No Please, After You: Detecting Fraud in Affiliate Marketing Networks

Peter Snyder <psyde2@uic.edu> and Chris Kanich <ckanich@uic.edu> University of Illinois at Chicago

Overview

- 1. Problem Area: affiliate marketing
- 2. Data Set: HTTP request records
- 3. **Methodology**: classification algorithm
- 4. Findings: numbers and stakeholder analysis
- 5. Big Picture: harm measurement and reduction

1. Problem Area

Affiliate Marketing

Online Retailers







Publishers









Web Users

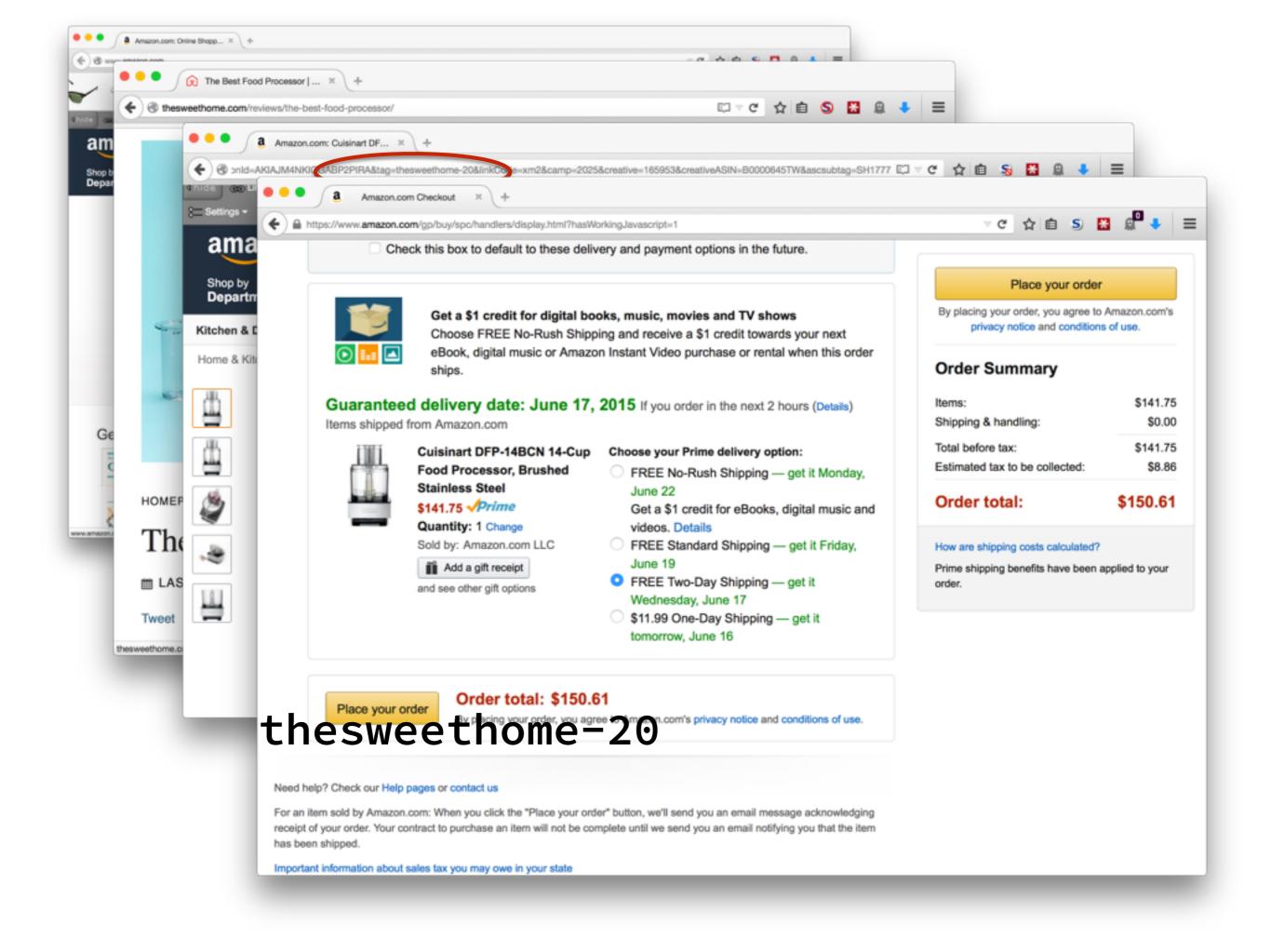






Affiliate Marketing

- Common method for funding "free" content
- Largest programs include Amazon, GoDaddy, eBay and WalMart
- Both direct programs and networks / middle-parties



Affiliate Marketing Terms

Affiliate Marketing ID:

Unique identifier that <u>Online Retailers</u> use to tie <u>Web Users</u> to <u>Publishers</u>

· Affiliate Marketing Cookie:

Cookie set by Online Retailer, tying Web User to the "delivering" Publisher

Cookie Setting URL:

End points, controlled by Online Retailers, that set an affiliate marketing cookie on a Web User

Affiliate Marketing Fraud

Assumption

Having an affiliate marketing cookie →
User intended to visit the online retailer →
Retailer helped sale

Exploit

Get your affiliate marketing cookie on as many browsers as possible

Methods

hidden iframes, plugins, malware, automatic redirects, etc.

2. Data Set

Affiliate Marketing Programs

- 164 affiliate marketing programs
 - Popular: Amazon, GoDaddy
 - Networks: ClickCash, MoreNiche
- Selection methods
 - Predictable URLs
 - HTTP / no encryption

Affiliate Marketing Programs

Data	Amazon	GoDaddy
Domains	(www\.)amazon\.com	^godaddy\.*
Cookie Setting URLs	^/(?:.*(dp gp)/.*)?	[&?]tag=(?:& \? ^ ;)isc=
Conversion URLs	*handle-buy-box*	*domains/domain-configuration\.aspx*
Affiliate ID Values	tag=(.*?)(?:& \$)	cvosrc=(.*?)(?:& \$)

HTTP Request Logs

- 660G of HTTP Requests (bro-log format)
- 2.3 billion records
- January and February 2014

Request Information	Response Information
Sender and destination IP	Mime type
Domain and path	HTTP response code
Referrer	
Timestamp	
Cookies	
User agent	

3. Methodology

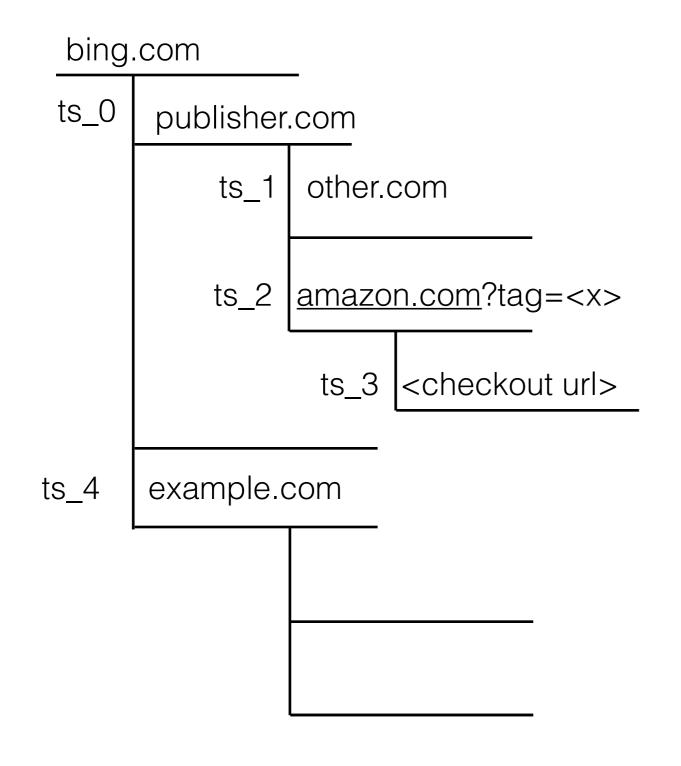
Data and Preprocessing

```
129.174.37.173 r14---sn-p5qlsnez.googlevideo.com
2014-01-01T18:00:00-0500
                                                                                                                                         /videoplayback?algorithm
                                                                                        http://www.youtube.com/watch?v=TabUnV7xoHM
2014-01-01T17:59:59-0500
                                192.5.215.222
                                                www.mmadsgadget.com
                                                                                /t?id=98394efb-7f7c-4ce4-0152-dee1df6de648&size=728x90
2014-01-01T17:59:59-0500
                                192.5.215.222
                                                www.mmadsgadget.com
                                                                        http://www.twitbridge.com/ads/300x250.cshtml?d=http://www.mangahere.com/manga/akuma_na_e
                                                                        /apple-touch-icon-72x72-precomposed.png
2014-01-01T18:00:00-0500
                               124.149.235.78
                                               wxmaps.org
2014-01-01T17:59:55-0500
                                                                                /announce.php?passkey=7776250d207e3bfce323b2107bbacf83&info_hash=b\xa2'M\x06;x\x
                               129.174.67.117 tracker.bitsoup.me
                                                                                /scrape.php?passkey=7776250d207e3bfce323b2107bbacf83&info_hash=b\xa2'M\x06;x\xae
2014-01-01T18:00:00-0500
                               129.174.67.117 tracker.bitsoup.me
                                                www.twitbridge.com
2014-01-01T18:00:00-0500
                                                                        http://www.mangahere.com/manga/akuma_na_eros/c003.2/4.html
                                                                                                                                         /ads/728x90.cshtml?d=htt
                               192.5.215.222
2014-01-01T18:00:00-0500
                               38.119.104.180
                                                edge.sharethis.com
                                                                        http://www.era.com/property/73716557-101024643/17600-e-fieldbrook-cir-e-boca-raton-fl-33
2014-01-01T18:00:00-0500
                               38.119.104.180
                                                b.scorecardresearch.com http://www.era.com/property/73716557-101024643/17600-e-fieldbrook-cir-e-boca-raton-fl-33
2014-01-01T17:59:59-0500
                               199.26.254.212
                                               scores.espn.go.com
                                                                        http://assets.espn.go.com/swf/caster/snapshotCasterClientv2.swf /ncf/caster/snapshot?ses
2014-01-01T18:00:00-0500
                                                                        http://www.mangahere.com/manga/akuma_na_eros/c003.2/4.html
                                                                                                                                         /ads/300x250.cshtml?d=ht
                                192.5.215.222
                                                www.twitbridge.com
2014-01-01T18:00:00-0500
                               199.26.254.69
                                                whois.freeshell.ustc.edu.cn
                                                                                         /whois/data
                                                                        http://www.mangahere.com/manga/akuma_na_eros/c003.2/4.html
                                                                                                                                         /ads/160x600.cshtml?d=ht
2014-01-01T18:00:00-0500
                                192.5.215.222
                                                www.twitbridge.com
                                                                                /case/1.0/pkg_version/default/1388616888070135
2014-01-01T18:00:00-0500
                               129.174.0.60
                                                updates.ironport.com
2014-01-01T18:00:00-0500
                               129.174.0.60
                                                updates.ironport.com
                                                                                /case/1.0/dfa/default/1388616888304604
2014-01-01T18:00:00-0500
                               5.10.83.24
                                                911digitalarchive.org
                                                                                /repository_object.php?object_id=111820
2014-01-01T18:00:00-0500
                                                                        http://russianhistoryblog.org/2011/02/russians-in-east-prussia-1914-pt-2/
                               23.238.207.60
                                                russianhistoryblog.org
2014-01-01T18:00:00-0500
                                                www.twitbridge.com
                                                                        http://www.mangahere.com/manga/akuma_na_eros/c003.2/4.html
                                                                                                                                         /ads/160x600.cshtml?d=ht
                                192.5.215.222
                                                                                         /announce?info_hash=\xc2\xf1\x81\x9bb$\xa7Cv\x12Q\x9f*~\x02Gh\x15s\xe6&p
2014-01-01T18:00:00-0500
                               111.85.195.23
                                                ipv4.tracker.harry.lu.nyud.net -
2014-01-01T18:00:00-0500
                                                updates.ironport.com
                                                                                /case/1.0/dfa_updates/default/1388616853644950
                               129.174.0.60
                               129.174.143.252 108.175.35.156
                                                                        /1654089019.ismv?c=us&n=11279&v=3&e=1388644580&t=-1aKqXqPhD01P0JZRz9VmZuRomQ&d=ps3&p=5.8
2014-01-01T17:59:59-0500
2014-01-01T18:00:00-0500
                                192.5.215.213
                                                108.175.34.103 http://movies.netflix.com/WiPlayer?movieid=70122768&trkid=7882979
                                                                                                                                         /1550506182.ismv/range/5
2014-01-01T18:00:00-0500
                               199.26.254.69
                                                whois.freeshell.ustc.edu.cn
                                                                                         /whois/registrant
2014-01-01T17:59:56-0500
                                                ns-streamingaudio.itunes.apple.com
                                                                                                 /us/r1000/017/Music4/v4/45/8c/60/458c60b4-da3b-2b07-8975-8824e2b
                                192.5.215.215
                                                                                         /announce?info_hash=nJ\xf4\xd1\x14\xd0\xbc|L\xd59\xb5\x93\x1f\x1eM\xabT\
2014-01-01T18:00:00-0500
                               124.120.221.183 ipv4.tracker.harry.lu.nyud.net -
2014-01-01T18:00:00-0500
                               124.120.221.183 ipv4.tracker.harry.lu.nyud.net -
                                                                                         /announce?info_hash=*o\xc7\xed\x1c\xd7\x192\x85qW\xca\x99^"\xcd0\x18m\x1
2014-01-01T18:00:01-0500
                               38.119.104.180
                                               media.newjobs.com
                                                                        http://jobsearch.monster.com/search/Systems-Engineer_5?where=Massachusetts
2014-01-01T18:00:01-0500
                               129.174.67.117 tracker.bitsoup.me
                                                                                /announce.php?passkey=7776250d207e3bfce323b2107bbacf83&info_hash=dl\xa2\xd2\xcaB
                                                                                /t?id=98394efb-7f7c-4ce4-0152-dee1df6de648&size=728x90
2014-01-01T18:00:01-0500
                               192.5.215.222
                                                www.mmadsgadget.com
                                129.174.143.252 108.175.35.156 -
                                                                        /1654089019.ismv?c=us&n=11279&v=3&e=1388644580&t=-1aKqXqPhD01POJZRz9VmZuRomQ&d=ps3&p=5.B
2014-01-01T17:59:58-0500
                                                                                                /nx/ac?_callback=window.__jindo_callback._$3361_0&q=are counterp
2014-01-01T18:00:01-0500
                               129.174.138.174 ac.search.naver.com
                                                                        http://www.naver.com/
2014-01-01T18:00:00-0500
                                                                                                 /announce?info_hash=\x057\x14,-\x7f\xbbp\xab\x12/\xc8\x0dC\xb5k]
                               173.64.103.239 calu-atrack.appspot.com.nyud.net
                                                                                         /depot/206421/chunk/c274f93a6b4428922bba93b3d727f8d2def1983b
2014-01-01T17:59:58-0500
                               192.5.215.217
                                                valve205.cs.steampowered.com
2014-01-01T17:59:01-0500
                                                whois.freeshell.ustc.edu.cn
                               199.26.254.69
                                                                                         /ip/getnext
2014-01-01T18:00:01-0500
                                                                        /url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0CDsQFjAB&url=http://engineering.stanford.
                                192.5.215.211
                                                www.google.com -
2014-01-01T18:00:01-0500
                               131.253.24.130
                                                publichistorycommons.org
                                                                                         /thoughts-on-completing-my-first-major-public-history-project/
2014-01-01T18:00:01-0500
                               129.174.70.208
                                               crl.microsoft.com
                                                                                /pki/crl/products/CodeSignPCA2.crl
2014-01-01T18:00:01-0500
                                                jdn.monster.com http://oas.monster.com/RealMedia/ads/adstream_sx.ads/us.monster.en/js20/powersearch.aspx/1860996
                               38.119.104.180
2014-01-01T18:00:01-0500
                               157.55.34.29
                                                teachinghistory.org
                                                                                /history-content/historical-site/10448
2014-01-01T18:00:01-0500
                                                googleads.g.doubleclick.net
```

/mads/static/sdk/native/sdk-core-v40.js

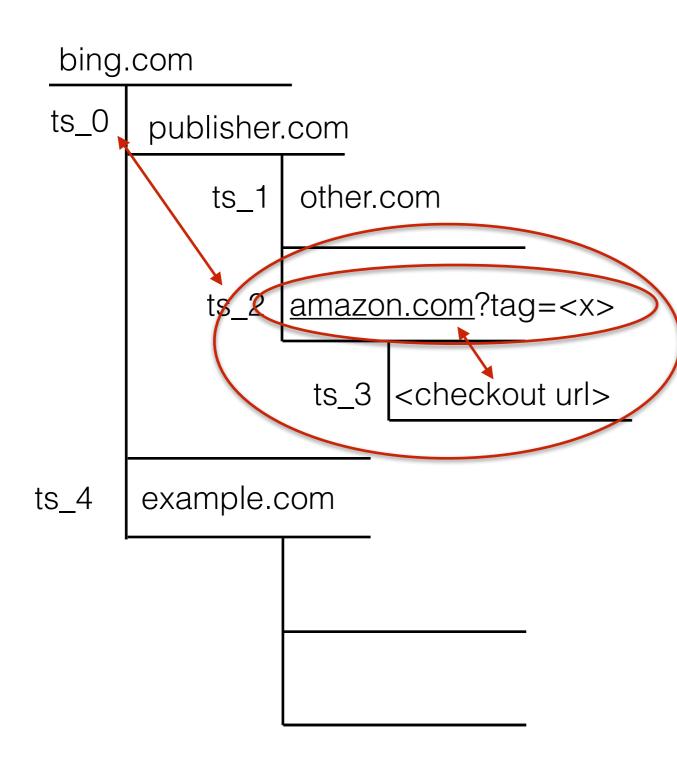
192.5.215.219

Browsing Session Trees



Xie, Guowu, et al. "Resurf: Reconstructing web-surfing activity from network traffic." *IFIP Networking Conference*, 2013. IEEE, 2013.

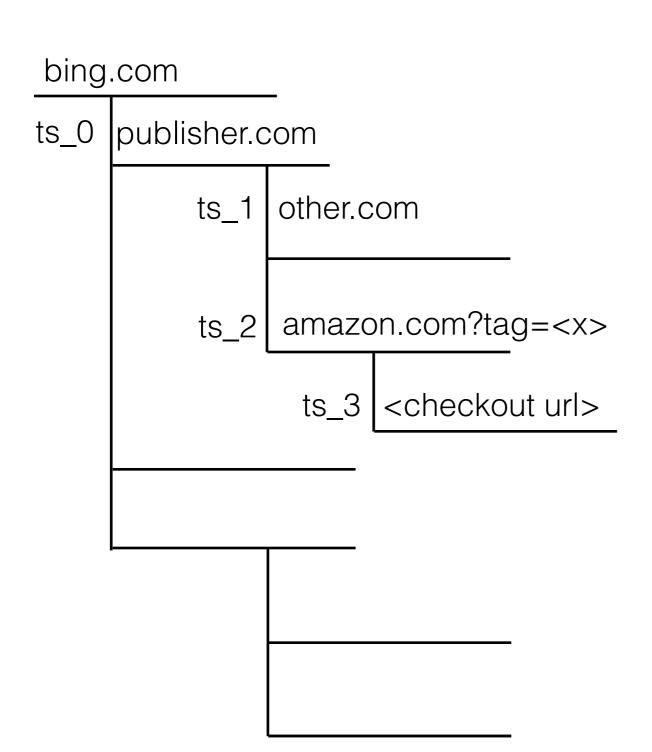
Browsing Session Trees



Simple Measurements

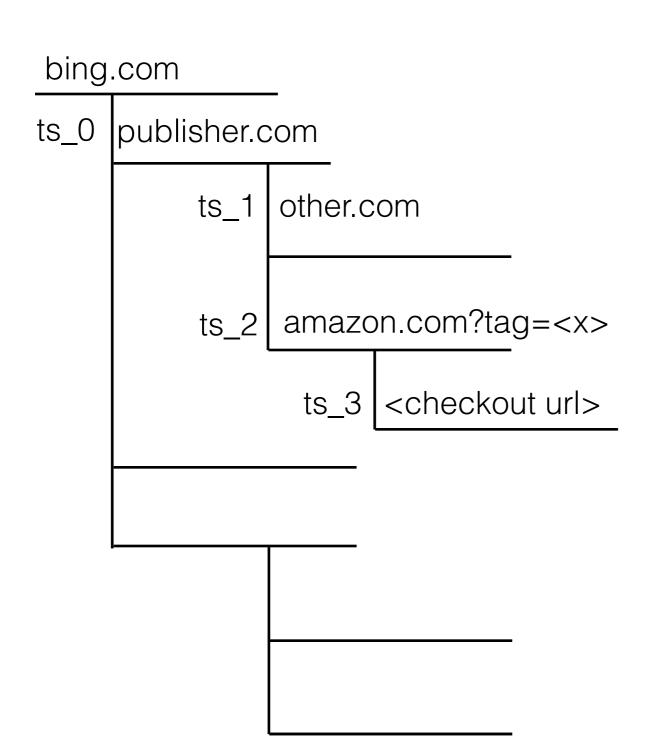
- Number of referrals in each program
- Number of publishers in each program
- Number of conversions / purchases in each program
- How long a user takes to be referred
- How long a user spent on site after being referred

Classifier: Training Set



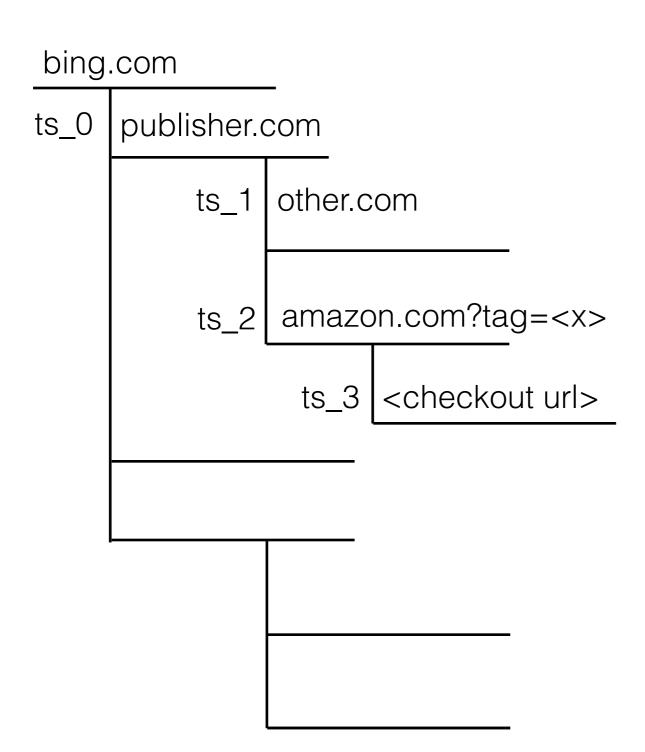
- Did the user intend to travel from some-referrer.com to amazon.com?
- Built training set of 1141 relevant trees (subset of January data)
- If referrer was still available, direct test
- If referrer was not available, infer from graph (log data)

Classifier: Features



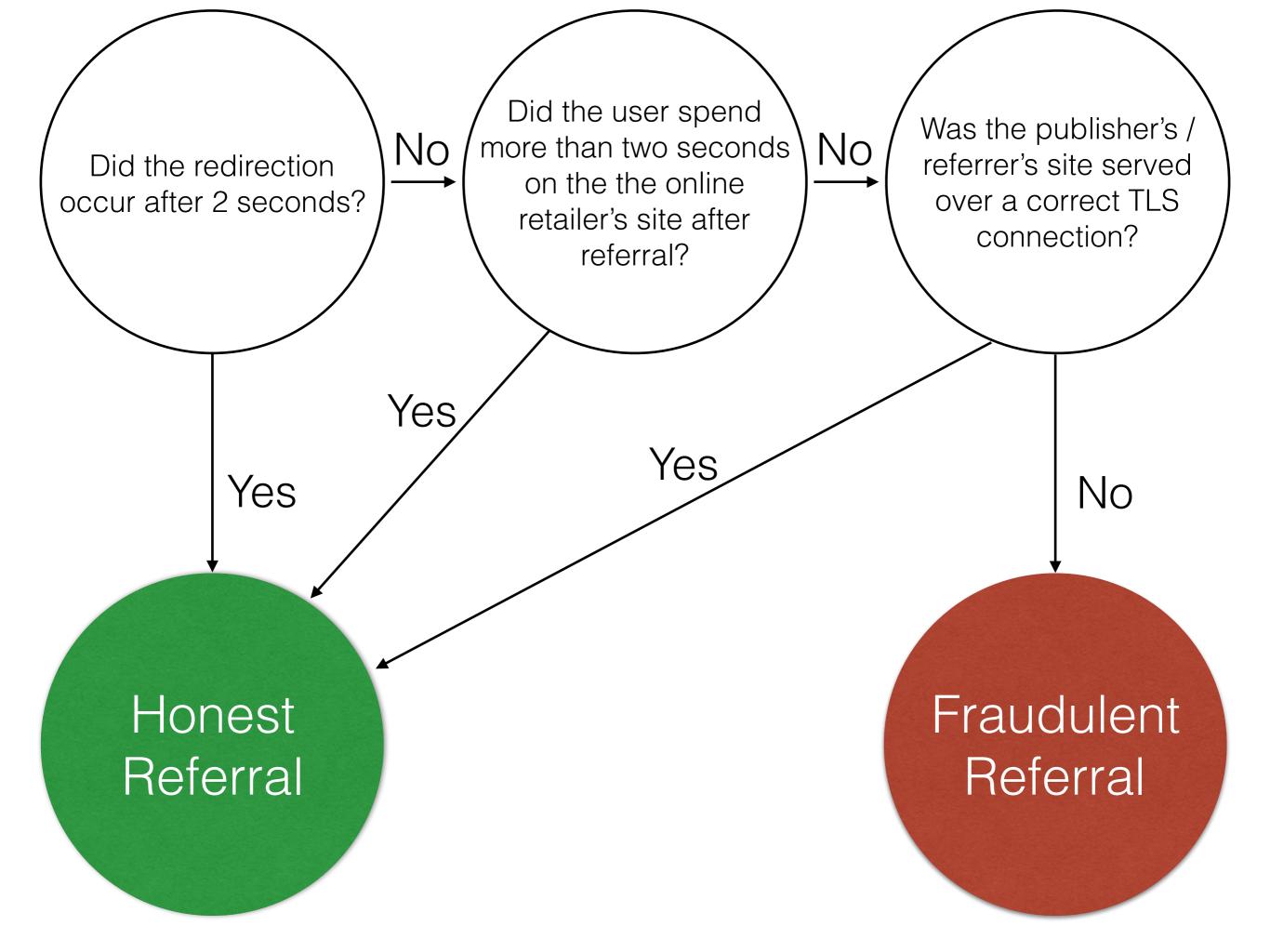
- Time before referral
- 2. Time after referral
- 3. Is referrer SSL?
- 4. Graph size
- 5. Is referrer reachable?
- 6. Google page rank of referrer
- 7. Alexia traffic rank
- 8. Is referrer domain registered?
- 9. # years domain is registered
- 10. Tag count

Classifier: Features



- 1. Time before referral
- 2. Time after referral
- 3. Is referrer SSL?
- 4. Graph size
- 5. Is referrer reachable?
- 6. Google page rank of referrer
- 7. Alexia traffic rank
- 8. Is referrer domain registered?
- 9. # years domain is registered
- 10. Tag count

93.3% accuracy



4. Findings

Online Retailer Popularity

Retailer	Requests	Unique Sessions
Amazon.com	2,663,574	87,654
GoDaddy	7,320	364
<u>ImLive.com</u>	731	194
wildmatch.com	3	1
Total (166 programs)	2,671,808	88,257

Publishers

Retailer	Honest	Fraudulent	Total
Amazon.com	2,268	1,396	3,664
GoDaddy	5	19	24
ImLive.com	4	7	11
wildmatch.com	0	1	1
Total (166 programs)	2,281	1,426	3,707

Affiliate Marketer Referrals

Retailer	Honest	Fraudulent	Total
Amazon.com	12,870	2,782	15,652
GoDaddy	399	98	497
ImLive.com	9	13	22
wildmatch.com	0	1	1
Total (166 programs)	13,283	2,897	16,180

Conversion Events

Retailer	Amazon.com	GoDaddy	Total (166 programs)
Conversion Events	15,624	26	15,650
Affiliate Conversions	955	8	693
Honest	781	8	789
Fradulent	174	0	174
"Stolen"	0	0	0

In The Paper...

- Session tree building algorithm
- Details of how we generated the classifier
- Stakeholder analysis of affiliate marketing fraud
- More numbers...

5. Big Picture

Harm Measurement and Reduction

- Not all harms hurt equally
- Technical inference + behavioral insights to find maximally beneficial strategies
- Focus treatment where users are affected most

Harm Measurement and Reduction

- 1. Affiliate marketing fraud
- 2. Typo squatting
- 3. Malware infections

Affiliate Marketing Fraud

Data Set:

HTTP request traces & affiliate marketing program details

Technical Inference:

Modeling intent from http logs & user web patterns

Behavioral Insights:

Detecting when cookie sets are because of user intent or cookie stuffing

· Conclusions:

Users are relatively unharmed by affiliate marketing fraud

Typosquatting

Data Set:

HTTP request traces & passive DNS logs

Technical Inference:

Time between requests to "incorrect" and desired domains

Behavioral Insights:

Time-and-opportunity costs to users, measures of harm to "correct domain holders"

Conclusions:

Typosquatting is not a substantial costs to most users, and sometimes is a convenience

Infection Analysis

Data Set:

Online computer support communities & malware research

Technical Inference:

Extrapolated costs of replaced equipment, lost information

Behavioral Insights:

Unseen harms, such as information loss or identity theft, opportunity costs

Conclusions:

TBD

Thanks!

Peter Snyder – <u>psnyde2@uic.edu</u> Chris Kanich – <u>ckanich@uic.edu</u> University of Illinois at Chicago

