Private Information Leakage on the Mobile Web

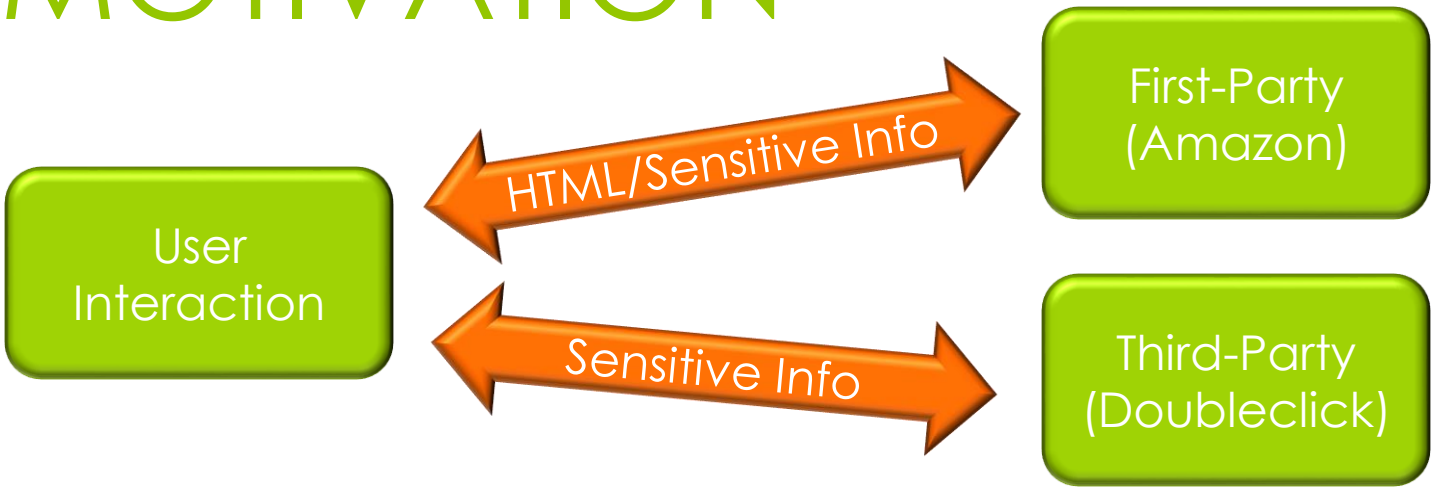
Zach Azar, Stephen Rice, Amanda Kirk, Yipu Wang

Introduction

- Goals:
  - Investigate the leakage of private information to third parties
  - Mobile Websites
  - Leakage via HTTP Requests
- <u>What</u> sensitive information leaks?
- <u>Who</u> does it leak to?
- <u>How</u> does it leak?

Introduction to Web Tracking

MOTIVATION



- "56% of 120 popular sites in our study (75% if you include user ids) directly leak sensitive and identification information to third party aggregators."

  Privacy Leakage vs. Protection Measures: The Growing Disconnect by Balachander Krishnamurthy, et al.

Third Party Leakage

GET	http://ad.doubleclick.net/adj/...radio;ag=30;gnd=1;zip=12201;artist=R53599;genre=rock;...
Referer	http://www.amazon.com/...

Example of information leakage through the GET URL

GET	http://ad.doubleclick.net/?l=7654&sz=200x250...
Referer	http://www.amazon.com/hserver/age=30/zip=12201/gender=M/...

Example of information leakage through the Referer

Krishnamurthy's Previous Work

	Desktop	Mobile
Online Social Network (OSN)	✓	✓
Non-OSN	✓	Our Study

Krishnamurthy's Results of Desktop Web Tracking

Leakage of Personal Information Via Web Sites Across Categories

Category	Sites w/ Direct Leakage	Action				
		Create Account	Account Login/ Navig.	View/ Edit Profile	Input Content	Sens. Search
Health	9	0	1	0	0	9
Travel	9	0	1	0	0	9
Employment	8	0	2	2	7	0
OSN	7	0	3	5	0	0
Arts	7	0	3	4	1	0
Relationships	7	0	3	2	2	0
News	5	0	5	0	0	0
PhotoShare	4	3	3	0	1	0
Sports	4	1	2	0	1	0
Shopping	3	0	2	0	2	0
AgeGroups	2	0	1	1	0	0
VideoGames	2	0	1	1	0	0
Tot. Sites/Cat.	67/12	4/2	27/12	15/6	14/6	18/2

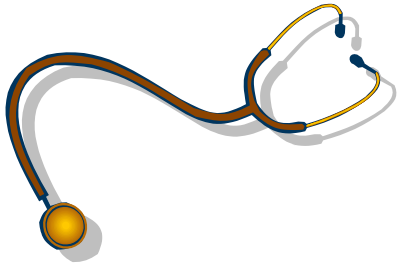
Foundational Work

Project Setup

- Data Gathering Methodology
  - What types of sites do we look at?
  - How do we choose specific sites?
  - How do we ensure consistency across our data set?

Categories

Health (Stephen)



Travel (Amanda)

Shopping (Yipu)

Relationships (Zach)

Choosing the Websites

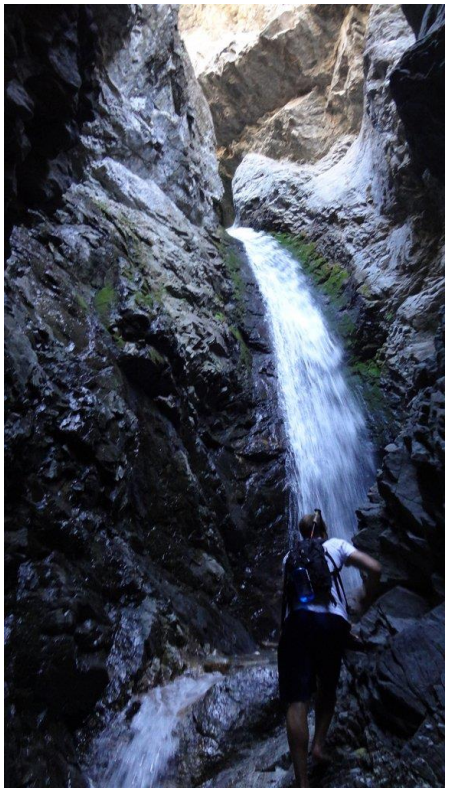
- Alexa
  - Ranks sites via popularity and use
- Our Criteria:
  - One of the top ranked sites
  - Does not require payment for registration
  - U.S. Based, Website in English
  - Not an Online Social Network (OSN)

Websites Chosen

Health	Relationships	Travel	Shopping
nih.gov	okcupid.com	agoda.com	amazon.com
webmd.com	pof.com	expedia.com	ebay.com
mayoclinic.com	kiss.com	booking.com	netflix.com
ncbi.nlm.nih.gov/ pubmed	datehookup.com	hotels.com	walmart.com
myfitnesspal.com	friendfinder.com	tripadvisor.com	cvs.com

Test User

- Meet Mathew Lamar Anderson
  - User Name: LambDUHyhn
  - DOB: 9/12/1988
  - E-mail: lambduhyhn@gmail.com
  - Likes: Batman, Eric Clapton, Sushi
  - Also contained information about social habits, drug and alcohol use, dating profiles and lyme disease



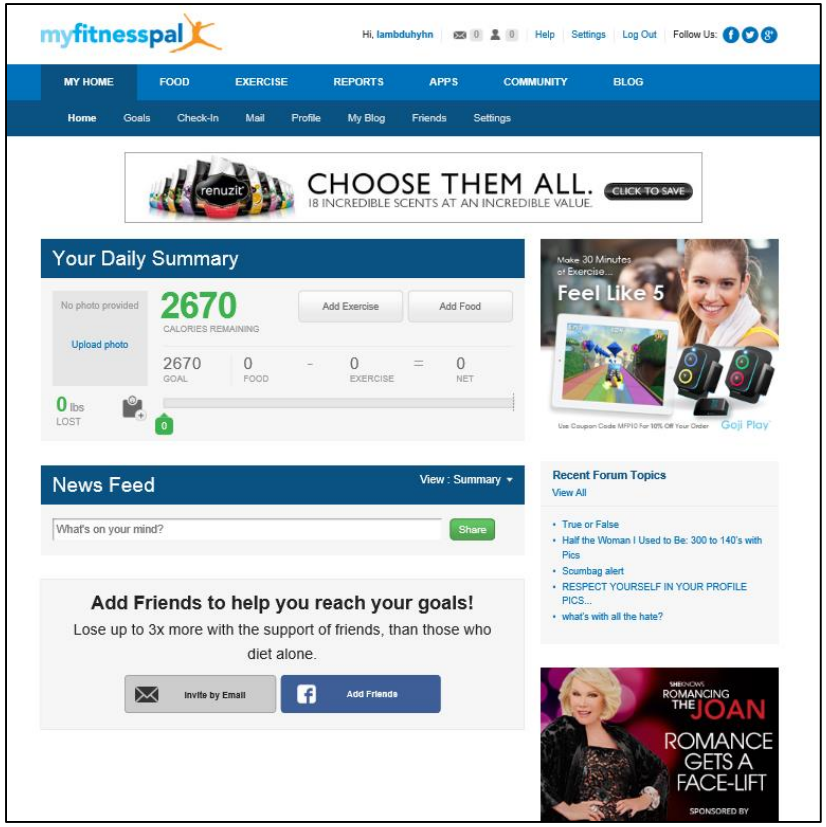
Data Collection & Analysis

Model / Experiment Setup / Tools Used

- Utilize the computer as a proxy server for the mobile device
- Fiddler allowed us to intercept, save and analyze the collected data

Web Crawls

- Each Crawl followed a set of procedures
  - Act like a normal user
    - Create Accounts
    - Search
    - Navigate Pages
    - Edit Profiles
  - Record actions in a roadmap document



Fiddler Example

Fiddler Web Debugger

File Edit Rules Tools View Help GET /book

Win8 Config | Replay | Go | Stream | Decode | Keep: All sessions | Any Process | Find | Save | Browse | Clear Cache | TextWizard | Tearoff | MSDN Search...

#	Result	Protocol	Host	URL
853	200	HTTP	www.myfitnesspal.com	/exercise/diary/lambd...
854	200	HTTP	secure-au.imrworldwide.com	/cgi-bin/m?ci=gorillana...
855	200	HTTP	aax-us-east.amazon-adsys...	/e/dtb/bid?src=1092&...
856	200	HTTP	www.google-analytics.com	/__utm.gif?utmwv=5....
857	200	HTTP	www.google-analytics.com	/__utm.gif?utmwv=5....
858	200	HTTP	pubads.g.doubleclick.net	/gampad/ads?gdfp_re...
859	204	HTTP	b.scorecardresearch.com	/b?c1=2&c2=1547633...
860	200	HTTP	d34yn14tavczy0.cloudfront...	/assets/jqueryui/ui-bg...
861	200	HTTP	pubads.g.doubleclick.net	/gampad/ads?gdfp_re...
862	200	HTTP	us-ads.openx.net	/w/1.0/jstag
863	304	HTTP	media.admob.com	/formats/templates.js
864	200	HTTP	googleads.g.doubleclick.net	/simgad/14965671017...
865	304	HTTP	pagead2.googlesyndication...	/pagead/images/adc-i-...
866	200	HTTP	cm.g.doubleclick.net	/push?client=ca-pub-9...
867	302	HTTP	connexity.net	/c/cse?a=Q&B=2d&go...
868	200	HTTP	go.realvu.net	/flip/2/c=IZ_r=RealVu...
869	200	HTTP	us-ads.openx.net	/w/1.0/acj?o=843984...
870	200	HTTP	Tunnel to	urs.microsoft.com:443
871	200	HTTP	cm.g.doubleclick.net	/pixel?google_nid=con...
872	204	HTTP	csi.gstatic.com	/csi?v=3&s=gmob&act...
873	200	HTTP	us-u.openx.net	/w/1.0/pd?plm=3&ph...
874	200	HTTP	us-ads.openx.net	/w/1.0/ri?ts=1fHJhaW...
875	200	HTTP	c1.rfihub.net	/creative/462317_634...
876	200	HTTP	secure-us.imrworldwide.com	/cgi-bin/m?ci=us-rocke...
877	200	HTTP	inw-655.inw-rtb1.rfihub.com	/bn/mbk.js?ai=462317...
878	302	HTTP	p.brilig.com	/contact/bct?pid=f9b6...
879	502	HTTP	syndication.mmismm.com	/mmtnt.php?mm_pub=...
880	200	HTTP	adadvisor.net	/adscores/g.js?sid=92...
881	200	HTTP	Tunnel to	urs.microsoft.com:443
882	304	HTTP	c1.rfihub.net	/adChoicesJs/rfacNew.js
883	302	HTTP	ox-m.d.chango.com	/m/ox
884	200	HTTP	openx.admailtiser.com	/match?mrbyus=true&...
885	302	HTTP	p.rfihub.com	/cm?in=1&pub=25
886	302	HTTP	a.rfihub.com	/ca.gif?rb=930&ca=2...
887	200	HTTP	go.realvu.net	/flip/2/c=IZ_r=RealVu...
888	302	HTTP	cm.g.doubleclick.net	/pixel?google_nid=rfi&

ALT+Q > type HELP...

Statistics | Inspectors | AutoResponder | Composer | Filters | Log | Timeline

Headers | TextView | WebForms | HexView | Auth | Cookies | Raw | JSON | XML

Request Headers [Raw] [Header Definitions]

GET /gampad/ads?gdfp_req=1&correlator=2525512039865770&output=json_html&callback=callbackProxy&impl=fif&iu=%2F1033141%2Fca-pub-8033683427063754%2Fexercisediary_72&x90&sz=72&x90&cookie=ID%3D2607f9cfb43472cf%3AT%3D1382396749%3AS%3DALNI_MZ6aRN3P4nXjrudtASt9gjazR0YyQ&lmt=1382396762&dt=1382396762899&cc=100&biw=1024&bih=1554&oid=3&adx=148&ady=193&adk=3439624703&gut=v2&oe=utf-8&fi=1&u_tz=-360&u_his=8&u_java=true&u_h=1280&u_w=768&u_ah=1280&u_aw=768&u_cd=24&flash=0&url=http%3A%2F%2Fwww.myfitnesspal.com%2Fexercise%2Fdiary%2Flambduhyhn&ref=http%3A%2F%2Fwww.myfitnesspal.com

Client

Accept: application/javascript, */*;q=0.8
Accept-Encoding: gzip, deflate
Accept-Language: en-US
UA-CPU: ARM
User-Agent: Mozilla/5.0 (compatible; MSIE 10.0; Windows Phone 8.0; Trident/6.0; IEMobile/10.0; ARM; Touch; NOKIA; Lumia 920)

Cookies / Login

Cookie
id
221c93dea70100a8|3660461/1275799/15940|t=1376779021|et=730|cs=002213fd489c29714ee96b4c2f
DNT: 1

Miscellaneous

Referer: http://www.myfitnesspal.com/exercise/diary/lambduhyhn

Transport

Connection: Keep-Alive
Host: pubads.g.doubleclick.net

Response is encoded and may need to be decoded before inspection. Click here to transform.

Get SyntaxView | Transformer | Headers | TextView | ImageView | HexView | WebView | Auth | Caching | Cookies | Raw | JSON | XML

The SyntaxView Inspector displays syntax-highlighted HTML, Script, CSS, and XML. If you're a web developer, you'll want this add-on.

Download and Install SyntaxView now...
Learn more about SyntaxView and other Inspector add-ons...

All Processes | 1 / 2,857 | http://pubads.g.doubleclick.net/gampad/ads?gdfp_req=1&correlator=2525512039865770&output=json_html&callback=callbackProxy&impl=fif&iu=%2F1033141%2Fca-pub-8033683427063754%2Fe

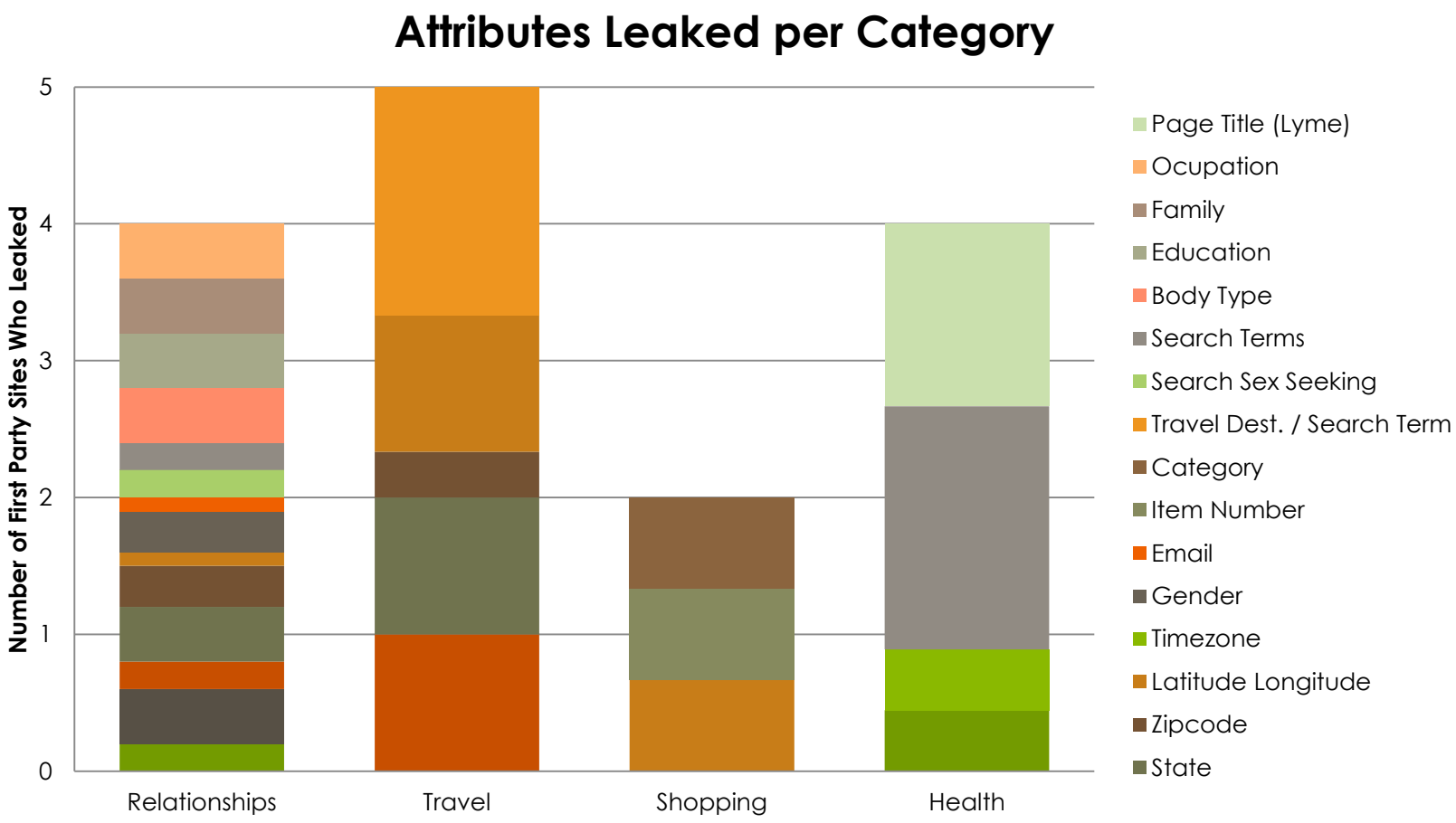
Method for Analyzing Data

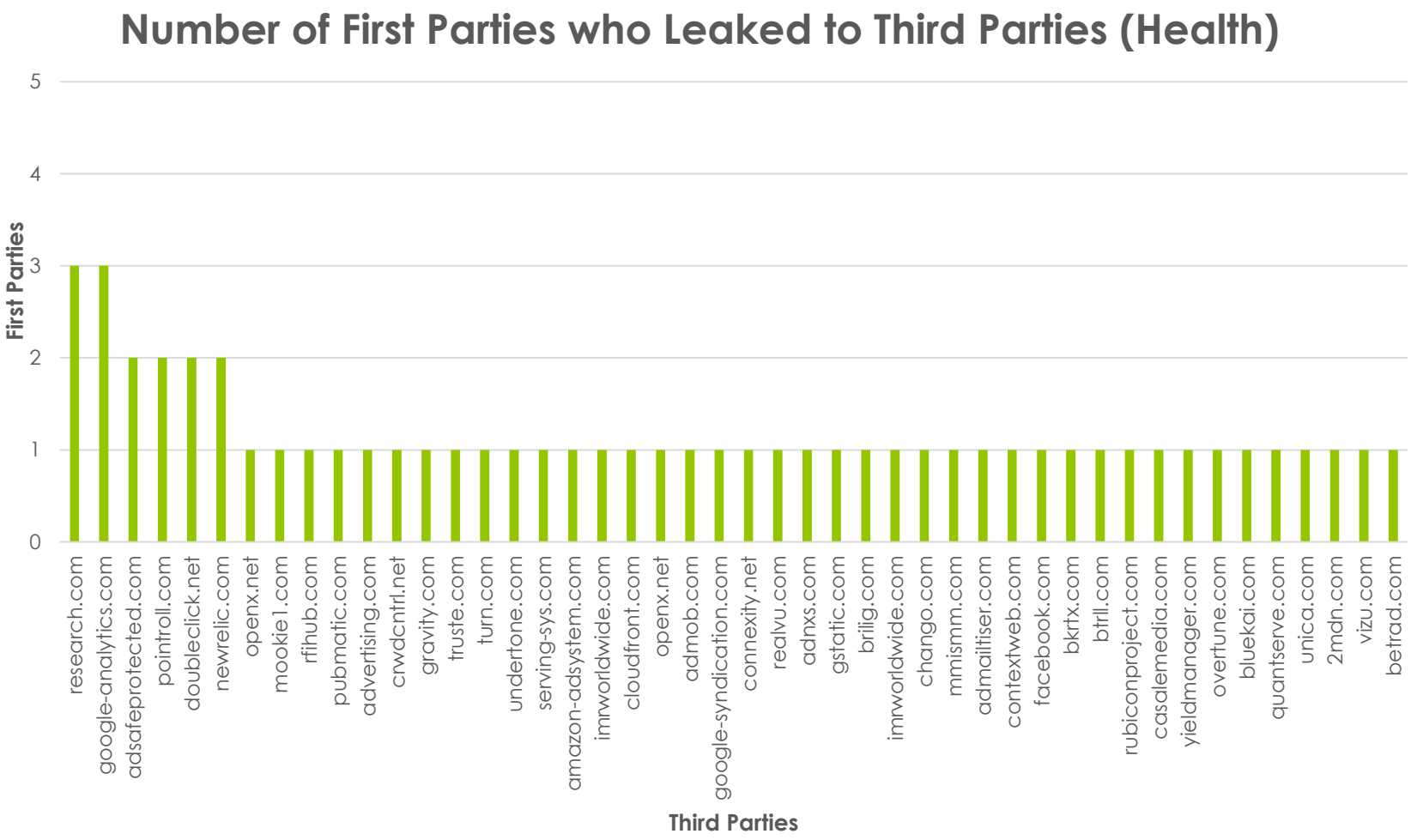
1. Create a list of possible search terms
2. Search all packets for leakage of each term
3. Investigate matches
   - Is it a third party?
   - Can we prove it is leakage?

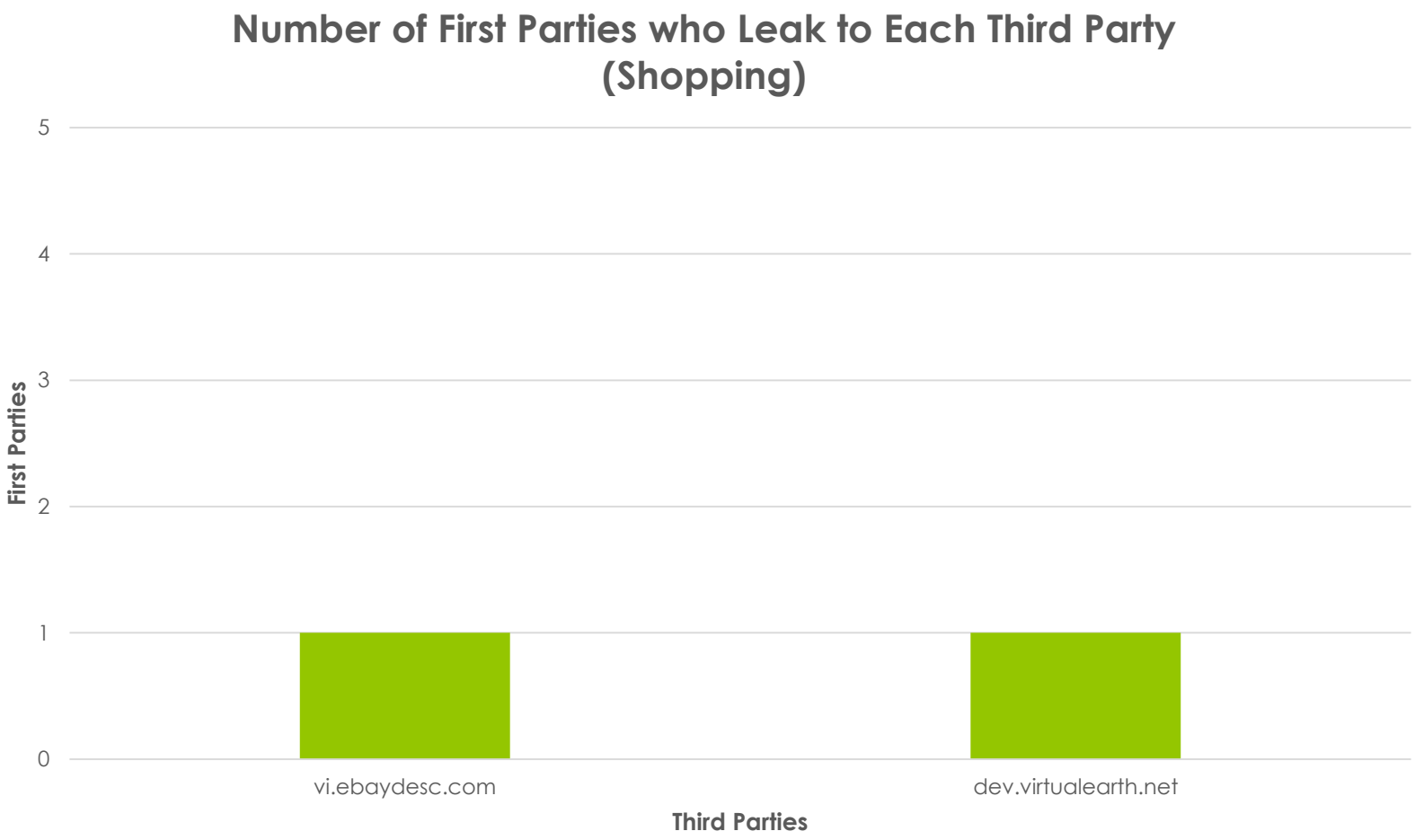
Search Terms:

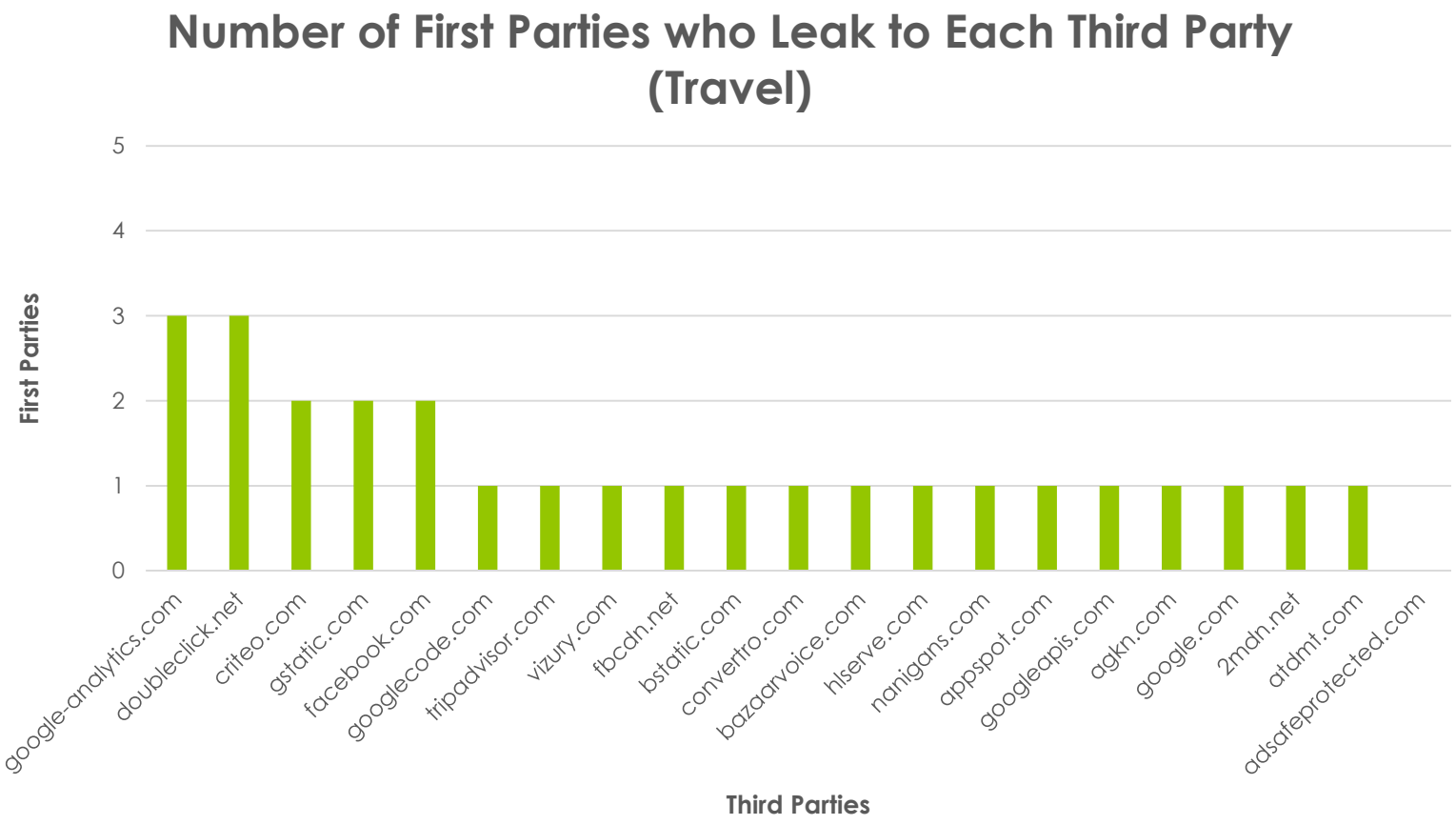
Username	Email	Zip Code
username	email	zipcode
un	em	zip
user	mail	postal
uid	lambduhyn@gmail.com	postalcode
LambDUHyhn		80203

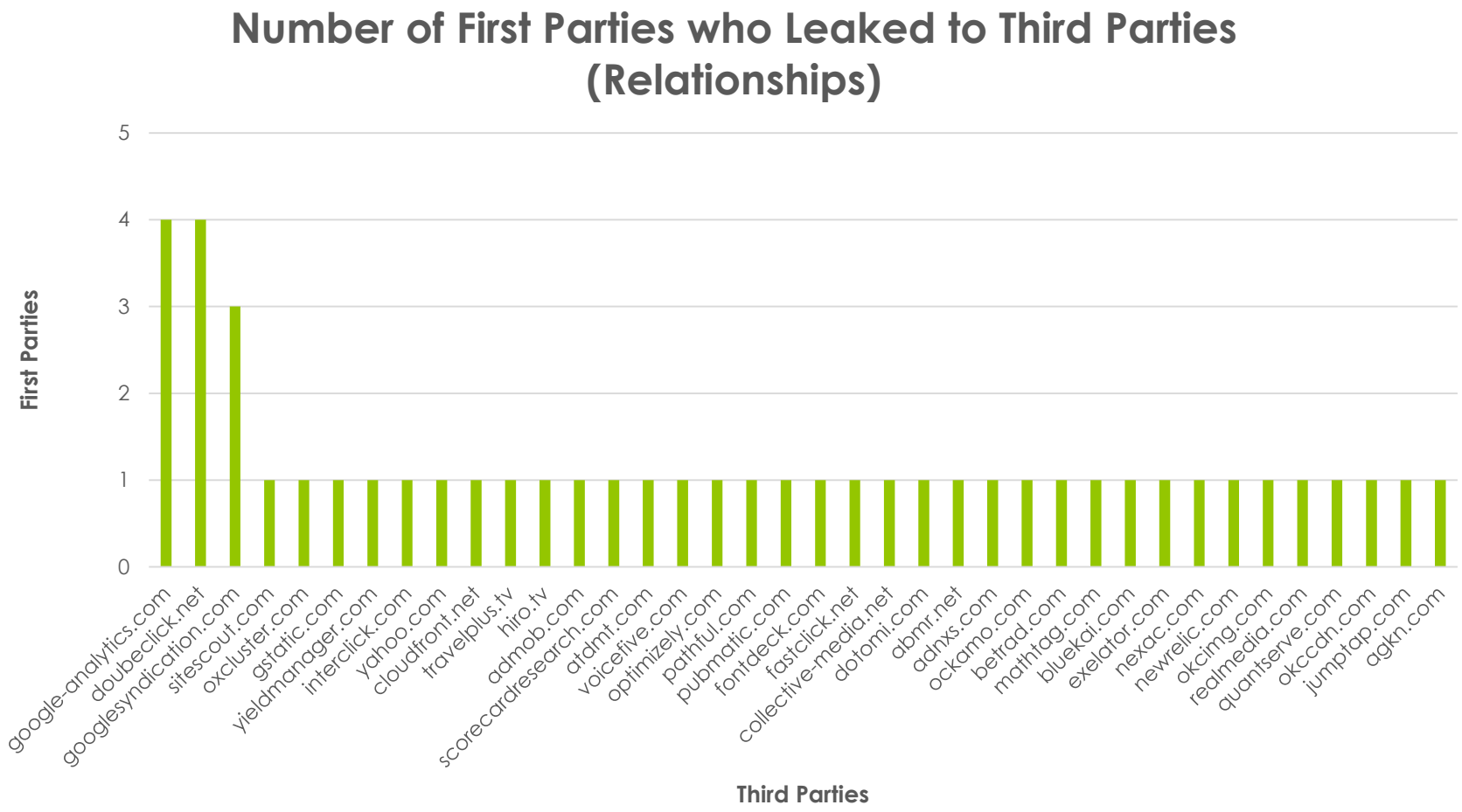
Results

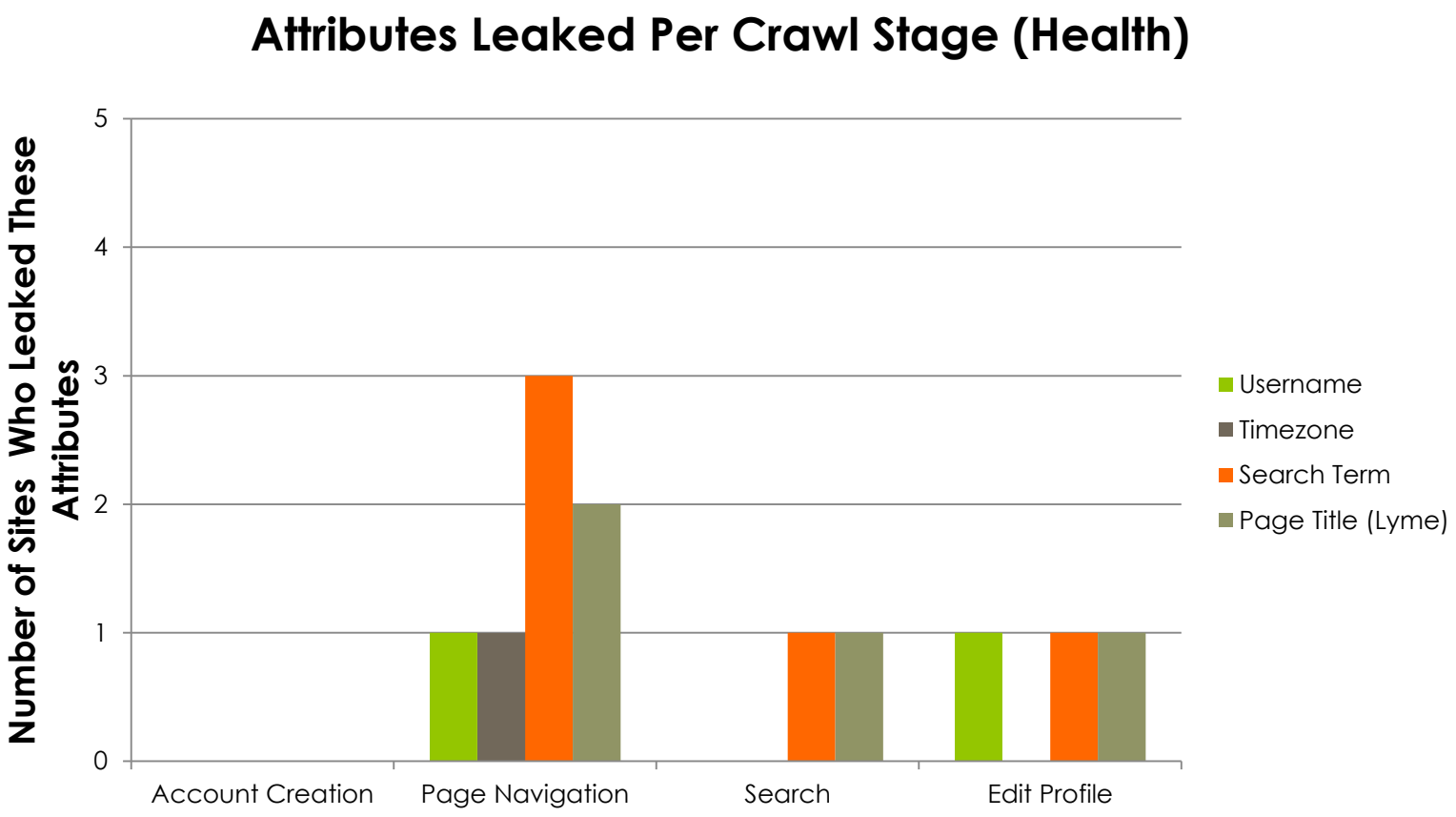

Attributes Leaked per Category
Number of First Party Sites Who Leaked
0
1
2
3
4
5
Relationships
Travel
Shopping
Health
Page Title (Lyme)
Ocupation
Family
Education
Body Type
Search Terms
Search Sex Seeking
Travel Dest. / Search Term
Category
Item Number
Email
Gender
Timezone
Latitude Longitude
Zipcode
State

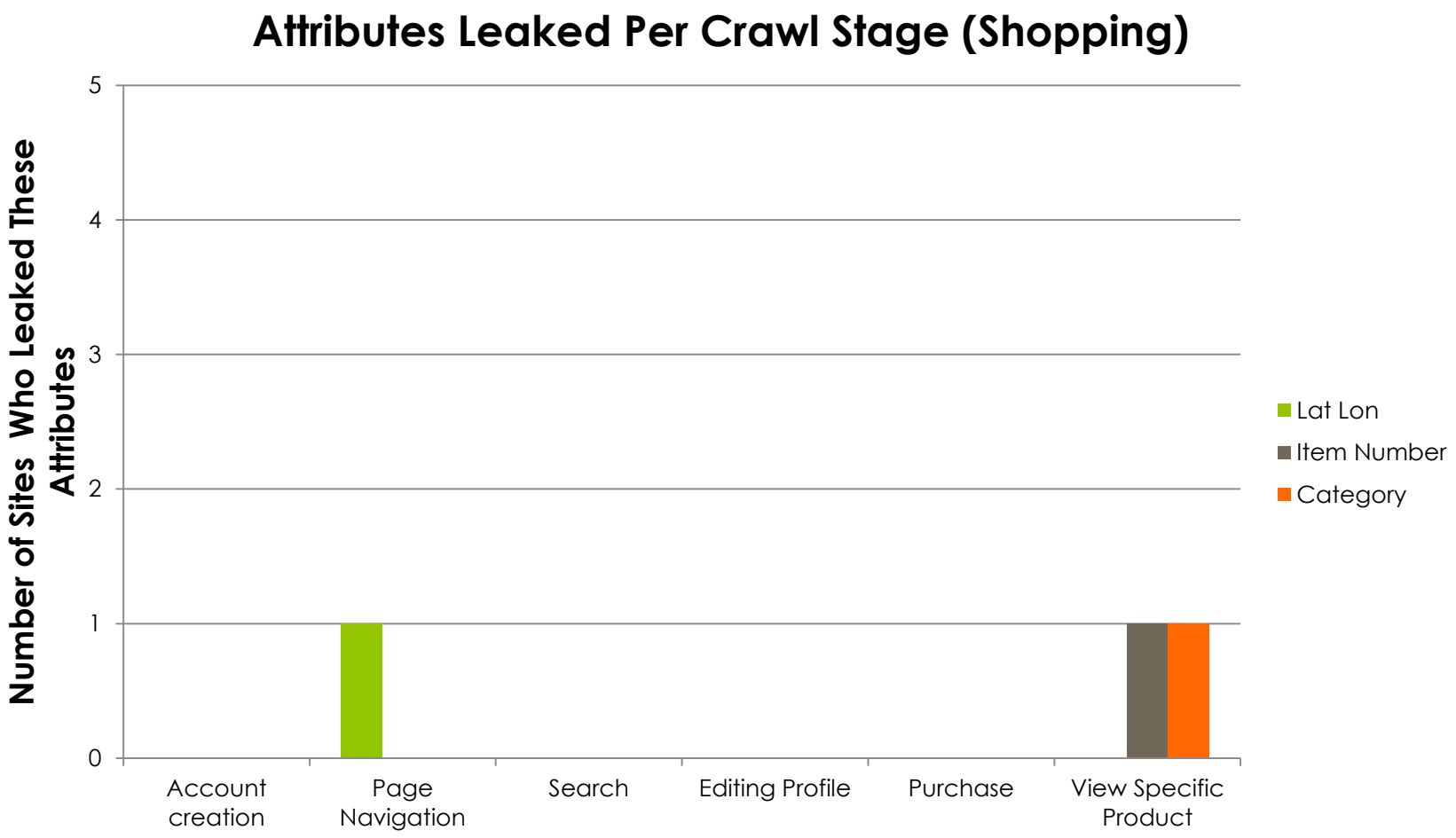

Number of First Parties who Leaked to Third Parties (Health)
First Parties
5
4
3
2
1
0
research.com
google-analytics.com
adsafeprotected.com
pointroll.com
doubleclick.net
newrelic.com
openx.net
mookie1.com
rfihub.com
pubmatic.com
advertising.com
crwdcntrl.net
gravity.com
truste.com
turn.com
undertone.com
serving-sys.com
amazon-adsystem.com
imrworldwide.com
cloudfront.com
openx.net
admob.com
google-syndication.com
connexity.net
realvu.com
adnxs.com
gstatic.com
brilig.com
imrworldwide.com
chango.com
mmismm.com
admailiser.com
contextweb.com
facebook.com
bkrtx.com
btrll.com
rubiconproject.com
casalemedia.com
yieldmanager.com
overtune.com
bluekai.com
quantserve.com
unica.com
2mdn.com
vizu.com
betrad.com
Third Parties

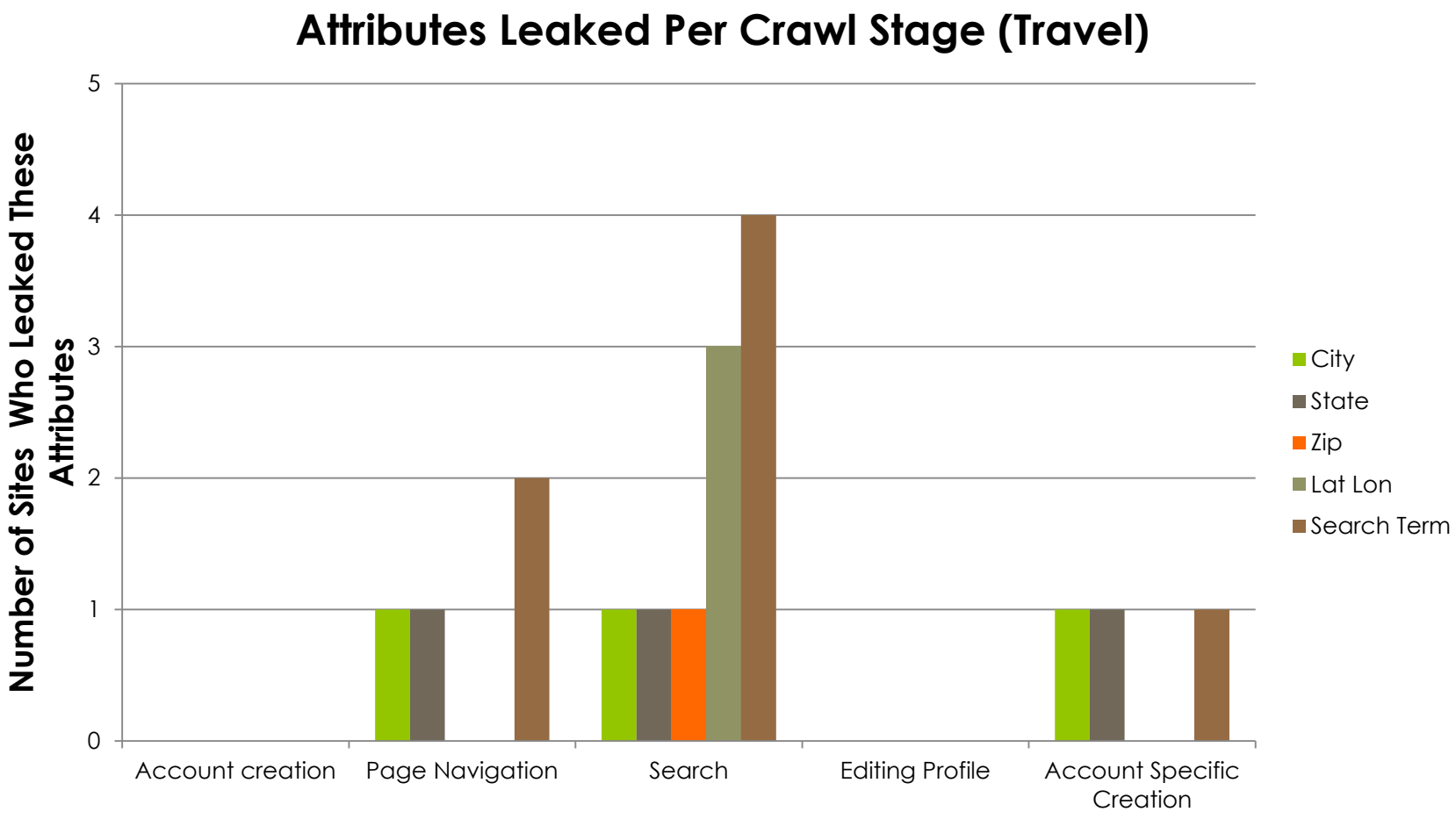

Number of First Parties who Leak to Each Third Party (Shopping)
5
4
3
2
1
0
First Parties
vi.ebaydesc.com
dev.virtualearth.net
Third Parties

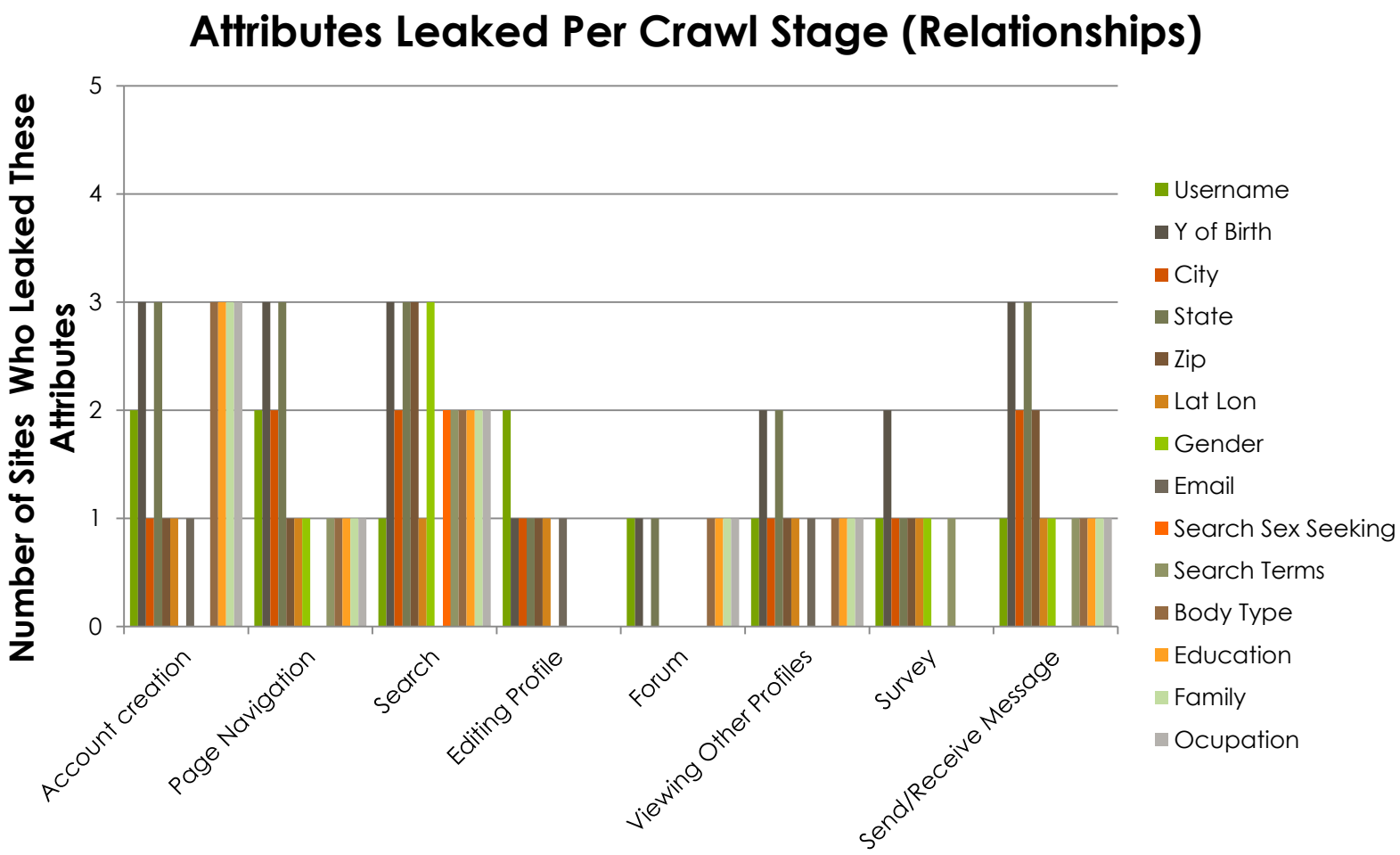

Number of First Parties who Leak to Each Third Party (Travel)
First Parties
5
4
3
2
1
0
google-analytics.com
doubleclick.net
criteo.com
gstatic.com
facebook.com
googlecode.com
tripadvisor.com
vizury.com
fbcdn.net
bstatic.com
convertro.com
bazaarvoice.com
hlserve.com
nanigans.com
appspot.com
googleapis.com
agkn.com
google.com
2mdn.net
atdmt.com
adsafeprotected.com
Third Parties

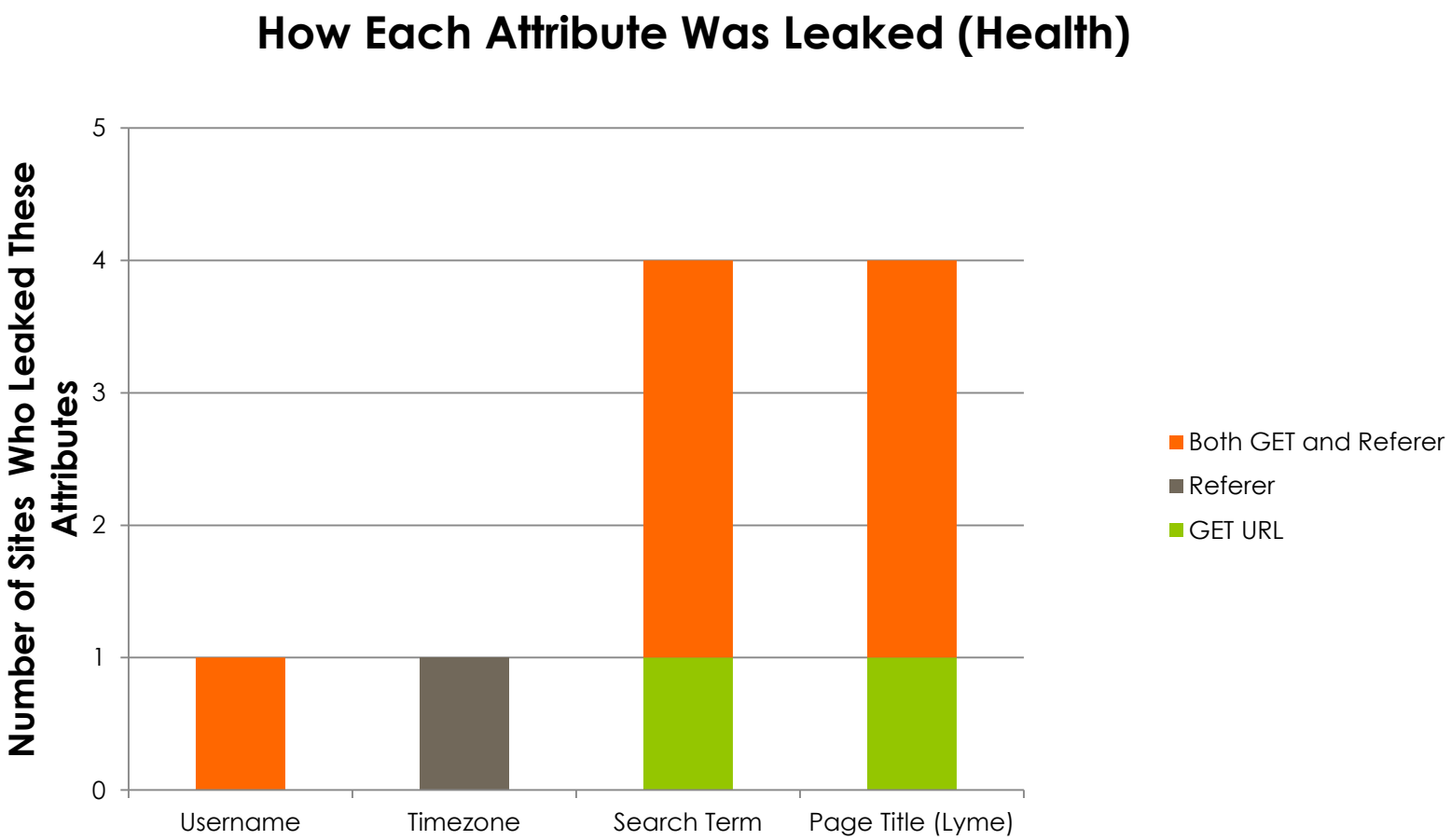

Number of First Parties who Leaked to Third Parties (Relationships)
First Parties
5
4
3
2
1
0
google-analytics.com
doubeclick.net
googlesyndication.com
sitescout.com
oxcluster.com
gstatic.com
yieldmanager.com
interclick.com
yahoo.com
cloudfront.net
travelplus.tv
hiro.tv
admob.com
scorecardresearch.com
atdmt.com
voicefive.com
optimizely.com
pathful.com
pubmatic.com
fontdeck.com
fastclick.net
collective-media.net
dotomi.com
abmr.net
adnxs.com
ockamo.com
betrad.com
mathtag.com
bluekai.com
exelator.com
nexac.com
newrelic.com
okcimg.com
realmedia.com
quantserve.com
okccdn.com
jumptap.com
agkn.com
Third Parties

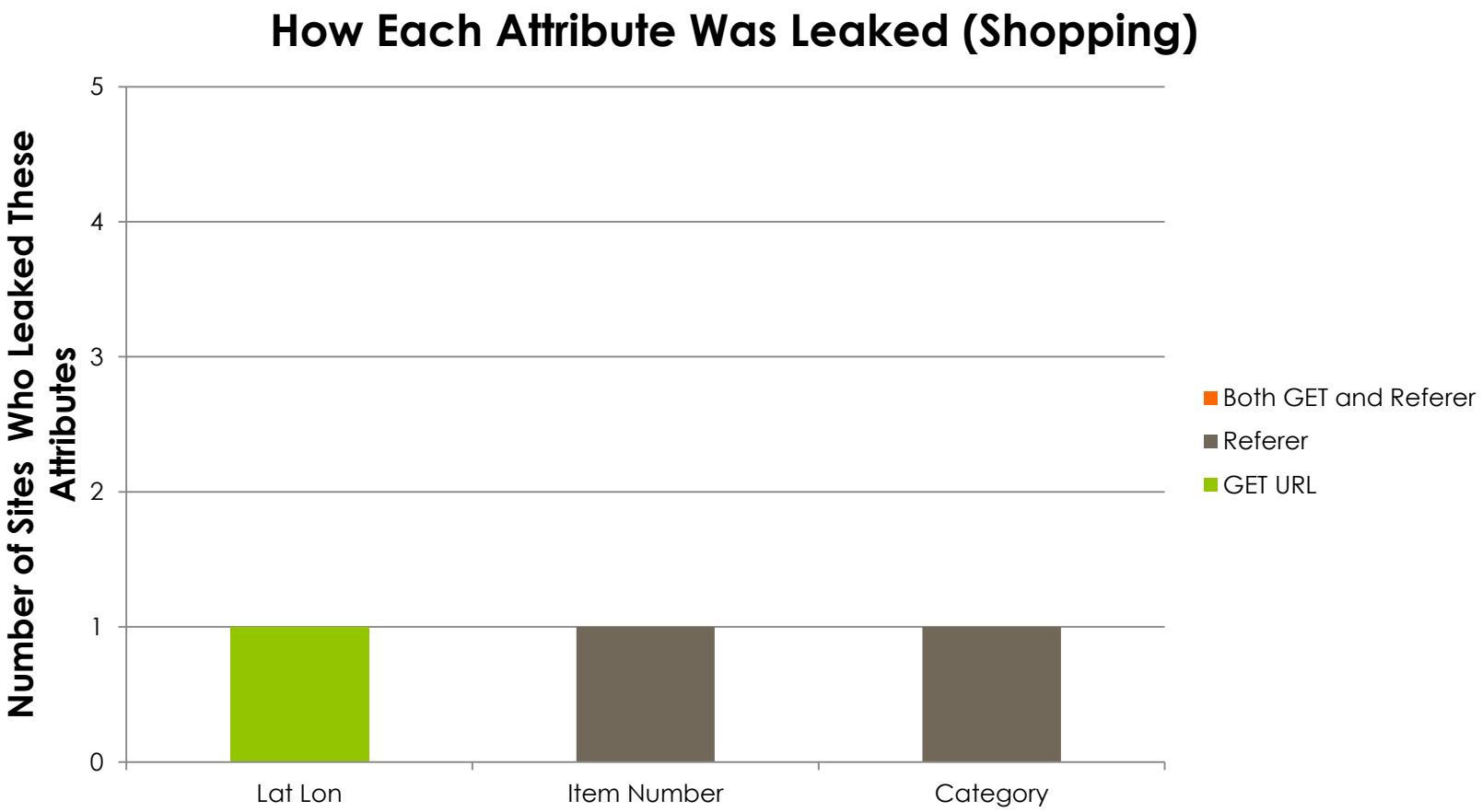

Attributes Leaked Per Crawl Stage (Health)
Number of Sites Who Leaked These Attributes
5
4
3
2
1
0
Account Creation
Page Navigation
Search
Edit Profile
Username
Timezone
Search Term
Page Title (Lyme)

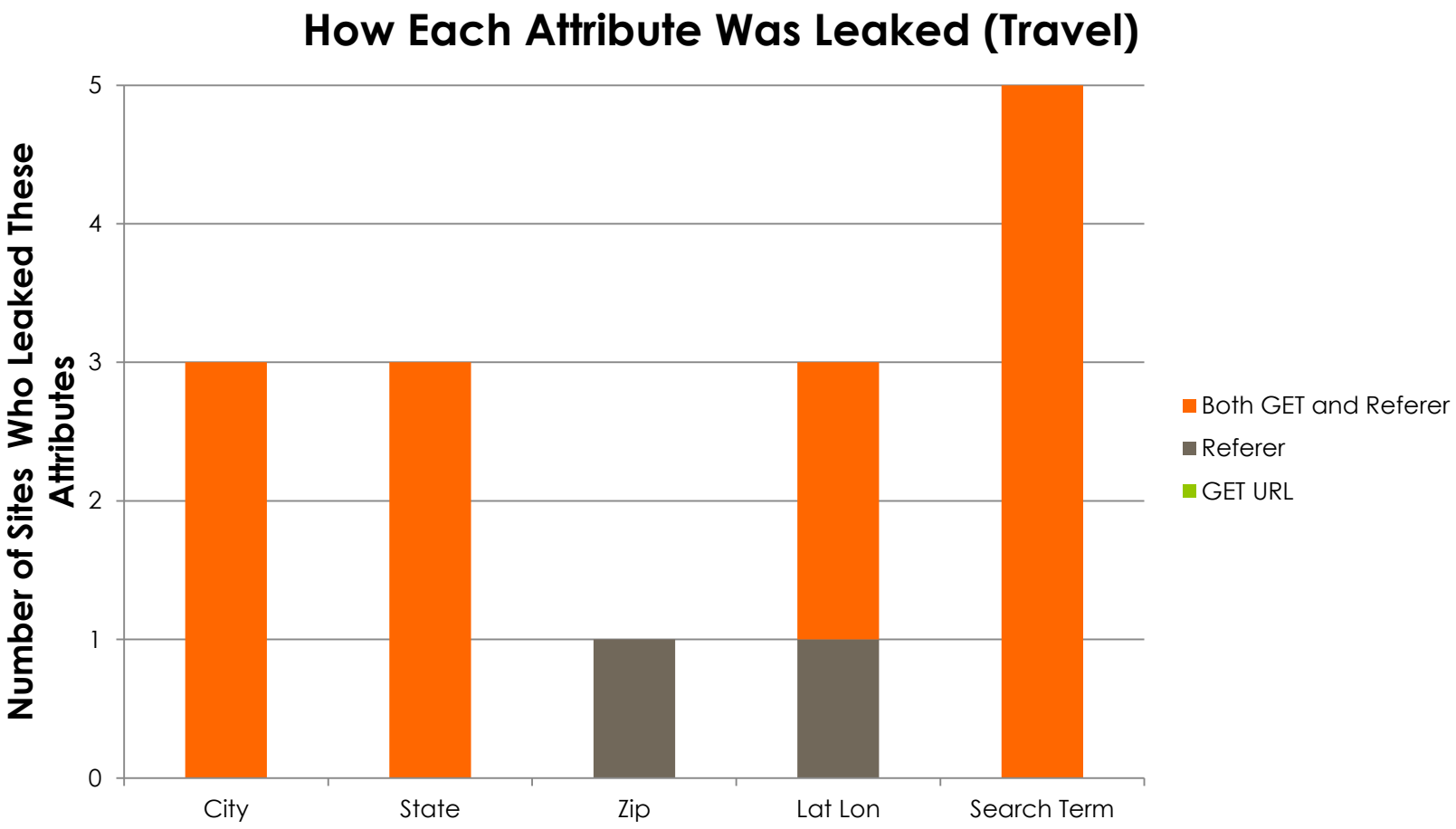

Attributes Leaked Per Crawl Stage (Shopping)
Number of Sites Who Leaked These Attributes
5
4
3
2
1
0
Account creation
Page Navigation
Search
Editing Profile
Purchase
View Specific Product
Lat Lon
Item Number
Category

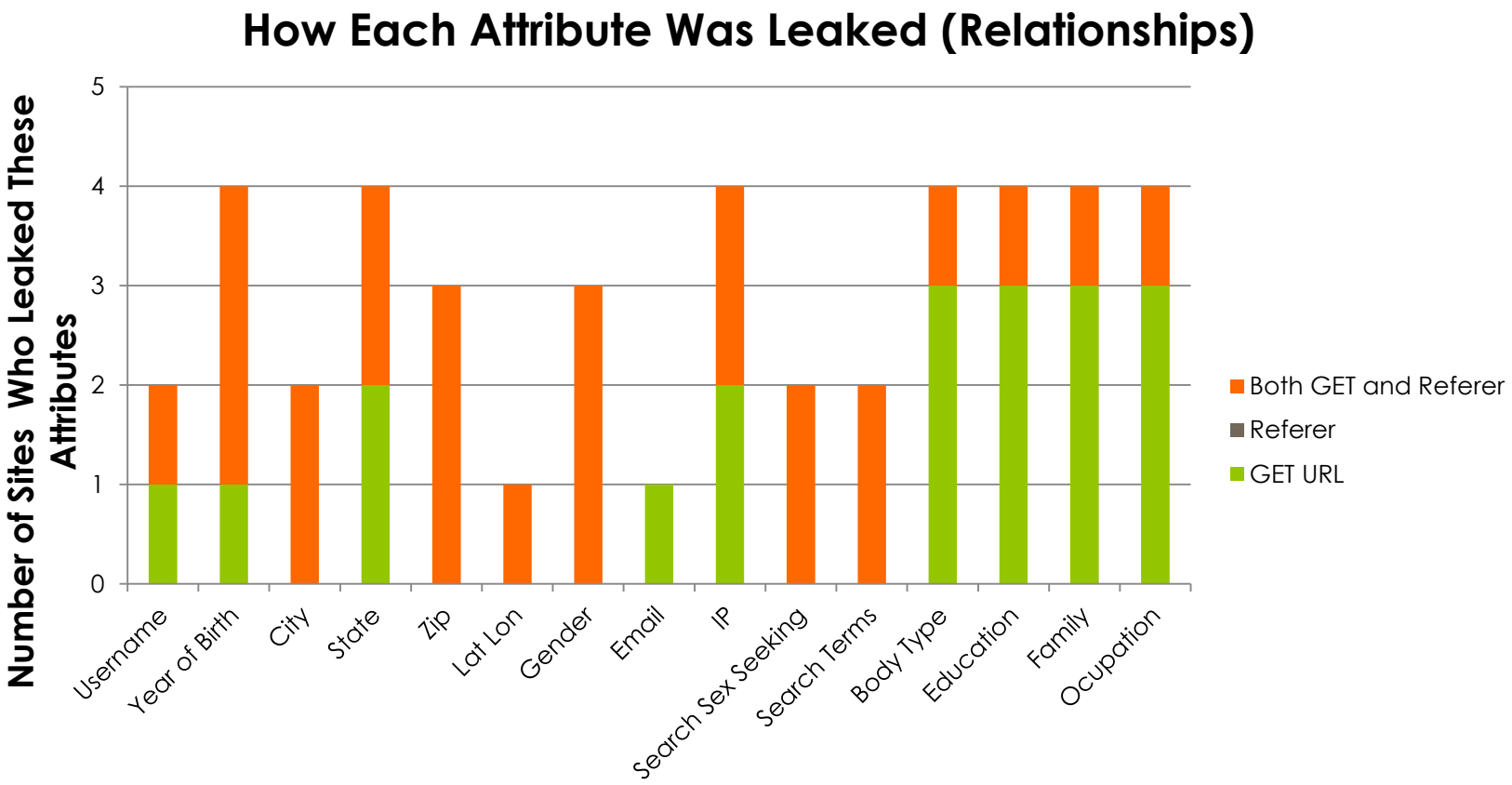

Attributes Leaked Per Crawl Stage (Travel)
Number of Sites Who Leaked These Attributes
5
4
3
2
1
0
Account creation
Page Navigation
Search
Editing Profile
Account Specific Creation
City
State
Zip
Lat Lon
Search Term


Attributes Leaked Per Crawl Stage (Relationships)
Number of Sites Who Leaked These Attributes
5
4
3
2
1
0
Account creation
Page Navigation
Search
Editing Profile
Forum
Viewing Other Profiles
Survey
Send/Receive Message
Username
Y of Birth
City
State
Zip
Lat Lon
Gender
Email
Search Sex Seeking
Search Terms
Body Type
Education
Family
Ocupation


How Each Attribute Was Leaked (Health)
Number of Sites Who Leaked These Attributes
5
4
3
2
1
0
Username
Timezone
Search Term
Page Title (Lyme)
Both GET and Referer
Referer
GET URL


How Each Attribute Was Leaked (Shopping)
Number of Sites Who Leaked These Attributes
5
4
3
2
1
0
Lat Lon
Item Number
Category
Both GET and Referer
Referer
GET URL

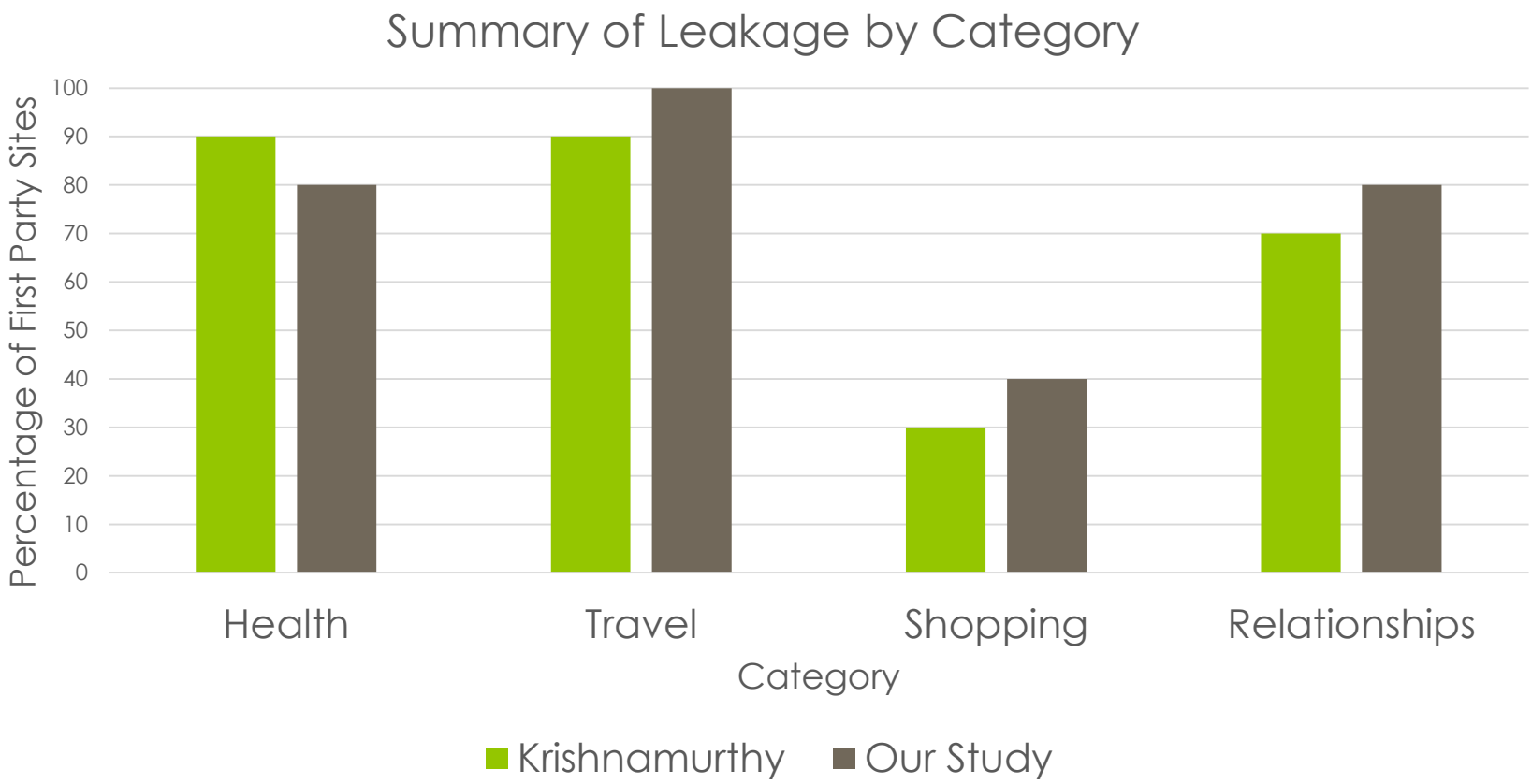

How Each Attribute Was Leaked (Travel)
Number of Sites Who Leaked These Attributes
5
4
3
2
1
0
City
State
Zip
Lat Lon
Search Term
Both GET and Referer
Referer
GET URL


How Each Attribute Was Leaked (Relationships)
Number of Sites Who Leaked These Attributes
0
1
2
3
4
5
Username
Year of Birth
City
State
Zip
Lat Lon
Gender
Email
IP
Search Sex Seeking
Search Terms
Body Type
Education
Family
Ocupation
Both GET and Referer
Referer
GET URL

Future Work

- More data, more sites, more crawls
  - Does the information leaked change for older user accounts?
- Encrypted arguments
  - Many packets contained encrypted data that we suspect has leakage but cannot prove
- Cookies
  - How do cookies interact with GET/Referer leakage?
- Linking between third parties
  - Do third parties share information about us?

Conclusion



Our results for private information leakage in a mobile platform are comparable to Krishnamurthy's results from his previous study on a desktop platform for various non-OSN categories.

Questions?

References

- *Third-Party Web Tracking: Policy and Technology* by Mayer and Mitchell
- *Privacy Leakage vs. Protection Measures* by Krishnamurthy and Wills
- *Tracking the Trackers: Where Everybody Knows Your Username* by Mayer
- *Privacy Leakage in Mobile Online Social Networks* by Krishnamurthy and Wills
- *Privacy Diffusion on the Web: a Longitudinal Study* by Krishnamurthy and Wills
- *On the Leakage of Personally Identifiable Information via OSN's* by Krishnamurthy and Wills