



DATA
61



Towards Seamless tracking-free web browsing: Improved detection of tracking JavaScripts via one-class learning

Muhammad Ikram, Hassan Asghar, Mohamed Ali Kaafar, Anirban Mahanti, Balachander Krishnamurthy

www.csiro.au



UNSW
THE UNIVERSITY OF NEW SOUTH WALES



The screenshot shows the homepage of the New York Times website. At the top, there is a navigation bar with a search bar containing "Search" and various icons. Below the search bar are buttons for "SUBSCRIBE NOW" and "LOG IN". The main header features the "The New York Times" logo in large, bold letters. To the left of the logo, a blue box displays the text "October 23–25, 2017 Dove Mountain, Arizona". Below the logo, the date "Saturday, July 8, 2017" and other links like "Today's Paper" and "Video" are visible. To the right, there is a promotional box for "The New York Times Corner Office MASTER CLASS". The bottom navigation bar lists categories such as World, U.S., Politics, N.Y., Business, Opinion, Tech, Science, Health, Sports, Arts, Style, Food, Travel, Magazine, T Magazine, Real Estate, and ALL.

What happens when a browser requests an HTML source from a web server (e.g., nytimes.com)?

≡ SECTIONS Q SEARCH

ENGLISH 中文 (CHINESE) ESPAÑOL

SUBSCRIBE NOW

LOG IN

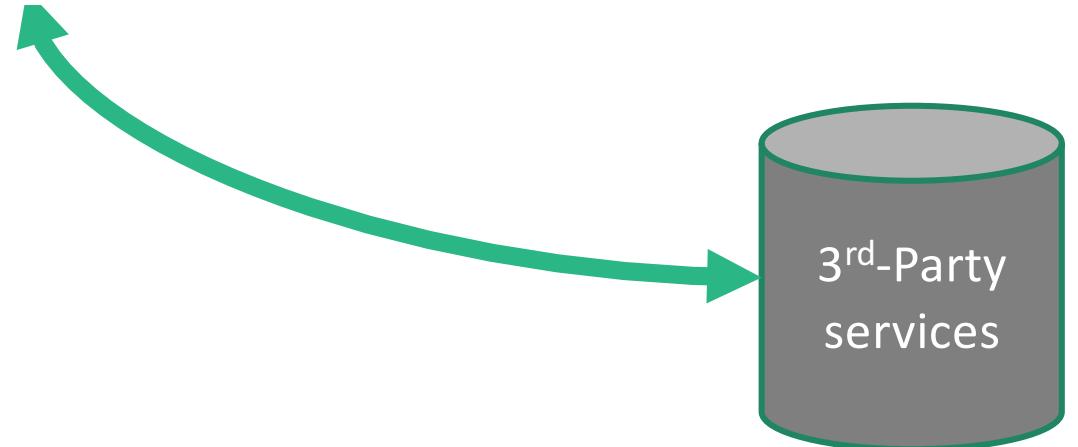


October 23–25, 2017

The New York Times

The New York Times
Corner Office

```
> <script async="" src="//cdn.krxd.net/ctjs/controltag.js.c3a35cae6beb84887ab6a83973d97fa9">
> <script src="https://s.yimg.com/wi/ytc.js" async="">
> <script type="text/javascript" async="" src="https://z.moatads.com/googleessencenyt485873431/moatcontent.js">
> <script async="" src="//www.google-analytics.com/analytics.js">
> <script type="text/javascript" async="" src="https://cdn.krxd.net/controltag/HrUwtkcl.js">
> <script async="" src="//static.ads-twitter.com/uwt.js">
> <script async="" src="https://s.pinimg.com/ct/core.js">
> <script src="https://connect.facebook.net/signals/config/592202027582499?v=2.7.18" async="">
> <script async="" src="https://connect.facebook.net/en_US/fbevents.js">
```



https://www.nytimes.com

SEARCH

SECTIONS

SUBSCRIBE NOW

LOG IN

ENGLISH 中文 (CHINESE) ESPAÑOL

October 23–25, 2017
Dove Mountain, Arizona

Saturday, July 8, 2017 | Today's Paper | Video | 76°F | Dow +0.44% ↑

> <script async="" src="//cdn.krxd.net/ctjs/controltag.js.c3a35cae6beb84887ab6a83973d97fa9">

> <script src="https://nyt.yimg.com/mi/ptc.js" async="">

U.S. Politics / NY Business Opinion Tech Science Health Sports Arts Style Food Travel Magazine T Magazine Real Estate ALL

The New York Times

The New York Times Corner Office MASTER CLASS

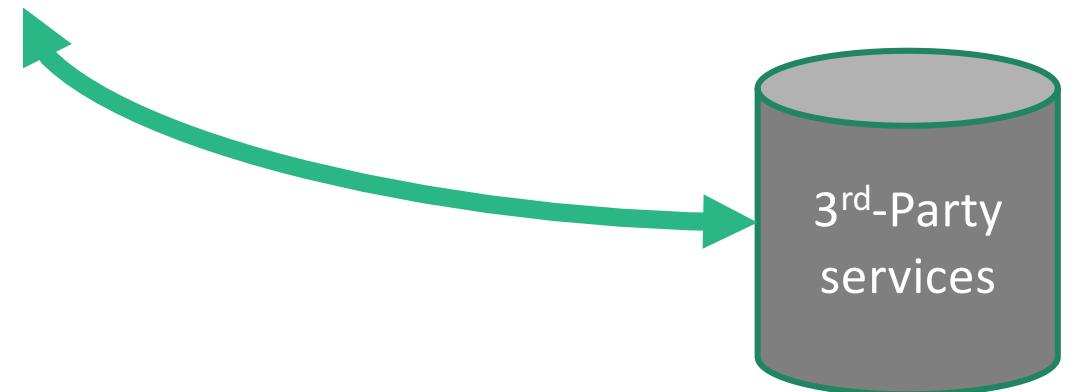
The Time to Switch Health Funds is Now. Here's Why

HEALTHINSURANCECOMPARISON.COM.AU



A red box highlights the tracking script in the header and the advertisement banner.

> <script async="" src="https://connect.facebook.net/en_US/fbevents.js">



https://www.nytimes.com

SEARCH

SUBSCRIBE NOW

LOG IN

ENGLISH 中文 (CHINESE) ESPAÑOL

October 23–25, 2017
Dove Mountain, Arizona

Saturday, July 8, 2017 | Today's Paper | Video | 76°F | Dow +0.44% ↑

The New York Times

Corner Office MASTER CLASS

<script async="" src="//cdn.krxn.net/ctjs/controltag.js.c3a35cae6beb84887ab6a83973d97fa9">

<script async="https://ev.vimgo.com/wi/xtc.js" src="">>

Arts Style Food Travel Magazine T Magazine Real Estate All

The Time to Switch Health Funds is Now.
Here's Why
HEALTHINSURANCECOMPARISON.COM.AU

Health Insurance Comparison

<script src="https://connect.facebook.net/en_US/fbevents.js">

Cookies DB

| | | |
|-------------|--|------------------|
| ► krux_segs | q6yvrtzf6 q8xmzq99j | .nytimes.com |
| ► idb | WET3iYNVWAAOi4xSWWGpSIKwryQ | .impdesk.com |
| ► id | 22ce55597d0b0045 t=1473...213fd480d0e6d98520841f8 | .doubleclick.net |
| ► et-a1 | %7B%22agentId%22%3A%2220...e%22%3A1499572551257%7D | .et.nytimes.com |

3rd-Party services

A green curved arrow originates from the 'Cookies DB' table and points towards a large grey cylinder labeled '3rd-Party services'. The cylinder has a small 'i' icon on its side.

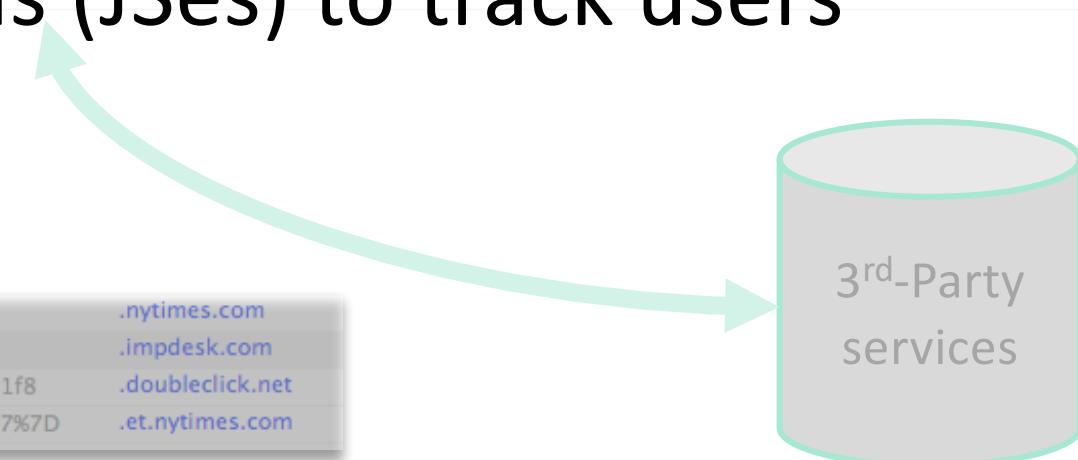
The screenshot shows the top navigation bar of The New York Times website. It includes a search bar, a 'SUBSCRIBE NOW' button, a 'LOG IN' button, and a 'DATA 61' icon. A green arrow points from the 'DATA 61' icon towards the main content area. On the left, there's a sidebar with the text 'October 23–25, 2017 Dove Mountain, Arizona'. The main header features 'The New York Times' logo. Below it, the date 'Saturday, July 8, 2017' and various news categories like 'World', 'U.S.', 'Politics', etc., are listed. To the right, there's a 'Corner Office MASTER CLASS' advertisement.

The Time to Switch Health Funds is Now.
HEALTHINSURANCECOMPARISON.COM.AU

Ads and cookies (and social widgets) are enabled by JavaScript programs (JSes) to track users

Cookies DB

| | | |
|-------------|--|------------------|
| ► krux_segs | q6yvrtzf6 q8xmzq99j | .nytimes.com |
| ► idb | WET3iYNVWAAOi4xSWWGpSIKwryQ | .impdesk.com |
| ► id | 22ce55597d0b0045 t=1473...213fd480d0e6d98520841f8 | .doubleclick.net |
| ► et-a1 | %7B%22agentId%22%3A%2220...e%22%3A1499572551257%7D | .et.nytimes.com |



<https://www.nytimes.com>

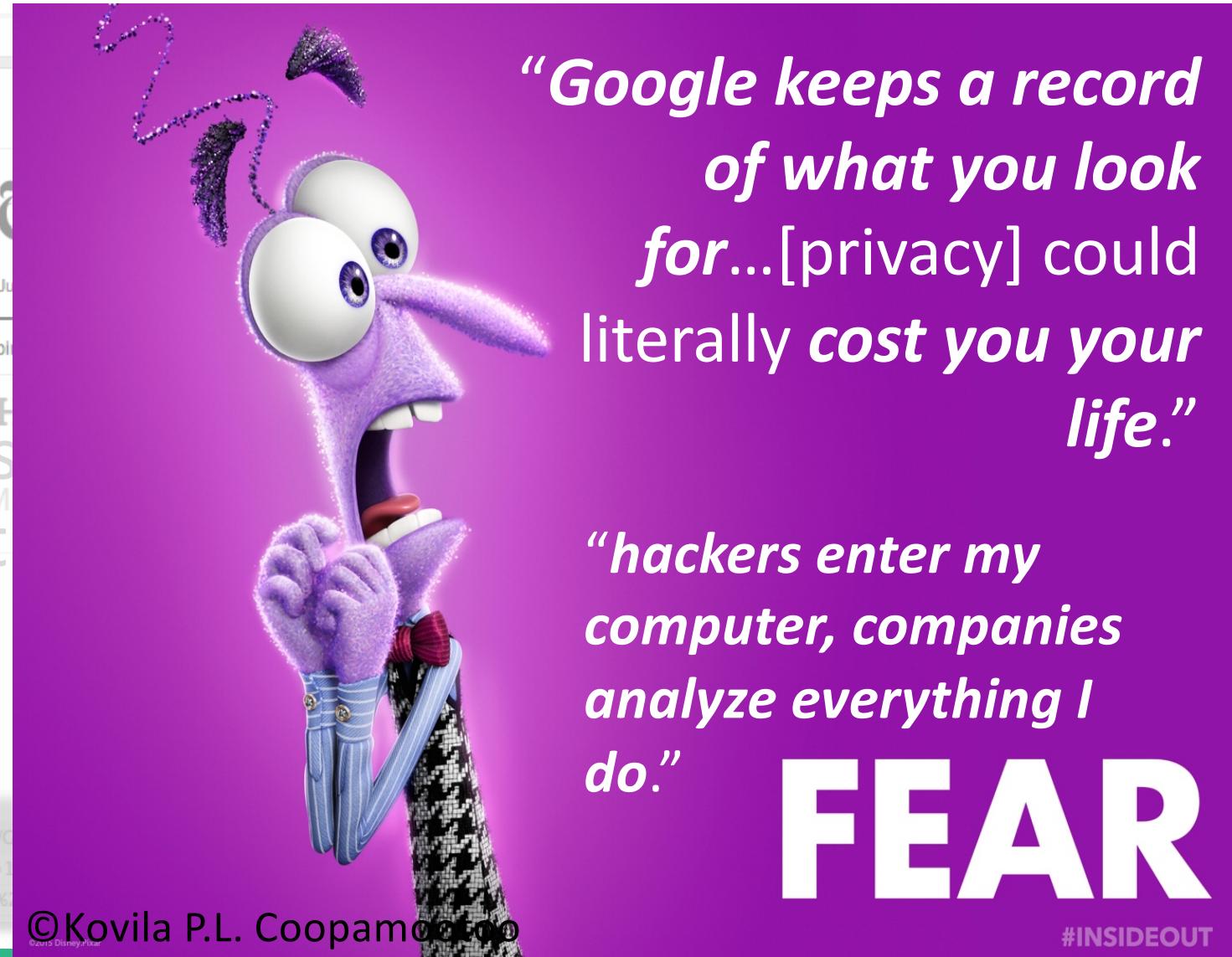
Privacy Concerns

World U.S. Politics N.Y. Business Opin

The Time to Switch H
Ads and cookies
Series Why
HEALTHINSURANCE.COM
JavaScript

Cookies DB

► krux_segs q6yvrtzf6|q8xmzq99j
► idb WET3iYNVWAAQi4xSWC
► id 22ce55597d0b0045||t=1
► et-a1 %7B%22agentId%22%3A%



Privacy Preserving Tools (PP-Tools)



- Blacklist of URLs (or patterns of URLs) to block 3rd-party *tracking* JSes

Privacy Preserving Tools (PP-Tools)



- Blacklist of URLs (or patterns of URLs) to block 3rd-party *tracking JSes*

A screenshot of the 9NEWS.com.au website. The header includes a red '9NEWS .com.au' logo and a blue navigation bar with tabs for NATIONAL, LOCAL, WORLD, 9RAW, and WEATHER. The main headline reads 'Teen's balloon letter to dead dad inspires community 700km away'. Below the headline is a byline 'By Nicholas McCallum' with a small profile picture. A text excerpt starts with 'A teenage girl in the US ...' and ends with a 'Continue reading >' link. A dark purple sidebar on the right lists various tracking scripts, with 'Brightcove' highlighted by a red box.

A screenshot of the 9NEWS.com.au website, identical to the one above but with a different headline: 'Teen's balloon letter to dead dad inspires community 700km away'. The layout is the same, with the 'Brightcove' tracking script highlighted in a red box in the sidebar.

Teen's balloon letter to dead dad inspires community
700km away



Teen's balloon letter to dead dad inspires community
700km away

Privacy Preserving Tools (PP-Tools)



- Blacklist of URL (or patterns of URLs) to block 3rd-party *tracking JSes*

The image shows two side-by-side screenshots of news website pages. Both pages have a similar header with sections for NEWS, NATIONAL, LOCAL, WORLD, 9RAW, and WEATHER.

Left Screenshot: The main headline reads "Teen's balloon letter to dead dad inspires community 700km away". Below the headline is a paragraph by Nicholas McCallum. A sidebar on the right lists various tracking scripts found in the page's source code, including AppNexus, Brightcove, ChartBeat, Facebook Social Plugins, Google AdSense, NetRatings SiteCensus, Post, ScoreCard Research Beacon, and Twitter Button.

Right Screenshot: The main headline reads "Teen's balloon letter to dead dad inspires community 700km away". Below the headline is a paragraph by Nicholas McCallum. A sidebar on the right lists various tracking scripts found in the page's source code, including AppNexus, Brightcove, ChartBeat, DoubleClick, Facebook Social Plugins, Google AdSense, NetRatings SiteCensus, Post, ScoreCard Research Beacon, and Twitter Button.

* Metwally et al., The Online Tracking Horde: a View from Passive Measurements, TMA'15

Privacy Preserving Tools (PP-Tools)



- Blacklist of URL (or patterns of URLs) to block 3rd-party *tracking* JSes

Two side-by-side screenshots of news website interfaces. Both show a main article about a teenage girl's balloon letter to her deceased father. A grey rectangular box is overlaid on both screenshots, containing a list of tracking scripts. The list includes: AppNexus, Brightcove, Doubleclick, Facebook Social Plugins, Google AdSense, NelRatings SiteCensus, Post, ScoreCard Research Beacon, and Twitter Button. The first screenshot shows the '9NEWS.com.au' logo at the top, while the second shows the '9RAW' logo.

* Metwally et al., The Online Tracking Horde: a View from Passive Measurements, TMA'15



**Tracking JSes are similar to each other and are distinct from
Functional JSes**

JavaScript Codes Similarity: Example



Tracker 1. Google Analytics Cookie Setting

```
var _gaq = _gaq || [];
_gaq.push(['_setAccount', 'UA-1627489-1']);
_gaq.push(['_setDomainName', 'geo.tv']);
_gaq.push(['_trackPageview']);
```

Tracker 2. Visual Revenue Cookie Setting

```
var _vrq = _vrq || [],
_vrqIsOnHP = (document.body.className ||
'').search('pg-section') >=0 ? true : false;
_vrq.push(['id', 396]);
_vrq.push(['automate', _vrqIsOnHP]);
_vrq.push(['track', function() {}]);
```

JavaScript Codes Similarity: Example



Tracker 1. Google Analytics Cookie Setting

```
var _gaq = _gaq || [];
_gaq.push(['_setAccount', 'UA-1627489-1']);
_gaq.push(['_setDomainName', 'geo.tv']);
_gaq.push(['_trackPageview']);
```

Tracker 2. Visual Revenue Cookie Setting

```
var _vrq = _vrq || [],
_vrqIsOnHP = (document.body.className ||
'').search('pg-section') >=0 ? true : false;
_vrq.push(['id', 396]);
_vrq.push(['automate', _vrqIsOnHP]);
_vrq.push(['track', function() {}]);
```

JavaScript Codes Similarity: Example



Tracker 1. Google Analytics Cookie Setting

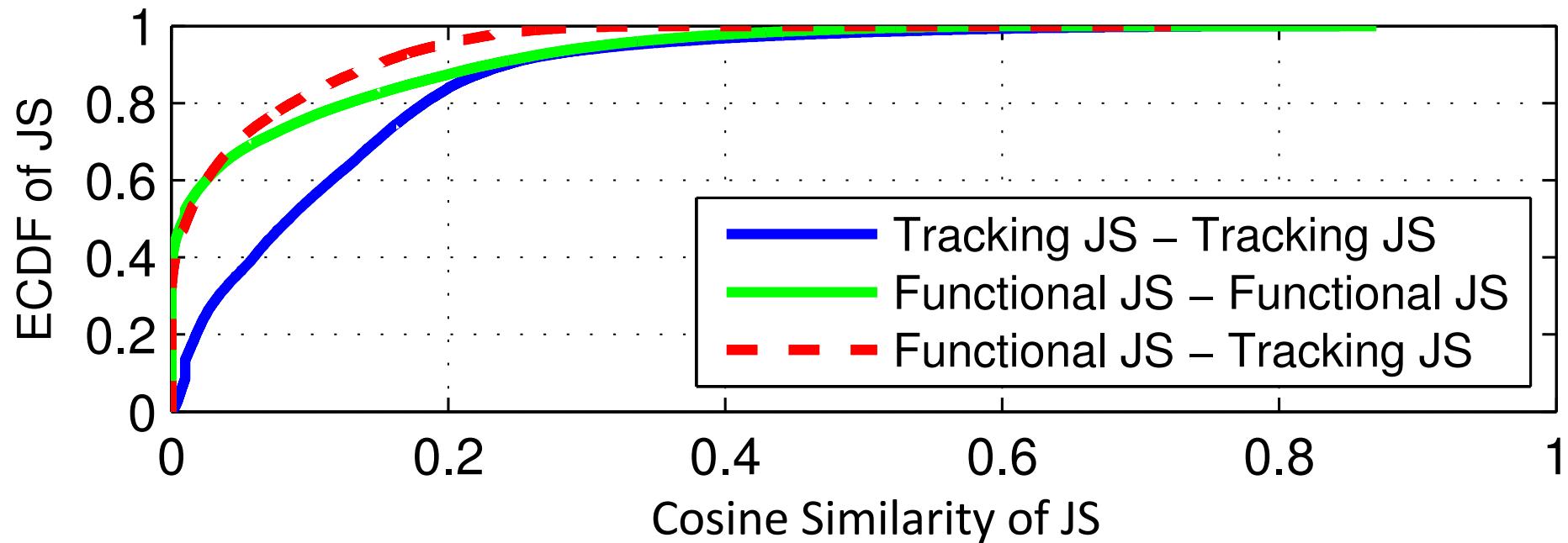
```
var _gaq = _gaq || [];
_gaq.push(['_setAccount', 'UA-1627489-1']);
_gaq.push(['_setDomainName', 'geo.tv']);
_gaq.push(['_trackPageview']);
```

Code snippets are **syntactically and semantically similar with**

~~Tracker 2. Vivaldi cookie setting~~
difference in variable names and values

```
_vrq = vrq || [];
_vrqIsOnHP = (document.body.className ||
'').search('pg-section') >=0 ? true : false;
_vrq.push(['id', 396]);
_vrq.push(['automate', _vrqIsOnHP]);
_vrq.push(['track', function() {}]);
```

JavaScript Codes Similarity: two sets of 500 Functional and 500 Tracking JSes

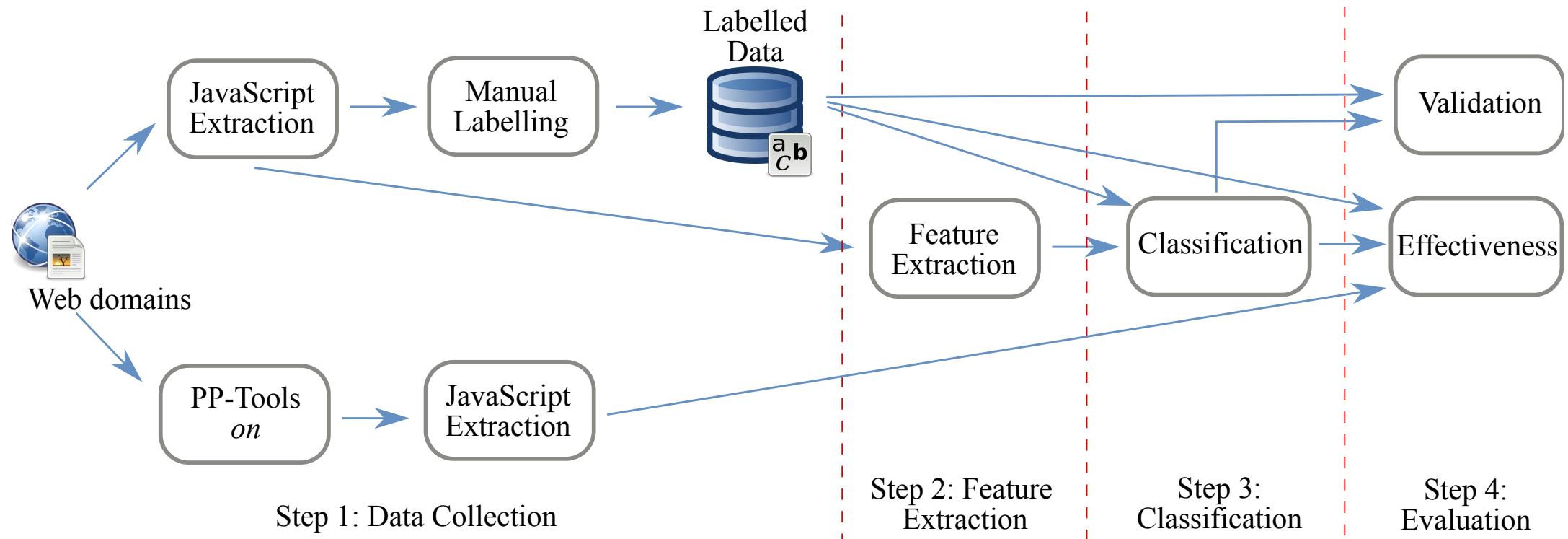


Overview of Our Contribution



- Analyze PP-Tools' performance
 - Regular expressions based on blacklists are ineffective
- Design a classification framework to separate *tracking* JSes from *functional* JSes
 - Training with partial single class of functional or tracking Jses
 - Partial view of tracking JSes from blacklists

Our Methodology



Data Collection: Rules for Labelling JSes



| Rule | JS | # | Description |
|------|----|-----|--|
| R1 | X | 216 | All JS that create panels and set margins for ads |
| R2 | X | 115 | All JS that access and display ads |
| R3 | X | 45 | All social media widgets |
| R4 | X | 324 | All in-page JS that include external JS from third-party analytics and advertisers |
| R5 | X | 353 | All external JS from third-party analytics and advertisers |
| R6 | X | 180 | All cookie enablers, readers or writers |
| R7 | ✓ | 542 | All external JS that provide useful functionality such as navigation menus, search and login |
| R8 | ✓ | 509 | All in-page JS that provide useful functionality |
| R9 | ✓ | 132 | All JS that fetch content from first-party content domains or third-party CDNs |
| R10 | X | 103 | All JS in hidden iframe that belong to third-party analytics, advertisers and social media |
| R11 | X | 40 | All JS in hidden iframe that enable, read or modify cookies |
| R12 | ✓ | 53 | All JS that track mouse or keyboard events |

X:Tracking JS, ✓ : Functional JS

Data Collection: Rules for Labelling JSes



| Rule | JS | # | Description |
|------|----|-----|--|
| R1 | X | 216 | All JS that create panels and set margins for ads |
| R2 | X | 115 | All JS that access and display ads |
| R3 | X | 45 | All social media widgets |
| R4 | X | 324 | All in-page JS that include external JS from third-party analytics and advertisers |
| R5 | X | 353 | All external JS from third-party analytics and advertisers |
| R6 | X | 180 | All cookie enablers, readers or writers |
| R7 | ✓ | 542 | All external JS that provide useful functionality such as navigation menus, search and login |
| R8 | ✓ | 509 | All in-page JS that provide useful functionality |
| R9 | ✓ | 132 | All JS that fetch content from first-party content domains or third-party CDNs |
| R10 | X | 103 | All JS in hidden iframe that belong to third-party analytics, advertisers and social media |
| R11 | X | 40 | All JS in hidden iframe that enable, read or modify cookies |
| R12 | ✓ | 53 | All JS that track mouse or keyboard events |

X:Tracking JS, ✓ : Functional JS

Analyzing PP-Tools



NoScript



Ghostry



Adblock Plus

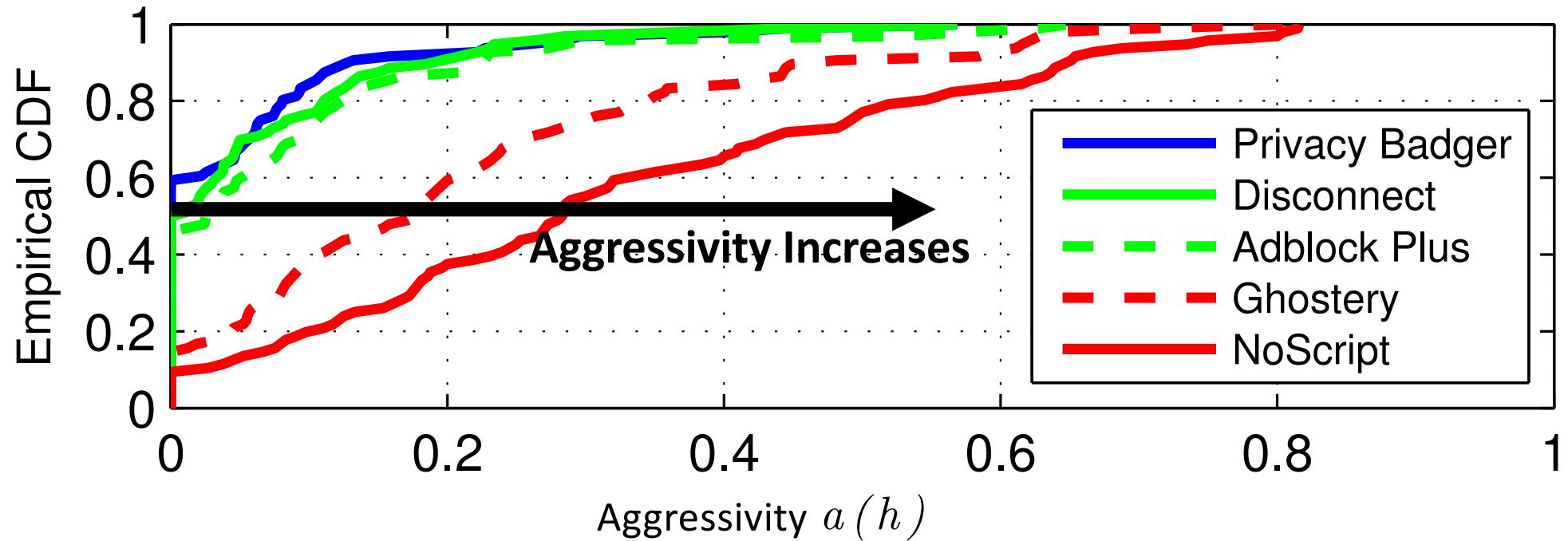


Disconnect



Privacy Badger

Aggressivity of PP-Tools



Aggressivity of a PP-Tool means the ratio of blocked JSes to the total number of JSes in a DOM tree

Aggressivity of PP-Tools

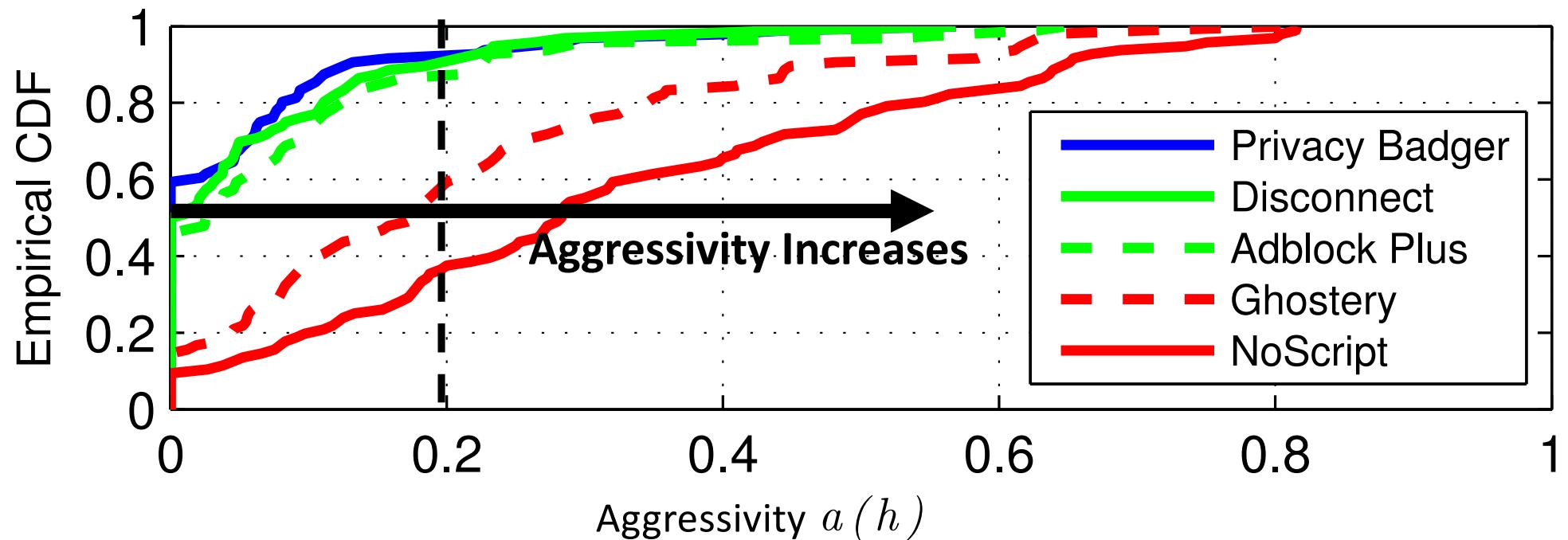


- Aggressivity – the ratio of blocked JSes by a PP-Tool to the total number of JSes in a DOM tree

Aggressivity of PP-Tools



- Aggressivity – the ratio of blocked JSes by a PP-Tool to the total number of JSes in a DOM tree



Effectiveness of PP-Tools



- Effectiveness – the balance between correctly blocking *tracking* JSes and incorrectly blocking *functional* JSes

Effectiveness of PP-Tools



- Effectiveness – the balance between correctly blocking tracking JSes and incorrectly blocking functional JSes

| PP-Tool | Tracking | | Functional | |
|----------------|----------|---------|------------|---------|
| | Blocked | Allowed | Blocked | Allowed |
| NoScript | 0.78 | 0.22 | 0.21 | 0.79 |
| Ghostery | 0.65 | 0.35 | 0.08 | 0.92 |
| Adblock Plus | 0.44 | 0.56 | 0.06 | 0.94 |
| Disconnect | 0.40 | 0.60 | 0.06 | 0.94 |
| Privacy Badger | 0.37 | 0.63 | 0.06 | 0.94 |

Effectiveness of PP-Tools



- Effectiveness – the balance between correctly blocking tracking JSes and incorrectly blocking functional JSes

| PP-Tool | Tracking | | Functional | |
|----------------|----------|---------|------------|---------|
| | Blocked | Allowed | Blocked | Allowed |
| NoScript | 0.78 | 0.22 | 0.21 | 0.79 |
| Ghostery | 0.65 | 0.35 | 0.08 | 0.92 |
| Adblock Plus | 0.44 | 0.56 | 0.06 | 0.94 |
| Disconnect | 0.40 | 0.60 | 0.06 | 0.94 |
| Privacy Badger | 0.37 | 0.63 | 0.06 | 0.94 |

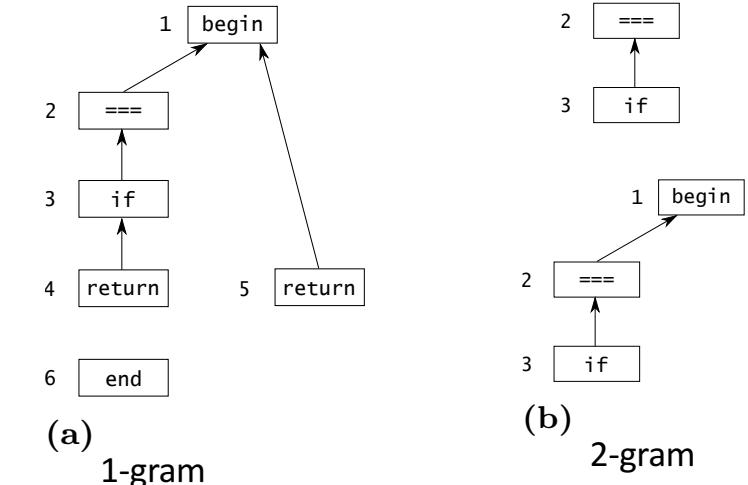
- NoScript** stops one out of five functional JSes while **Privacy Badger** overlooks tracking JSes

Feature Extraction: Constructing Program Dependency Graphs (PDGs)



JavaScript Program → Canonical Form → PDGs n-grams

```
function equalTest(a, b){  
    if(a == b){  
        return true;  
    }  
    return false;  
}  
  
function equalTest(a, b){  
1:    begin;  
2:    $0 = a === b;  
3:    if($0){  
4:        return true;  
5:        return false;  
6:    end;  
}
```



Term Frequency-Inverse Document Frequency (TF-IDF) of n-gram ($2 \leq n \leq 7$)

Hsiao et al., Using Web Corpus Statistics for Program Analysis. OOPSLA, 2014

One-Class Learning



One-Class SVM* and Positive and Unlabeled (PU) Learning**

Training requires only single class

Classifying new data as similar or different to the training set

* Schölkopf et al., Estimating the Support of a High-Dimensional Distribution. *NC*, 2001

** Elkan et al., Learning Classifiers from Only Positive and Unlabeled Data. *KDD*, 2008.

Validation: Our Classifiers' Performance



| Feature Model | Classifier | Tracking | | Functional | |
|----------------------|------------|----------|---------|------------|---------|
| | | Blocked | Allowed | Blocked | Allowed |
| Syntactic | SSVM | 0.93 | 0.07 | 0.01 | 0.99 |
| | OCSVM | 0.88 | 0.12 | 0.02 | 0.98 |
| | PU | 0.86 | 0.14 | 0.02 | 0.98 |
| PDG
4-gram | SSVM | 0.96 | 0.04 | 0.03 | 0.97 |
| | OCSVM | 0.95 | 0.05 | 0.03 | 0.97 |
| | PU | 0.93 | 0.07 | 0.04 | 0.96 |
| Sequential
4-gram | SSVM | 0.98 | 0.02 | 0.01 | 0.99 |
| | OCSVM | 0.98 | 0.02 | 0.02 | 0.98 |
| | PU | 0.96 | 0.04 | 0.03 | 0.97 |
| PDG
7-gram | SSVM | 0.99 | 0.01 | 0.01 | 0.99 |
| | OCSVM | 0.99 | 0.01 | 0.01 | 0.99 |
| | PU | 0.98 | 0.02 | 0.02 | 0.98 |
| Sequential
7-gram | SSVM | 0.99 | 0.01 | 0.01 | 0.99 |
| | OCSVM | 0.99 | 0.01 | 0.01 | 0.99 |
| | PU | 0.98 | 0.02 | 0.02 | 0.98 |

Validation: Our Classifiers' Performance

| Feature Model | Classifier | Tracking | | Functional | |
|---------------|------------|----------|---------|------------|---------|
| | | Blocked | Allowed | Blocked | Allowed |
| Syntactic | SSVM | 0.93 | 0.07 | 0.01 | 0.99 |
| | OCSVM | 0.88 | 0.12 | 0.02 | 0.98 |
| | PU | 0.86 | 0.14 | 0.02 | 0.98 |
| 4-gram | SSVM | 0.99 | 0.01 | 0.05 | 0.95 |
| | OCSVM | 0.95 | 0.05 | 0.03 | 0.97 |
| | PU | 0.93 | 0.07 | 0.04 | 0.96 |
| Sequential | SSVM | 0.98 | 0.02 | 0.01 | 0.99 |
| | OCSVM | 0.98 | 0.02 | 0.02 | 0.98 |
| | PU | 0.99 | 0.01 | 0.03 | 0.97 |
| PDG | SSVM | 0.99 | 0.01 | 0.01 | 0.99 |
| | OCSVM | 0.99 | 0.01 | 0.01 | 0.99 |
| | PU | 0.98 | 0.02 | 0.02 | 0.98 |
| 7-gram | SSVM | 0.99 | 0.01 | 0.01 | 0.99 |
| | OCSVM | 0.99 | 0.01 | 0.01 | 0.99 |
| | PU | 0.98 | 0.02 | 0.02 | 0.98 |
| Sequential | SSVM | 0.99 | 0.01 | 0.01 | 0.99 |
| | OCSVM | 0.99 | 0.01 | 0.01 | 0.99 |
| | PU | 0.98 | 0.02 | 0.02 | 0.98 |

Our best classifier has **99% Accuracy** and less than **1% False Positive rates**

Improves PP-Tools' Accuracy by 21% to 63% and reduces False Positive rates by 5% to 20%

Our Classifiers and PP-Tools in the Wild (4084 Websites)

Our Classifiers and PP-Tools in the Wild



Agreements between our classifier's and a PP-Tool's output:

$$T_c \cap T_p$$

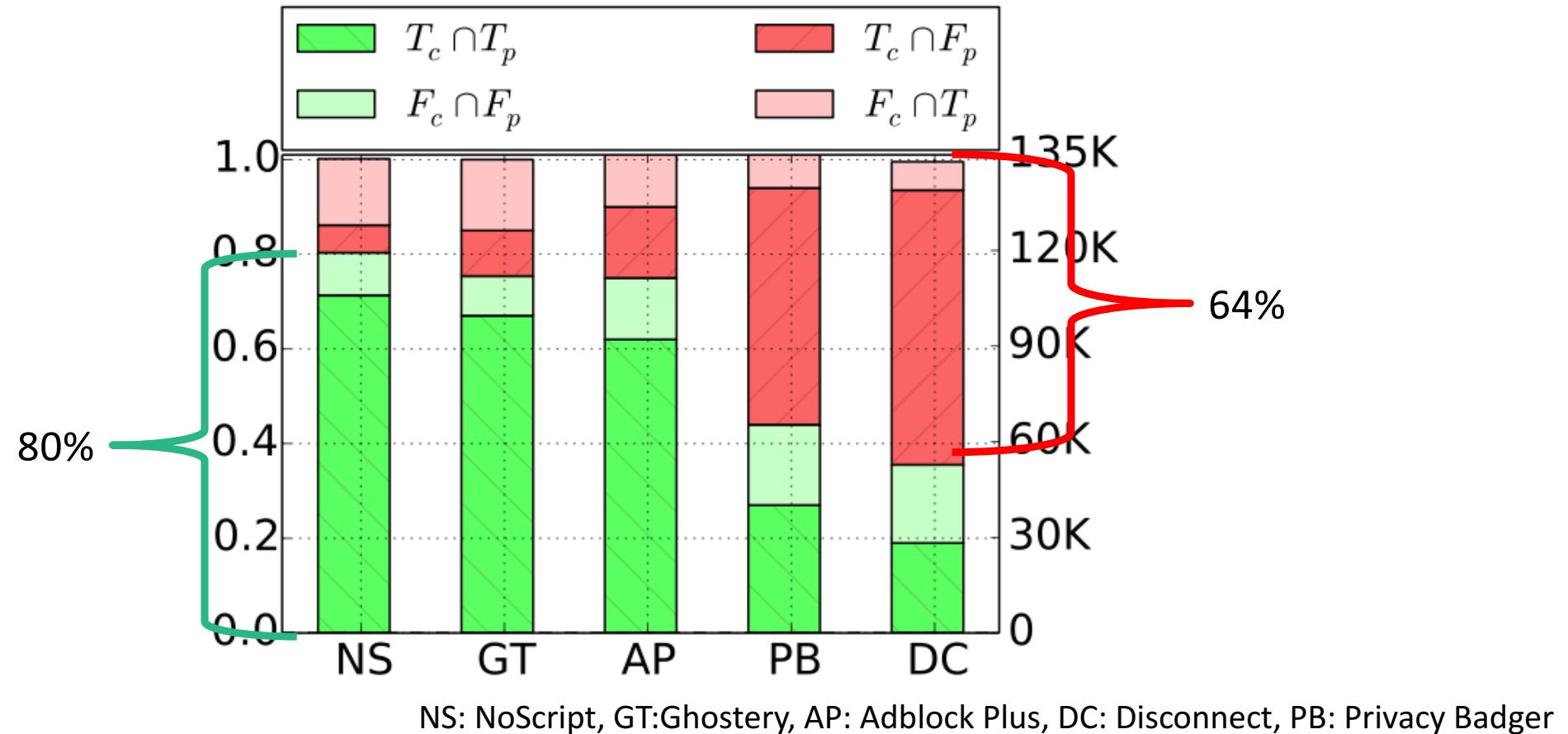
$$F_c \cap F_p$$

Disagreements between our classifier's and a PP-Tool's output:

$$T_c \cap F_p$$

$$F_c \cap T_p$$

Our Classifiers and PP-Tools in the Wild



Our Classifiers and PP-Tools in the Wild



Disagreement

| | |
|------------------|--|
| $T_c \cap_p F_p$ | JSes that our classifier labels as tracking while all PP-Tools consider functional |
| $F_c \cap_p T_p$ | JSes that our classifier labels as functional while all PP-Tools consider tracking |

c: classifier, p: PP-Tool

Our Classifiers and PP-Tools in the Wild



| Disagreement | Total | Sample |
|------------------|-------|--------|
| $T_c \cap_p F_p$ | 4,610 | 100 |
| $F_c \cap_p T_p$ | 4,461 | 100 |

c: classifier, p: PP-Tool

Our Classifiers and PP-Tools in the Wild



| Disagreement | Total | Sample | Manual Labelling | |
|------------------|-------|--------|------------------|------------|
| | | | Tracking | Functional |
| $T_c \cap_p F_p$ | 4,610 | 100 | 75 | 25 |
| $F_c \cap_p T_p$ | 4,461 | 100 | 19 | 81 |

c: classifier, p: PP-Tool

Our Classifiers and PP-Tools in the Wild

- 75% of the case, our classifier is right

| Disagreement | Total | Sample | Manual Labelling | |
|------------------|-------|--------|------------------|------------|
| | | | Tracking | Functional |
| $T_c \cap_p F_p$ | 4,610 | 100 | 75 | 25 |
| $F_c \cap_p T_p$ | 4,461 | 100 | 19 | 81 |

- PP-Tools perform RE matching on the URL in <script> and fail to block tracking JSes that are not in the blacklists

c: classifier, p: PP-Tool

Our Classifiers and PP-Tools in the Wild



- 81% of the case, our classifier is right

| Disagreement | Total | Sample | Manual Labelling | |
|------------------|-------|--------|------------------|------------|
| | | | Tracking | Functional |
| $T_c \cap_p F_p$ | 4,610 | 100 | 75 | 25 |
| $F_c \cap_p T_p$ | 4,461 | 100 | 19 | 81 |

- PP-Tools block JSes from a tracking domain even though JSes perform useful functionality

c: classifier, p: PP-Tool

New Trackers Found



Discover more than 4K+ previously unknown tracking services

Discardment | Total | Sample | Manual Labelling

| # | Website | JavaScript Program | Referred Domain | Function Performed |
|----|-------------------|---|---------------------------|-------------------------|
| 1 | examiner.com | cdn2-b.examiner.com/.../ex_omniture/s_code.js | omniture.com | Analytics |
| 2 | bbc.com | static.bbci.co.uk/bbcdotcom/.../adverts.js | pubads.g.doubleclick.net | Analytics + Ads |
| 3 | telegraph.co.uk | telegraph.co.uk/template/ver1-0/js/gpt.js | pubads.g.doubleclick.net | Analytics + Ads |
| 4 | vesti.ru | s.i-vengo.com/js/ivengo.min.js | www.i-vengo.com | Analytics + Ads |
| 5 | climatempo.com.br | http://s1.trrsf.com/metrics/inc/br/201411250000d.js | scorecardresearch.com | Analytics |
| 6 | amc.com | amc.com/wp-content/plugins/amcn-common-analytics/js/common-analytics.js | omniture.com | Track user activities |
| 7 | lancer.com | static.lancers.jp/js/ga_social_tracking.js | google.com | Tracker user activities |
| 8 | iqiyi.com | static.iqiyi.com/js/pingback/qa.js | pps.tv, baidu.com, 71.com | Tracker user activities |
| 9 | babyblog.ru | act.babyblog.ru/static844/likes.js | babyblog.ru | Social widgets |
| 10 | autoscout.de | s.autoscout24.net/unifiedtracking/gtm.js | autoscout.de | Tracks user activities |

More Results, Analyses and Discussions



DE GRUYTER OPEN

Proceedings on Privacy Enhancing Technologies 2017; 2017 (1):1–21

Muhammad Ikram*, Hassan Jameel Asghar, Mohamed Ali Kaafar, Anirban Mahanti, and Balachander Krishnamurthy

Towards Seamless Tracking-Free Web: Improved Detection of Trackers via One-class Learning

Abstract: Numerous tools have been developed to aggressively block the execution of popular JavaScript programs in Web browsers. Such blocking also affects func-

sites with 135,656 JavaScript programs. The output of our best classifier on this data is between 20 to 64% different from the tools under study. We manually anal-

Summary of Contributions



- Analyze PP-Tools' performance
 - Regular expressions based on blacklists are ineffective
 - **Accuracy** ranges from 37% to 78% while **false positive rates** range from 6% to 21%.
- Design a classification framework to separate *tracking* JSes from *functional* JSes:
 - Training with small single class of tracking (or functional) JSes
 - Our validated classifiers achieve 99% accuracy
 - Discover more than 4K+ previously unknown tracking services



Questions?

Big Thanks to PETS Reviewers

Muhammad.Ikram@data61.csiro.au

<https://research.csiro.au/ng/about-us/students/muhammad-ikram/>