The Ad Wars: Retrospective Measurement and Analysis of Anti-Adblock Filter Lists

Umar Iqbal*, Zubair Shafiq*, and Zhiyun Qian†

The University of Iowa*
University of California-Riverside†

Agenda

The Ad Wars

- Online ads
- Adblocking
- Anti-Adblocking
- Anti-Anti-Adblocking

Contributions

- Anti-Adblock filter list analysis
- Retrospective coverage analysis
- Detecting Anti-Adblock Scripts

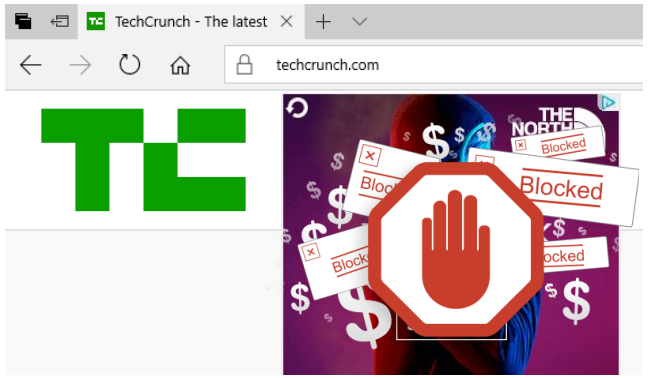
Conclusion

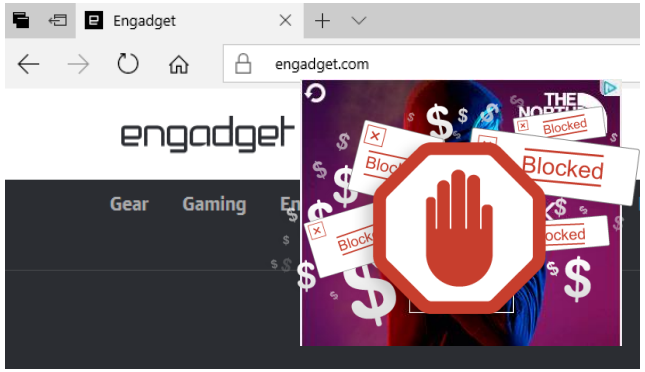
Online Advertising

Advertising enables free content
- Publishers show free content
- Earn revenue with ads

Problems with ads
- Privacy
- Intrusive
- Malware
- Performance

Solution
- Adblocking





Ad/Tracker Blocking Solutions

Trackers Blocking Extensions

Privacy Badger & Ghostery


Adblocking Extensions

Adblock Plus & Adblock

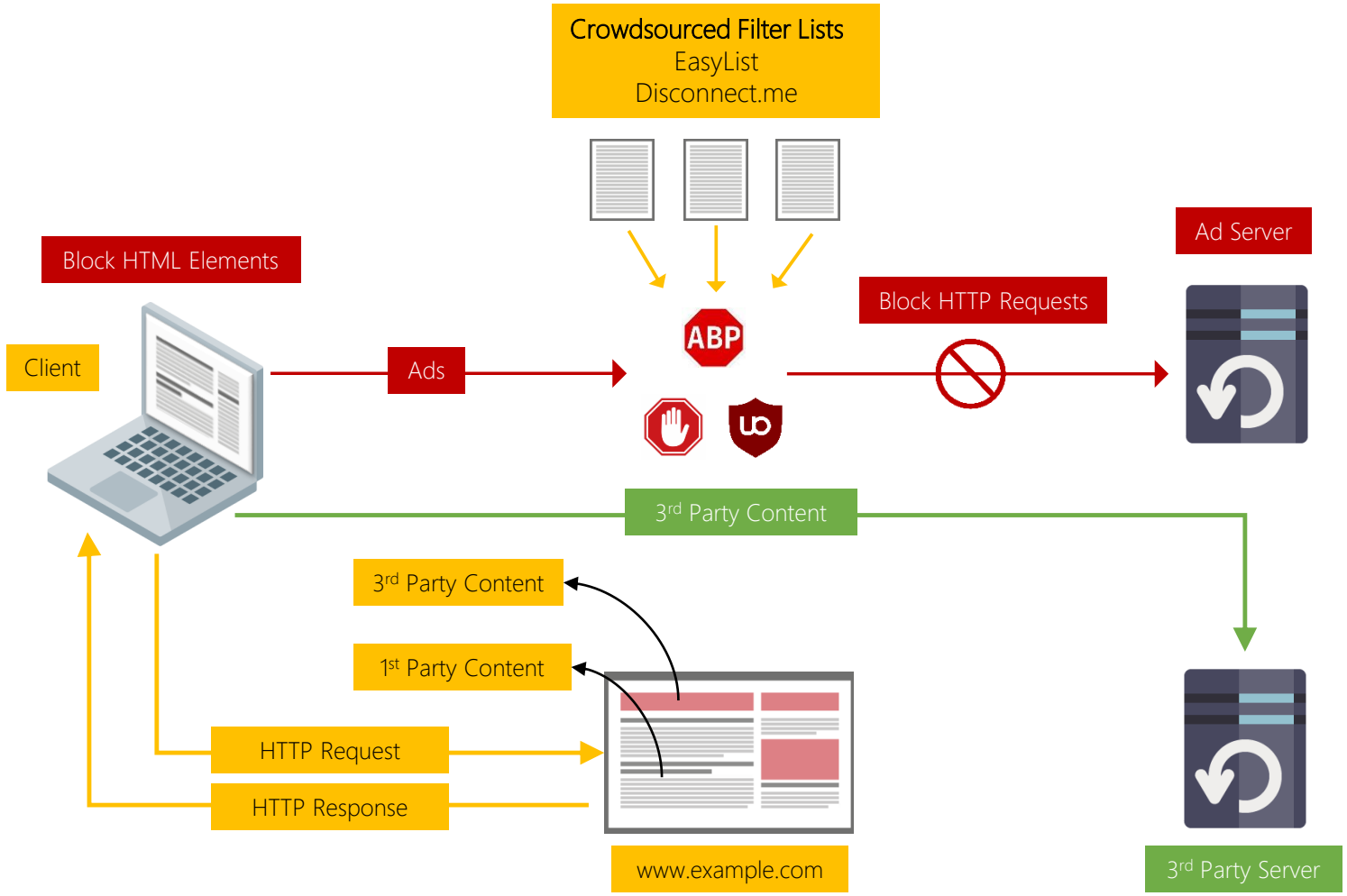
Ad/Tracker Blocking Browsers

Brave browser & Cliqz browser


Mainstream Ad/Tracker Blocking Browsers

Apple Safari & Google Chrome

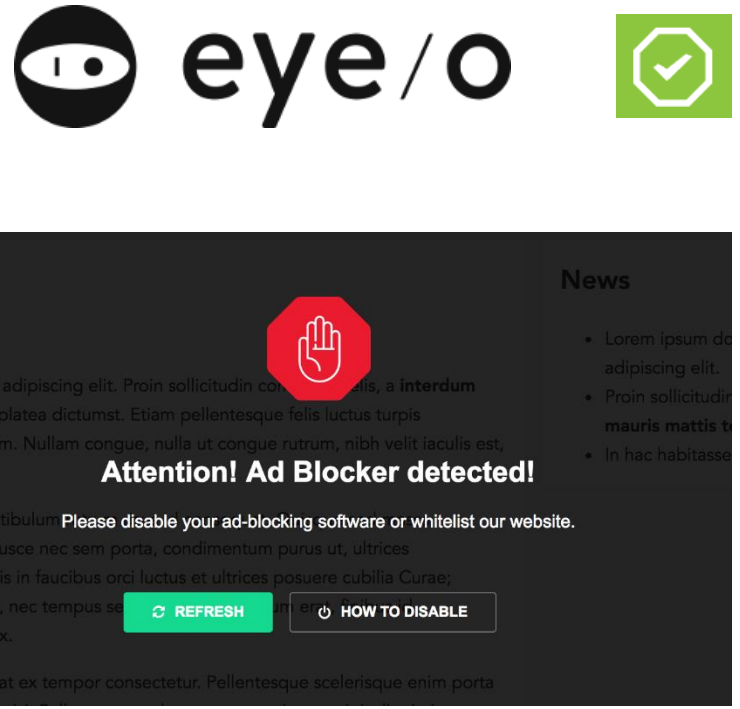

How do Adblockers Work?



Publishers vs Adblockers

Acceptable Ads Program

- Whitelisting fee
- Transparency concerns
- Enabled by default in major adblockers



Use of Anti-Adblockers

- Insert bait elements
- Detect adblockers
- Prompt to disable adblockers/whitelist website

Anti-Anti-Adblocking

Block/allow bait HTTP requests

Hide/allow bait HTML elements

Use anti-adblocking filter lists

- Anti-Adblock Killer
- EasyList

Agenda

The Ad Wars

- Online ads
- Adblocking
- Anti-Adblocking
- Anti-Anti-Adblocking

Contributions

- Anti-Adblock filter list analysis
- Retrospective coverage analysis
- Detecting Anti-Adblock Scripts

Conclusion

Filter List Rules

HTTP Request Filter Rules

- Domain anchor ||
- Domain tag domain=

! Rule with domain anchor
|| example.com

! Rule with domain tag
/example.js $script, **domain** = example.com

HTML Element Filter Rules

- Domain restriction
- Without domain restriction

! Rule with domain restriction
example.com###examplebanner

! Rule without domain restriction
###examplebanner

Exception Rules

- HTTP exception rules
- HTML exception rules

! Exception rule for HTTP request
@@/example.js $script domain = example1.com

! Exception rule for HTML element
example.com#@##examplebanner

Popular Filter Lists

Anti-Adblock Killer (2014)

353 to 1,811 filter rules

6.2 filter rules for every revision

EasyList (2011)

Anti-Adblock sections

67 to 1,317 filter rules

0.6 filter rules per day

Warning Removal List (2013)

4 to 167 filter rules

0.2 filter rules per day

} EasyList + Warning Removal List

Combined EasyList

Anti-Adblock Killer vs Combined EasyList

Number of domains

- Anti-Adblock Killer 1,415
- Combined EasyList 1,394
- Common domains 282

Similar distr

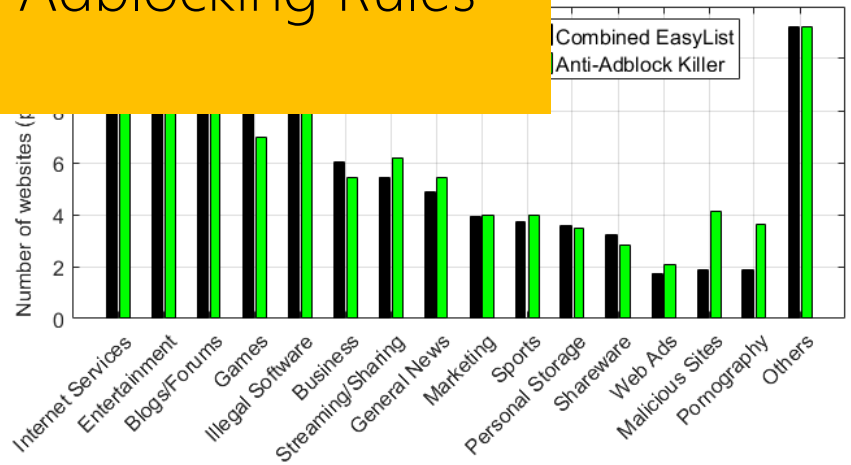
Similar distribution for categories

Exception vs Non-Exception domains

- Combined EasyList 4:1
- Anti-Adblock Killer 1:1

Alexa Rank	Anti-Adblock Killer List	Combined EasyList
1-5K	112	124
5K-10K	49	69
	0	312
	4	359
	0	530

Different Strategies of Crafting Anti-Adblocking Rules

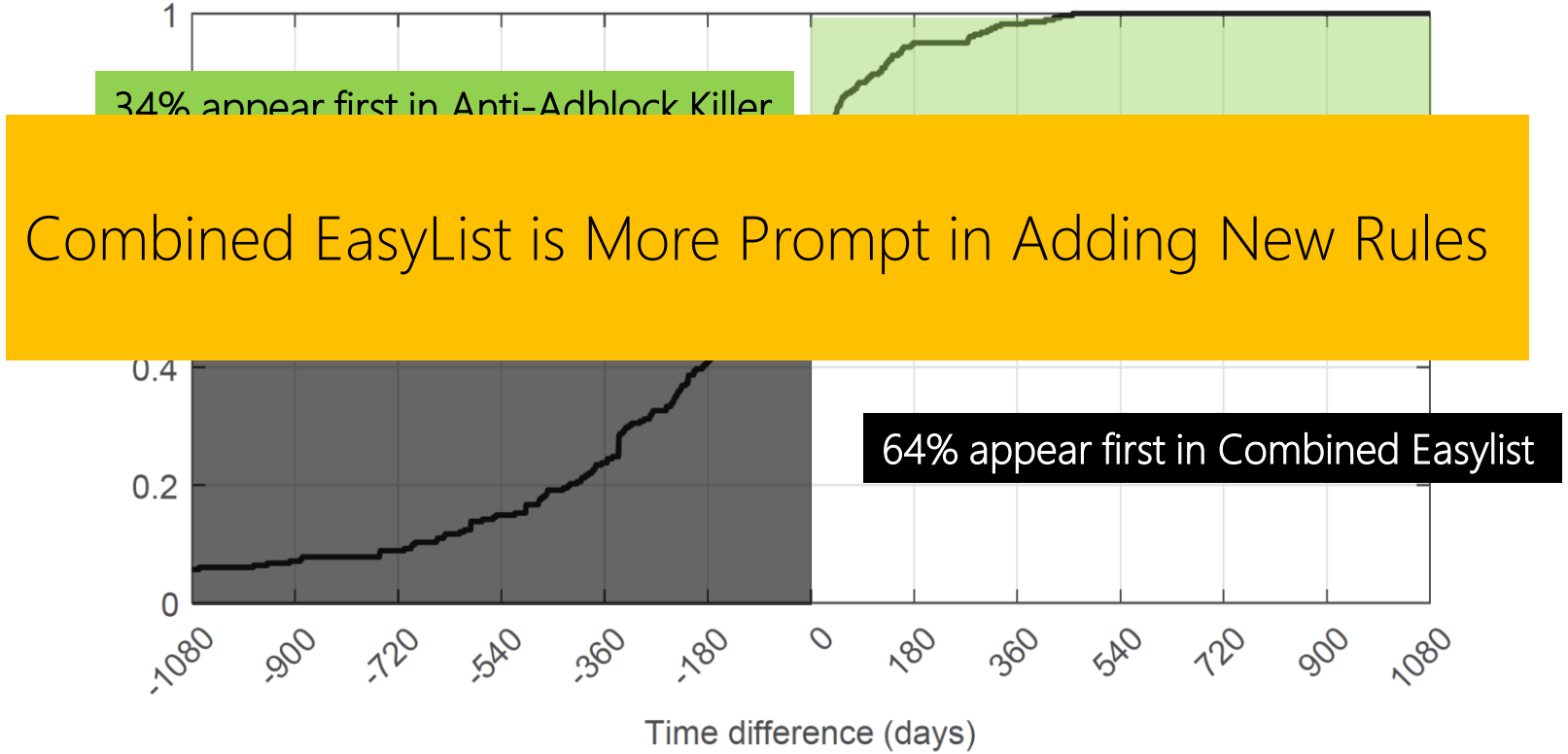


Domain Categorization

Anti-Adblock Killer vs Combined EasyList

282 common domains

Prompt in adding new rules



Agenda

The Ad Wars

- Online ads
- Adblocking
- Anti-Adblocking
- Anti-Anti-Adblocking

Contributions

- Anti-Adblock filter list analysis
- Retrospective coverage analysis
- Detecting Anti-Adblock Scripts

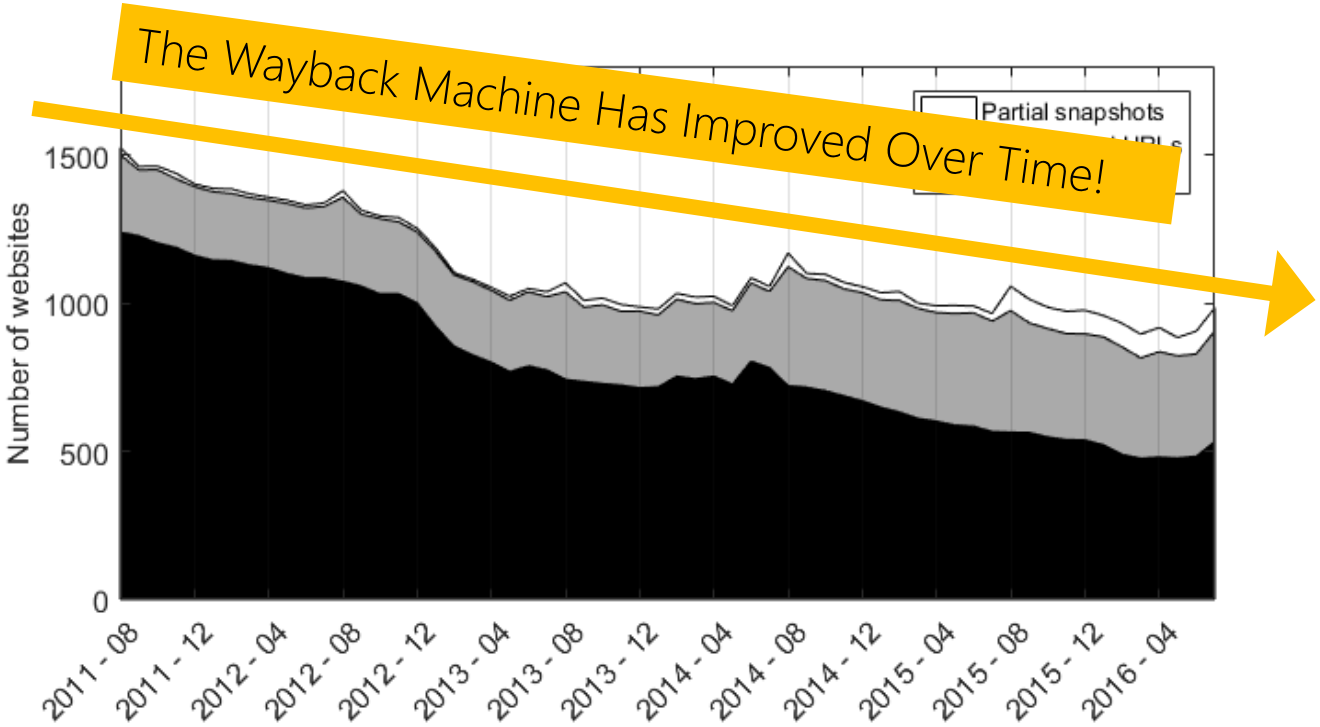
Conclusion

The Internet Archive's Wayback Machine

Archives web pages
- 279 billion webpages
- Archives webpage resources as well
- Used in prior literature [USENIX Security '16]
- API to retrieve content

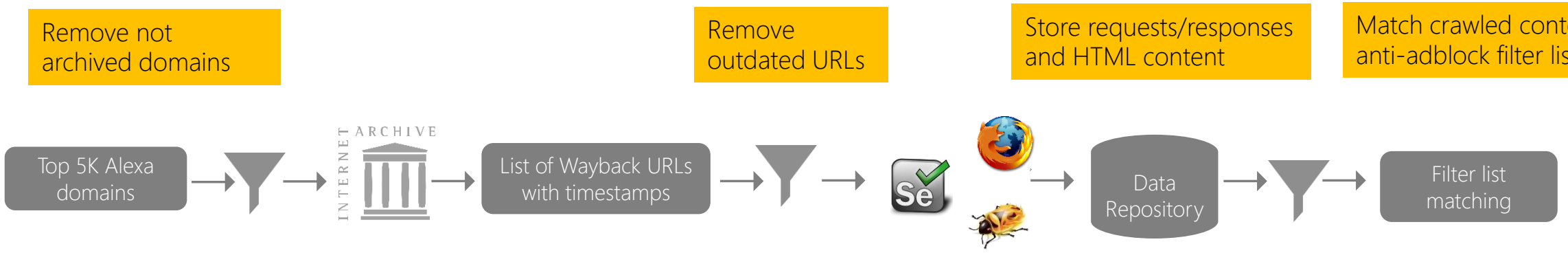
Alexa top 5K websites
- 5 years (2011 – 2016)

Wayback Machine is incomplete!
- robots.txt permissions
- Partial snapshots
- Outdated URLs
- Not archived URLs



Missing Snaphots

Analysis Workflow



Anti-Adblock Filter Lists Coverage

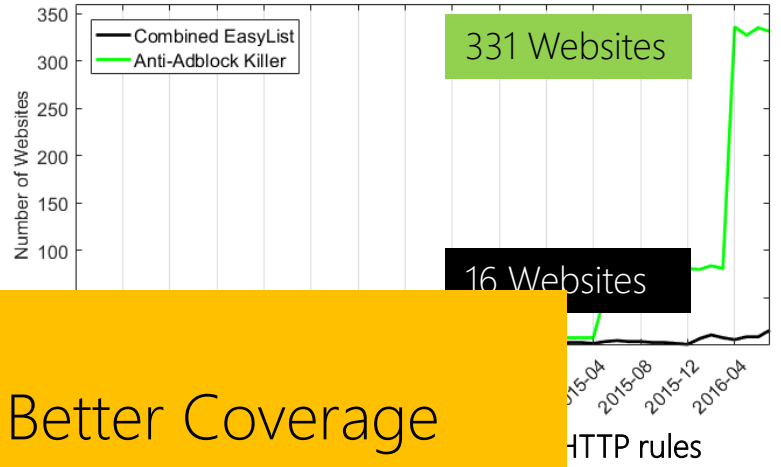
HTTP matching

HTML matching

Use resp

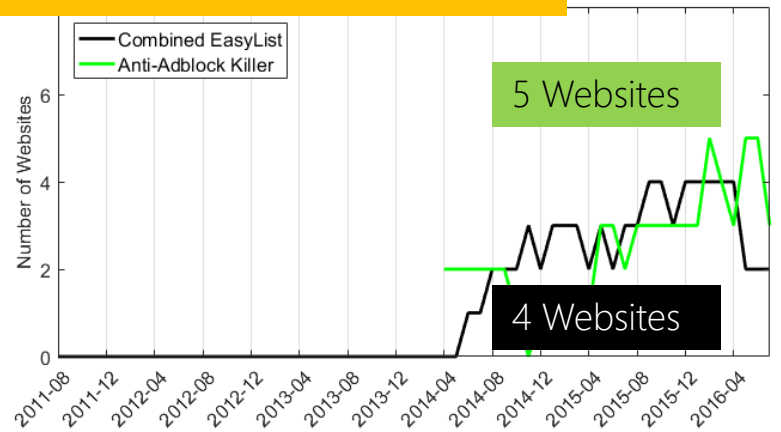
Anti-Adblock Killer filter list

Combined EasyList filter list



HTTP rules

Anti-Adblock Killer Filter List Has Better Coverage



Number of websites that trigger HTML rules

Anti-Adblock Filter Lists Coverage

Detection on the Live Web

Alexa top 100K

Anti-Ac

4

Anti-Adblock Killer Filter List Has Better Coverage on the Live Web

Combined EasyList

195 websites

Anti-Adblock Filter Lists Lag

Crowdsourced

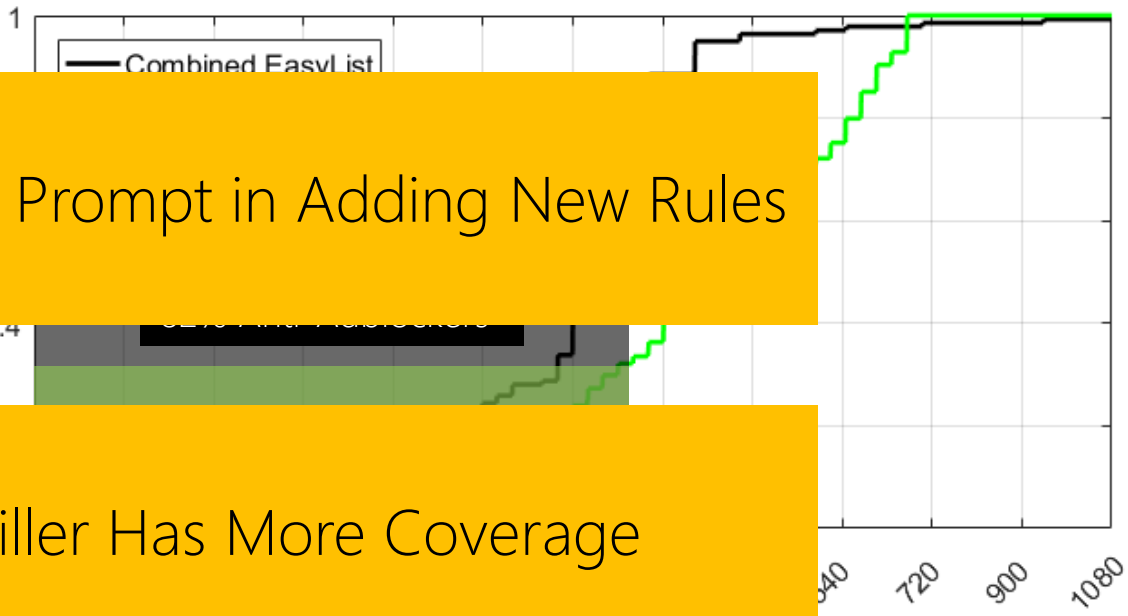
Manually

Challeng

New rule

Combi

Anti-A



Combined EasyList is More Prompt in Adding New Rules

While Anti-Adblock Killer Has More Coverage

Agenda

The Ad Wars

- Online ads
- Adblocking
- Anti-Adblocking
- Anti-Anti-Adblocking

Contributions

- Anti-Adblock filter list analysis
- Retrospective coverage analysis
- Detecting Anti-Adblock Scripts

Conclusion

Static Code Analysis

Anti-Adblocking code from 3rd party vendors

Anti-Adblocking code have structural similarities

Static analysis to capture code structure

Fingerprint anti-adblocking JavaScript

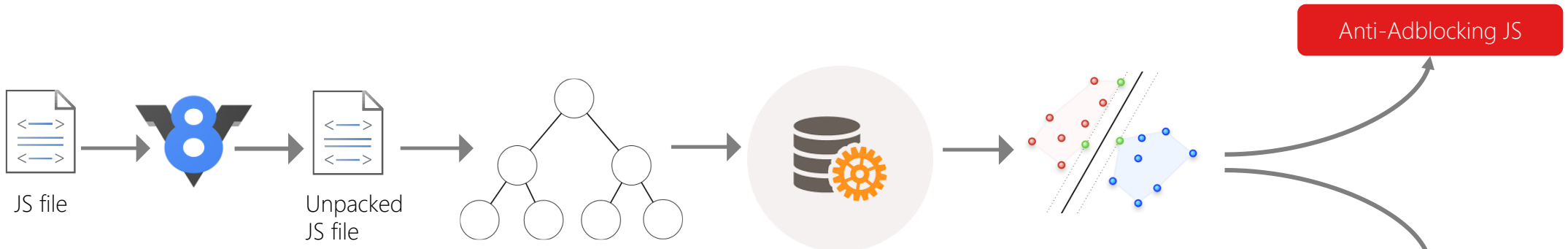
- Curtsinger [USENIX Security '11]
- Ikram [PETS '17]

Anti-Adblock Detection Workflow

Unpack packed JavaScript files with V8 engine

Extract features from ASTs and filter features with low correlation

Classify Anti-Adblocking and Non Anti-Adblocking JavaScripts



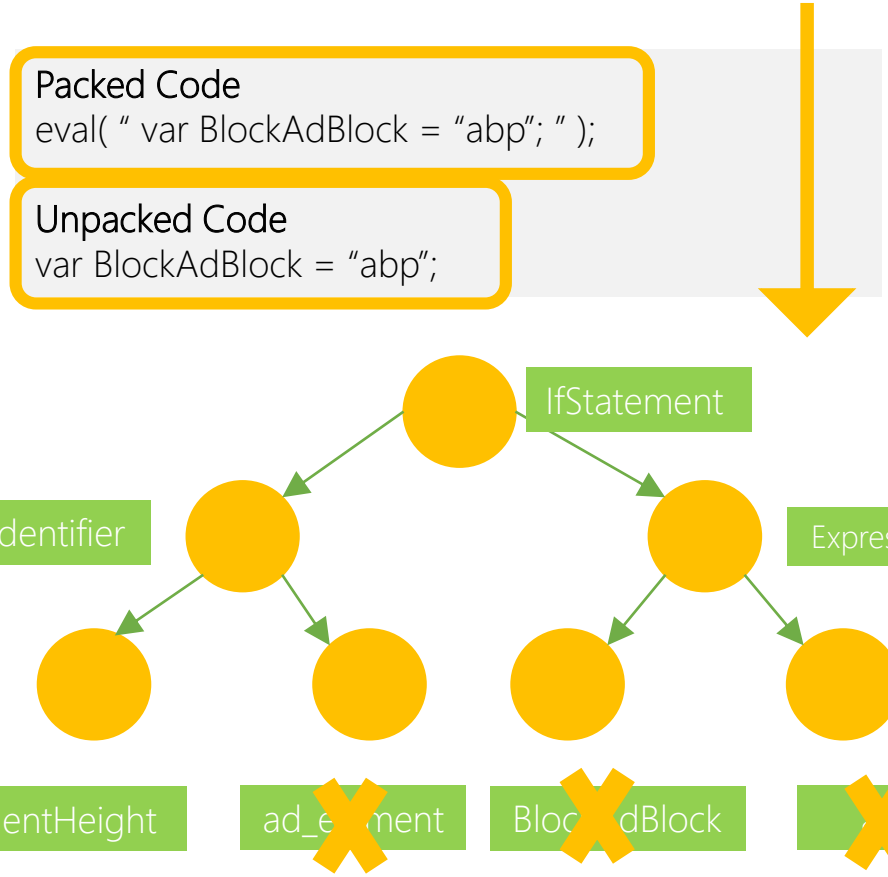
Construct ASTs from Unpacked JavaScript Code

Train AdaBoost using SVM as base classifier

Feature Extraction

Preprocessing

- Unpack eval() using V8 Engine
- Construct Abstract Syntax Tree (AST)



Features (context : text)

All	(AssignmentExpression:BlockAdBlock)
Literal	(Literal:abp)
Keyword	(Identifier:clientHeight)

Map scripts to a vector space

$$\emptyset : x \rightarrow (\emptyset_s(x))_{s \in S}$$

$$\emptyset_s(x) = \begin{cases} 1, & if\ x\ contains\ the\ feature\ s \\ 0, & otherwise \end{cases}$$

Feature Selection & Training

Labeled Data

- 372 anti-adblocking
- 4021 non anti-adblocking

Feature selection

- Filter using χ^2 correlation
- Reduce features

Classifier training

- AdaBoost + SVM
- 10 fold cross validation

Results & Evaluation

Results in term of

- True Positive (TP) rate
  - Correctly classified anti-adblocking scripts
- False Positive (FP) rate
  - Incorrectly classified anti-adblocking scripts

Feature Set	Classifier	Number of Features	TP rate (%)	FP rate (%)
all	AdaBoost + SVM	10K	99.6	3.9
literal	AdaBoost + SVM	10K	99.6	3.9
keyword	AdaBoost + SVM	1K	99.7	3.2

Test in the wild on Alexa top 100K websites

- 2,701 detected anti-adblockers
- TP rate of 92.5%

Complement manual analysis

- Periodic crawl to expedite manual process
- Substantial reduction of manual effort

Key Takeaways

Comprehensive measurement study of anti-adblocking filter lists
- Retrospective analysis on Alexa top 5K websites from 2011 to 2016
- Effectiveness and evolution

Lightweight machine learning approach
- Static analysis to detect anti-adblocking scripts
- Complement filter lists rules creation

The Wayback Machine enables retrospective analysis
- Can be used to study similar filter lists
- Malware, Tracking, Censorship

Questions?

Umar Iqbal

 www.umariqbal.com

 @umaarr6

 The University of Iowa

References

A. Lerner, A. K. Simpson, T. Kohno, and F. Roesner. Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016. In USENIX Security Symposium, 2016.

C. Curtsinger, B. Livshits, B. Zorn, and C. Seifert. ZOZZLE: Fast and Precise In-Browser JavaScript Malware Detection. In USENIX Security Symposium, 2011.

M. Ikram, H. J. Asghar, M. A. Kaafar, A. Mahanti, and B. Krishnamurthy. Towards Seamless Tracking-FreeWeb:Improved Detection of Trackers via One-class Learning . In Privacy Enhancing Technologies Symposium (PETS), 2017.