

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

Peter Snyder <psnyde2@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

Amazon.com: Online Shopp... x

The Best Food Processor | ... x


thesweethome.com/reviews/the-best-food-processor/

Amazon.com: Cuisinart DF... x


Amazon.com Checkout x

https://www.amazon.com/gp/buy/spc/handlers/display.html?hasWorkingJavascript=1

Check this box to default to these delivery and payment options in the future.

 **Get a \$1 credit for digital books, music, movies and TV shows**
Choose FREE No-Rush Shipping and receive a \$1 credit towards your next eBook, digital music or Amazon Instant Video purchase or rental when this order ships.

Guaranteed delivery date: June 17, 2015 If you order in the next 2 hours ([Details](#))
Items shipped from Amazon.com

 **Cuisinart DFP-14BCN 14-Cup Food Processor, Brushed Stainless Steel**
\$141.75 ✓ Prime
Quantity: 1 [Change](#)
Sold by: Amazon.com LLC
[Add a gift receipt](#) and see other gift options

Choose your Prime delivery option:

- FREE No-Rush Shipping — get it **Monday, June 22**
Get a \$1 credit for eBooks, digital music and videos. [Details](#)
- FREE Standard Shipping — get it **Friday, June 19**
- FREE Two-Day Shipping — get it **Wednesday, June 17**
- \$11.99 One-Day Shipping — get it **tomorrow, June 16**

Place your order

By placing your order, you agree to Amazon.com's [privacy notice](#) and [conditions of use](#).

Order Summary

Items:	\$141.75
Shipping & handling:	\$0.00
Total before tax:	\$141.75
Estimated tax to be collected:	\$8.86
Order total:	\$150.61

[How are shipping costs calculated?](#)

Prime shipping benefits have been applied to your order.

Place your order **Order total: \$150.61**

By placing your order, you agree to Amazon.com's [privacy notice](#) and [conditions of use](#).

Need help? Check our [Help pages](#) or [contact us](#)

For an item sold by Amazon.com: When you click the "Place your order" button, we'll send you an email message acknowledging receipt of your order. Your contract to purchase an item will not be complete until we send you an email notifying you that the item has been shipped.

[Important information about sales tax you may owe in your state](#)

thesweethome-20

Affiliate Marketing Terms

- **Affiliate Marketing ID:**
Unique identifier that Online Retailers use to tie Web Users to Publishers
- **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	<code>(www\.)amazon\.com</code>	<code>^godaddy\.*</code>
Cookie Setting URLs	<code>^/(?:.*(dp gp)/.*)?</code>	<code>[&?]tag=(?:& \? ^ ;)isc=</code>
Conversion URLs	<code>*handle-buy-box*</code>	<code>*domains/domain-configuration\.aspx*</code>
Affiliate ID Values	<code>tag=(.*?)(?:& \$)</code>	<code>cvosrc=(.*?)(?:& \$)</code>

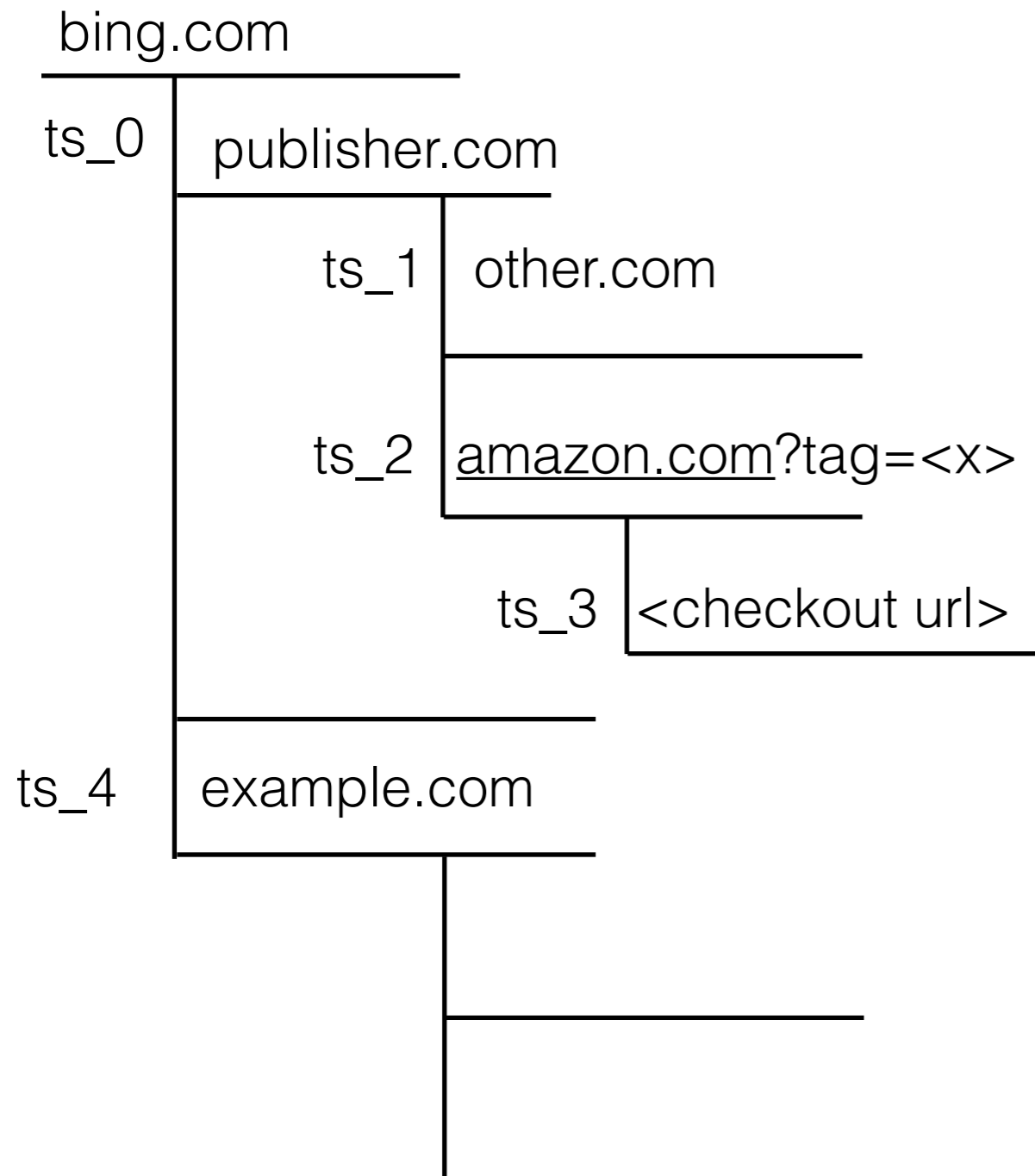
HTTP Request Logs

- 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

Browsing Session Trees



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

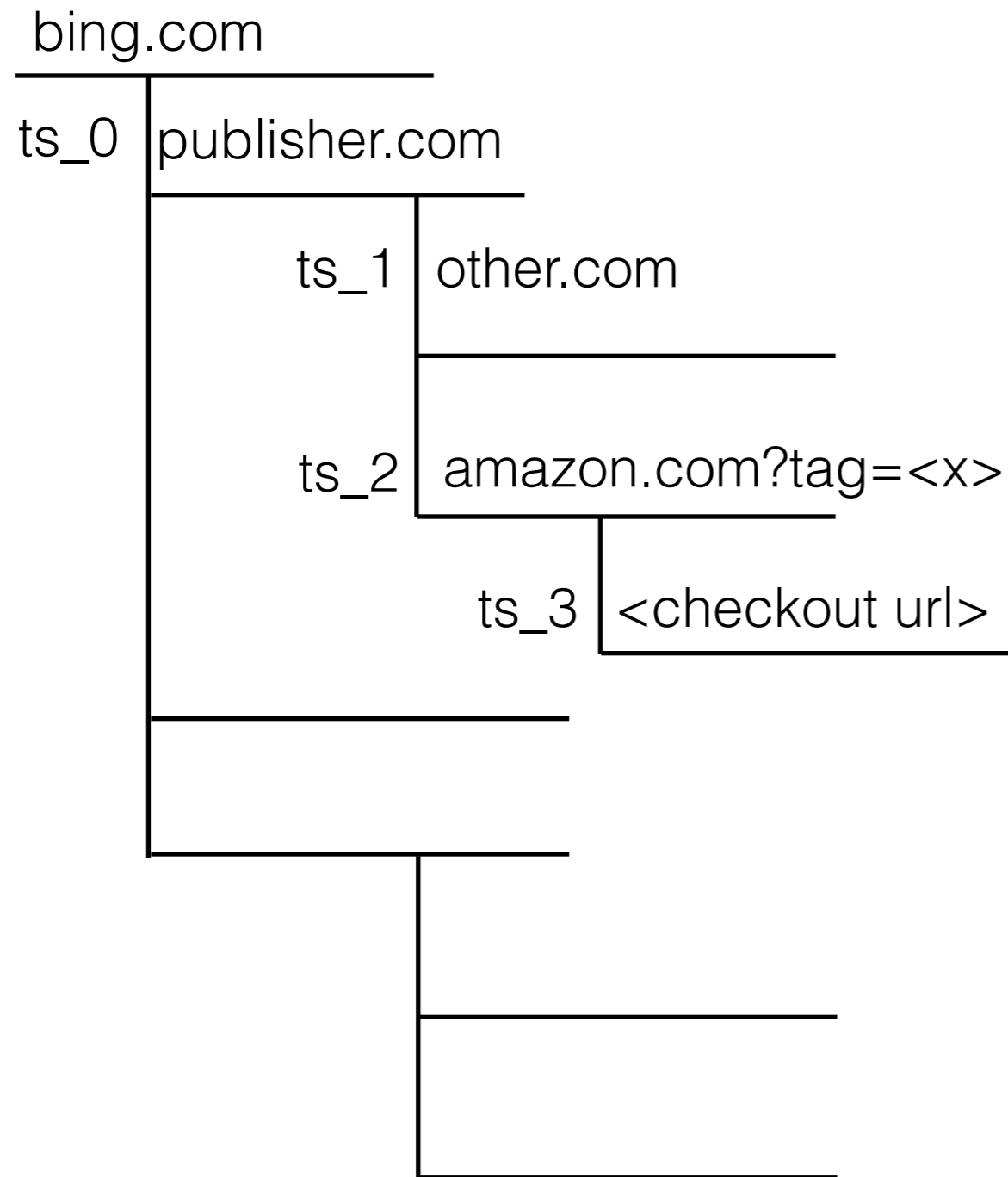
Browsing Session Trees

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

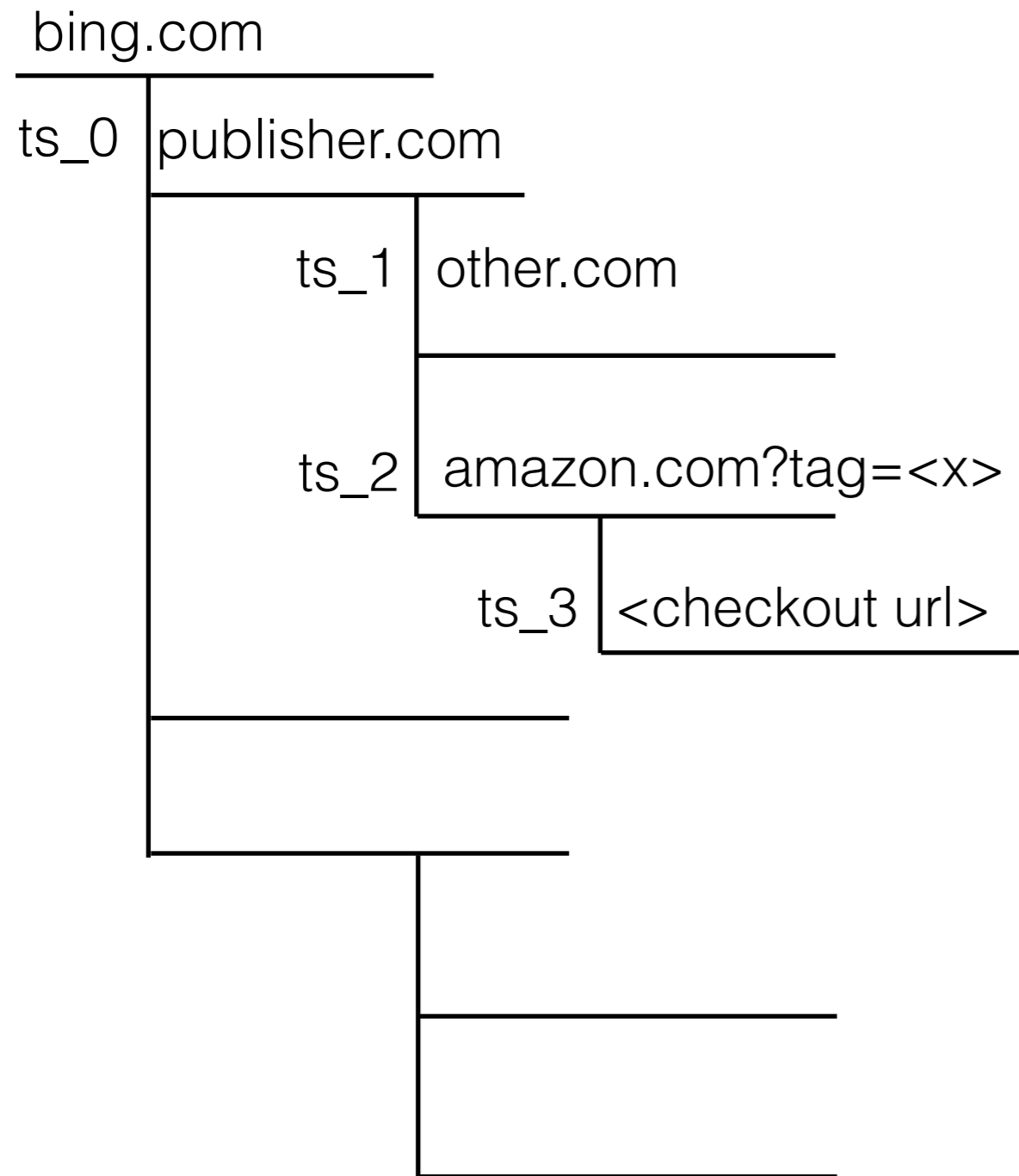


Classifier: Training Set



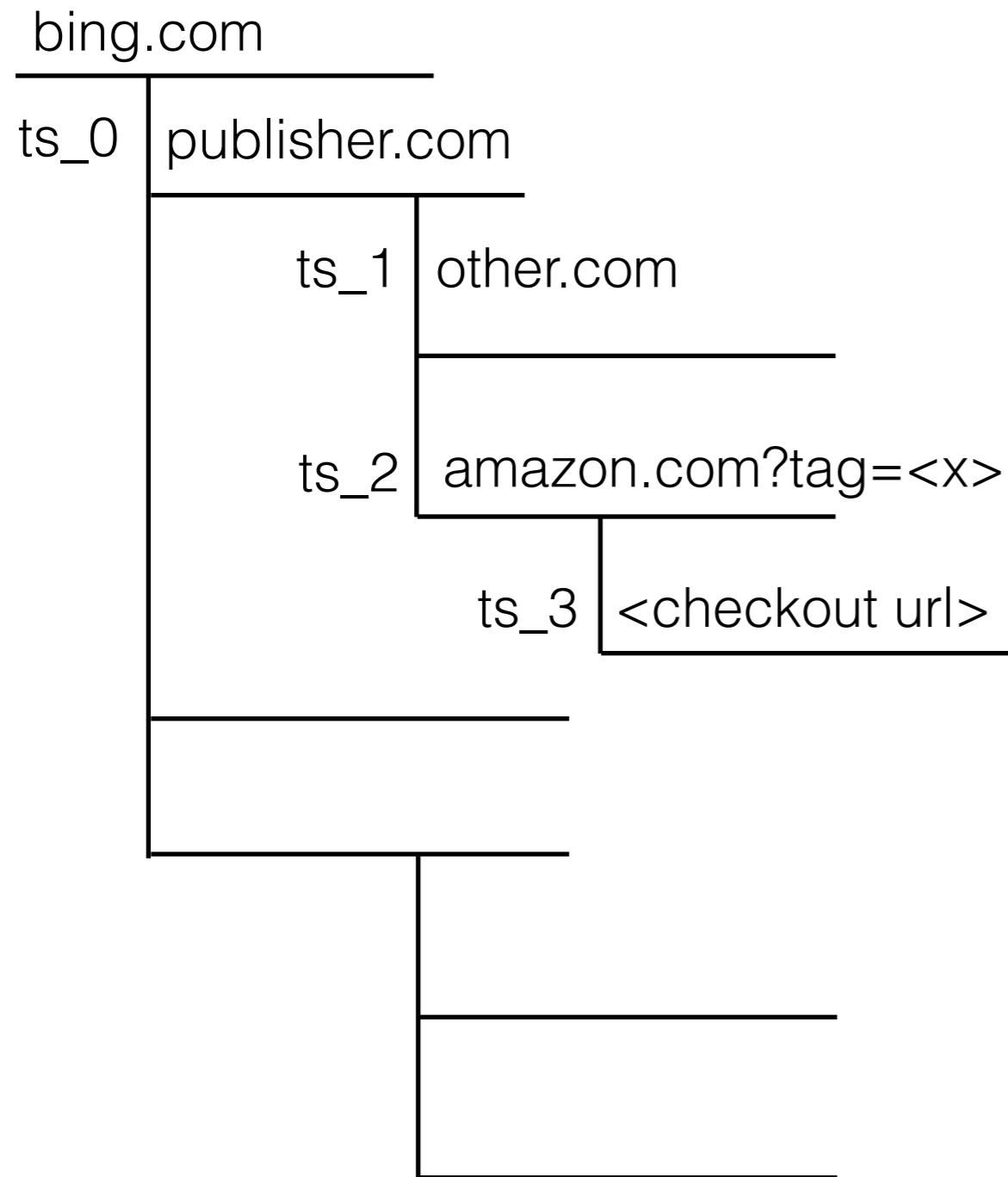
- Did the user intend to travel from some-referrer.com to amazon.com?
- Built training set of 1141 relevant trees (subset of January data)
- If referrer was still available, **direct test**
- If referrer was not available, **infer from graph** (log data)

Classifier: Features



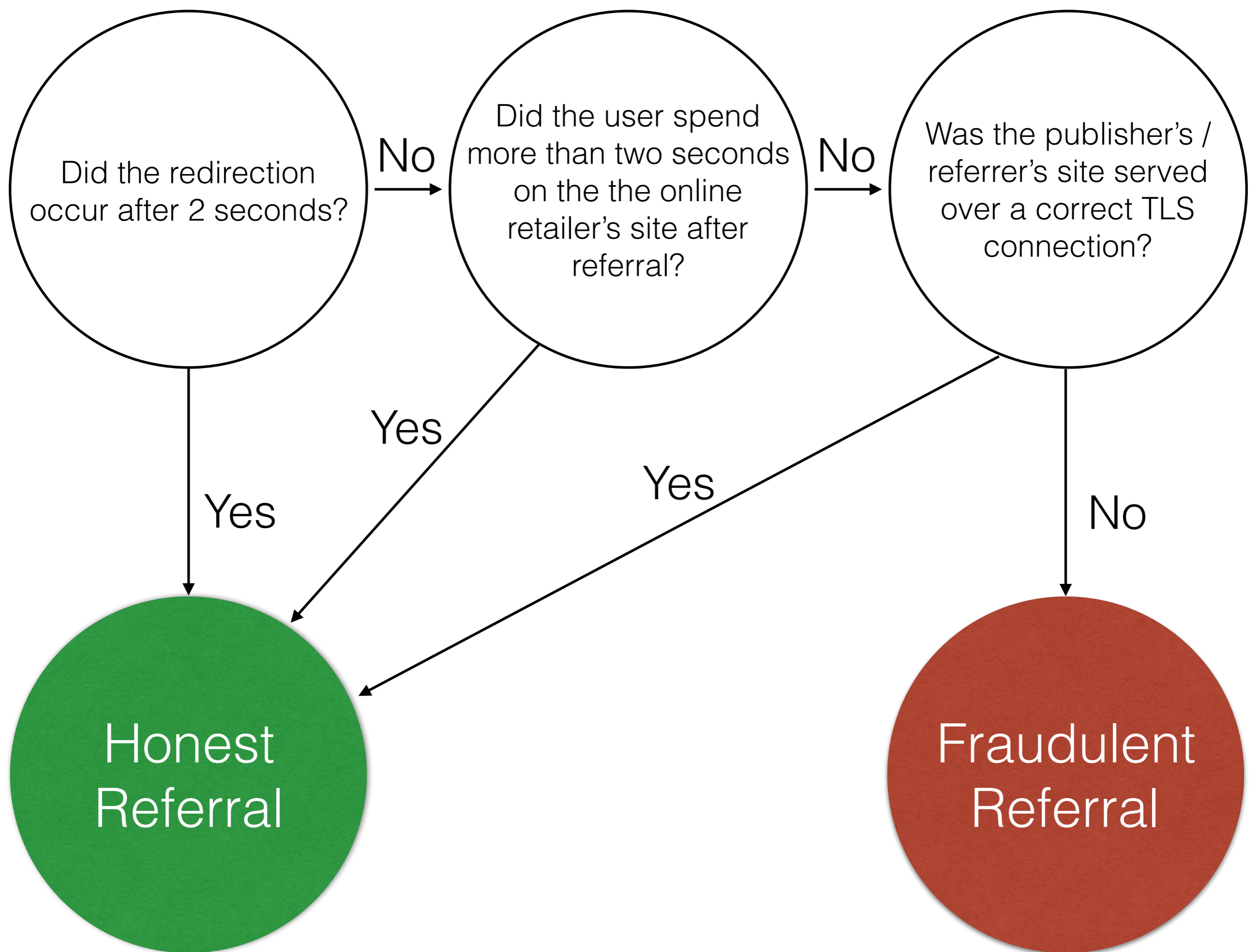
1. Time before referral
2. Time after referral
3. Is referrer SSL?
4. Graph size
5. Is referrer reachable?
6. Google page rank of referrer
7. Alexia traffic rank
8. Is referrer domain registered?
9. # years domain is registered
10. Tag count

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
Fraudulent	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 – psnyde2@uic.edu
Chris Kanich – ckanich@uic.edu
University of Illinois at Chicago

