

UNCOVERING AND ANALYZING OBFUSCATED WEB TRACKING

INTRODUCTION

- With the prevalence and pervasiveness of web fingerprinting and tracking, Internet users do not know when and how they are tracked in general,
- Most advanced tracking is based on JavaScript, we are validating our hypothesis on web tracking and studying the phenomenon of obfuscated web tracking,
- The hypothesis is that web tracking has started to use obfuscated programming.

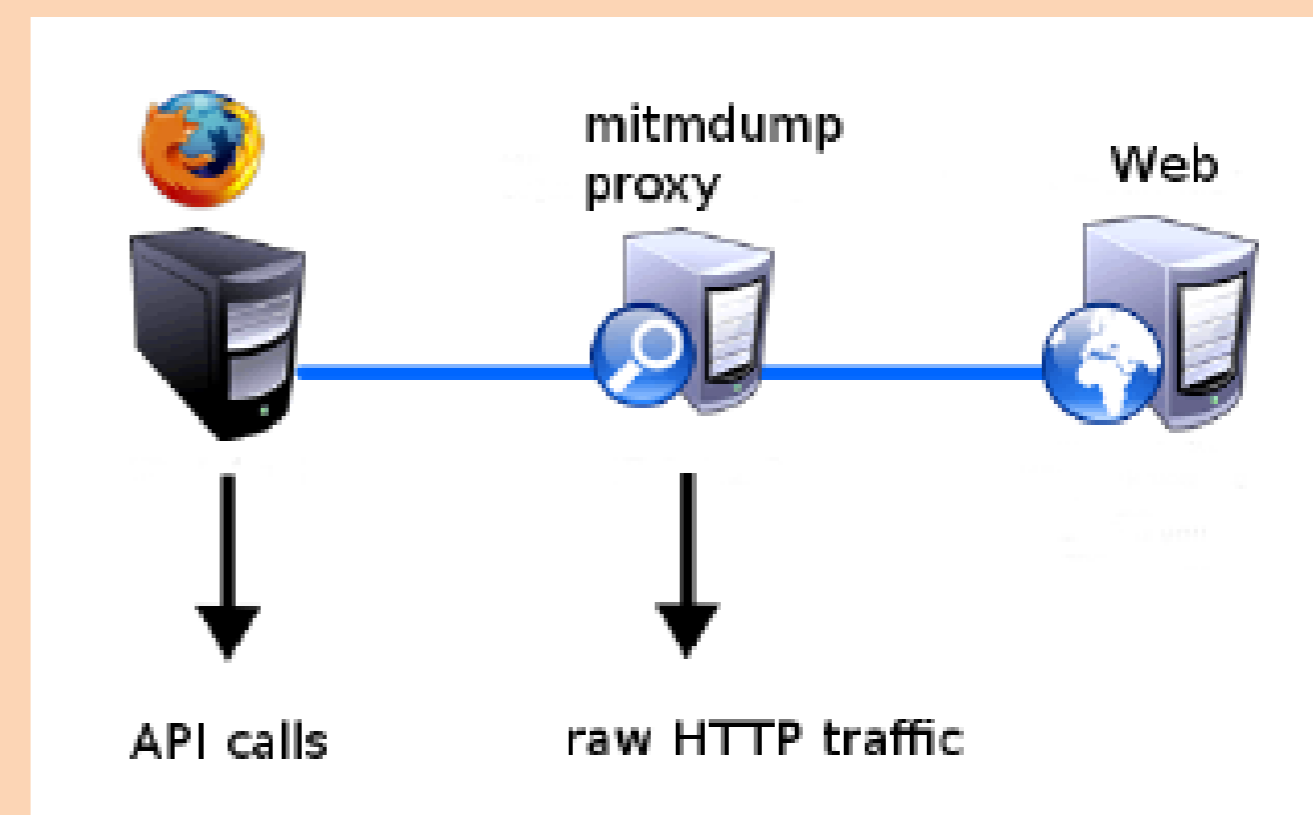
TRACKING MECHANISMS - TECHNOLOGIES

Session-only	Web-server session, HTML5, JavaScript
Storage-based	HTTP headers, JavaScript, Flash, Java, Silverlight, HTML5
Cache-based	HTML5, JavaScript, Server-side measurements, HTTP headers, Web-server session
Fingerprinting-based	IP address, Server-based geolocation techniques, HTTP headers, HTML5, JavaScript, Flash, Java, TCP headers
Other trackings	HTTP headers, Smartphone malware, HTML5, JavaScript, CSS, Web-server session, Silverlight, Java

GOAL

- Develop a browser extension,
- Collect the necessary datasets,
- Analyze the obfuscated tracking phenomenon.

OUR METHODOLOGY



- Intercepting API calls in the browser,
- Collecting HTTP traffic in a proxy,
- Analyzing of the intercepted API calls and the mitmdump log, JavaScript codes and functions,
- Applying a machine learning solution and distributed measurement to our problem.



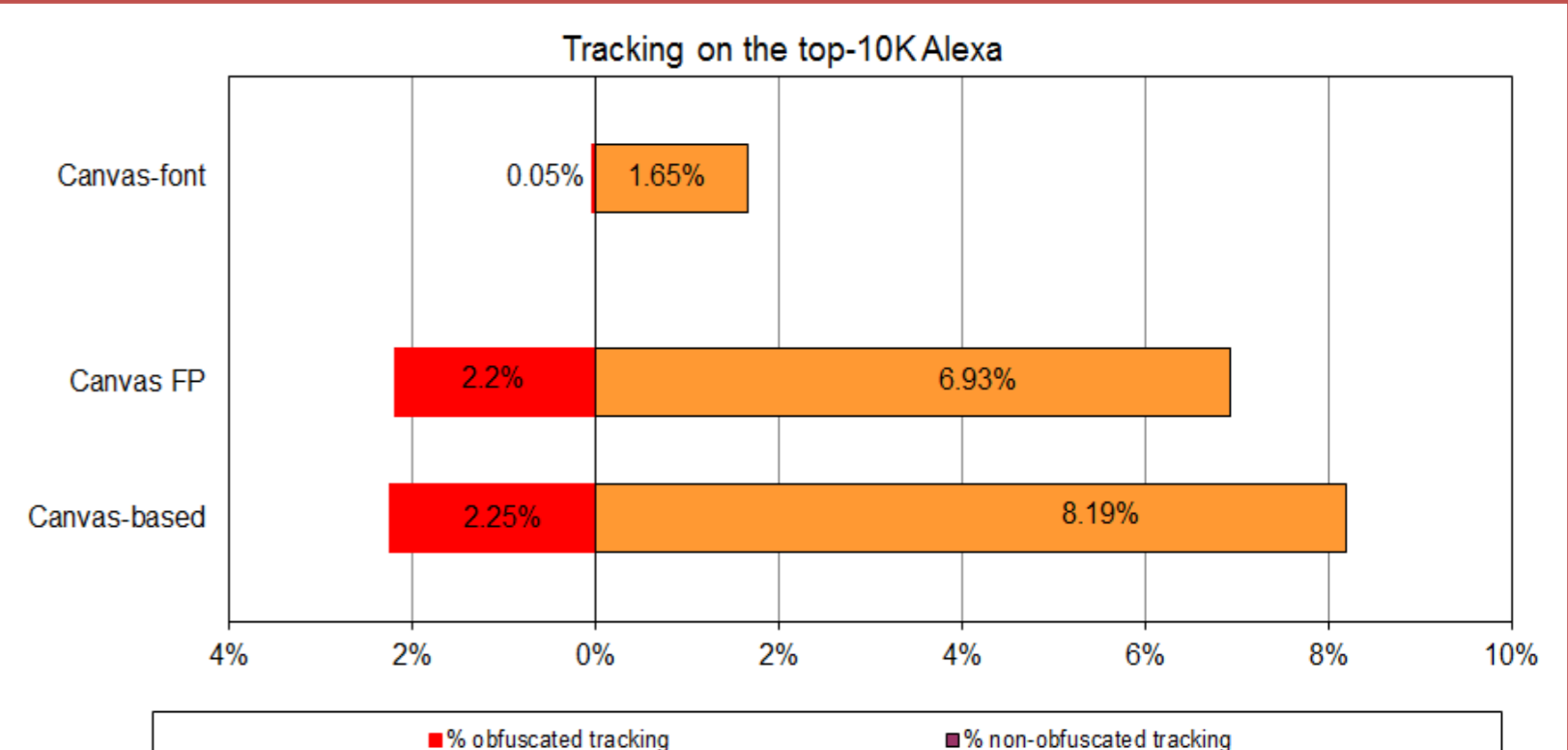
A script on alicdn.com from capturing the following 400px x 60px canvas (via toDataURL)

Cwm fjordbank glyphs vext quiz, ☺

A script on tmall.com from capturing the following 400px x 60px canvas (via toDataURL)

Cwm fjordbank glyphs vext quiz, ☺

RESULTS



Tracking	Top-100	Top-1K	Top-10K
Non-obfuscated	7%	5%	6.93%
Obfuscated	1%	1%	2.2%

	Canvas tracking	Obfuscated canvas tracking
Top-100 homepages only	8/100	1/100
Top-100 websites	47/100	24/100

CONCLUSIONS

- Uncovered obfuscated canvas-based tracking and efficiently detected canvas fingerprinting on the Internet,
- Confirmed the existence of obfuscated canvas-based tracking on the top-10K most visited Alexa's ranking websites,
- At least 10.44% of the analyzed websites used canvas-based tracking and 2.25% obfuscated it on the top-10K homepages,
- Tracking methods can be more present on sub pages than on landing pages.

Canvas/Obfuscation	Name websites	Canvas by OpenWPM
Yes/No	taobao.com	Yes
Yes/No	qq.com	No
Yes/No	sina.com.cn	No
Yes/No	hao123.com	No
Yes/No	sohu.com	No
Yes/No	tmall.com	Yes
Yes/Yes	microsoft.com	No
Yes/No	espn.com	No

Tracking on the top-100 Alexa

