Web Security Research at Indiana University

Dr. XiaoFeng Wang
Associate Professor
School of Informatics and Computing
Indiana University

Our Adventures on the Web

- Privacy: get your health records, salary, investment secret from Web apps' side channel
  - Microsoft Buddies: Shuo Chen and Rui Wang
- Flaws: shop for free, log into your web account through 3rd-party Web APIs
  - Microsoft Buddies: Shuo Chen and Rui Wang
- Misdeeds: how to advertise infections and click frauds through the New York Times
  - Microsoft Buddies: Yinglian Xie and Fang Yu

Knowing Your Enemy: Understanding and Detecting Malicious Web Advertising

Zhou Li, Kehuan Zhang, Yinglian Xie,
Fang Yu and XiaoFeng Wang

The New York Times

Tuesday, November 29, 2011 Last Update: 1:49 AM ET

Switch to Global Edition ›

JOBS
REAL ESTATE
AUTOS
ALL CLASSIFIEDS

WORLD
U.S.
POLITICS
NEW YORK
BUSINESS
DEALBOOK
TECHNOLOGY
SPORTS
SCIENCE
HEALTH
OPINION
ARTS
Books
Movies
Music
Television
Theater
STYLE
Dining & Wine

Businesses Scramble as Credit Tightens Across Europe

By ERIC DASH and NELSON D. SCHWARTZ

As European banks pull back on lending, companies around the globe are finding it harder to borrow, edging the world economy toward another slump.

- Obama Meets Leaders of the European Union
- Dire Warnings Are Building on European Debt Crisis

CAMPAIGN 2012

Barney Frank, a Top Liberal, Won't Seek Re-election

By ABBY GOODNOUGH

Mr. Frank, who first

Monica Lopossay for The New York Times

Black Workers Struggle in Public Sector

By TIMOTHY WILLIAMS

Since the recession's declared end, middle-class blacks like Pamela Sparks, a postal worker in Baltimore, above, have seen a traditionally prosperous job source run dry.

DEALBOOK

Money Found in Britain May Belong to MF Global

OPINION »

EDITORIAL

What Happened on the Border?

A transparent investigation of the NATO strikes, with Pakistan's participation, is essential.

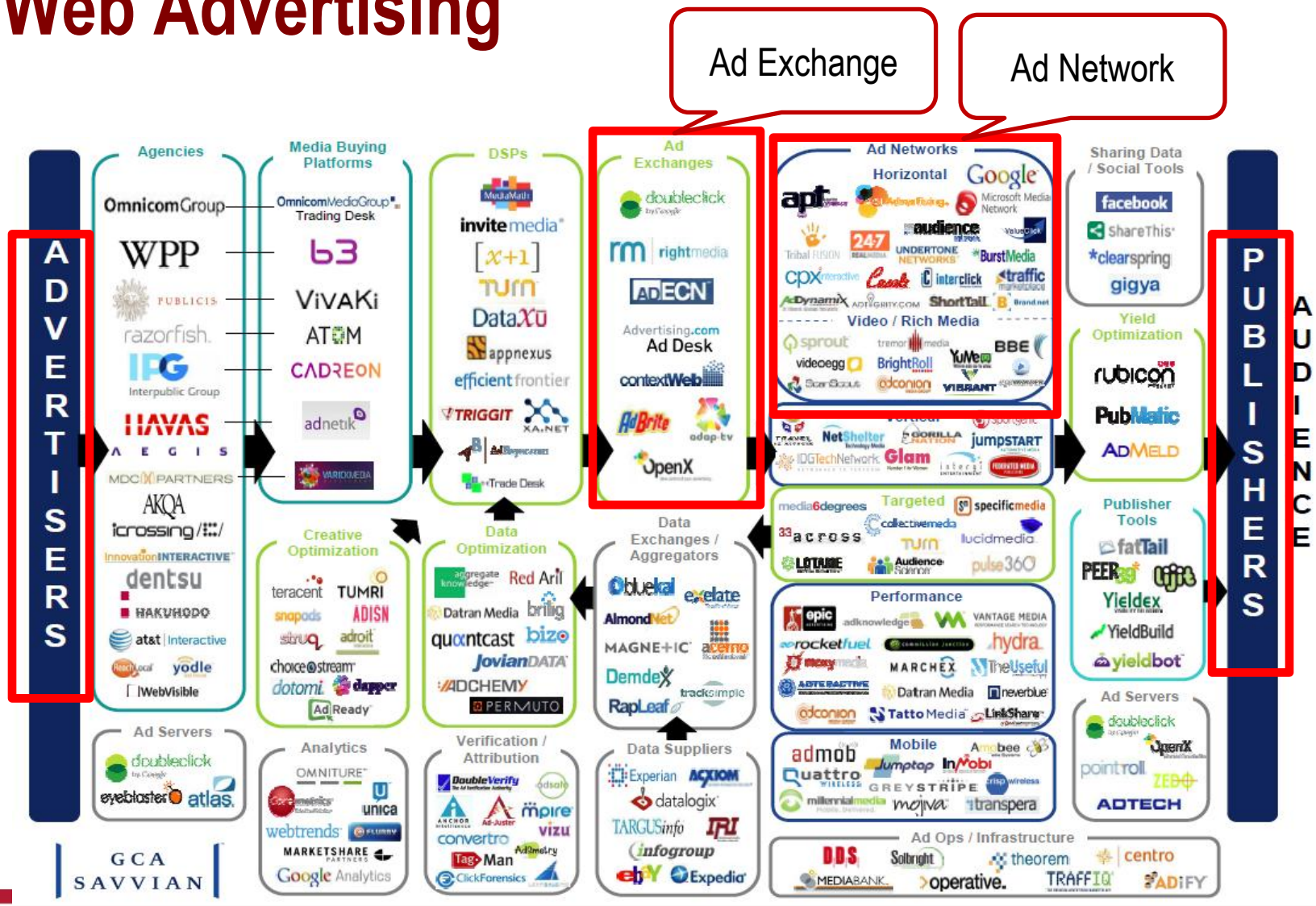
- Brooks: The Life Reports II
- Bruni: Silvio's Postscript
- Nocera: Germany Cuts Off Its Nose
- Cohen: Doctrine of Silence
- Fish: Dogs and Cars
- Op-Ed: Optimism in Egypt
- Room for Debate: Legal Rights for the Obese?

MARKETS » At 2:04 AM ET

JAPAN		CHINA
Nikkei	HangSeng	Shanghai
8,477.82	18,302.43	2,411.01
+190.33	+264.62	+27.98
+2.30%	+1.47%	+1.17%

Data delayed at least 15 minutes

Web Advertising



[http://www.adexchanger.com/pdf/Display-Advertising-Technology-Landscape-2010-05-03.pdf]

Credit Tightens Across Europe

By ERIC DASH and NELSON D. SCHWARTZ

As European banks pull back on lending, companies around the globe are finding it harder to borrow, edging the world economy toward another slump.

- Obama Meets Leaders of the European Union
- Dire Warnings Are Building on European Debt Crisis

CAMPAIGN **2012**

Barney Frank, a Top Liberal, Won't Seek Re-election

By ABBY GOODNOUGH

Mr. Frank, who first

Monica Lopossay for The New York Times

Black Workers Struggle in Public Sector

By TIMOTHY WILLIAMS

Since the recession's declared end, middle-class blacks like Pamela Sparks, a postal worker in Baltimore, above, have seen a traditionally prosperous job source run dry.

DEALBOOK

Money Found in Britain May Belong to MF Global

What Happened on the Border?

A transparent investigation of the NATO strikes, with Pakistan's participation, is essential.

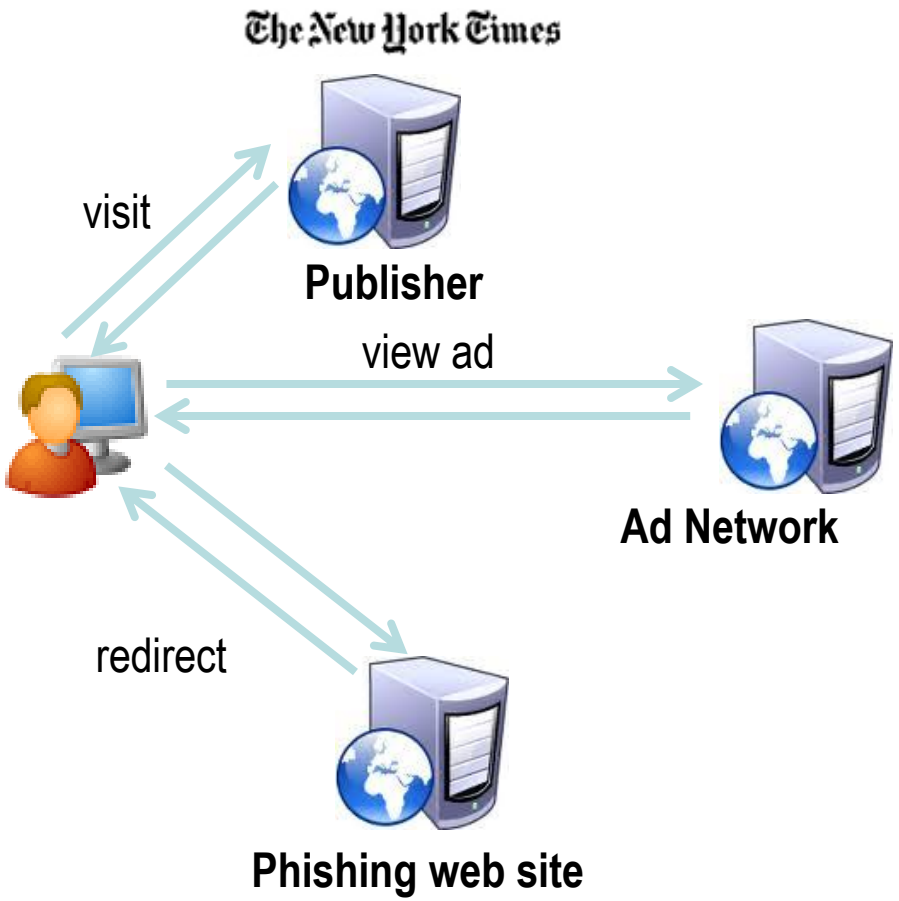
- Brooks: The Life Reports II
- Bruni: Silvio's Postscript
- Nocera: Germany Cuts Off Its Nose
- Cohen: Doctrine of Silence
- Fish: Dogs and Cars
- Op-Ed: Optimism in Egypt
- Room for Debate: Legal Rights for the Obese?

MARKETS » At 2:04 AM ET

JAPAN		CHINA
Nikkei	HangSeng	Shanghai
8,477.82	18,302.43	2,411.01
+190.33	+264.62	+27.98
+2.30%	+1.47%	+1.17%

Data delayed at least 15 minutes

Malvertising via Nytimes



Defense: What has been done

- Ad behavior restriction [AdSafe, Finifter'10, Louw'10]
  - Limited applicability to different forms of attacks: drive-by-download, phishing, click-fraud
- URL features and domain reputations [Zhang'11, John'11]
  - URLs can be easily modified by attackers
  - Domains can be hijacked
- Code analysis [Cova'10]
  - Obfuscate code to evade detection
  - Leverage ad syndication to bypass the code checking

What hasn't been done

- Understanding:
  - How serious is the problem?
  - What does malcontent delivery path look like?
  - Roles of ad nodes? Topologies?
- Detection:
  - Complement existing techniques with infrastructure information?

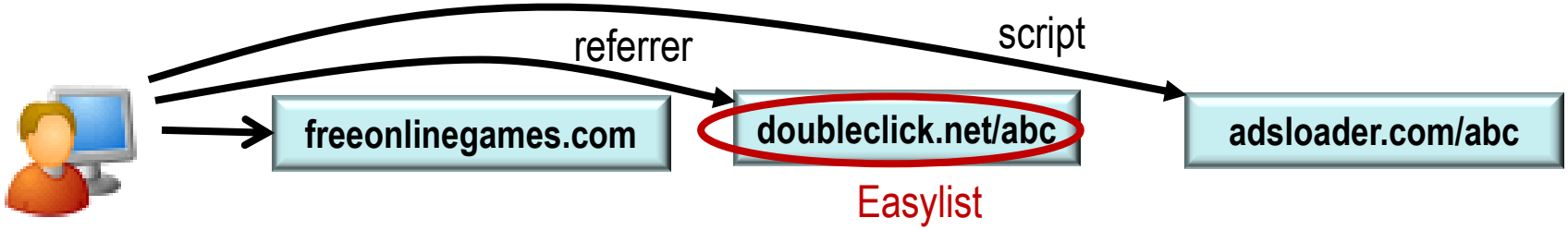
The First Step We Made

- Measurement of Malvertising
  - Malvertising: drive-by download, Scam, Click fraud
  - Scale, node/path features of malcontent delivery infrastructures
- Detection of Malvertising
  - Path segment based detection
  - It works: caught 15 times more cases than popular blacklist/malware scanners (Safe Browsing, Forefront)

Get down to the specifics: Data Collection

- From June 21st to September 30th
- 12 Virtual Machines with instrumented browser
- Alexa top 90,000 web sites visited regularly
- Extract ad redirection paths

Node: freeonlinegames.com doubleclick.net/abc, advertising.com/abc



Path: freeonlinegames.com -> doubleclick.net/abc -> adsloader.com/abc

Case: freeonlinegames.com -> doubleclick.net -> adsloader.com

Some Statistics

- 24 million ad paths
- 22 million nodes, >90% ad nodes

- Scanned with Forefront and Google Safebrowsing
- 543 malicious nodes, 263 domains
- 938 malicious cases
- 286 infected publishers (Ranked from 314 to 89184)
- Long-lived campaign (2 months) , short-lived domain (3 days)

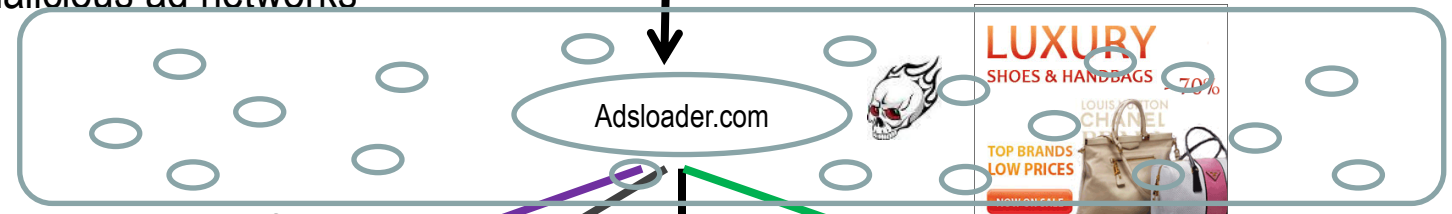
Example: a Fake AV Campaign

65 infected publishers
(highest ranked 400)

24 malicious ad networks

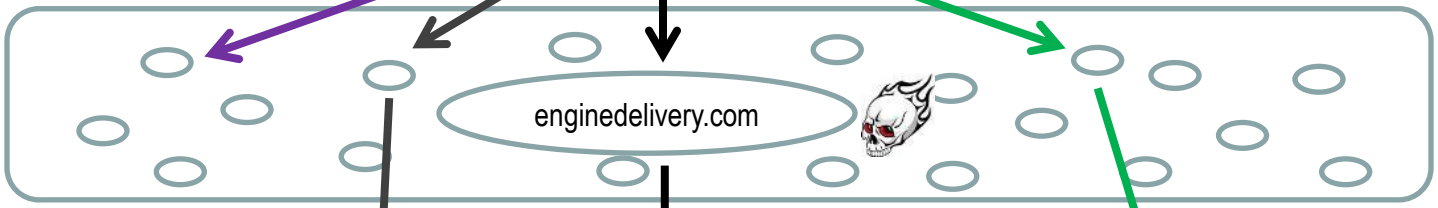
Attack Strategy:

- Set up malicious ad network
- Penetrate big ad network
- Multi-layers
- Rotation
- Cloaking

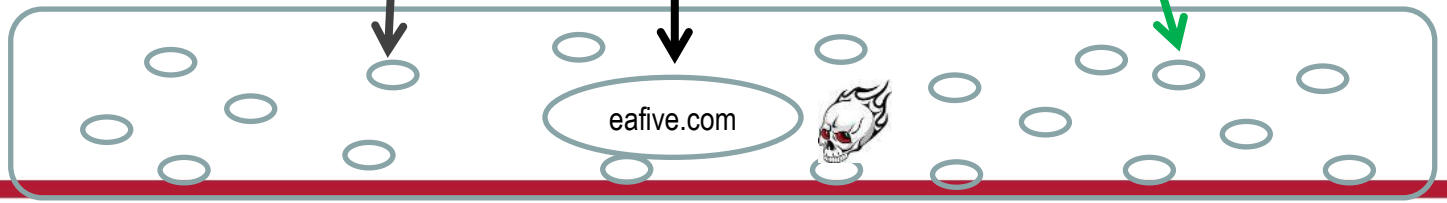


16 Redirectors

Cloaking



84 Scam sites

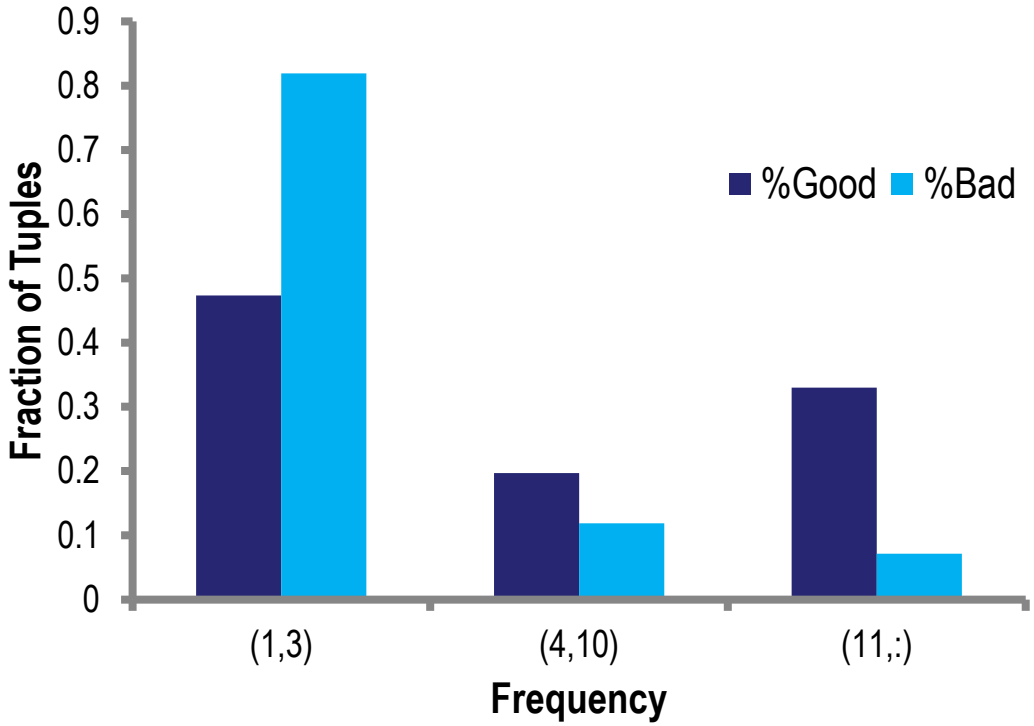


Node Features

- Most of malicious nodes have unknown roles
  - >90% malicious nodes, <8% legitimate nodes
- Registered within a year, expire in a year
  - >70% malicious domains, <20% legitimate domains
- Free domain providers like .co.cc used widely
- Follow URL patterns
  - /showthread.php\.php\?t=\d{8} matches 34 domains

Properties of Malicious Pairs

- Malicious node pairs appear less frequently

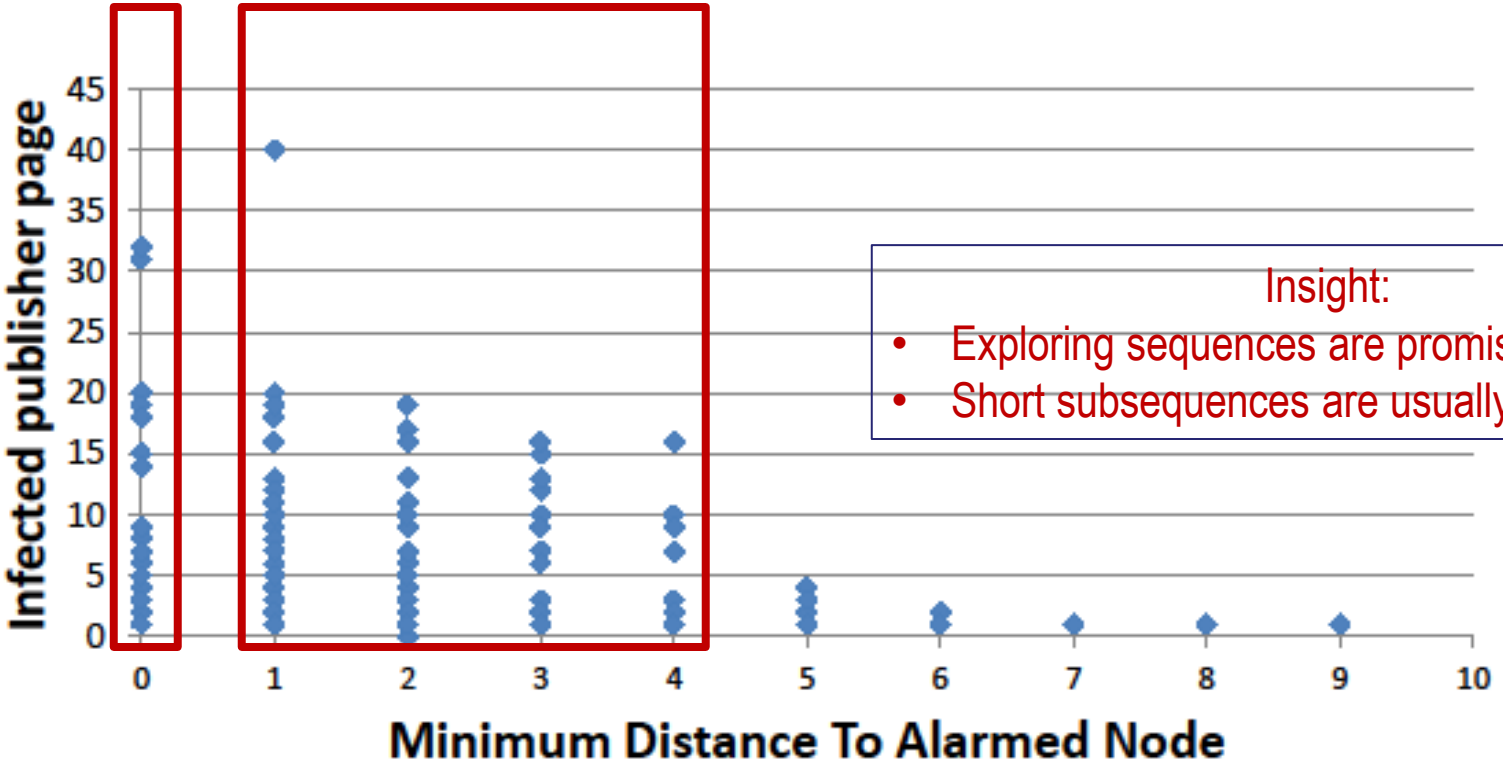


Insight:

- The relationships with other entities are not stable

Properties of Malicious Paths

- Longer path length (8.11 > 3.59 of legitimate paths)
- Ad syndication is the major problem (>60% cases)
- The closer to bad nodes, the more suspicious

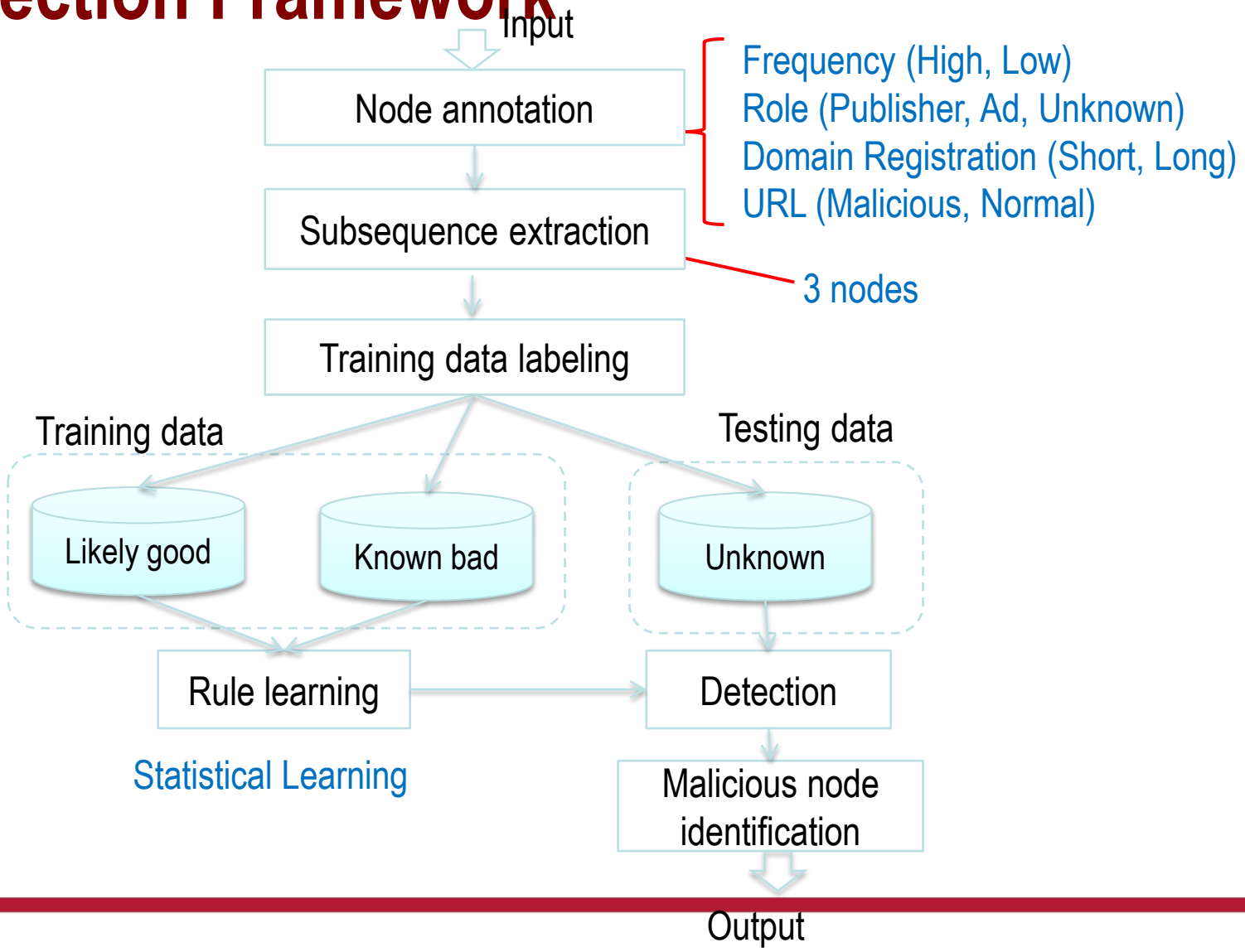


Insight:
- Exploring sequences are promising
- Short subsequences are usually good enough

Our Ideas for Detection

- Analyze ad-delivery sequences
  - Focus on short subsequences
  - Annotate nodes with rich attributes
- Use statistical learning to generate detection rules
  - Adapt to new, ever changing attacker strategies

Detection Framework



Results

Testing
June - September

phishing pages	56	0.00%
drive-by-download pages	172	9.88%
click-fraud pages	155	10.97%
all pages	326	8.90%
phishing cases	104	0.00%
drive-by-download cases	1171	6.23%
click-fraud cases	4221	4.10%
all cases	5496	4.48%

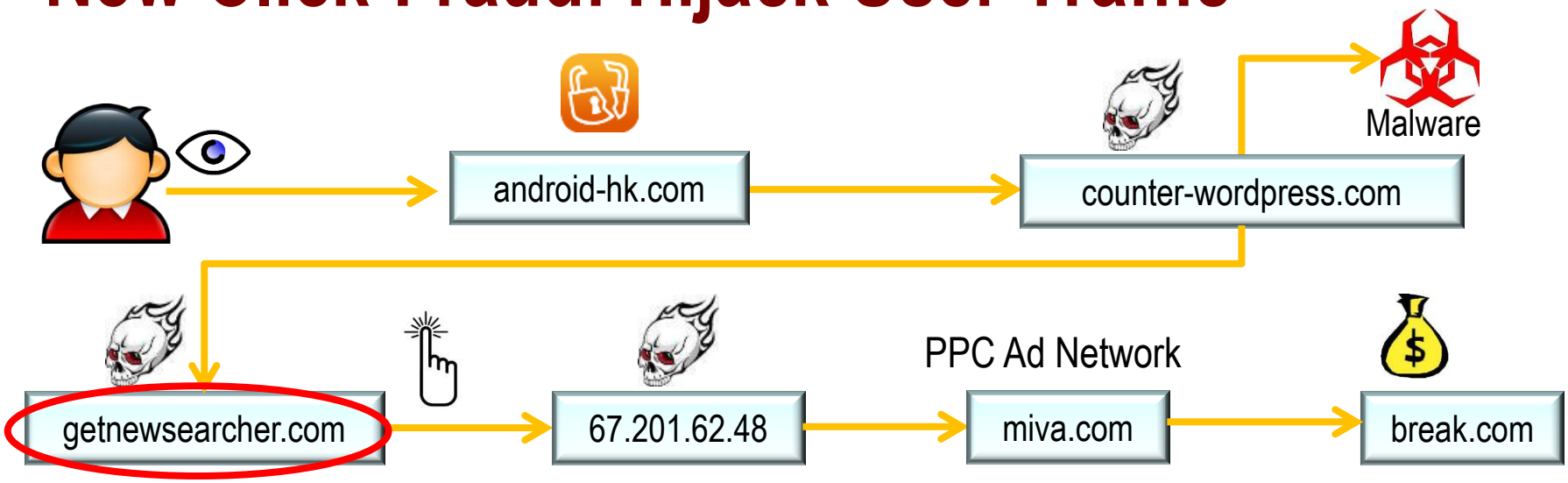
- 5% FDR
- 15x new findings
- 10.5 days early detection than safebrowsing

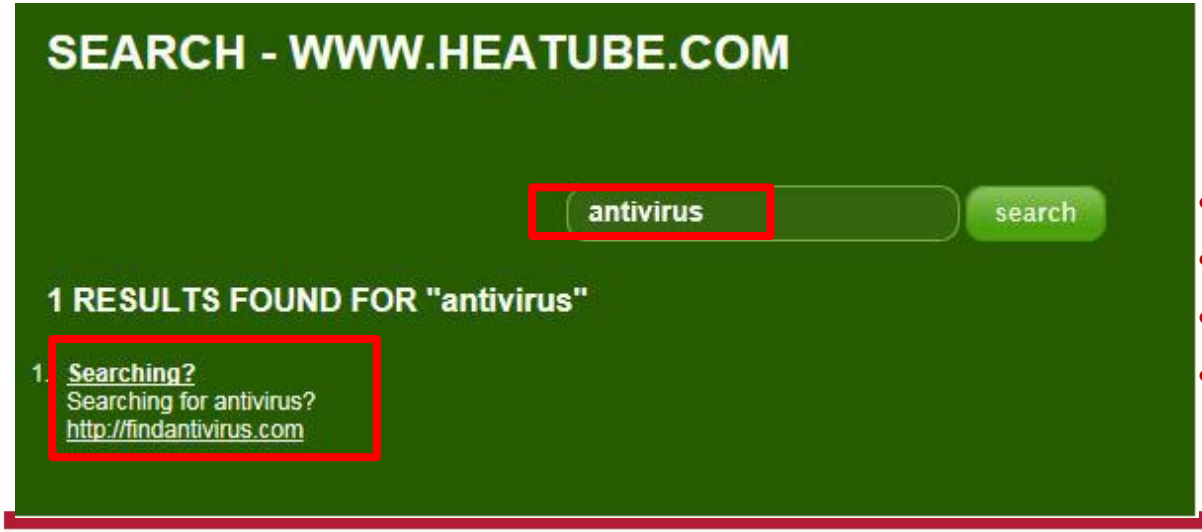
Testing
October

Safebrowsing & Forefront

	#MadTracer	#S&F	%FP	%New Findings
phishing pages	12	0	0.00%	100.00%
drive-by-download pages	216	104	9.26%	51.85%
click-fraud pages	89	7	14.61%	92.13%
all pages	291	111	11.00%	61.86%
phishing cases	23	0	0.00%	100.00%
drive-by-download cases	627	216	13.88%	65.55%
click-fraud cases	3422	42	3.65%	98.77%
all cases	4072	258	5.21%	93.66%

New Click-Fraud: Hijack User Traffic





Findings:

- Do not require botnets
- Use of doggy search engine
- Target 2nd-tier PPC ad networks
- High successful rate (72.5%)

Conclusion

- Malvertising is a big issue
  - 1% top publishers are infected
- Study on infrastructures can lead to a promising new direction
  - 15x more coverage, 5% false positive
  - Discover new attacks
- Usage and deployment
  - Ad exchange service (e.g., Ad-center): capture malicious and fraudulent ad entities
  - Anti-virus (e.g., ForeFront): provide new malware signatures
  - End users (you and me): detect and stop ongoing exploits


The End