Best-of-Breed Content Blocking in Brave:

Three Projects to Improve the Depth, Breath, and Usefulness of Blocking at Scale

Pete Snyder
Senior Privacy Researcher
pes@brave.com
@pes10k

Hi, I’m Pete 👋

- **Grew up in Chicago**
  …actual Chicago

- **Law school -> freelance web stuff**
  Started: Anchorage, AK
  Ended: Judge Judy Show invitation

- **PhD in Computer Science**
  University of Illinois at Chicago

Me at Brave

- **Researcher at Brave**
  …privacy, blocking, reliability
- **Co-Chair of PING**
  Privacy committee on W3C
- **Research <-> Engineering**
  Web compat, filter lists, fingerprinting, etc.

Brave in a Slide

- **Privacy focused**
- **Alternative web funding model**
  Fix incentive problems
- **Research + Engineering**
- **Browsers and infrastructure now, more to come…**

Overview

- **PageGraph**
  Novel, open source, DOM attribution system
- **Behavioral Blocking**
  Protecting Privacy Beyond URLs
- **Regional Filter List Generation**
  Privacy protection for users in underserved regions
- **Automated Web Compatibility Testing**
  Making sure we’re only breaking bad stuff…

Overview

- **PageGraph**
  Novel, open source, DOM attribution system

- **Behavioral Blocking**
  Protecting Privacy Beyond URLs

- **Regional Filter List Generation**
  Privacy protection for users in underserved regions

- **Automated Web Compatibility Testing**
  Making sure we're only breaking bad stuff…

Anton Lazarev - alazarev@brave.com
Northeastern University

PageGraph: Motivation

- **Lots of research requires understanding what script does what…**
  Forensics, privacy, performance prediction, debugging…

- **Seems simple…**
  Doesn't devtools do this already? Projects in research?

- **Online and offline analysis**
  Ordered, correct causal analysis

PageGraph: Challenges

- **Scripts**
  - Script injects another script?
  - Inline scripts?
  - Script in HTML attributes? JS URLs?
  - Eval, new Function(), callback functions, micro tasks…

- **Network**
  - Attribution for injected images / videos
  - Fetch? Ajax? Remote frames?

- **Web API use? Storage? DOM modifications? Etc…**

PageGraph: High Fidelity Cause Attribution

- **Modification Cause Attribution**
  Record cause of all DOM modifications, network requests, privacy-relevant WebAPI calls

- **Graph Representation**
  events -> edges, page elements -> nodes

- **Online and Offline Analysis**
  In memory graph, GraphML Export

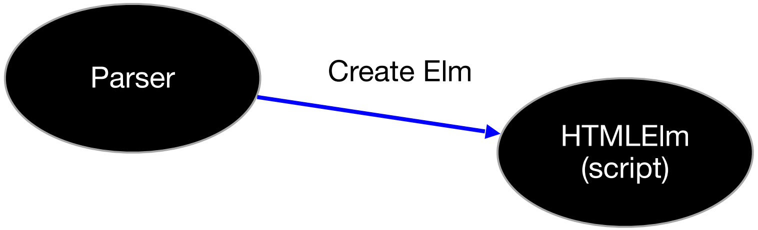
- https://github.com/brave/brave-browser/wiki/PageGraph

PageGraph: Toy Example

```
<script>
const elm = document.createElement("img");
elm.src = "/pixel.gif";
document.body.appendChild(elm);
document.getElementsByTagName("script")[0]
    …(rest removed)
</script>
```

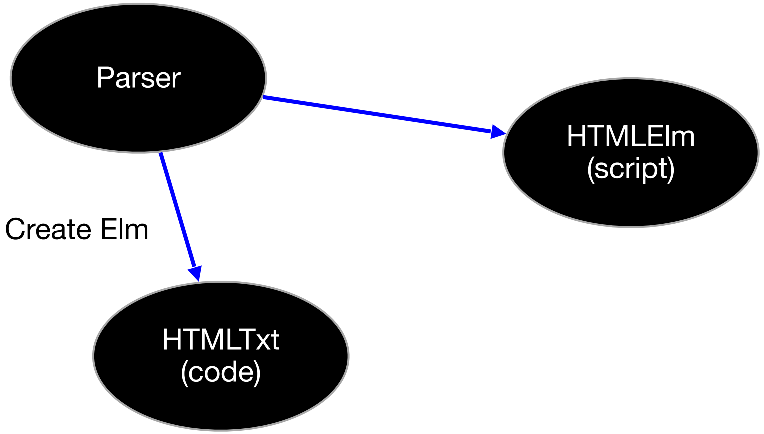
PageGraph: Toy Example

```
<script>
const elm = document.createElement("img");
elm.src = "/pixel.gif";
document.body.appendChild(elm);
document.getElementsByTagName(“script”)[0]
    …(rest removed)
</script>
```



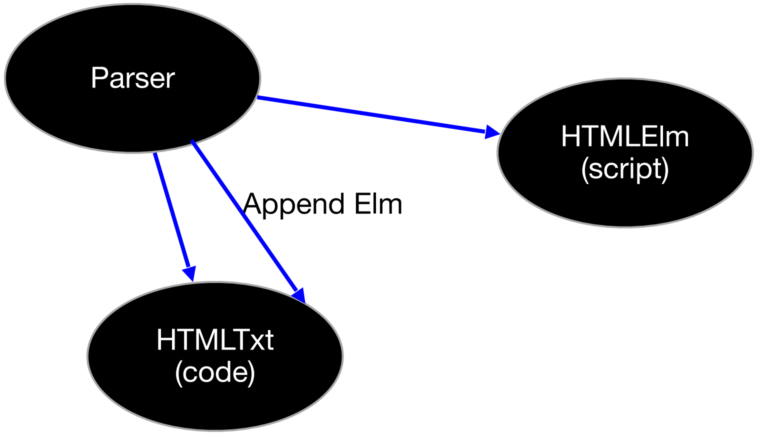
PageGraph: Toy Example

```
<script>
const elm = document.createElement("img");
elm.src = "/pixel.gif";
document.body.appendChild(elm);
document.getElementsByTagName("script")[0]
    …(rest removed)
</script>
```



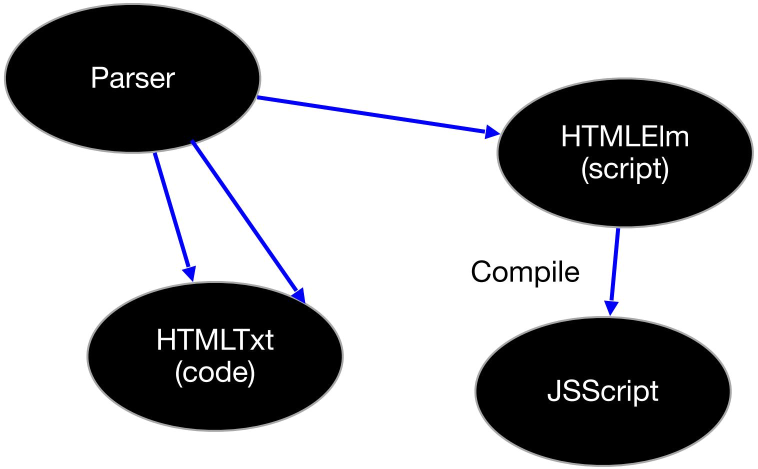
PageGraph: Toy Example

```
<script>
const elm = document.createElement("img");
elm.src = "/pixel.gif";
document.body.appendChild(elm);
document.getElementsByTagName(“script”)[0]
    …(rest removed)
</script>
```



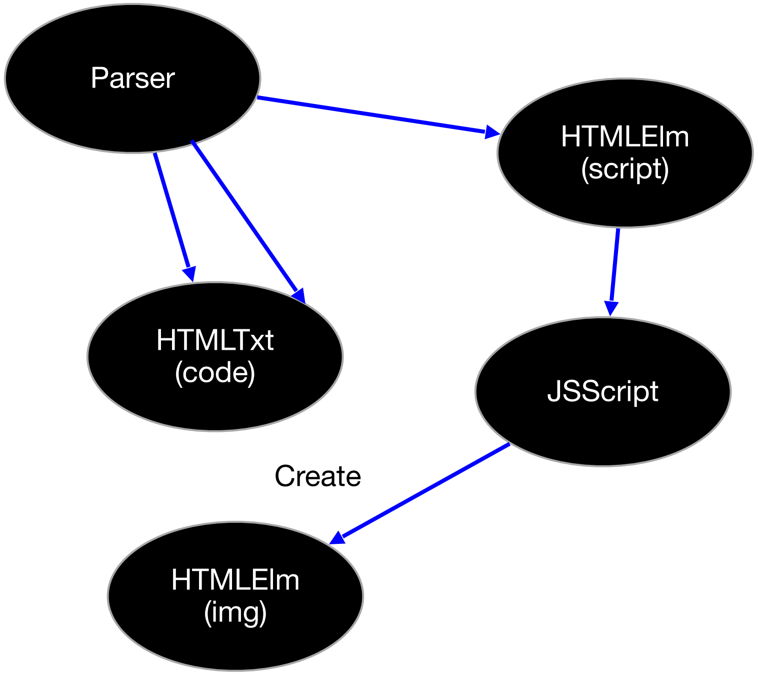
PageGraph: Toy Example

```
<script>
const elm = document.createElement("img");
elm.src = "/pixel.gif";
document.body.appendChild(elm);
document.getElementsByTagName("script")[0]
    …(rest removed)
</script>
```



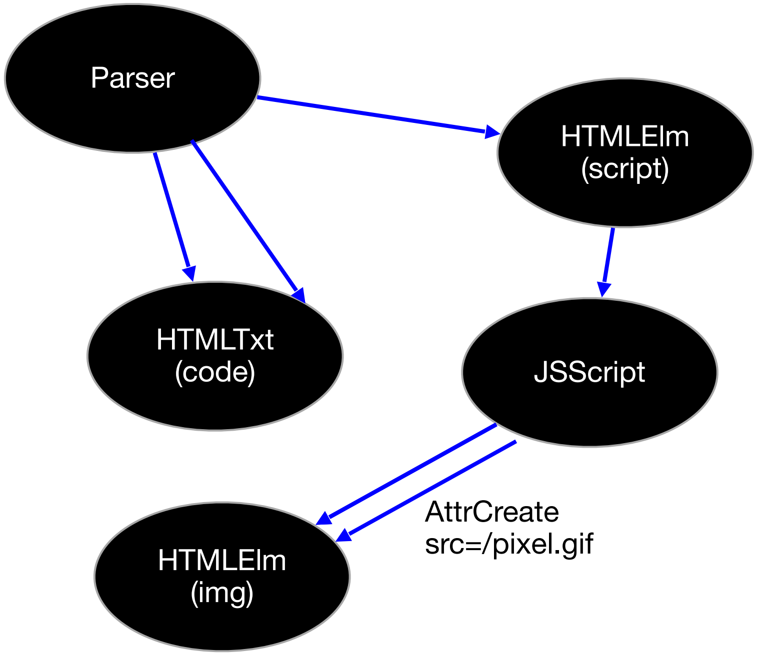
PageGraph: Toy Example

```
<script>
const elm = document.createElement("img");
elm.src = "/pixel.gif";
document.body.appendChild(elm);
document.getElementsByTagName(“script”)[0]
    …(rest removed)
</script>
```



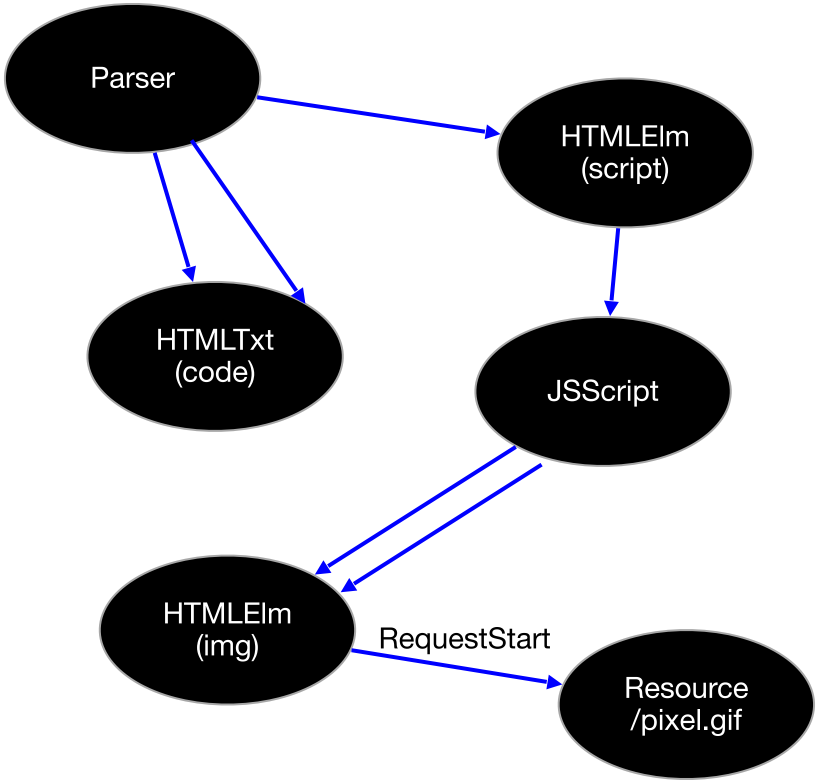
PageGraph: Toy Example

```
<script>
const elm = document.createElement("img");
elm.src = "/pixel.gif";
document.body.appendChild(elm);
document.getElementsByTagName("script")[0]
    …(rest removed)
</script>
```



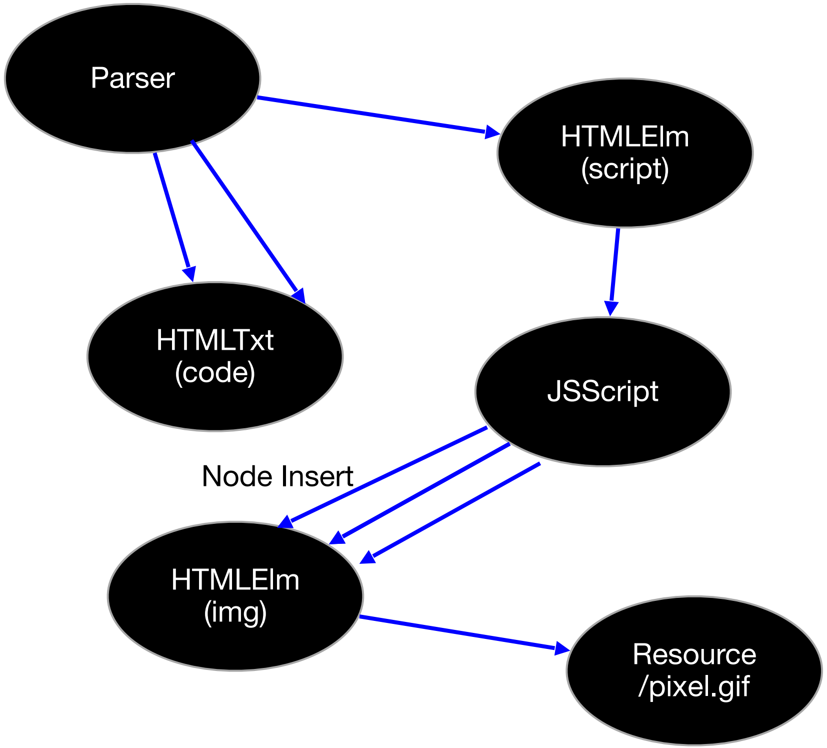
PageGraph: Toy Example

```
<script>
const elm = document.createElement("img");
elm.src = "/pixel.gif";
document.body.appendChild(elm);
document.getElementsByTagName("script")[0]
    …(rest removed)
</script>
```



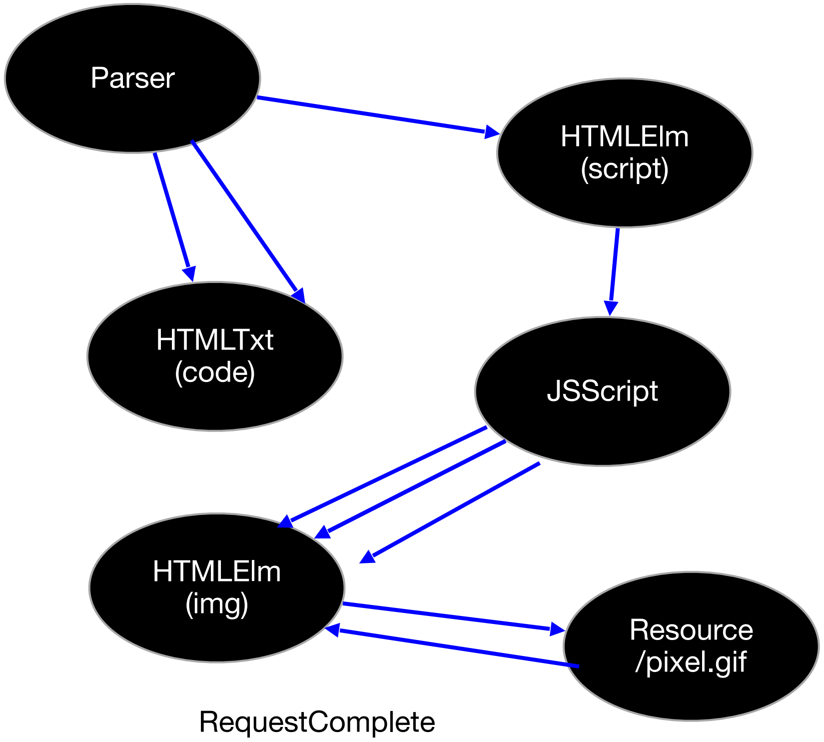
PageGraph: Toy Example

```
<script>
const elm = document.createElement("img");
elm.src = "/pixel.gif";
document.body.appendChild(elm);
document.getElementsByTagName(“script”)[0]
    …(rest removed)
</script>
```



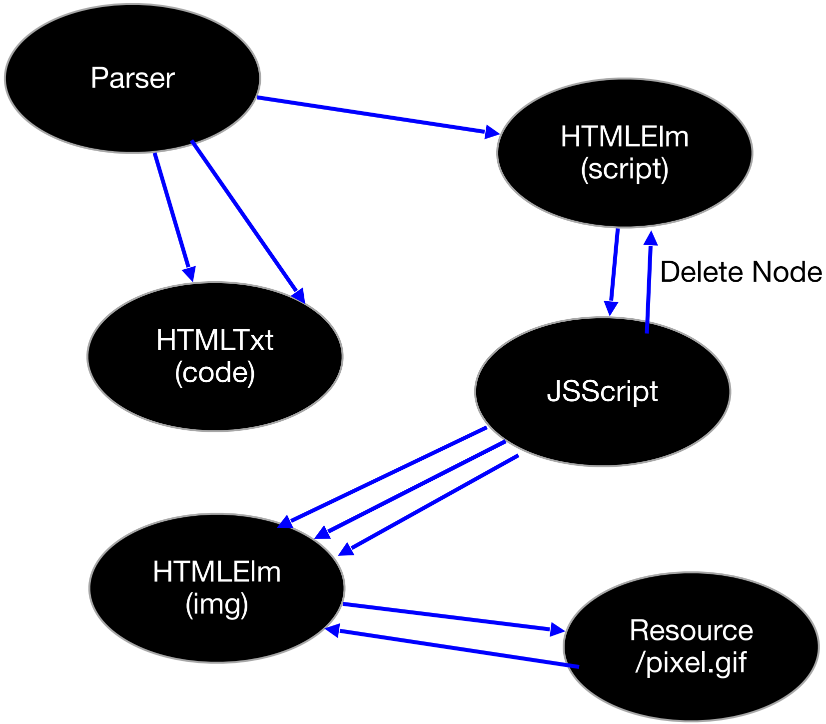
PageGraph: Toy Example

```
<script>
const elm = document.createElement("img");
elm.src = "/pixel.gif";
document.body.appendChild(elm);
document.getElementsByTagName(“script”)[0]
    …(rest removed)
</script>
```

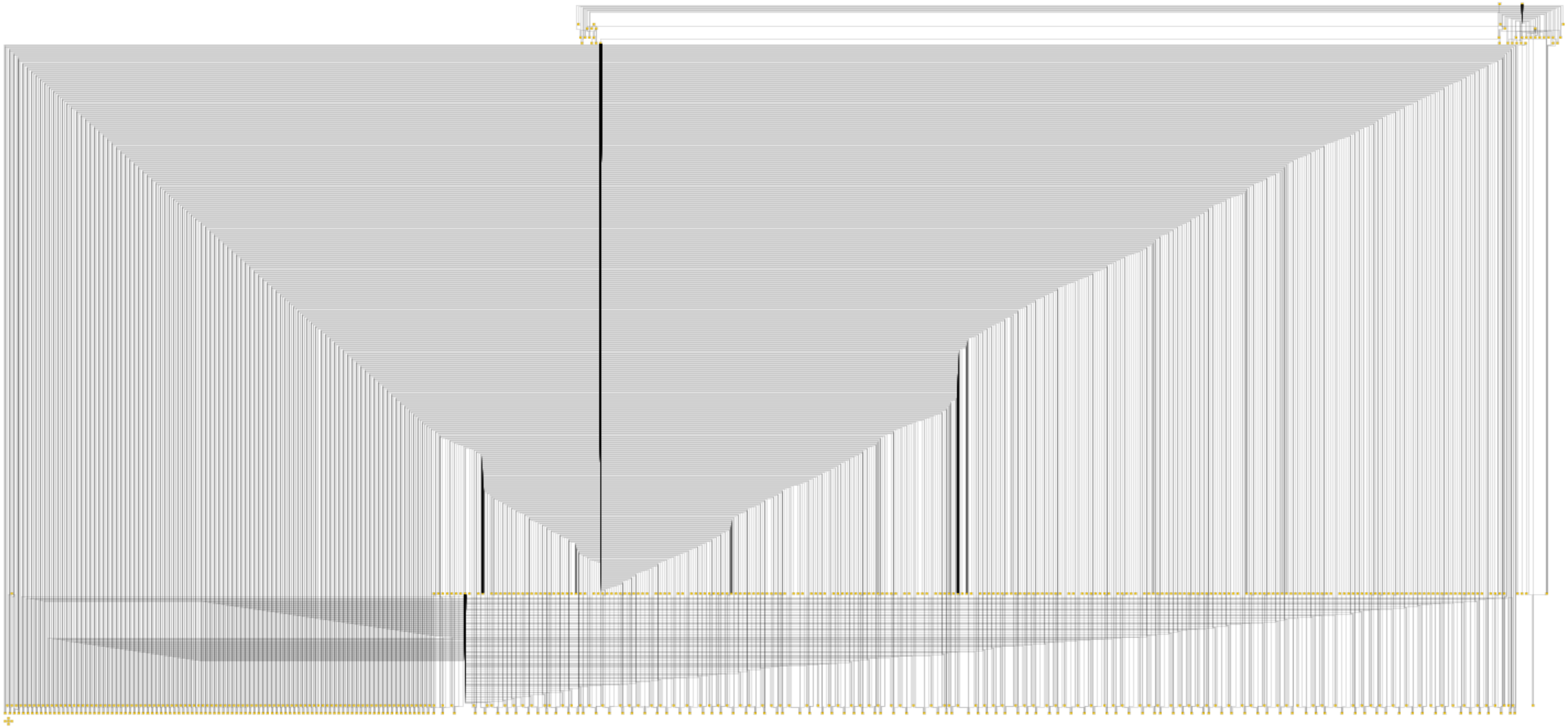


PageGraph: Toy Example

```
<script>
const elm = document.createElement("img");
elm.src = "/pixel.gif";
document.body.appendChild(elm);
document.getElementsByTagName("script")[0]
    …(rest removed)
</script>
```



Fingerprint2.js example…



All of CNN.com (for 30 sec…)

Limitations

- **Still stuff to instrument…**
  e.g. Module scripts

- **Bugs (but, few!)**
  < 5% of pages will have an attribution error

- **Dense graphs**
  Pages do a lot of things…

PageGraph: Summary

- **Comprehensive causal attribution**
  Attribute every request, DOM modification, compilation and privacy event to responsible script

- **In modern chromium based browser**
  Up to date Brave, up to date Chromium

- **Open source and For Use Now**
  Build it today!

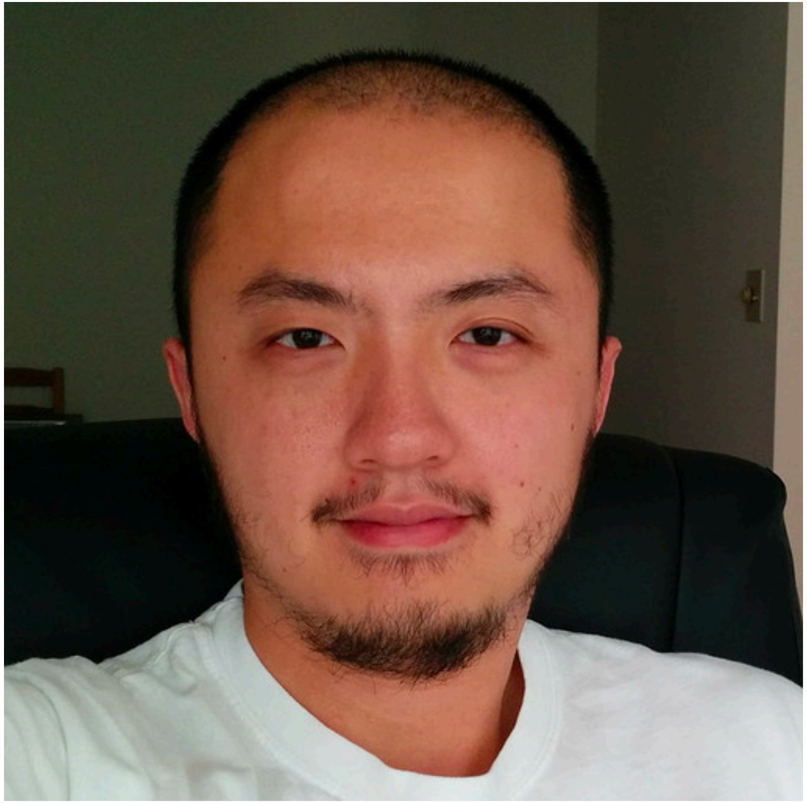
Overview

- **PageGraph**
  Novel, open source, DOM attribution system

- **Behavioral Blocking**
  Protecting Privacy Beyond URLs

- **Regional Filter List Generation**
  Privacy protection for users in underserved regions

- **Automated Web Compatibility Testing**
  Making sure we're only breaking bad stuff…



Quan Chen - qchen10@ncsu.edu
North Carolina State University

Blocking Arms Race

Tracking Domains → /etc/hosts, Safari ITP DNS filtering, etc… → Move “bad” code to “good” domains

Blocking Arms Race

Tracking Domains → /etc/hosts, Safari ITP DNS filtering, etc… → Move “bad” code to “good” domains

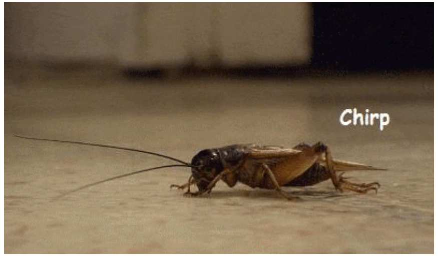
Mixed Domains → Filter Lists, ITP2.0 → Mix “bad” code with “good” code

Blocking Arms Race

Tracking Domains → /etc/hosts, Safari ITP DNS filtering, etc… → Move “bad” code to “good” domains

Mixed Domains → Filter Lists, ITP2.0 → Mix “bad” code with “good” code

Mixed Code →



Blocking Arms Race

Tracking Domains → /etc/hosts, Safari ITP DNS filtering, etc… → Move “bad” code to “good” domains

Mixed Domains → Filter Lists, ITP2.0 → Mix “bad” code with “good” code

Mixed Code →

Behavioral Blocking Roadmap

1. Motivation
2. Methodology
3. Results
4. Whats next

Behavioral Blocking Roadmap

1. **Motivation**
2. Methodology
3. Results
4. Whats next

Motivating Example

```
<html>
 <head>
   <script src="(google-analytics.js)">
   <script src="(jQuery)">
   <script src="(Site stuff)">
 </head>
 <body>...</body>
</html>
```

Motivating Example

```
<html>
 <head>
   <script src="(google-analytics.js)">
   <script src="(jQuery)">
   <script src="(Site stuff)">
 </head>
 <body>...</body>
</html>
```

Motivating Example

```
<html>
 <head>
   <script src="(
       google-analytics.js + jQuery + site stuff
    ).min.js">
 </head>
 <body>...</body>
</html>
```

Motivating Example

```
<html>
 <head>
   <script src="(
       google-analytics.js + jQuery + site stuff
    ).min.js"> ¯\_(ツ)_/¯
 </head>
 <body>...</body>
</html>
```

Core Problem: URLs

- Moved code
- Changing URLs
- Inlining code
- Bundling code
- etc

Project Goals

- Deliver the same privacy protections, independent of code delivery…
- Independent of URL…
- w/r/t Network requests, storage access, fingerprinting endpoints

```html
<html>
 <head>
  <script src="(google-analytics.js)">
  <script src="(jQuery)">
  <script src="(Site stuff)">
 </head>
 <body>...</body>
</html>
```

≡

```html
<html>
 <head>
  <script src="(google-analytics.js +
   jQuery + site stuff).min.js">
 </head>
 <body>...</body>
</html>
```

Straw Proposals

Straw Proposals

Matching hashes? → Trivial to circumvent

ASTs? → Code bundlers / mungers

Straw Proposals

Matching hashes?	Trivial to circumvent
ASTs?	Code bundlers / mungers
Identifying tokens	Minification / obfuscation / etc

Behavioral Blocking Roadmap

1. Motivation
2. **Methodology**
3. Results
4. Deployment

Methodology

1. **Behavioral signatures of scripts**
   Page graph, event loop turns

2. **Build set of known privacy-harming behaviors**
   EasyList + EasyPrivacy as ground truth

3. **Check to see if new scripts do known bad things**
   e.g. apply extracted signatures

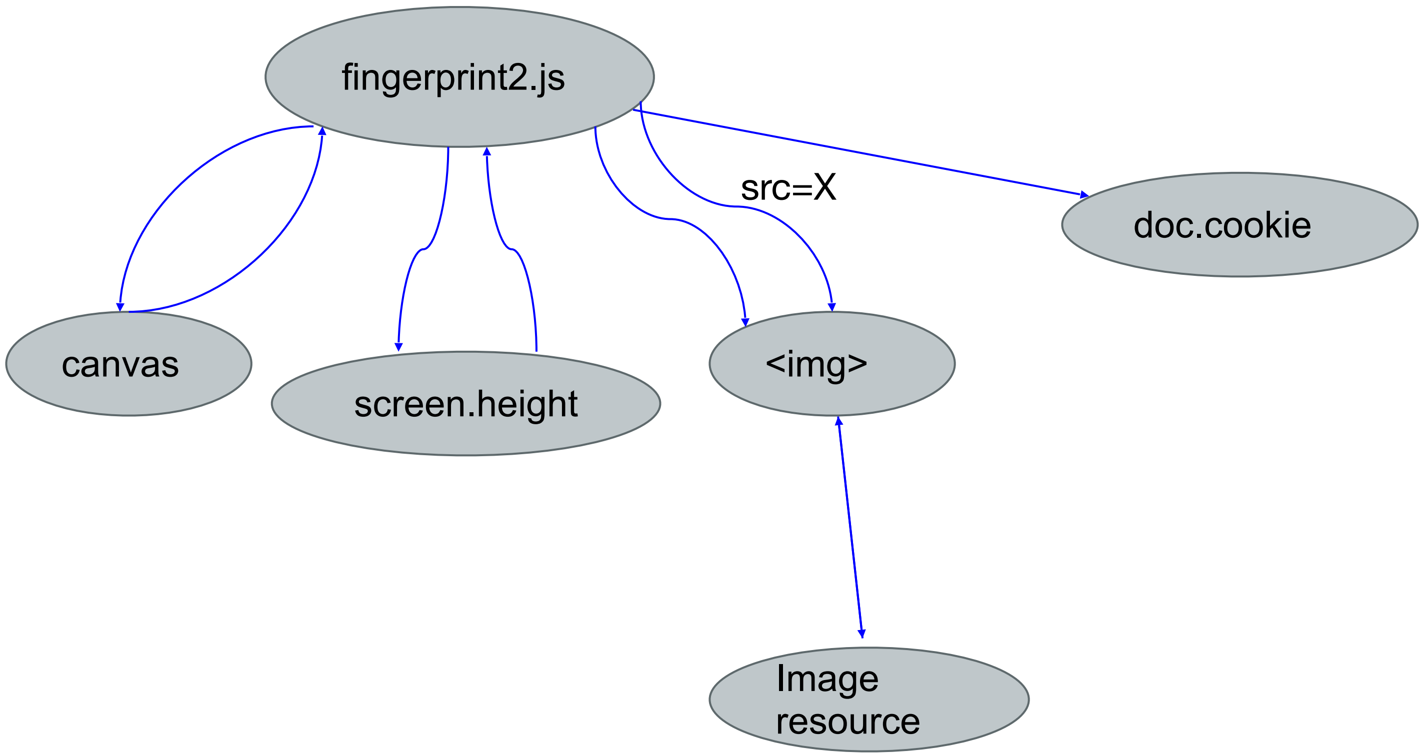
Signature Granularity

- **Entire Script / Library?**
  Too non-deterministic, too much variation between sites
  e.g. false negatives

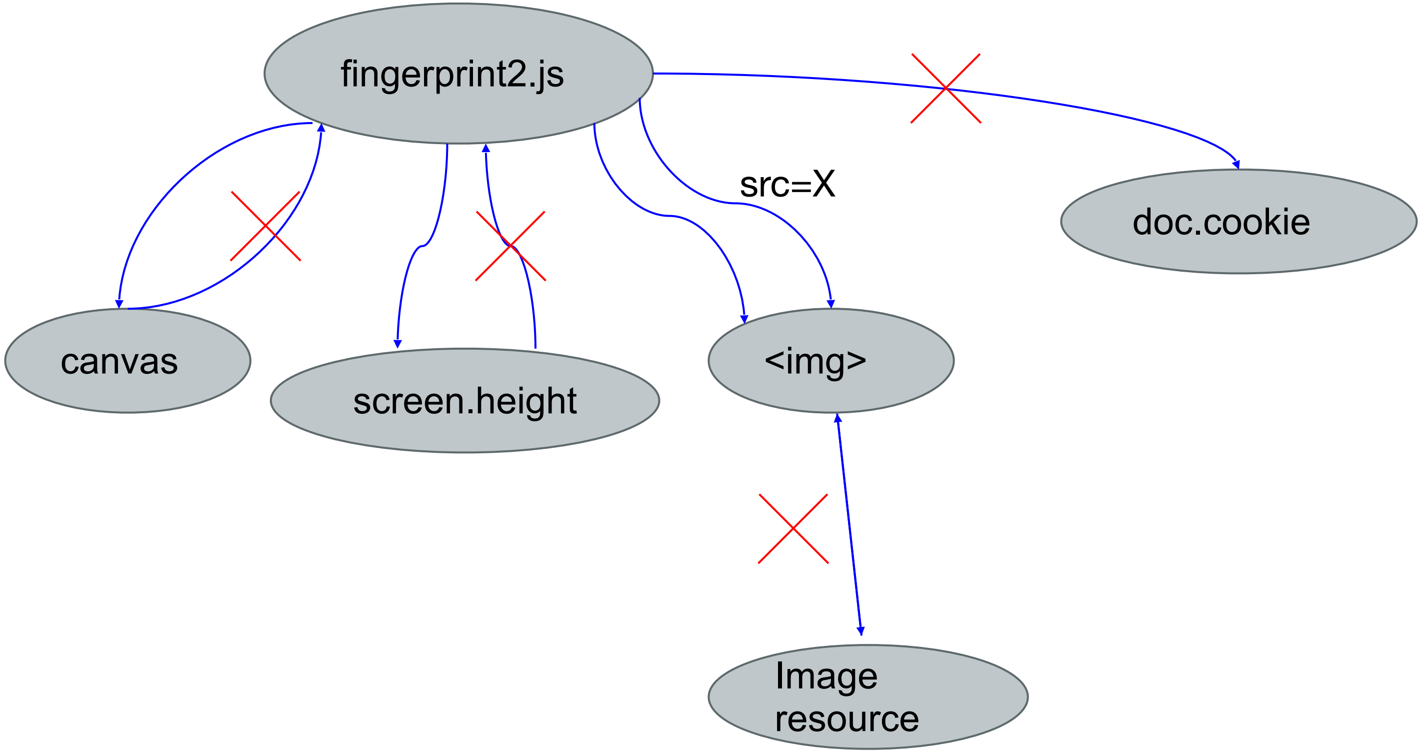
- **Individual API Calls?**
  Not enough information
  e.g. false positives

- **Event Loop Turns**
  Good balance
  Allows some variation between library use

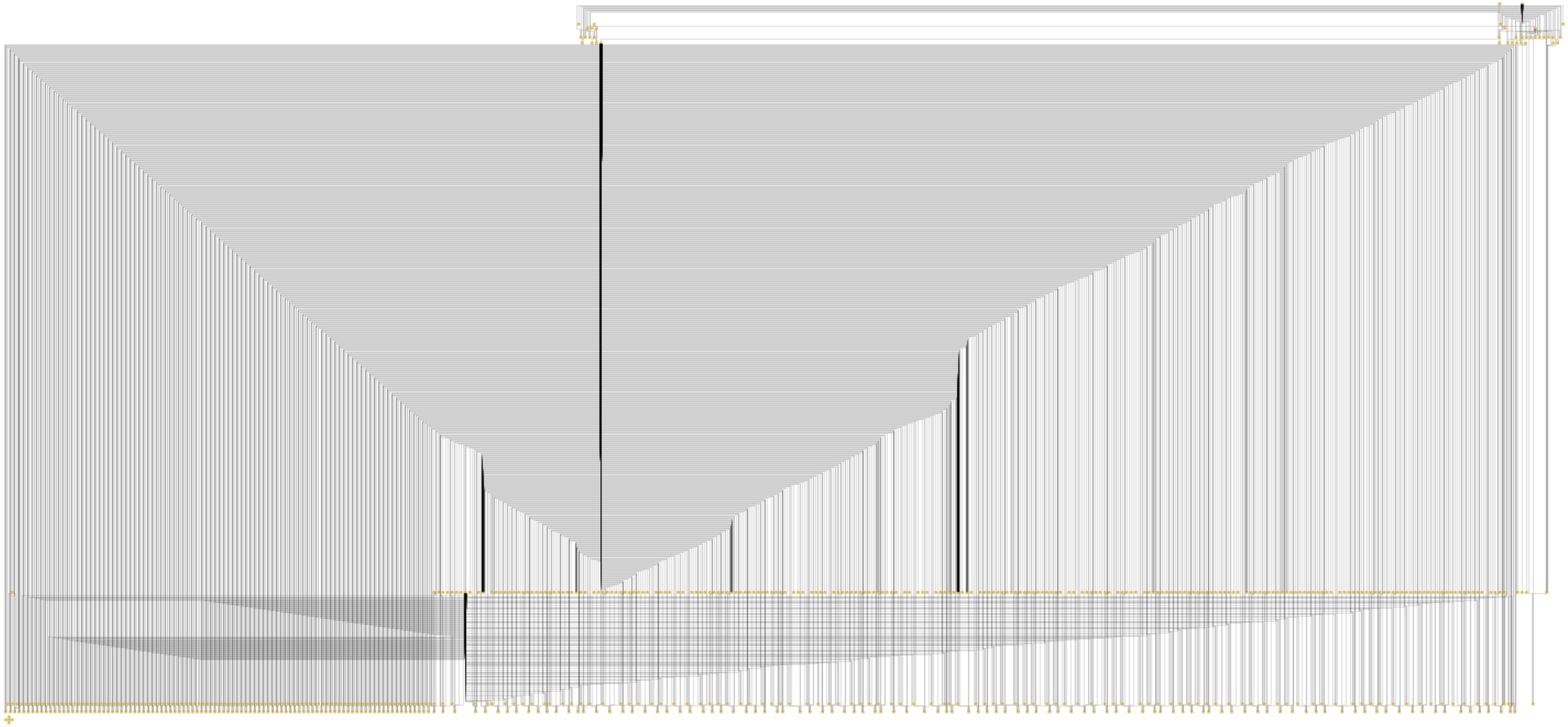
Fingerprint2.js (e.g. bad)



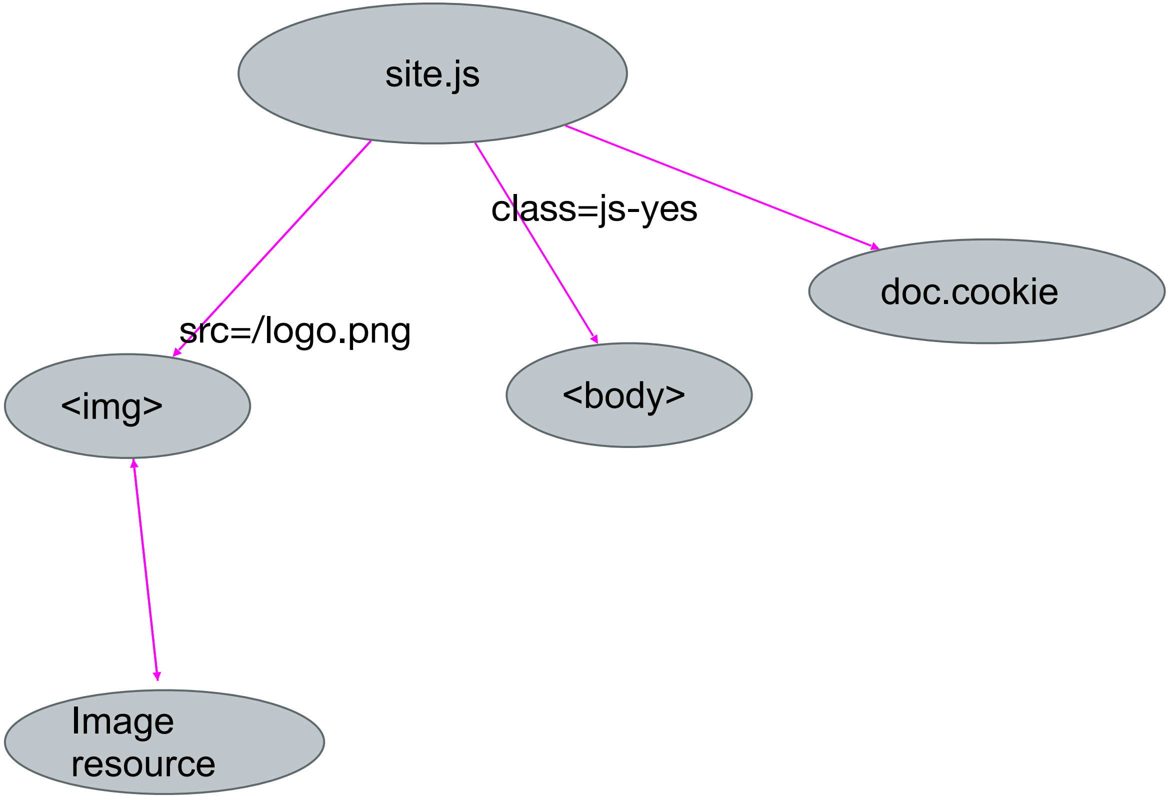
Fingerprint2.js (e.g. bad)



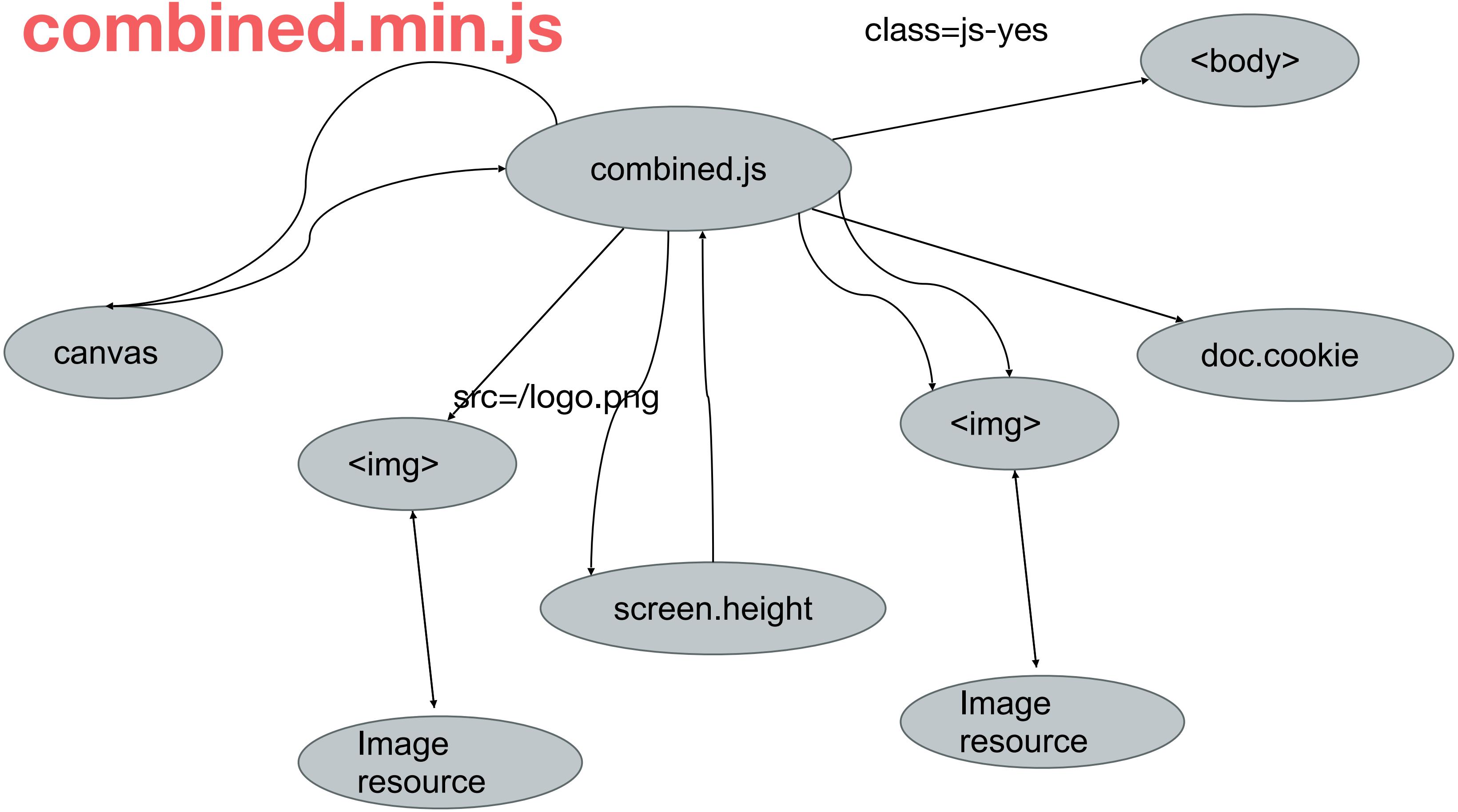
Obvious Fingerprint2.js is… obvious



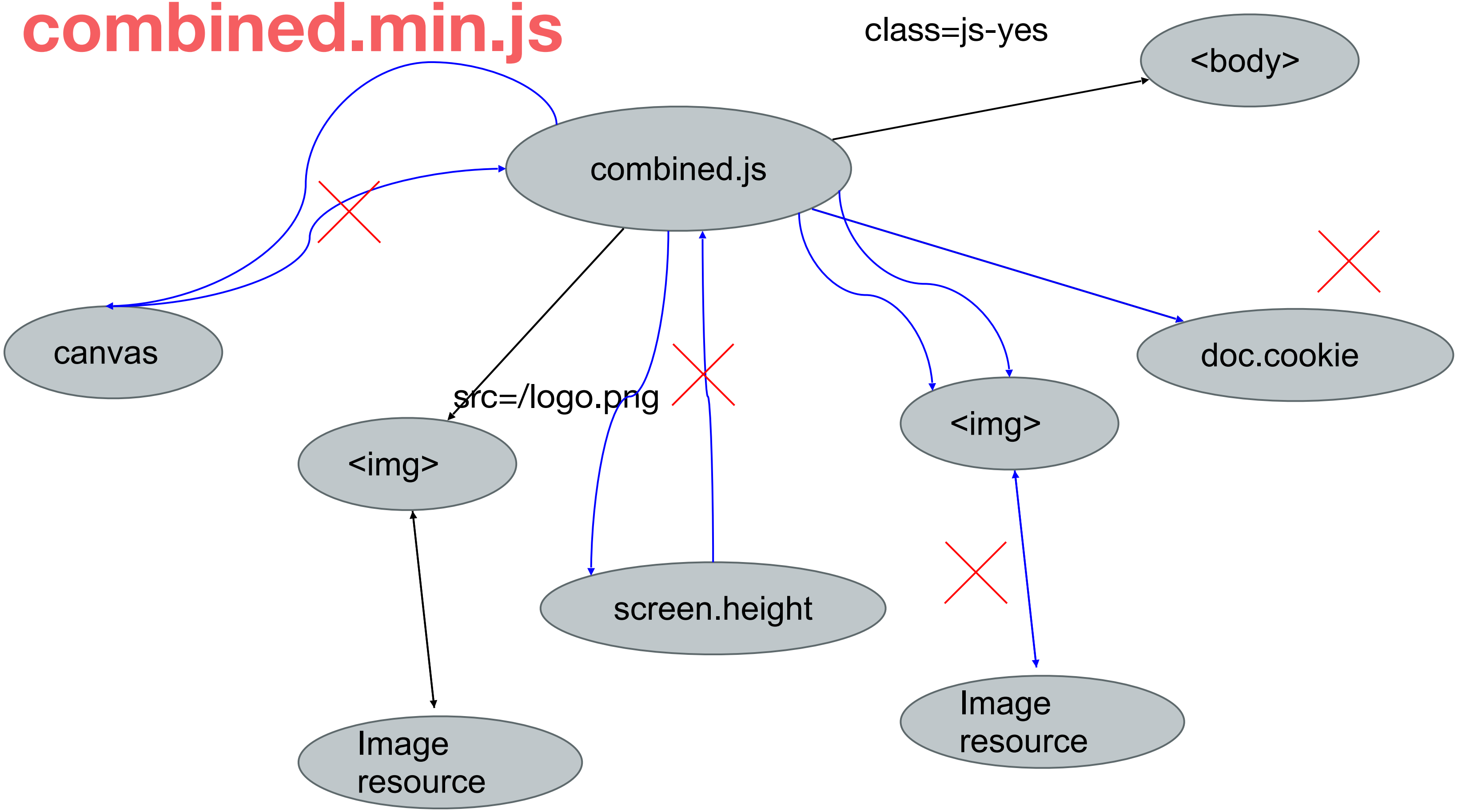
site.js (e.g. benign)



combined.min.js



combined.min.js



Putting it together…

- Catch script activities between event loop yields
- Graph represents as sequential edges and nodes
- Extract EL+EP subgraphs using the above algo
- Look for these recurring patterns in non-labeled JS units
- Enforce protections as if it came from original labeled script

Behavioral Blocking Roadmap

1. Motivation
2. Methodology
3. **Results**
4. Whats next

Raw data (partial)

Crawl set	Alexa 100k
Generated signatures	1,996,093
Privacy affecting, “bad” behaviors	400,287
Num sites with >= 1 missed behavior	11,443

Taxonomy: Moving Scripts

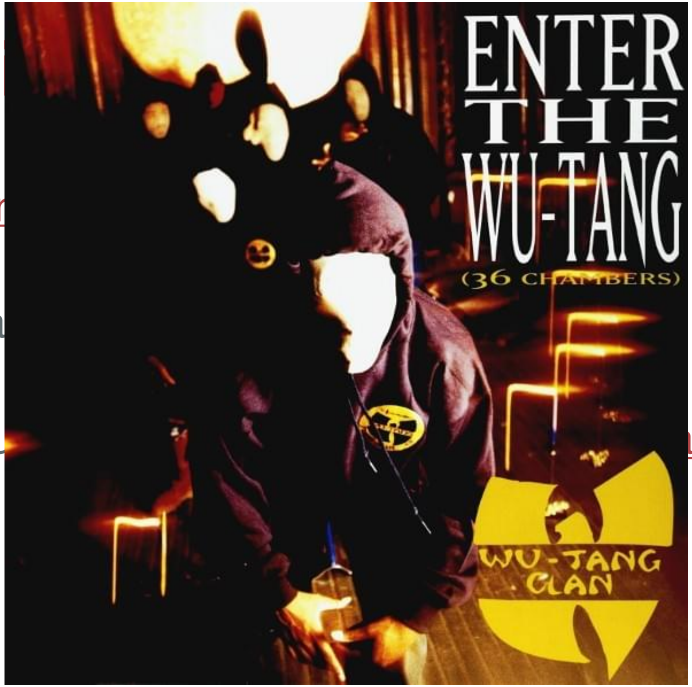
- Moving script file from blocked code to new URL
- https://example.org/tracker.js -> https://helpful.org/site.js
- Long tail, hard to crowdsource, etc.
- 4,730 sites

Google Analytics

- https://www.google-analytics.com/analytics.js
- Blocked by `||google-analytics.com/analytics.js`
- https://messari.io includes https://messari.io/js/wutangtrack.js

Google Analyti

- https://www.google-ar
- Blocked by `||google-a
- https://messari.io inclu ack.js



- 🤔🤔🤔🤔🤔🤔🤔🤔🤔🤔🤔🤔

Taxonomy: Inlining

- Moving script from a URL to txt
- <script src="X"> to <script>(code)</script>
- Difficult for most tools to address at all
- Impossible for current tools to address well
- 223 sites

Dynatrace Example

- https://www.dynatrace.com/*
  (e.g. https://js-cdn.dynatrace.com/jstag/157944990f8/fsf84414/…)
- Blocked by `||dynatrace.com^$third-party`
- But look at https://www.dynatrace.com…

Taxonomy: Bundling

- Including harmful code with benign / user-serving code
- jQuery + fingerprint2.js + site.js -> combined.min.js
- No win choice for blockers…
- 122 sites

Microsoft Insights

- E.g. https://az416426.vo.msecnd.net/scripts/a/ai.0.js
- Blocked by `||msecnd.net/scripts/a/ai.0.js`
- https://lindex.com -> https://www.lindex.com/web-assets/js/vendors.8035c13832ab6bb90a46.js

Taxonomy: Common Library

- Third party libraries used by larger libraries
- Original library URL may never (rarely?) be fetched directly
- Not *anything* to block off
- 6,141 sites

Adobe Visitor API

- Adobe has many tracking libraries (e.g. Adobe Target) that all use the Adobe Visitor API
- Visitor API is the tracking part
- Much is blocked with `||adobedtm.com^$third-party`
- Other libs suck in Visitor API too
- E.g. https://tags-eu.tiqcdn.com/utag/intel/profile-cq.emea/prod/utag.js

Technique	# Instances	Unique Scripts
Moving	4,730	586
Inlining	450	450
Bundling	122	92
Common Code	6,141	2,342

Behavioral Blocking Roadmap

1. Motivation
2. Methodology
3. Results
4. **Whats next**

What To Do Next

- **Defenses**
  Moved scripts -> new filter rules
  Others -> maybe runtime fingerprint enforcement?

- **Is the problem getting worse?**
  Longitudinal measurements (and WayBack Machine)
  Measurements going forward

- **Other places too**
  JS urls, HTML attributes, service workers, etc.

Take Aways…

- Filter lists are porous
- Focusing on URLs is the problem
- Behavior >> delivery
- Brave has plans to protect users in a novel, more comprehensive way

Overview

- **PageGraph**
  Novel, open source, DOM attribution system

- **Behavioral Blocking**
  Protecting Privacy Beyond URLs

- **Regional Filter List Generation**
  Privacy protection for users in underserved regions

- **Automated Web Compatibility Testing**
  Making sure we're only breaking bad stuff…

Alexander Sjösten - sjosten@chalmers.se
Chalmers University of Technology

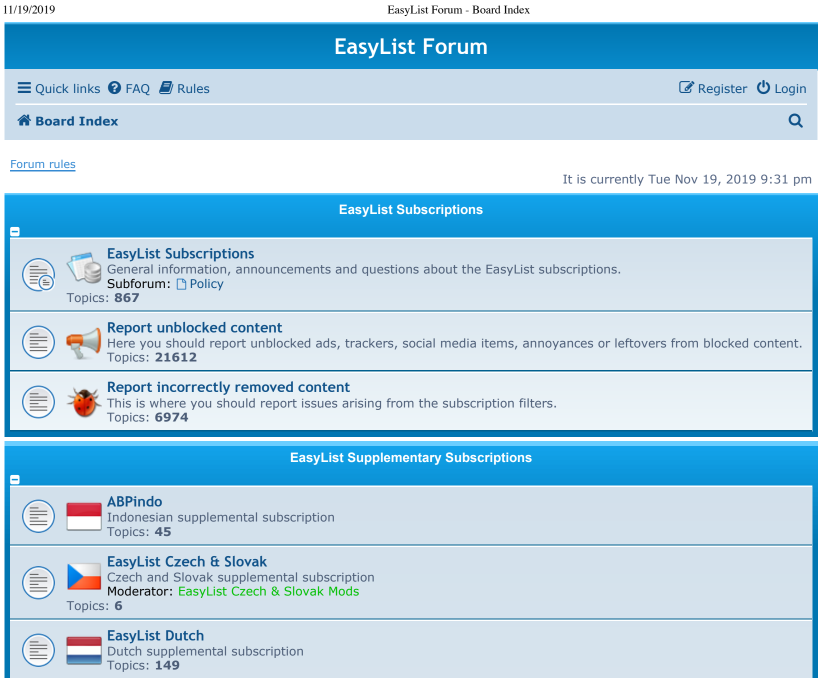
Filter Lists, Unsung Heros…

- **Crowdsourced Lists**
  Network Rules: Requests
  Cosmetic Rules: Display
  Exceptions: Reverse the Above

- **EasyList: Blocks Ads**
  73,079 rules

- **EasyPrivacy: Block Trackers**
  17,024 rules

Crowdsourcing Is Tricky



11/19/2019 EasyList Forum - Board Index

EasyList Forum

Quick links FAQ Rules Register Login

Board Index

Forum rules

It is currently Tue Nov 19, 2019 9:31 pm

EasyList Subscriptions

EasyList Subscriptions
General information, announcements and questions about the EasyList subscriptions.
Subforum: Policy
Topics: **867**

Report unblocked content
Here you should report unblocked ads, trackers, social media items, annoyances or leftovers from blocked content.
Topics: **21612**

Report incorrectly removed content
This is where you should report issues arising from the subscription filters.
Topics: **6974**

EasyList Supplementary Subscriptions

ABPindo
Indonesian supplemental subscription
Topics: **45**

EasyList Czech & Slovak
Czech and Slovak supplemental subscription
Moderator: EasyList Czech & Slovak Mods
Topics: **6**

EasyList Dutch
Dutch supplemental subscription
Topics: **149**

EasyList Dutch
Dutch supplemental subscription
Topics: **149**

EasyList Finland
Finish supplemental subscription
Moderator: EasyList Finland Mods
Topics: **72**

EasyList Germany
German supplemental subscription
Topics: **1507**

EasyList Italy
Italian supplemental subscription
Topics: **261**

EasyList Lithuania
Lithuanian supplemental subscription
Moderator: EasyList Lithuania Mods
Topics: **9**

EasyList Polish
Polish supplemental subscription
Moderator: EasyList Polish Mods
Topics: **18**

EasyList Portuguese
Portuguese supplemental subscription
Moderator: EasyList Spanish Mods
Topics: **12**

https://forums.lanik.us 1/3

What About Everyone Else

- Regions with Fewer Speakers?
- Regions with Less Affluent Speakers?
- Less Internet Users?
- Higher Data Plans?
- etc…

Project In a Nutshell

- **Image and Frame Classifier**
  Heavy, offline, contextual
- **Block High in Request Chain**
  e.g. PageGraph to determine request causes
- **Don’t Break Pages**
  e.g. PageGraph to understand impact of blocking
- **Generate AdBlock Compatible Rules**
  e.g. shippable and shareable

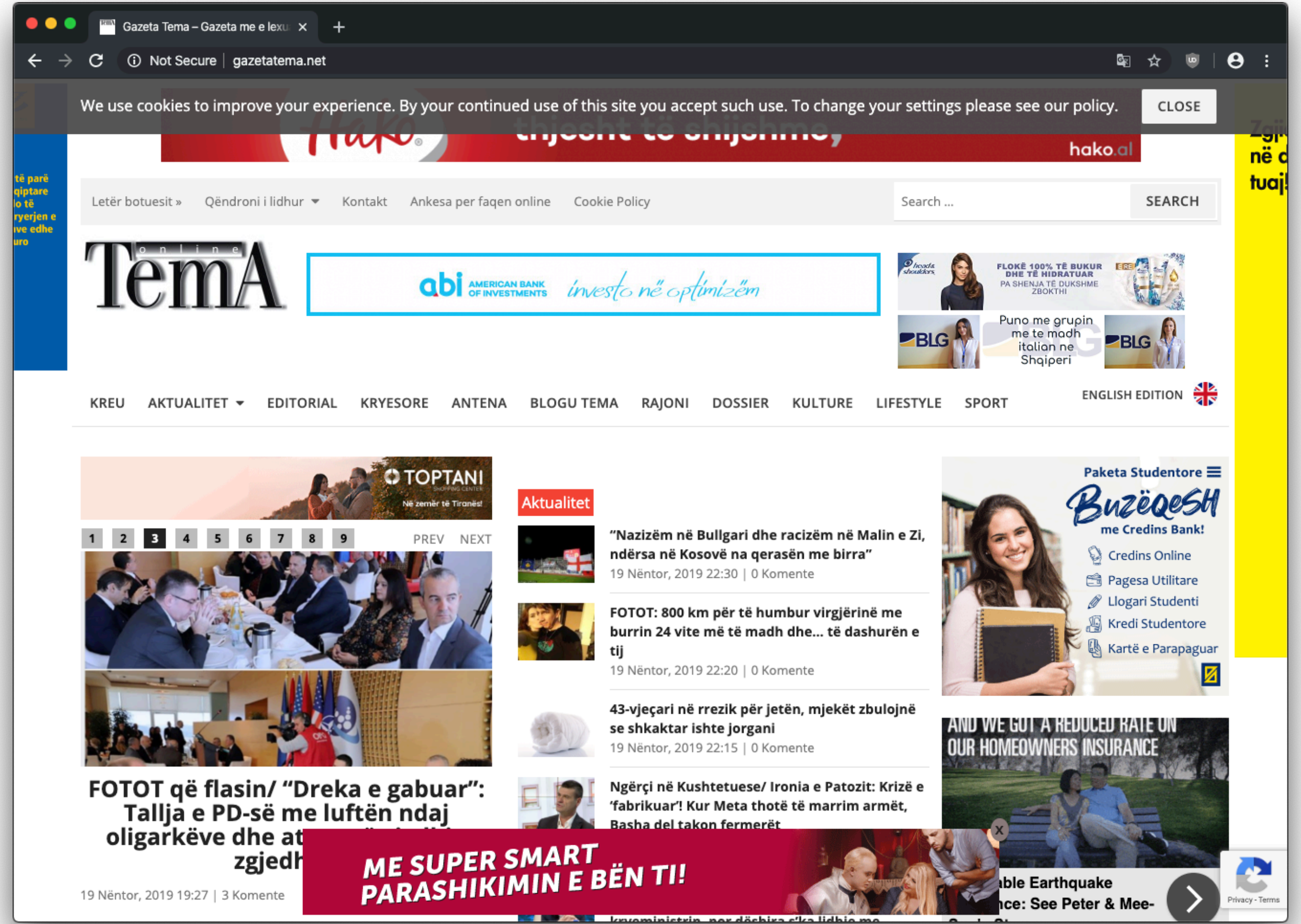
Project In a Nutshell

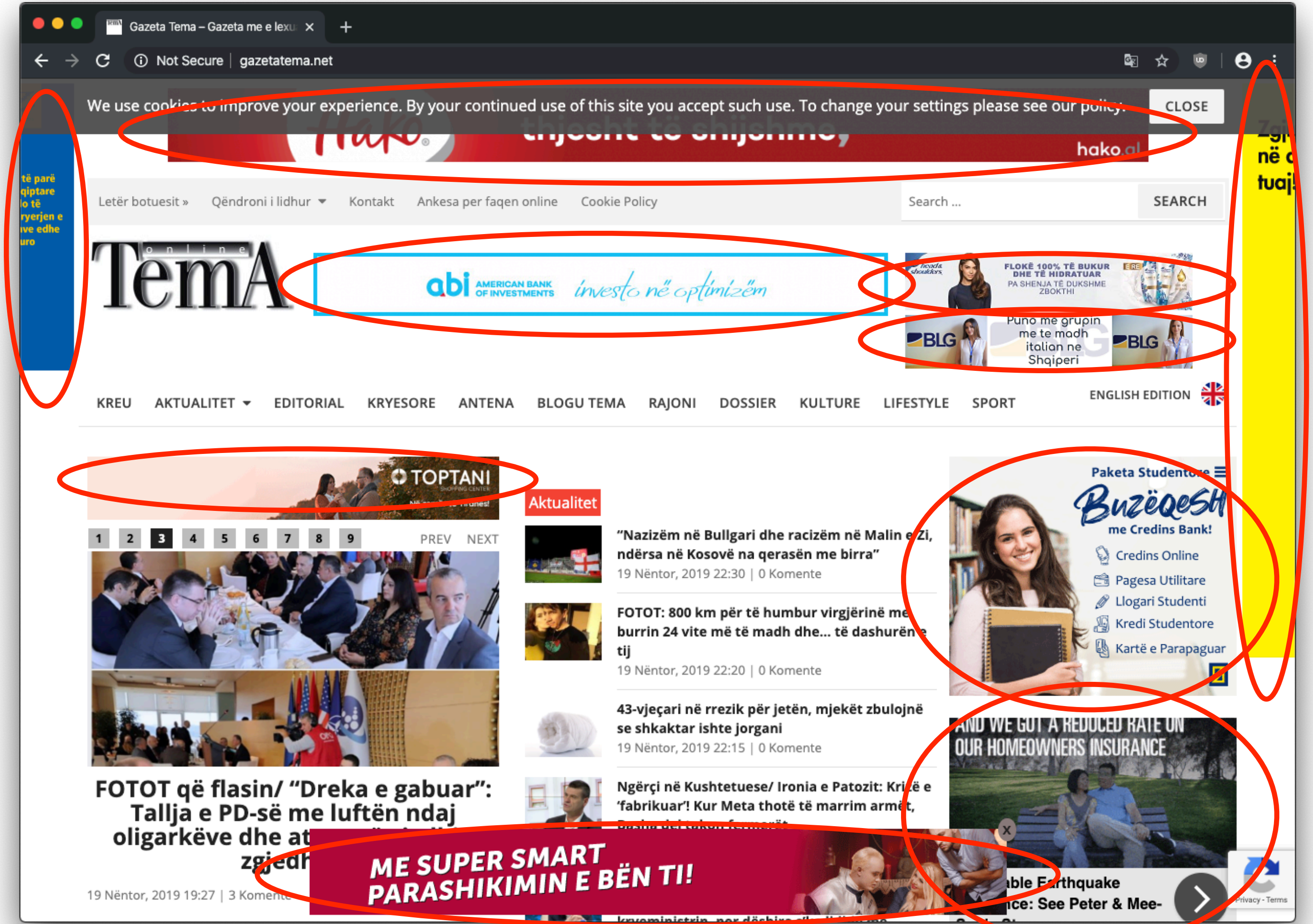
- **Image and Frame Classifier**
  Heavy, offline, contextual

- **Block High in Request Chain**
  e.g. PageGraph to determine request causes

- **Don’t Break Pages**
  e.g. PageGraph to understand impact of blocking

- **Generate AdBlock Compatible Rules**
  e.g. shippable and shareable


Gazeta Tema – Gazeta me e lexu
Not Secure | gazetatema.net
We use cookies to improve your experience. By your continued use of this site you accept such use. To change your settings please see our policy.
CLOSE
hako.al
Letër botuesit »
Qëndroni i lidhur
Kontakt
Ankesa per faqen online
Cookie Policy
Search ...
SEARCH
online
TemA
abi AMERICAN BANK OF INVESTMENTS
investo në optimizëm
FLOKË 100% TË BUKUR DHE TË HIDRATUAR
PA SHENJA TË DUKSHME ZBOKTHI
BLG
Puno me grupin me te madh italian ne Shqiperi
KREU
AKTUALITET
EDITORIAL
KRYESORE
ANTENA
BLOGU TEMA
RAJONI
DOSSIER
KULTURE
LIFESTYLE
SPORT
ENGLISH EDITION
TOPTANI
Në zemër të Tiranës!
1 2 3 4 5 6 7 8 9
PREV NEXT
FOTOT që flasin/ "Dreka e gabuar": Tallja e PD-së me luftën ndaj oligarkëve dhe at
zgjedh
19 Nëntor, 2019 19:27 | 3 Komente
Aktualitet
"Nazizëm në Bullgari dhe racizëm në Malin e Zi, ndërsa në Kosovë na qerasën me birra"
19 Nëntor, 2019 22:30 | 0 Komente
FOTOT: 800 km për të humbur virgjërinë me burrin 24 vite më të madh dhe... të dashurën e tij
19 Nëntor, 2019 22:20 | 0 Komente
43-vjeçari në rrezik për jetën, mjekët zbulojnë se shkaktar ishte jorgani
19 Nëntor, 2019 22:15 | 0 Komente
Ngërçi në Kushtetuese/ Ironia e Patozit: Krizë e 'fabrikuar'! Kur Meta thotë të marrim armët, Basha del takon fermerët
Paketa Studentore
BuzëqeSH
me Credins Bank!
Credins Online
Pagesa Utilitare
Llogari Studenti
Kredi Studentore
Kartë e Parapaguar
AND WE GOT A REDUCED RATE ON OUR HOMEOWNERS INSURANCE
ME SUPER SMART PARASHIKIMIN E BËN TI!
Privacy - Terms



Gazeta Tema – Gazeta me e lexu
Not Secure | gazetatema.net
We use cookies to improve your experience. By your continued use of this site you accept such use. To change your settings please see our policy.
CLOSE
hako.al
Letër botuesit »
Qëndroni i lidhur
Kontakt
Ankesa per faqen online
Cookie Policy
Search ...
SEARCH
TemA
abi AMERICAN BANK OF INVESTMENTS
investo në optimizëm
FLOKË 100% TË BUKUR DHE TË HIDRATUAR
PA SHENJA TË DUKSHME ZBOKTHI
Puno me grupin me te madh italian ne Shqiperi
BLG
KREU
AKTUALITET
EDITORIAL
KRYESORE
ANTENA
BLOGU TEMA
RAJONI
DOSSIER
KULTURE
LIFESTYLE
SPORT
ENGLISH EDITION
TOPTANI
1 2 3 4 5 6 7 8 9 PREV NEXT
FOTOT që flasin/ “Dreka e gabuar”: Tallja e PD-së me luftën ndaj oligarkëve dhe at
19 Nëntor, 2019 19:27 | 3 Komente
Aktualitet
“Nazizëm në Bullgari dhe racizëm në Malin e Zi, ndërsa në Kosovë na qerasën me birra”
19 Nëntor, 2019 22:30 | 0 Komente
FOTOT: 800 km për të humbur virgjërinë me burrin 24 vite më të madh dhe... të dashurën e tij
19 Nëntor, 2019 22:20 | 0 Komente
43-vjeçari në rrezik për jetën, mjekët zbulojnë se shkaktar ishte jorgani
19 Nëntor, 2019 22:15 | 0 Komente
Ngërçi në Kushtetuese/ Ironia e Patozit: Krizë e ‘fabrikuar’! Kur Meta thotë të marrim armët,
Paketa Studentore
Buzëqesh me Credins Bank!
Credins Online
Pagesa Utilitare
Llogari Studenti
Kredi Studentore
Kartë e Parapaguar
AND WE GOT A REDUCED RATE ON OUR HOMEOWNERS INSURANCE
ME SUPER SMART PARASHIKIMIN E BËN TI!
Privacy - Terms

Nike Air Zoom Fencing Shoes Black and Pink

Existing Work Doesn't Generalize

	Accuracy	Precision	Recall
Trained Data	95.9%	95.5%	96.4%
New Data	77.0%	48.8%	87.4%

Contextual Classifier

- **Existing Image Classifier**
  Tigas, Panagiotis, Samuel T. King, and Benjamin Livshits.
  "Percival: Making In-Browser Perceptual Ad Blocking Practical With Deep Learning."

- **PageGraph Contextual Information**
  First or third party URL?
  Injected by script?
  Parent node degree?

Context Helps a Lot…

	Accuracy	Precision	Recall
Trained Data	95.9%	95.5%	96.4%
New Data	77.0%	48.8%	87.4%
Contextual	97.6%	92.0%	75.0%

Project In a Nutshell

- **Image and Frame Classifier**
  Heavy, offline, contextual

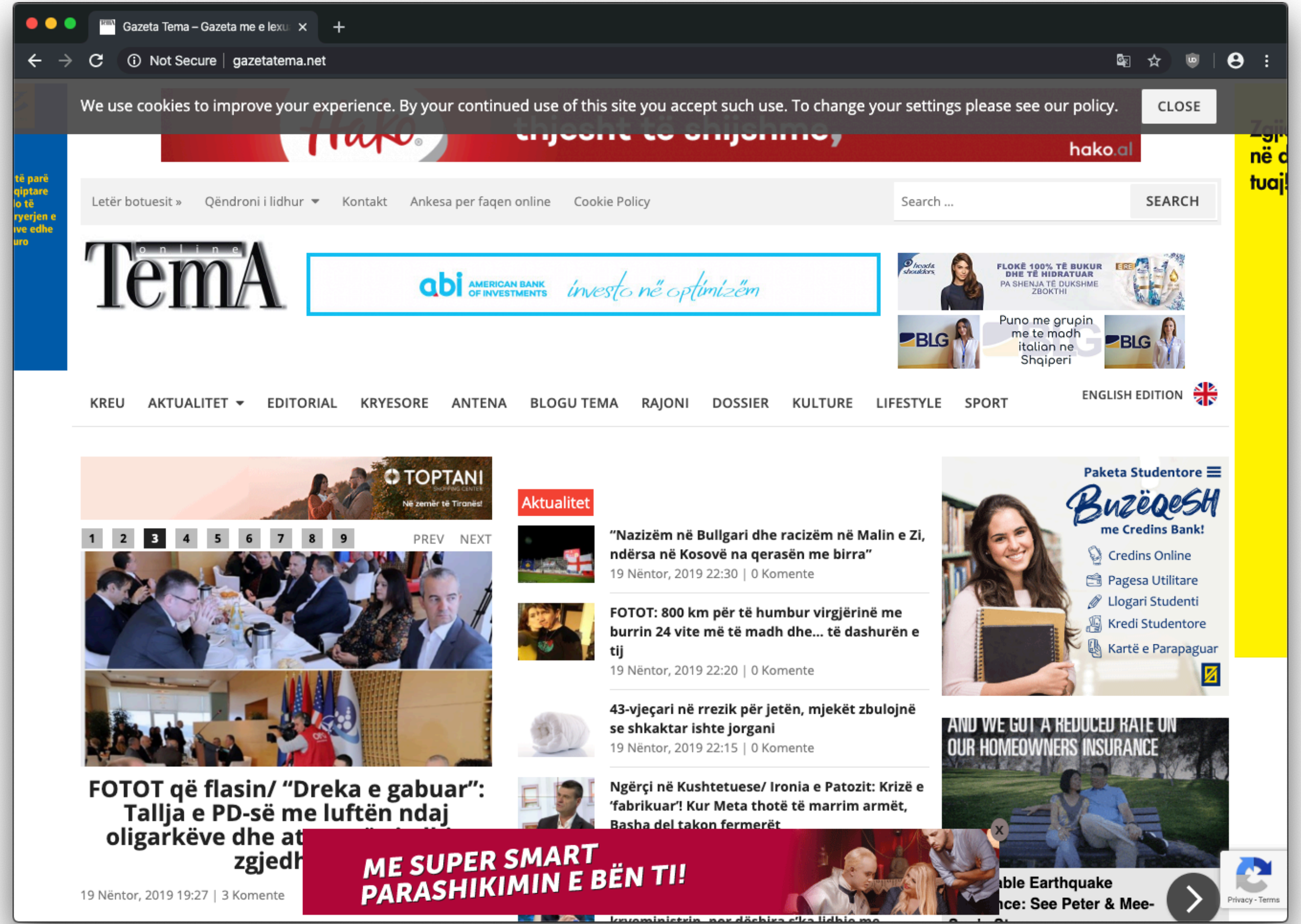
- **Block High in Request Chain**
  e.g. PageGraph to determine request causes

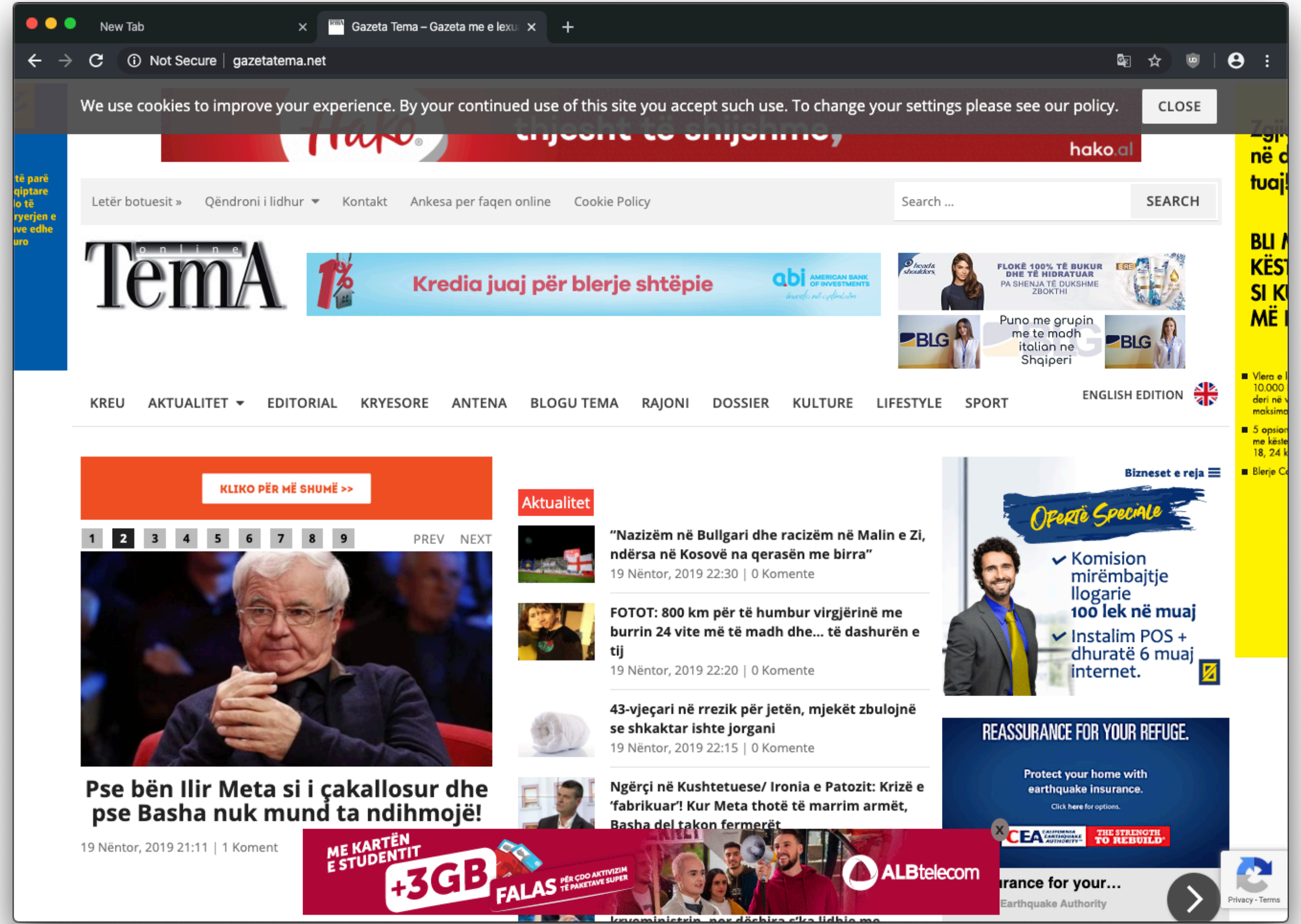
- **Don’t Break Pages**
  e.g. PageGraph to understand impact of blocking

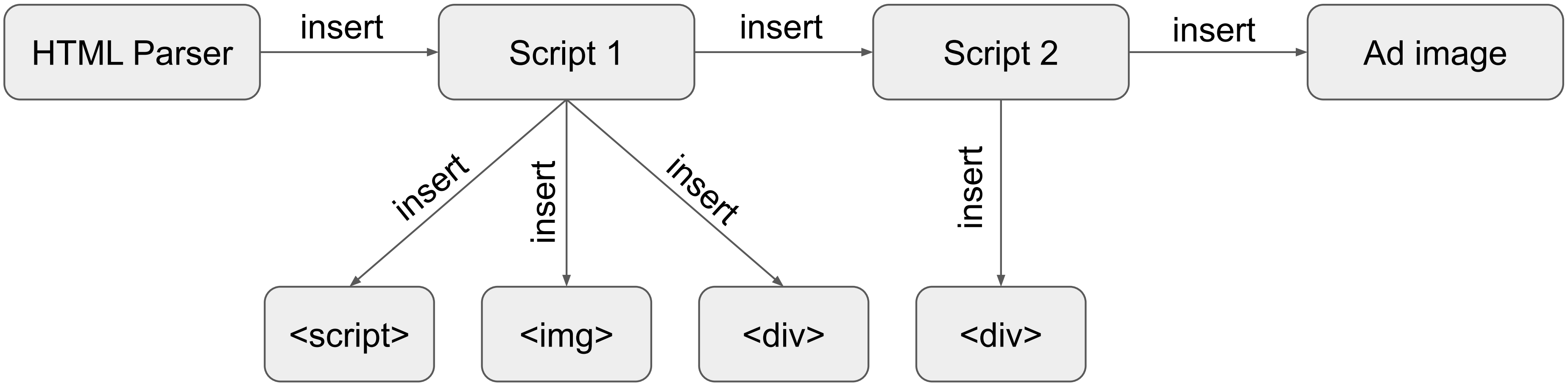
- **Generate AdBlock Compatible Rules**
  e.g. shippable and shareable

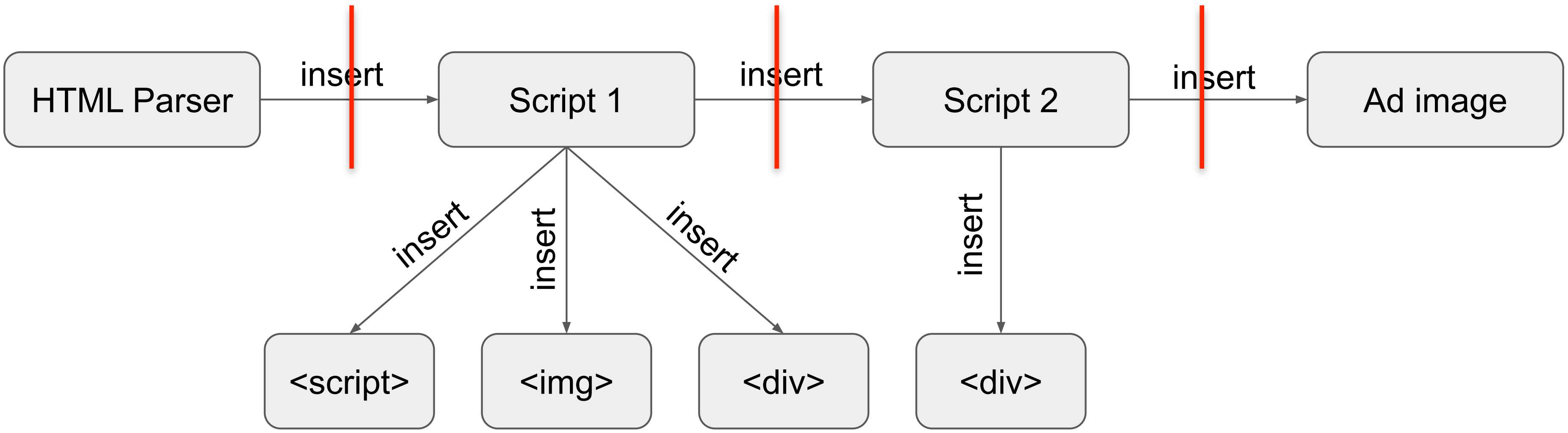
Ads Aren’t (Usually) Static URLs

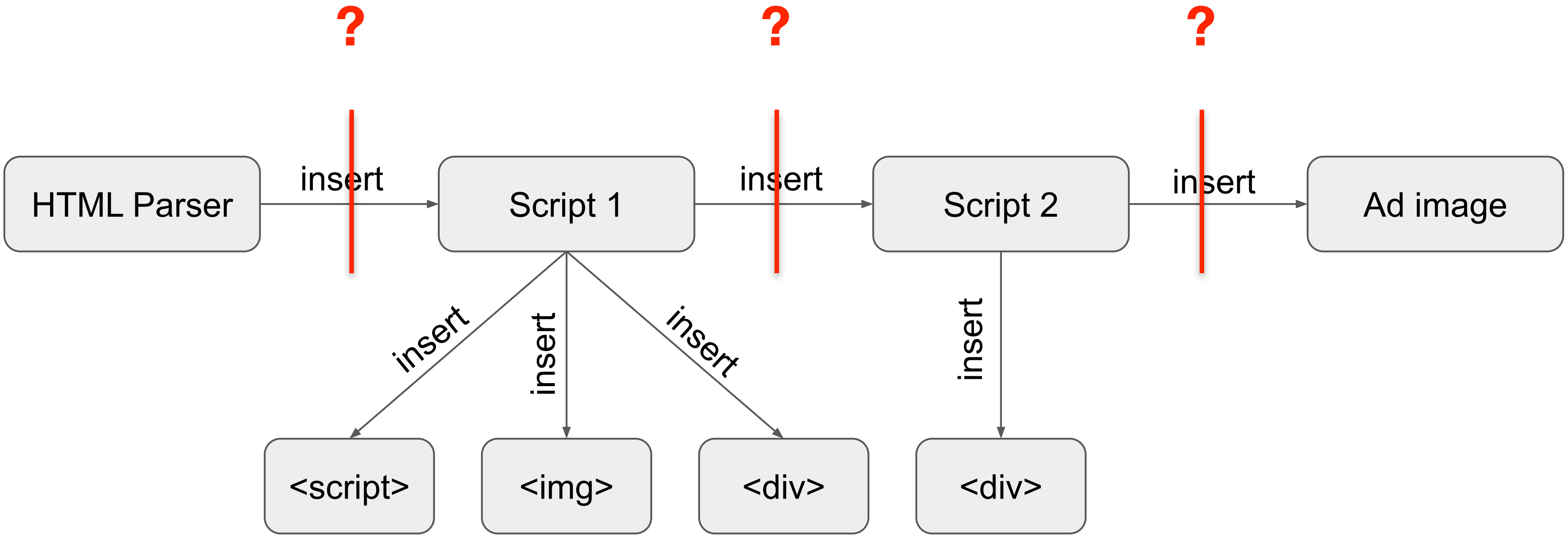
- Programmatically Inserted Scripts
- Real Time Bidding
- Header Bidding
- Targeting Final URLs -> Stale Rules


Gazeta Tema – Gazeta me e lexu
Not Secure | gazetatema.net
We use cookies to improve your experience. By your continued use of this site you accept such use. To change your settings please see our policy.
CLOSE
hako.al
Letër botuesit »
Qëndroni i lidhur
Kontakt
Ankesa per faqen online
Cookie Policy
Search ...
SEARCH
TemA
abi AMERICAN BANK OF INVESTMENTS
investo në optimizëm
FLOKË 100% TË BUKUR DHE TË HIDRATUAR
PA SHENJA TË DUKSHME ZBOKTHI
BLG
Puno me grupin me te madh italian ne Shqiperi
KREU
AKTUALITET
EDITORIAL
KRYESORE
ANTENA
BLOGU TEMA
RAJONI
DOSSIER
KULTURE
LIFESTYLE
SPORT
ENGLISH EDITION
TOPTANI
Në zemër të Tiranës!
1 2 3 4 5 6 7 8 9
PREV NEXT
FOTOT që flasin/ "Dreka e gabuar": Tallja e PD-së me luftën ndaj oligarkëve dhe at
zgjedh
19 Nëntor, 2019 19:27 | 3 Komente
Aktualitet
"Nazizëm në Bullgari dhe racizëm në Malin e Zi, ndërsa në Kosovë na qerasën me birra"
19 Nëntor, 2019 22:30 | 0 Komente
FOTOT: 800 km për të humbur virgjërinë me burrin 24 vite më të madh dhe... të dashurën e tij
19 Nëntor, 2019 22:20 | 0 Komente
43-vjeçari në rrezik për jetën, mjekët zbulojnë se shkaktar ishte jorgani
19 Nëntor, 2019 22:15 | 0 Komente
Ngërçi në Kushtetuese/ Ironia e Patozit: Krizë e 'fabrikuar'! Kur Meta thotë të marrim armët, Basha del takon fermerët
Paketa Studentore
BuzëqesH
me Credins Bank!
Credins Online
Pagesa Utilitare
Llogari Studenti
Kredi Studentore
Kartë e Parapaguar
AND WE GOT A REDUCED RATE ON OUR HOMEOWNERS INSURANCE
ME SUPER SMART PARASHIKIMIN E BËN TI!
Privacy - Terms


New Tab
Gazeta Tema – Gazeta me e lexu
Not Secure
gazetatema.net
We use cookies to improve your experience. By your continued use of this site you accept such use. To change your settings please see our policy.
CLOSE
hako.al
Letër botuesit »
Qëndroni i lidhur
Kontakt
Ankesa per faqen online
Cookie Policy
Search ...
SEARCH
TemA
online
Kredia juaj për blerje shtëpie
abi
Puno me grupin me te madh italian ne Shqiperi
BLG
KREU
AKTUALITET
EDITORIAL
KRYESORE
ANTENA
BLOGU TEMA
RAJONI
DOSSIER
KULTURE
LIFESTYLE
SPORT
ENGLISH EDITION
KLIKO PËR MË SHUMË >>
1 2 3 4 5 6 7 8 9
PREV NEXT
Pse bën Ilir Meta si i çakallosur dhe pse Basha nuk mund ta ndihmojë!
19 Nëntor, 2019 21:11 | 1 Koment
Aktualitet
"Nazizëm në Bullgari dhe racizëm në Malin e Zi, ndërsa në Kosovë na qerasën me birra"
19 Nëntor, 2019 22:30 | 0 Komente
FOTOT: 800 km për të humbur virgjërinë me burrin 24 vite më të madh dhe… të dashurën e tij
19 Nëntor, 2019 22:20 | 0 Komente
43-vjeçari në rrezik për jetën, mjekët zbulojnë se shkaktar ishte jorgani
19 Nëntor, 2019 22:15 | 0 Komente
Ngërçi në Kushtetuese/ Ironia e Patozit: Krizë e 'fabrikuar'! Kur Meta thotë të marrim armët, Basha del takon fermerët
Bizneset e reja
Ofertë Speciale
Komision mirëmbajtje llogarie 100 lek në muaj
Instalim POS + dhuratë 6 muaj internet.
REASSURANCE FOR YOUR REFUGE.
Protect your home with earthquake insurance.
Click here for options.
CEA
THE STRENGTH TO REBUILD
Earthquake Authority
ME KARTËN E STUDENTIT
+3GB
FALAS
ALBtelecom
Privacy - Terms


HTML Parser
insert
Script 1
insert
Script 2
insert
Ad image
insert
insert
insert
insert
<script>
<img>
<div>
<div>


HTML Parser
insert
Script 1
insert
Script 2
insert
Ad image
insert
insert
insert
insert
<script>
<img>
<div>
<div>


?
?
?
HTML Parser
insert
Script 1
insert
Script 2
insert
Ad image
insert
insert
insert
insert
<script>
<img>
<div>
<div>

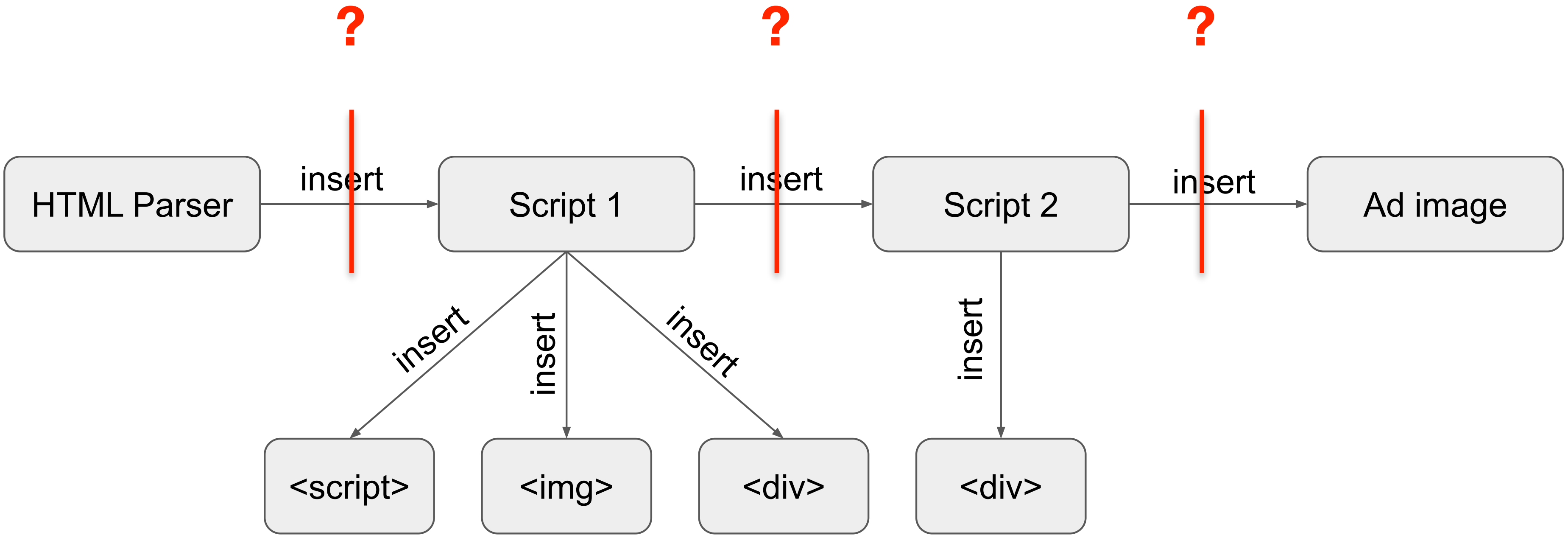
Project In a Nutshell

- **Image and Frame Classifier**
  Heavy, offline, contextual

- **Block High in Request Chain**
  e.g. PageGraph to determine request causes

- **Don't Break Pages**
  e.g. PageGraph to understand impact of blocking

- **Generate AdBlock Compatible Rules**
  e.g. shippable and shareable


?
?
?
HTML Parser
insert
Script 1
insert
Script 2
insert
Ad image
insert
insert
insert
insert
<script>
<img>
<div>
<div>

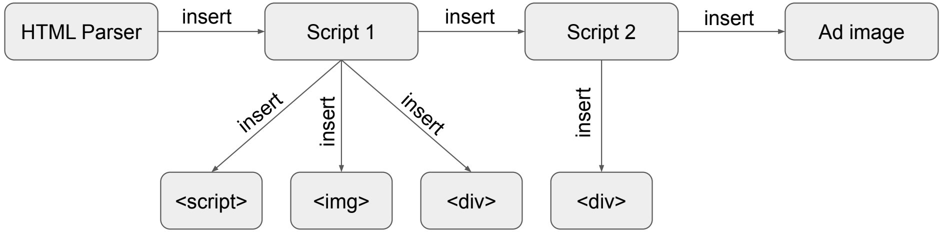
Where to Safely Cut?

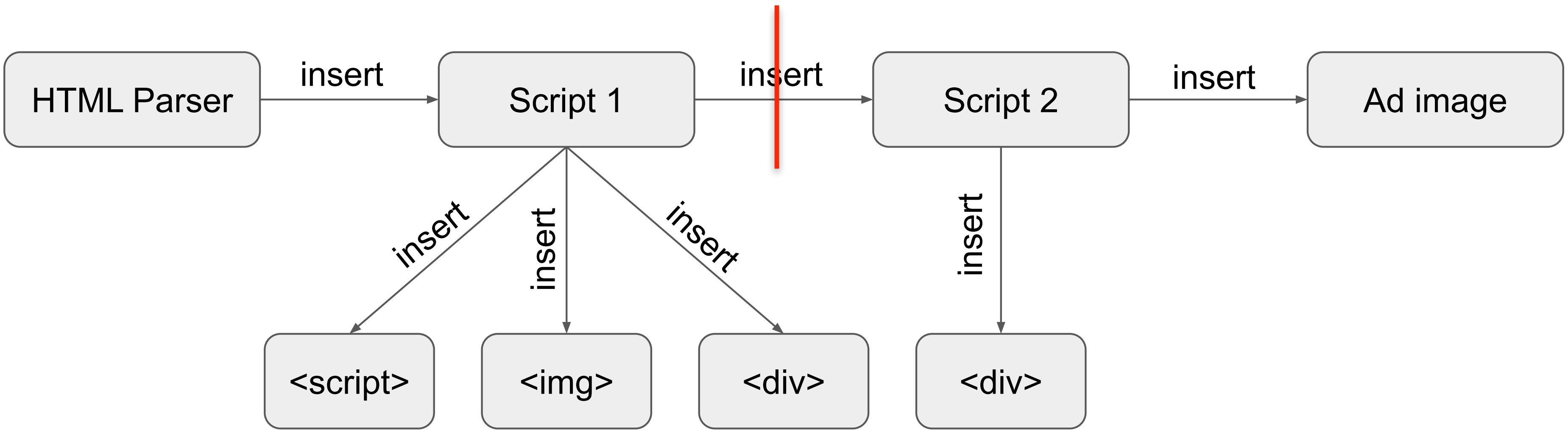
1. Start with identified ad…

2. If a script creates more then two subtrees in the document, **stop**

3. If script inserts a script that that does #2, **stop**

4. Otherwise, **block** and continue up




HTML Parser
insert
Script 1
insert
Script 2
insert
Ad image
insert
insert
insert
insert
<script>
<img>
<div>
<div>

Project In a Nutshell

- **Image and Frame Classifier**
  Heavy, offline, contextual
- **Block High in Request Chain**
  e.g. PageGraph to determine request causes
- **Don’t Break Pages**
  e.g. PageGraph to understand impact of blocking
- <u>**Generate AdBlock Compatible Rules**</u>
  <u>e.g. shippable and shareable</u>

URLs → Filter Rules

1. Take URL
2. Reduce to eTLD+1
3. Remove query params and fragment
4. Remove protocol

1. https://a.good.example.com/ad.html#banner?id=3
2. https://example.com/ad.html#banner?id=3
3. https://example.com/ad.html
4. ||example.com/ad.html

Country	Current Lists	Classifier	∪ Chains	Δ
Albania	3,021	451	521	17.2%
Hungary	2,619	549	736	28.1%
Sri Lanka	2,204	512	644	29.2%
Total	7,844	1,512	1,901	24.2%

Take Aways…

- Filter lists are good (if not great)
- Right now they fail folks who need the most
- ML + serious consideration can help
- Brave will be shipping soon

Overview

- **PageGraph**
  Novel, open source, DOM attribution system

- **Behavioral Blocking**
  Protecting Privacy Beyond URLs

- **Regional Filter List Generation**
  Privacy protection for users in underserved regions

- **Automated Web Compatibility Testing**
  Making sure we’re only breaking bad stuff…

Michael Smith - mds009@ucsd.edu
University of California at San Diego

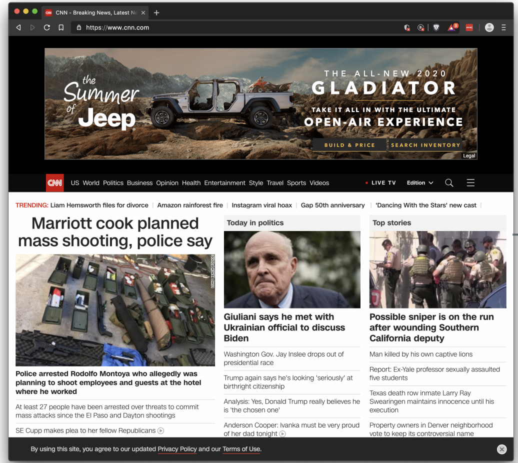
Automated WebCompat: In A Slide…

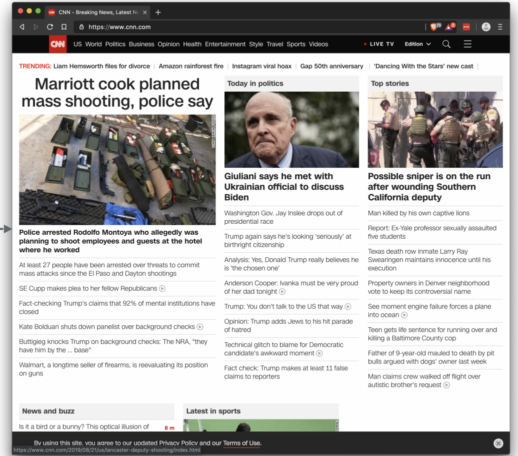
- **Brave makes aggressive modifications**
  And not just Brave…

- **We're Flying Blind**
  Privacy community only has vague intuitions if / when we're breaking things

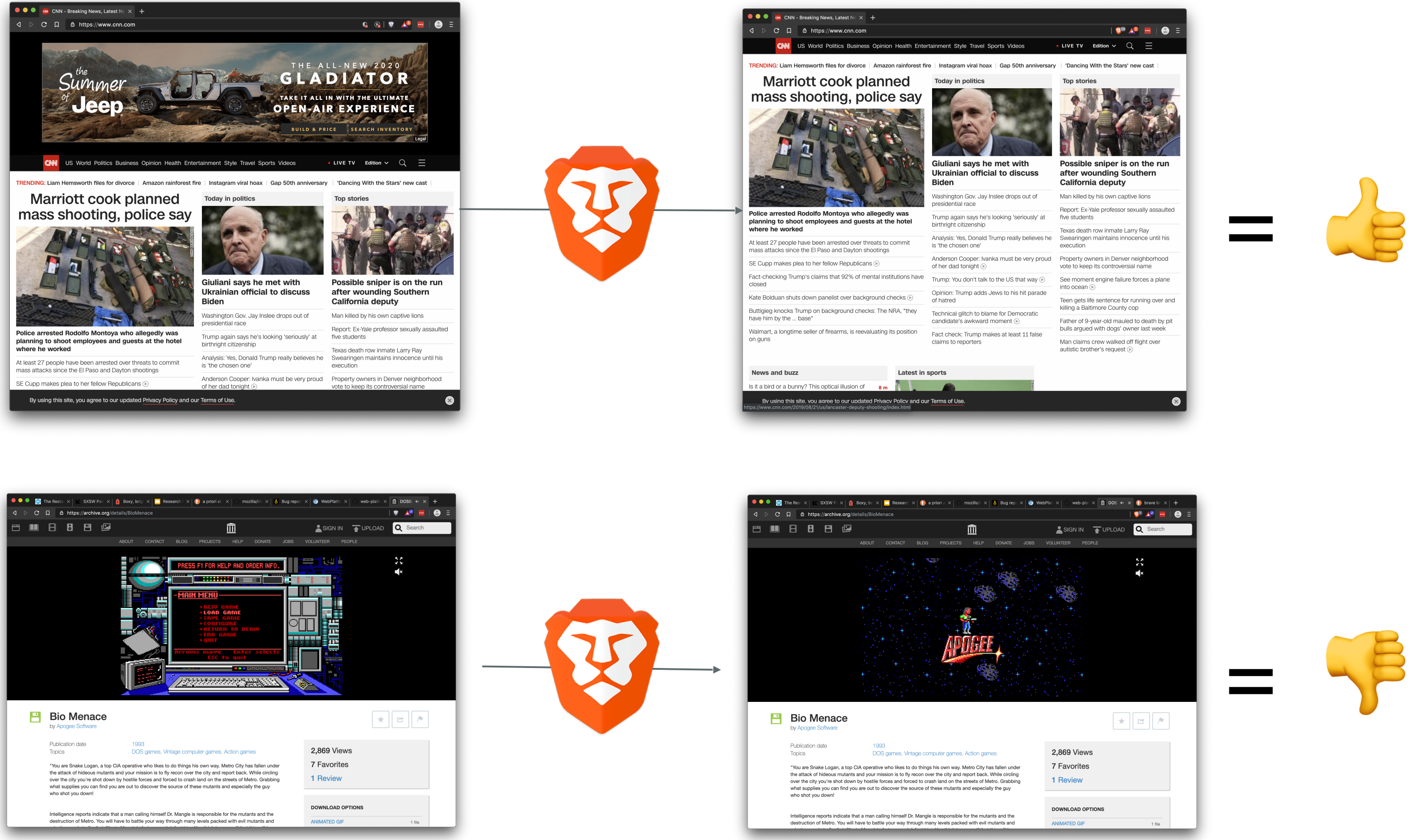
- **This Hurts Privacy**
  Uncertainty -> risk averseness, hard to iterate to improvements

- **This Makes People Want Other Problems**
  Brenden's Twitter blows up, I work late, my plants die 😥


CNN - Breaking News, Latest N
https://www.cnn.com
the Summer of Jeep
THE ALL-NEW 2020
GLADIATOR
TAKE IT ALL IN WITH THE ULTIMATE
OPEN-AIR EXPERIENCE
BUILD & PRICE
SEARCH INVENTORY
Legal
US World Politics Business Opinion Health Entertainment Style Travel Sports Videos
LIVE TV
Edition
TRENDING: Liam Hemsworth files for divorce | Amazon rainforest fire | Instagram viral hoax | Gap 50th anniversary | 'Dancing With the Stars' new cast
Marriott cook planned mass shooting, police say
Today in politics
Top stories
Police arrested Rodolfo Montoya who allegedly was planning to shoot employees and guests at the hotel where he worked
Giuliani says he met with Ukrainian official to discuss Biden
Possible sniper is on the run after wounding Southern California deputy
At least 27 people have been arrested over threats to commit mass attacks since the El Paso and Dayton shootings
SE Cupp makes plea to her fellow Republicans
Washington Gov. Jay Inslee drops out of presidential race
Trump again says he's looking 'seriously' at birthright citizenship
Analysis: Yes, Donald Trump really believes he is 'the chosen one'
Anderson Cooper: Ivanka must be very proud of her dad tonight
Man killed by his own captive lions
Report: Ex-Yale professor sexually assaulted five students
Texas death row inmate Larry Ray Swearingen maintains innocence until his execution
Property owners in Denver neighborhood vote to keep its controversial name
By using this site, you agree to our updated Privacy Policy and our Terms of Use.


CNN - Breaking News, Latest N
https://www.cnn.com
US World Politics Business Opinion Health Entertainment Style Travel Sports Videos
LIVE TV
Edition
TRENDING: Liam Hemsworth files for divorce | Amazon rainforest fire | Instagram viral hoax | Gap 50th anniversary | 'Dancing With the Stars' new cast
Marriott cook planned mass shooting, police say
Today in politics
Top stories
Police arrested Rodolfo Montoya who allegedly was planning to shoot employees and guests at the hotel where he worked
Giuliani says he met with Ukrainian official to discuss Biden
Possible sniper is on the run after wounding Southern California deputy
At least 27 people have been arrested over threats to commit mass attacks since the El Paso and Dayton shootings
SE Cupp makes plea to her fellow Republicans
Fact-checking Trump's claims that 92% of mental institutions have closed
Kate Bolduan shuts down panelist over background checks
Buttigieg knocks Trump on background checks: The NRA, "they have him by the ... base"
Walmart, a longtime seller of firearms, is reevaluating its position on guns
Washington Gov. Jay Inslee drops out of presidential race
Trump again says he's looking 'seriously' at birthright citizenship
Analysis: Yes, Donald Trump really believes he is 'the chosen one'
Anderson Cooper: Ivanka must be very proud of her dad tonight
Trump: You don't talk to the US that way
Opinion: Trump adds Jews to his hit parade of hatred
Technical glitch to blame for Democratic candidate's awkward moment
Fact check: Trump makes at least 11 false claims to reporters
Man killed by his own captive lions
Report: Ex-Yale professor sexually assaulted five students
Texas death row inmate Larry Ray Swearingen maintains innocence until his execution
Property owners in Denver neighborhood vote to keep its controversial name
See moment engine failure forces a plane into ocean
Teen gets life sentence for running over and killing a Baltimore County cop
Father of 9-year-old mauled to death by pit bulls argued with dogs' owner last week
Man claims crew walked off flight over autistic brother's request
News and buzz
Latest in sports
Is it a bird or a bunny? This optical illusion of
By using this site, you agree to our updated Privacy Policy and our Terms of Use.
https://www.cnn.com/2019/08/21/us/lancaster-deputy-shooting/index.html

= 👍


Marriott cook planned mass shooting, police say
Today in politics
Top stories
Giuliani says he met with Ukrainian official to discuss Biden
Possible sniper is on the run after wounding Southern California deputy
Police arrested Rodolfo Montoya who allegedly was planning to shoot employees and guests at the hotel where he worked
Bio Menace
by Apogee Software
2,869 Views
7 Favorites
1 Review
DOWNLOAD OPTIONS

WebCompat Problem Cavalcade

- Missing desired images
- Missing / broken event registrations
- Media playback
- Navigation problems
- Unstyled content
- Authentication
- Broken redirections
- Missing embedded content
- Form submissions
- Etc etc etc etc etc (etc!)

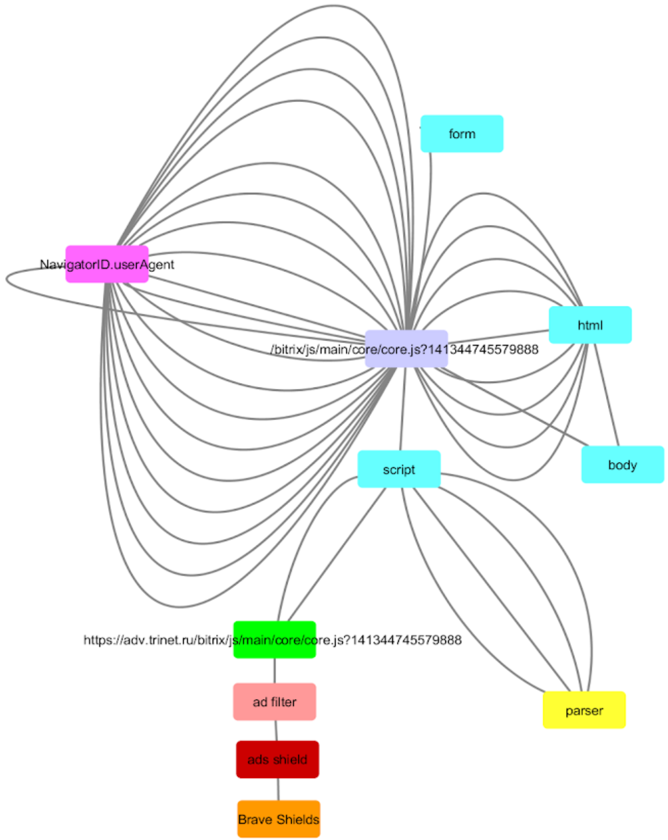
What Makes Web Compat Difficult

- **Existing work focuses on fulfilling promises to site owners**
  - Web Platform Tests: https://github.com/web-platform-tests/wpt
  - WebCompat.org: https://webcompat.com/
  - etc…
- **We want to measure subjective user experience**
  - “If I break standard X, will users get upset”
- **Limited existing work**
  - Academic work is useful but naive
  - Remaining, subjective work has to do with accessibility

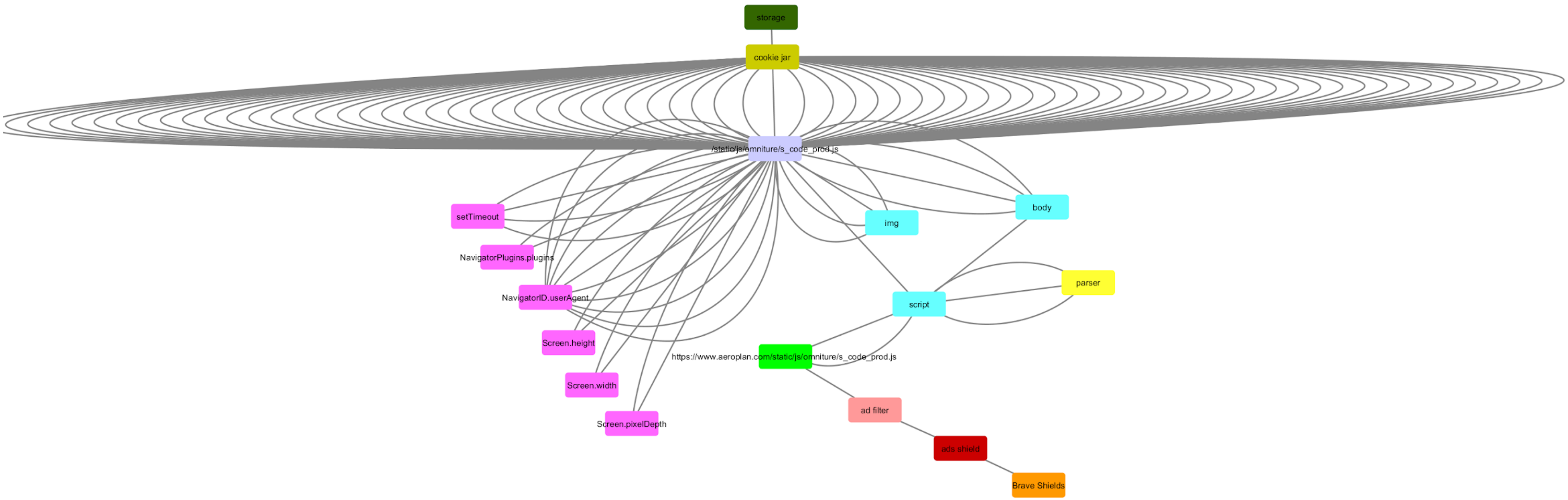
Bounding the Problem (First Cut)

- Recall > Precision
- Top level frame and local frames
- Target things that are broken from the get go

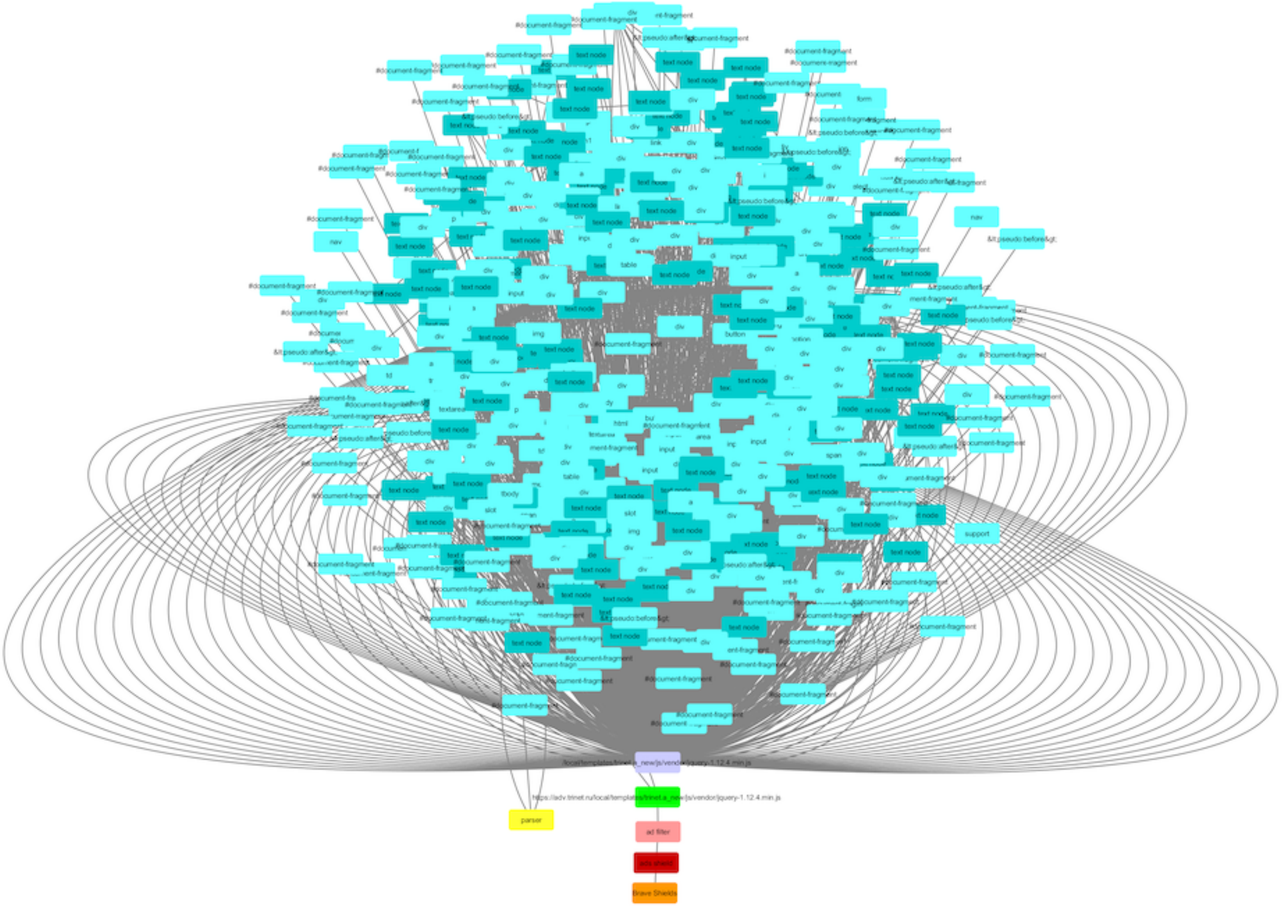
Random script



Tracking script



jQuery



Take Aways…

- Privacy community is drowning in suggestions
- We ignore half the ledger
- Useable privacy -> More privacy

Wrapping Up

- **PageGraph**
  Novel, open source, DOM attribution system

- **Behavioral Blocking**
  Protecting Privacy Beyond URLs

- **Regional Filter List Generation**
  Privacy protection for users in underserved regions

- **Automated Web Compatibility Testing**
  Making sure we’re only breaking bad stuff…

Internships

- **Research focused**
  Papers published at S&P, WWW, USENIX, MADWeb, etc.

- **Engineering focused**
  Shipping features to related to measurement, improved blocking, detection, and more

- **San Francisco and London offices**
  https://brave.com/careers/?gh_jid=896018

Thank You

- **Brave is Uniquely Doing Deployment-Minded Privacy Research**
- **Short Term: Better Regional Filter Lists**
- **Medium Term: Behavioral Based Blocking**
- **Long(ish) Term: Automatic WebCompat**

Pete Snyder
Privacy Researcher
pes@brave.com
@pes10k