	6611437.1.10.1472811...	2016-09-02 10:42:15Z
sueddeutsche.de	__utma	6611437.1069439662.1...	2018-09-02 10:12:15Z
sueddeutsche.de	__utmt	1	2016-09-02 10:22:15Z
sueddeutsche.de	creid	1544354409074455532	2037-12-31 23:55:55Z
sueddeutsche.de	BiGipServerIb-pay_http	1107959468.20480.000...	session
sueddeutsche.de	BiGipServerIb-pol-web_varnish	3087670956.20480.000...	session
www.sueddeutsche.de	__chartbeat2	.1471983093734.14728...	2017-10-02 10:12:16Z
www.sueddeutsche.de	__cb	CmhbfAFD6ujQ00bZa47	2017-10-02 10:12:16Z
www.sueddeutsche.de	__cb_is	1	2017-10-02 10:12:16Z
www.sueddeutsche.de	BiGipServerIb-play-prod	3288997548.20480.000...	session
www.sueddeutsche.de	BiGipServerIb-pay_http	1107959468.20480.000...	session
www.sueddeutsche.de	BiGipServerIb-pol-web_varnish	3087670956.20480.000...	session



## Third-party cookies

13 third-party cookies.

Domain	Name	Value	Expires on
.adnxs.com	uuid2	3443694693045707732	2016-12-01 10:12:17Z
.adnxs.com	sess	1	2016-09-03 10:12:17Z
.doubleclick.net	DSID	HO_DATA	2016-09-02 11:12:18Z
.doubleclick.net	IDE	AHWqTUALDxIB19WJx7b...	2018-09-02 10:12:16Z
.doubleclick.net	id	22ef317133060027[[#...	2018-09-02 10:12:16Z
.oam.de	i00	0023908aa1ee2a0e557c...	2017-05-22 18:46:09Z
.t4t.de	ftgnetid	MjYsNjASMDIsMjE5yMDAw...	2017-09-02 10:12:16Z
.theadex.com	axd	1000810802211330000	2017-09-02 10:12:16Z
.theadex.com	tis	EP1%3A1104	2026-08-31 10:12:16Z
.twitter.com	pid	"v3:1472811137859900...	2018-03-03 10:12:17Z
.w55c.net	matchgoogle	2	2016-10-02 10:12:15Z
.w55c.net	wfivefivec	ocslpPa3lBFLsj2	2018-09-02 10:12:15Z
.de.sitestat.com	s1	41.162.557C9507F002A...	2021-09-01 10:12:15Z

## Third-party requests

98 requests (34 secure, 64 insecure) to 40 unique hosts.

A third-party request is a request to a domain that's not `sueddeutsche.de` or one of its subdomains.

Host	Classification
ad.yieldlab.net	Advertising (Yieldlab)
c14lt.de	
cdn.emetrix.de	
cdn.lqcontentplatform.de	
cdn.m-pathy.com	
cdn.syndication.twimg.com	Disconnect (Twitter)
cdnjs.cloudflare.com	
cm.g.doubleclick.net	Disconnect (Google)
connect.facebook.net	Disconnect (Facebook)
de.ioam.de	Analytics (INFOline)
de.sitestat.com	Analytics (comScore)
dnsptheadex.com	
dyn.emetrix.de	
fonts.googleapis.com	Content (Google)
glaring-torch-3314.firebaseio.com	
googleads.g.doubleclick.net	Disconnect (Google)
ib.adnxs.com	Advertising (AppNexus)
js.moataads.com	Advertising (Moat)
p.lp4.io	
pagead2.googlesyndication.com	Disconnect (Google)
pde.lp4.io	
ping.chartbeat.net	Analytics (Chartbeat)
platform.twitter.com	Disconnect (Twitter)

## HTTP headers

Header	Set?
<b>Content-Security-Policy</b>	<b>✗ NO</b> <a href="#">Content Security Policy</a> is an effective measure to protect your site from <a href="#">XSS</a> attacks. By whitelisting sources of approved content, you can prevent the browser from loading malicious assets. It can also help prevent information leakage.
<b>Public-Key-Pins</b>	<b>✗ NO</b> <a href="#">HTTP Public Key Pinning</a> protects your site from <a href="#">MITM attacks</a> using rogue X.509 certificates. By whitelisting only the identities that the browser should trust, your users are protected in the event a certificate authority is compromised.
<b>Strict-Transport-Security</b>	<b>✗ NO</b> <a href="#">HTTP Strict Transport Security</a> is an excellent feature to support on your site and strengthens your implementation of TLS by getting the User Agent to enforce the use of HTTPS.
<b>X-Content-Type-Options</b>	<b>✗ NO</b> <a href="#">X-Content-Type-Options</a> stops a browser from trying to MIME-sniff the content type and forces it to stick with the declared content-type. This helps to reduce the danger of drive-by downloads. The only valid value for this header is "X-Content-Type-Options: nosniff".
<b>X-Frame-Options</b>	<b>✗ NO</b> <a href="#">X-Frame-Options</a> tells the browser whether you want to allow your site to be framed or not. By preventing a browser from framing your site you can defend against attacks like clickjacking.
<b>X-Xss-Protection</b>	<b>✗ NO</b> <a href="#">X-XSS-Protection</a> sets the configuration for the cross-site scripting filters built into most browsers. The best configuration is "X-XSS-Protection: 1; mode=block".

```

defp get_cookies(cookies, registerable_domain) do
  {first, third} =
    Enum.partition(cookies, fn(x) ->
      (x["domain"] |> String.trim(".")) |> get_registerable_domain) == registerable_domain
    end)
  %{"first_party" => first, "third_party" => third}
end

defp get_cookie_count(cookies) do
  %{"first_party" => Enum.count(cookies["first_party"]),
    "third_party" => Enum.count(cookies["third_party"])}
end

defp get_meta_referrer(content) do
  content
  |> Floki.find("meta[name='referrer']")
  |> Floki.attribute("content")
  |> List.to_string
end

```

Try it: <https://webbkoll.dataskydd.net/>

(English/Swedish) (*no cookies!*)

Code: <https://github.com/andersju/webbkoll>

(MIT license)

# What else?

☞ Advocacy. Strategy: mimic legal texts.

*Ex: government hacking, incident reporting*

☞ Separating “normal” from “crisis”!

*Ex: privacy, security and economics*

☞ Data protection requires data security

but the reverse is **not** true.

*E.g. security is for the agent with money and this is frequently not the consumer or the citizen*

# *Thank you!*

Slides, links, etc.: <https://dataskydd.net/nightlybuild2016>

Amelia Andersdotter  
@teirdes  
[amelia.andersdotter@dataskydd.net](mailto:amelia.andersdotter@dataskydd.net)

Anders Jensen-Urstad  
@ndrsju  
[anders.jensen-urstad@dataskydd.net](mailto:anders.jensen-urstad@dataskydd.net)  
[anders.unix.se](mailto:anders.unix.se)

**dataskydd.net**