

Unsupervised discovery of malware redirection campaigns from fake news sites

Zhouhan Chen, PhD student

New York University, Center for Data Science

zc1245@nyu.edu

Last updated: 04/11/2020

Executive Summary

1. I developed an unsupervised detection system that identified large-scale redirection campaigns, some with 4000+ domains. Those campaigns use cloaking and fast flux to evade Google safe browsing's detection.
2. I analyzed the final landing URLs and found at least 10+ adware Chrome extensions that overwrite default Chrome search URLs with aggressive permissions (e.g.: access to all http*, https* sites)
3. This research will benefit both social science community and cybersecurity community.

Roadmap

1. Method
2. Challenge and crawling architecture
3. Entry point (seed suspicious domains) analysis
4. Results: Discovered redirection campaigns
5. Results: Fast flux evidence
6. Results: Malicious chrome extension analysis
7. Conclusion

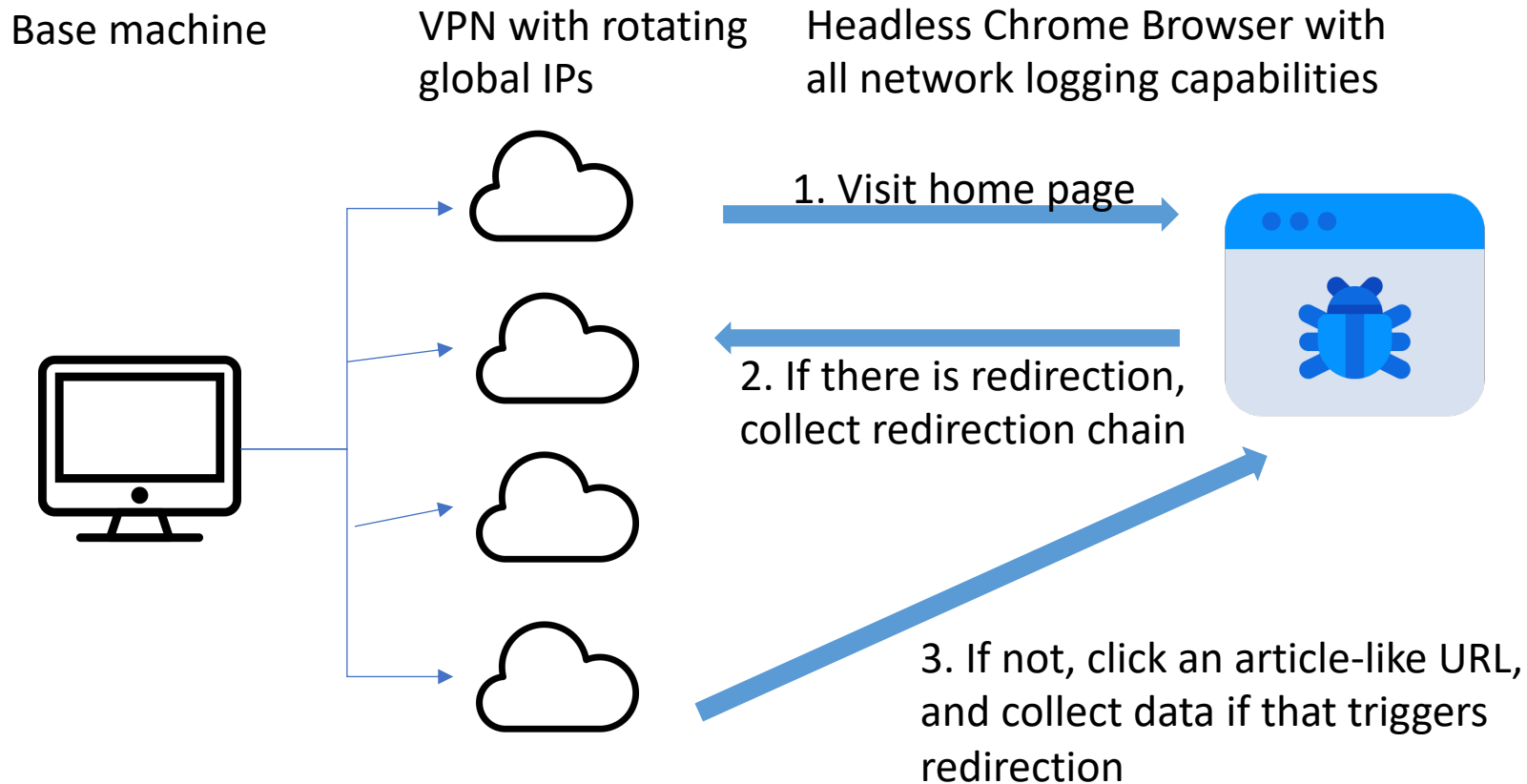
Method

1. My discovery **entry point** is a list of fake news domains
2. Identify suspicious fake news sites (seeds) that redirect (13% of total)
3. Double reverse search:
 1. Get all IPs a domain is hosted on
 2. Get all domains hosted on those IPs
4. Cluster domains from #3.2 based on common redirection paths
5. Visualize malware campaigns

Challenge

- This field is very dynamic
- Detection (evidence collection) is very difficult
- Abusers use anti-crawling techniques to evade detection
 - IP ban
 - Require javascript execution
 - Require user interaction (click a link)
 - Fast flux (domains change IP frequently)

My crawling architecture



Roadmap

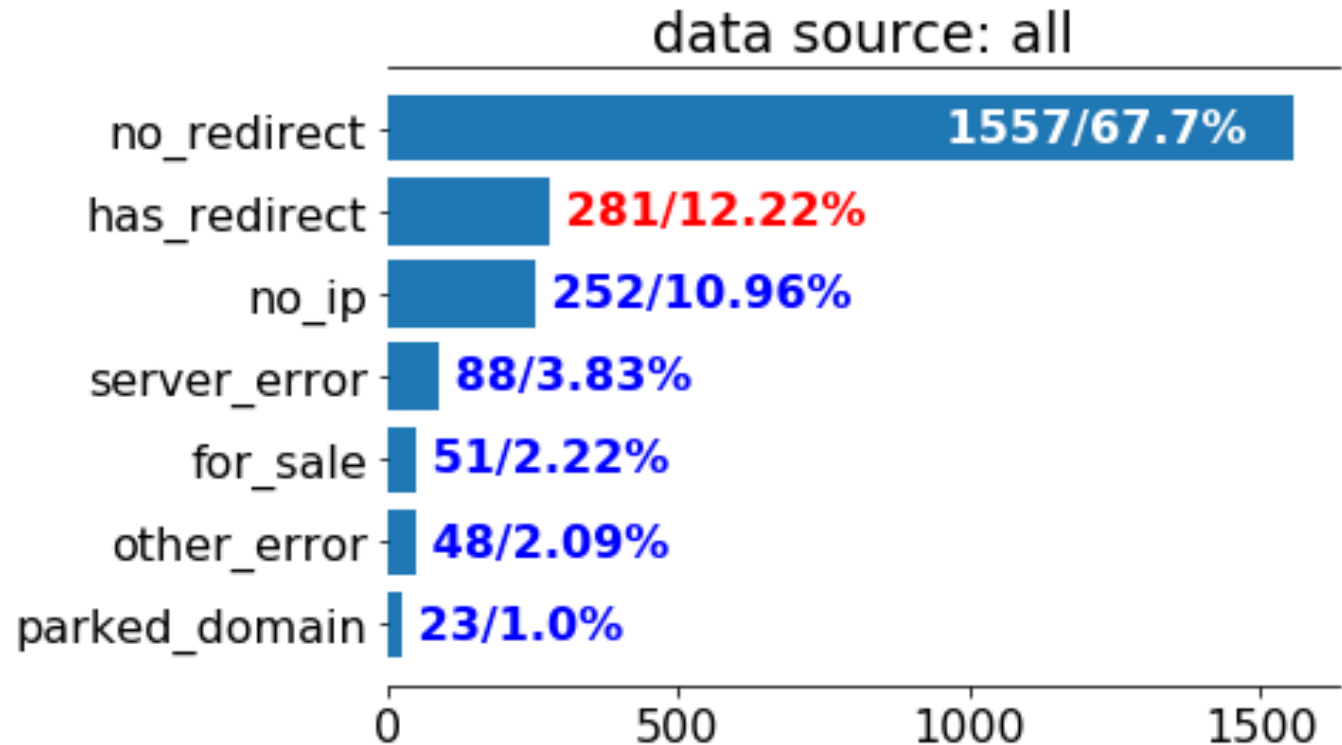
1. Method
2. Challenge and crawling architecture
- 3. Entry point (seed suspicious domains) analysis**
4. Results: Discovered redirection campaigns
5. Results: Fast flux evidence
6. Results: Malicious chrome extension analysis
7. Conclusion

Status of fake news domains from five popular sources

Status vs Source	Total # domains	No redirect	Has redirect	No IP/domain not exist	Server error	For sale	Other error	Parked domain
MediaBias-FactCheck	1395	77.99%	8.96%	6.38%	2.87%	1.43%	1.36%	1.00%
Politifact	325	44.00%	22.15%	19.38%	6.46%	3.38%	3.08%	1.54%
Opensources	992	67.24%	14.62%	8.06%	4.23%	2.32%	2.12%	1.41%
Buzzfeed	129	55.81%	17.83%	17.83%	3.88%	2.33%	1.55%	0.78%
Allcott (MIT)	375	57.33%	16.00%	14.67%	3.47%	3.20%	2.93%	2.40%
All	2300	67.70%	12.22%	10.96%	3.83%	2.22%	2.09%	1.00%

Status of fake news domains

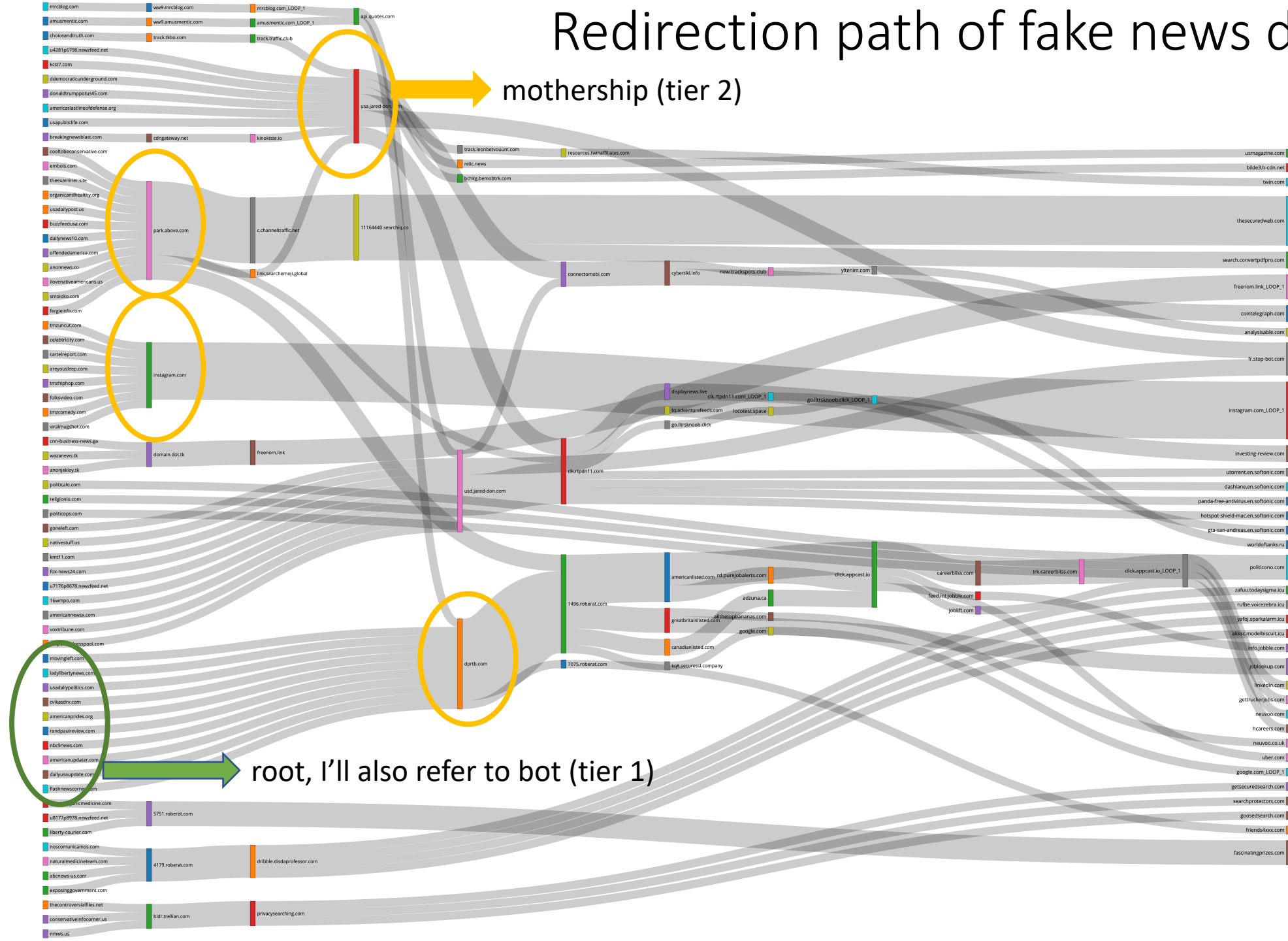
1. Less than 70% domains are normal
2. **Redirection** is the most common abnormal behavior
3. No IP, Server error, For sale indicate the dynamic nature of the field



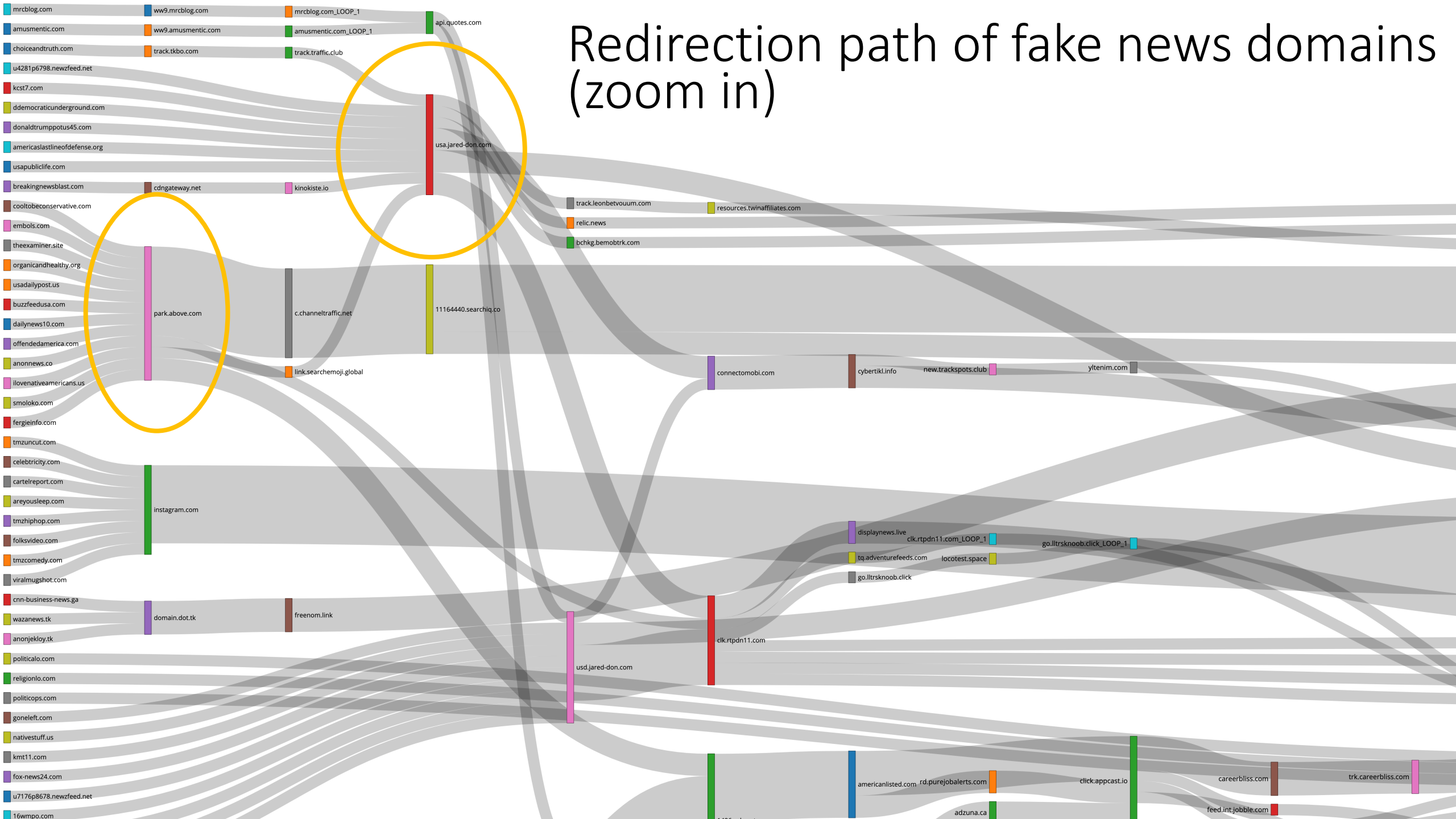
How to discover the initial seed of suspicious domains?

1. Convert redirection paths to tree structures, where roots are fake news domains, and leaves are final landing domains
2. Identify *mothership* domains that connect to multiple root domains
3. In another word, identify nodes with high in-degree and high out-degree
4. Extract root domains connected to nodes from #3

Redirection path of fake news domains



Redirection path of fake news domains (zoom in)



Summary: a list of seed domains connected to malware campaigns

Last resolved IP: 9:22PM, March 26, 2020, Eastern Time

Intermediate domain	# related fake news domains	IP	Reverse DNS
park.above.com	12	103.224.212.241	lb-212-241.above.com
dprtb.com	10	209.15.13.136	Not found
usd.jared-don.com	9	52.207.32.96	ec2-52-207-32-96.compute-1.amazonaws.com
usa.jared-don.com	6	100.24.94.176	ec2-100-24-94-176.compute-1.amazonaws.com
4179.roberat.com	4	198.54.112.216	Not found
domain.dot.tk	3	88.198.252.121	static.88-198-252-121.clients.your-server.de
5751.roberat.com	3	198.54.112.216	Not found
bidr.trellian.com	3	103.224.182.206	bidr.trellian.com

Roadmap

1. Method
2. Challenge and crawling architecture
3. Entry point (seed suspicious domains) analysis
- 4. Results: Discovered redirection campaigns**
5. Results: Fast flux evidence
6. Results: Malicious chrome extension analysis
7. Conclusion

Overview of three discovered redirection campaigns

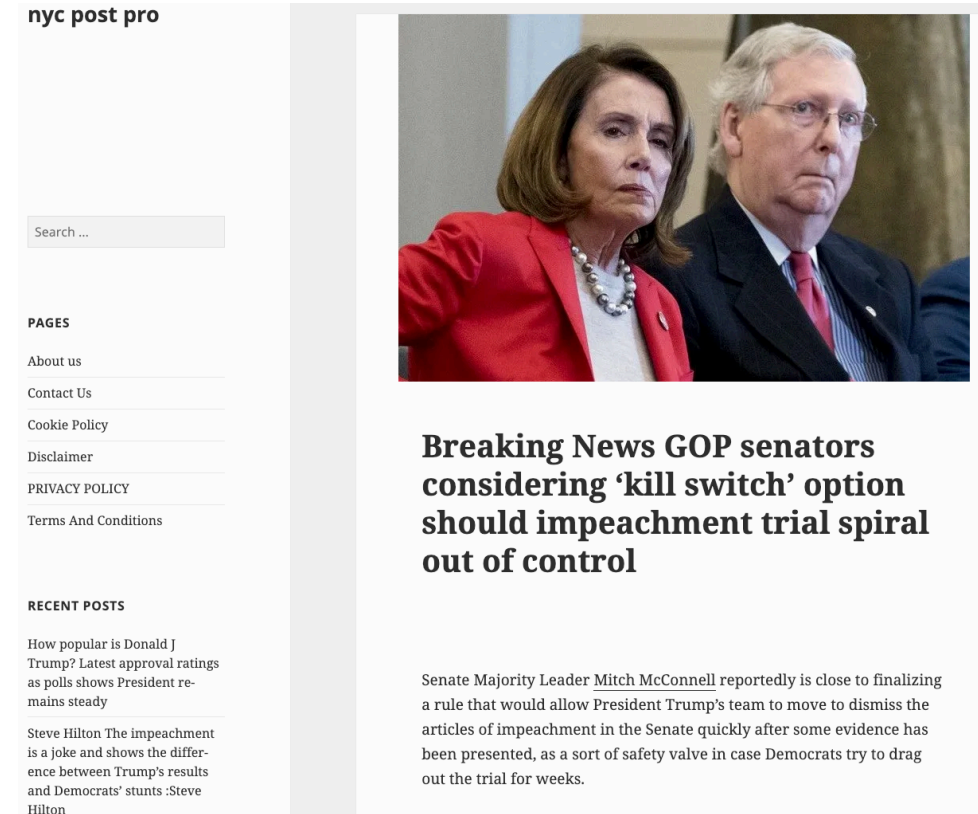
campaign index	example domain	number of domains	network resiliency	require click	cloaking	fast flux
1	nycpost.pro	30+	low	yes	no	no
2	cnnews3.com	700+	high	no	yes	no
3	16wpsm.com	4500+	very high	no	yes	yes

Campaign type 1: click and redirect

1. Seed domain is *nycpost.pro*

2. There are 1000+ other domains hosted on the same IP, 30+ are malicious

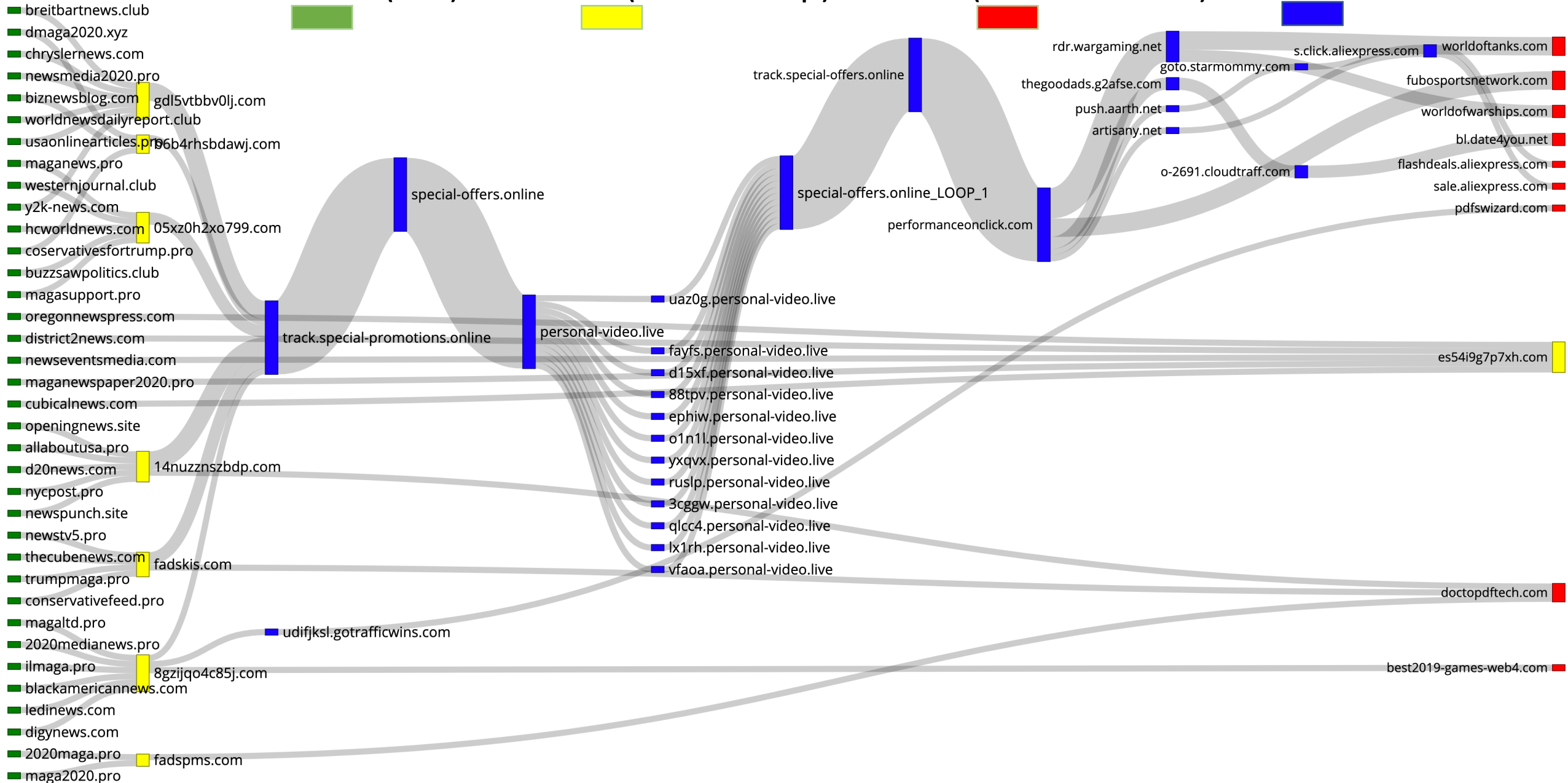
3. Collect all redirection paths and visualize them using Sankey diagrams



Screenshot of nycpost.pro, when a user clicks an article, he/she will be redirected

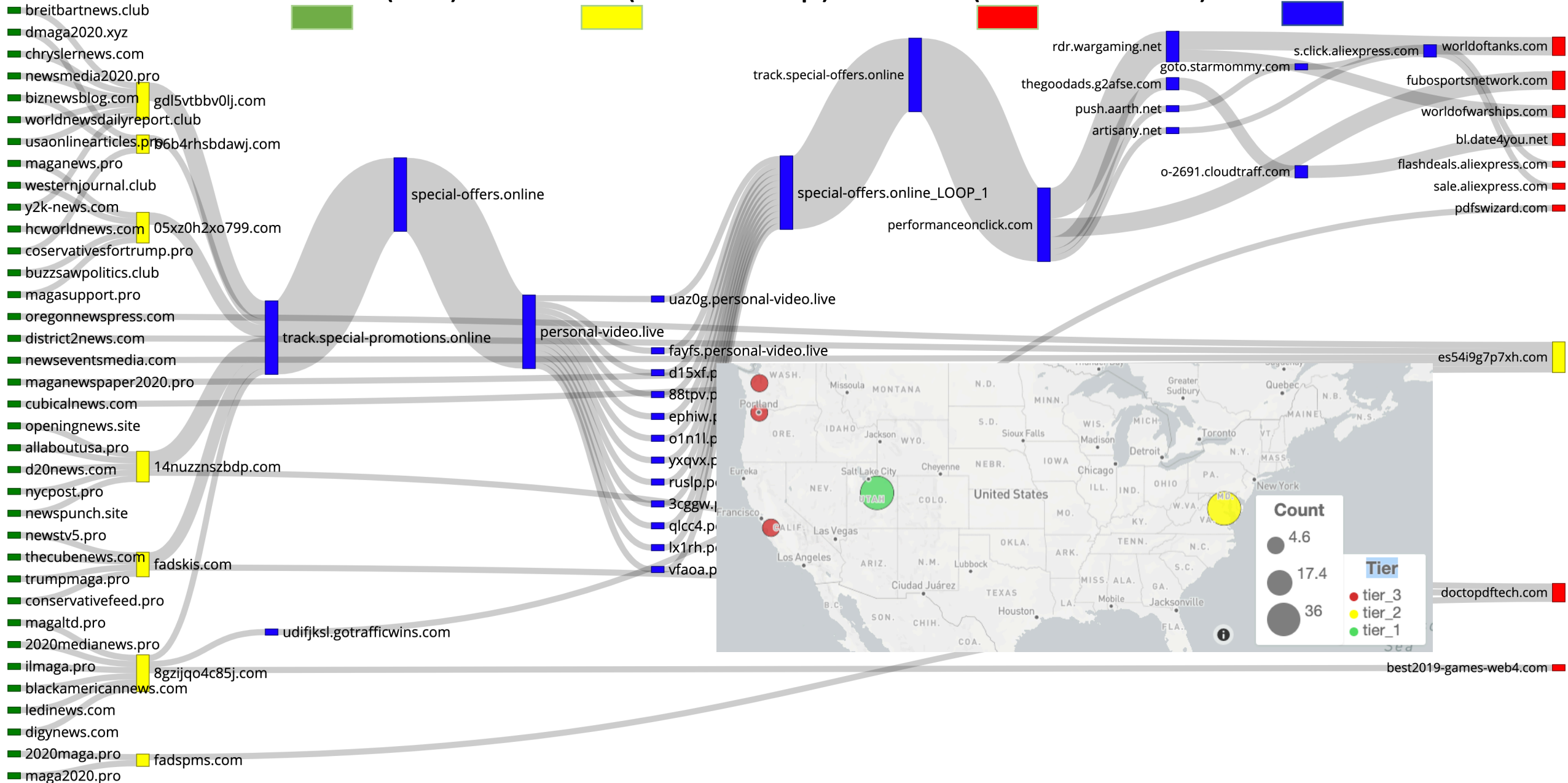
Visualizing the redirection path

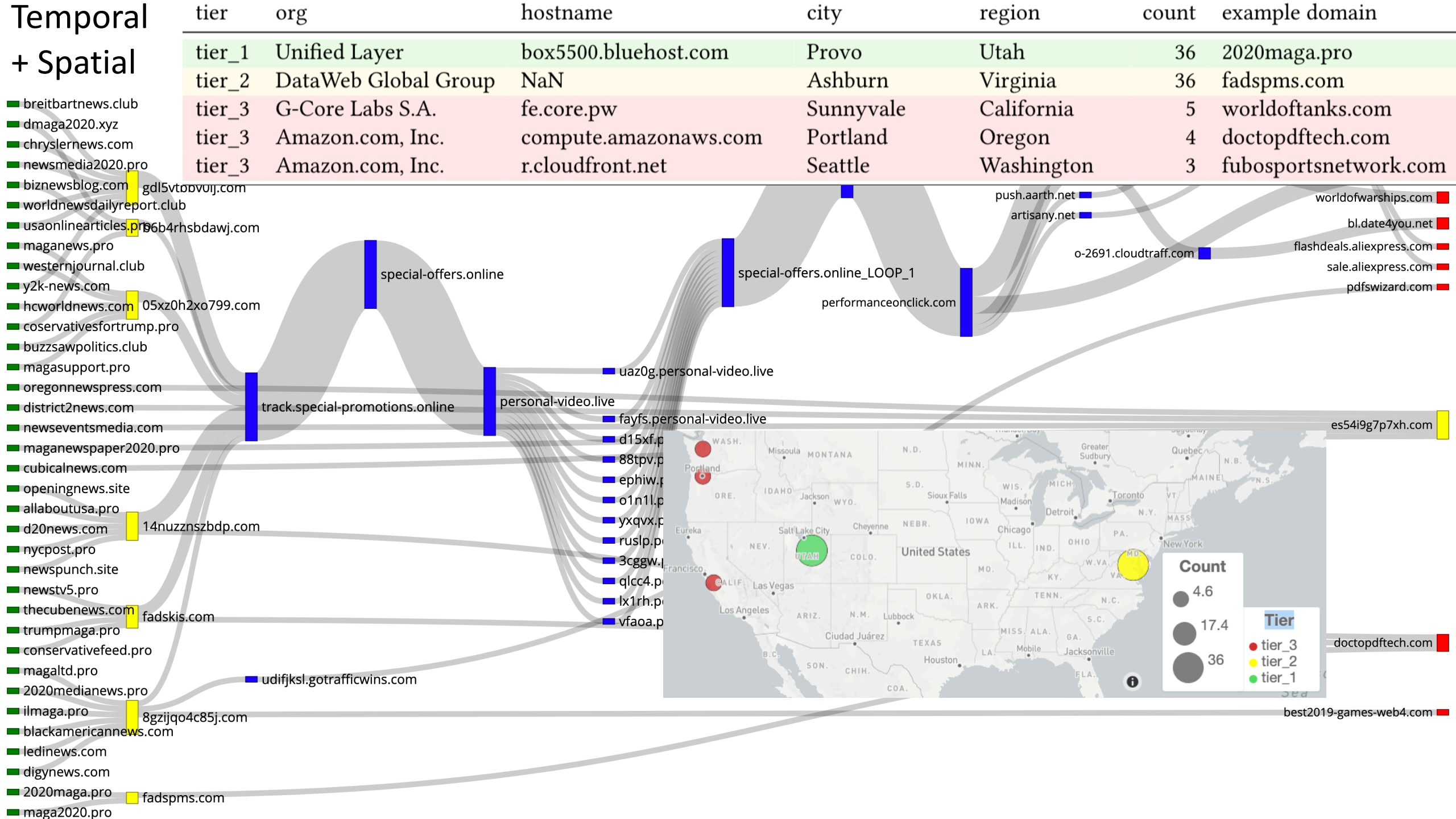
Color code: tier 1 (bot) tier 2 (mothership) tier 3 (final malware) other



Visualizing the redirection path + IP location

Color code: tier 1 (bot) tier 2 (mothership) tier 3 (final malware) other



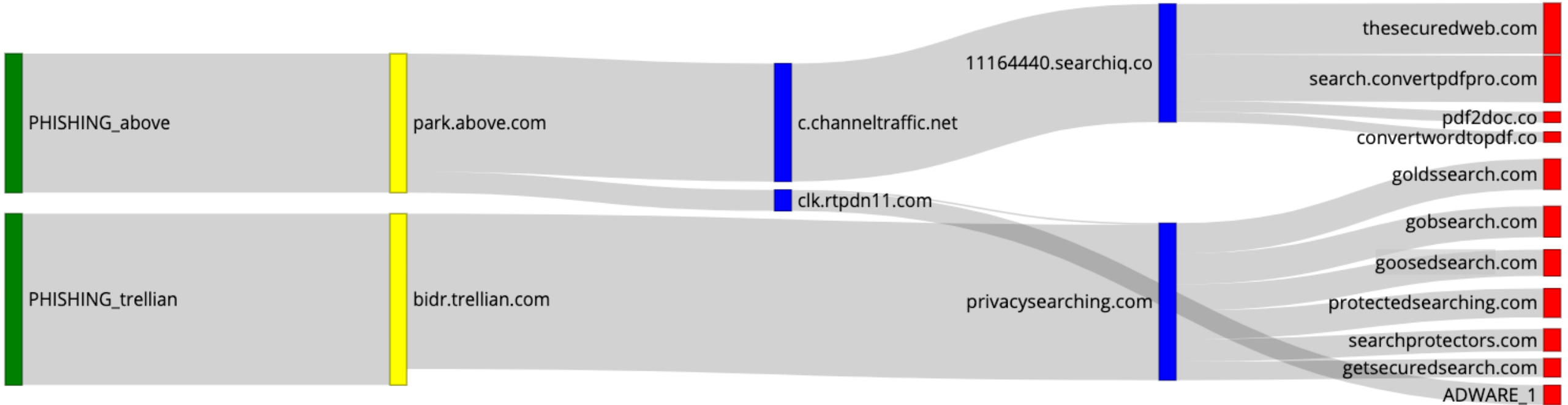


IP analysis

1. All tier 1 domains are hosted on 162.241.217.177 (bluehost.com)
2. Tier 2 domains are hosted on a potential bullet-proof hosting provider (DataWeb Global)
3. Tier 3 domains are hosted on large content distribution networks (AWS, Cloudfront)



Campaign type 2: cloaking



An evasive campaign of 750+ domains, aggregated redirection path

PHISHING_XXX refers to all domains that redirect to XXX



tier	org	hostname	city	region	count	example domain	country
tier_1	Trellian Pty. Limited	lb-182-207.above.com	Beaumaris	Victoria	760	reportexample.com	AU
tier_2	Trellian Pty. Limited	lb-212-241.above.com	Beaumaris	Victoria	721	park.above.com	AU
tier_2	Trellian Pty. Limited	bidr.trellian.com	Beaumaris	Victoria	39	bidr.trellian.com	AU
tier_3	Google LLC	bc.googleusercontent.com	Mountain View	California	47	utorrent.en.softonic.com	US
tier_3	Cloudflare, Inc.	NaN	San Francisco	California	287	search.convertpdfpro.com	US
tier_3	DigitalOcean, LLC	NaN	Clifton	New Jersey	335	goldsearch.com	US
tier_3	Amazon.com, Inc.	compute.amazonaws.com	Portland	Oregon	41	getsecuredsearch.com	US

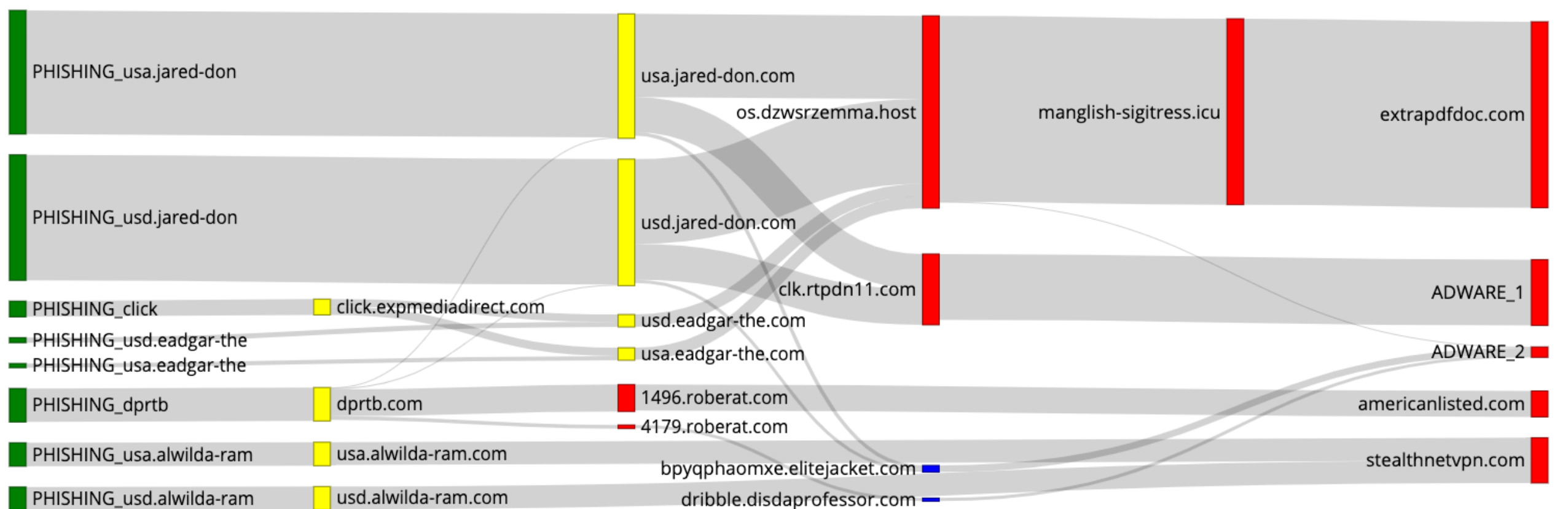
Count refers to the total number of redirection paths that go through this host

Tier 1 and 2 domains are hosted in Australia

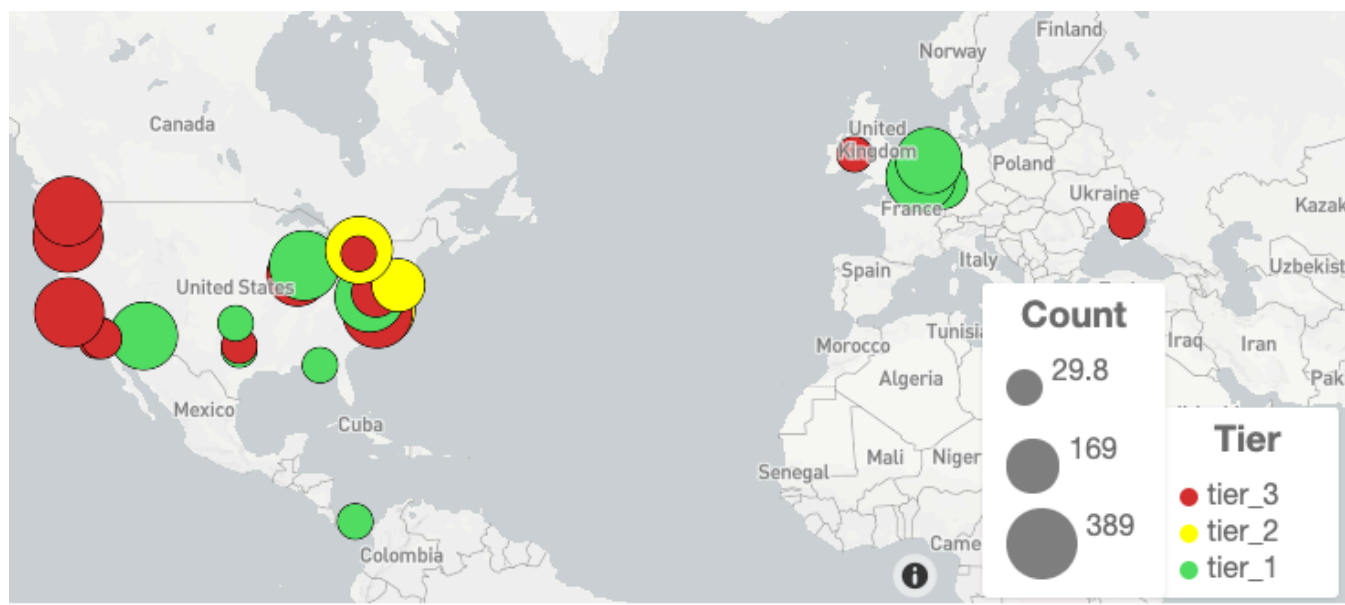
Tier 3 domains are hosted on Google, AWS, Digital Ocean



Campaign type 3: cloaking + fast flux

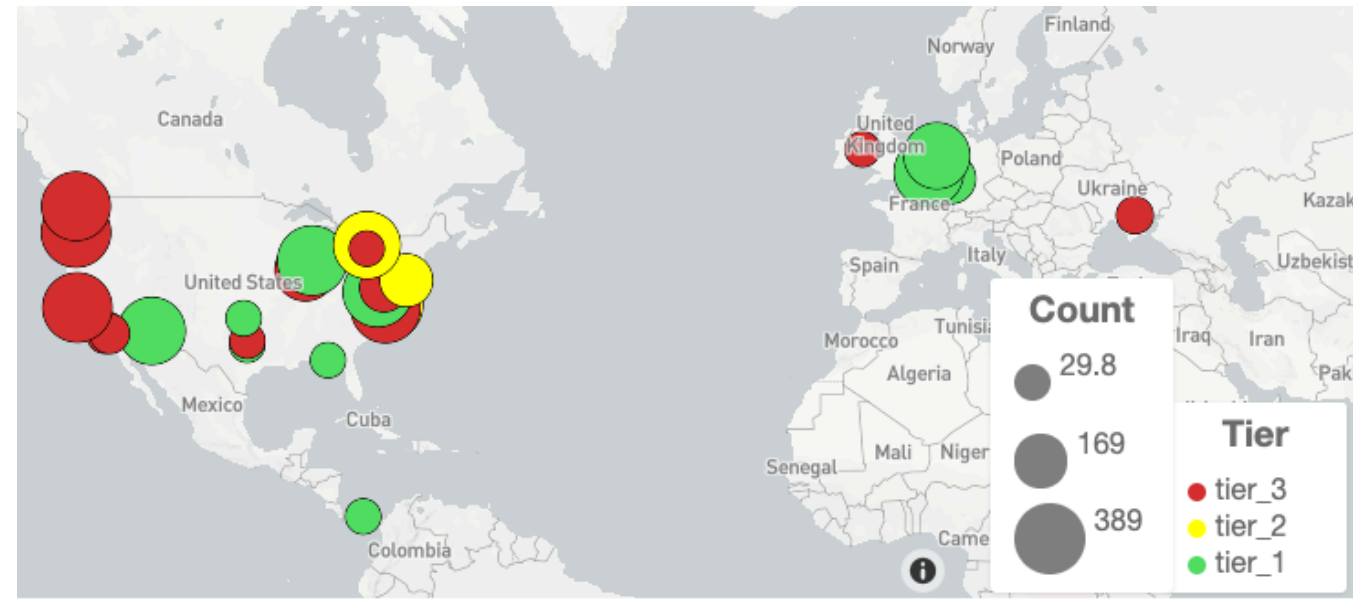


Evasive networks of 4000+ domains
from multiple IP addresses,
aggregated view

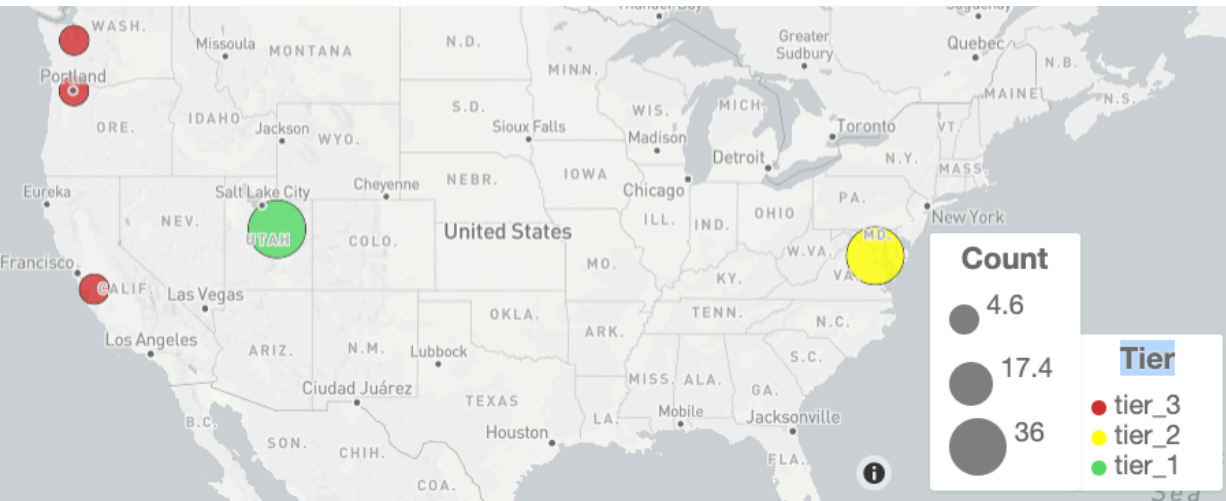


tier	org	hostname	city	region	count	domain	country
tier_1	LeaseWeb Netherlands	NaN	Alkmaar	North Holland	346	wikramahotspot.net	NL
tier_1	NForce Entertainment	NaN	Roosendaal	North Brabant	395	porkybeauties.com	NL
tier_1	Sharktech	customer.sharktech.net	Chicago	Illinois	1159	instagramchief.com	US
tier_1	Leaseweb USA, Inc.	NaN	Manassas	Virginia	1398	prettyteenpictures.com	US
tier_1	Host Europe GmbH	NaN	Scottsdale	Arizona	361	milve.com	US
tier_2	Amazon.com, Inc.	compute-1.amazonaws.com	Virginia Beach	Virginia	3276	usa.jared-don.com	US
tier_2	Aptum Technologies	NaN	Toronto	Ontario	357	dprtb.com	CA
tier_3	Google LLC	googleusercontent.com	Mountain View	California	723	tik-tok.en.softonic.com	US
tier_3	Amazon.com, Inc.	cloudfront.net	Seattle	Washington	380	extrapdfdoc.com	US
tier_3	Amazon.com, Inc.	compute-1.amazonaws.com	Virginia Beach	Virginia	612	usa.jared-don.com	US
tier_3	Amazon.com, Inc.	compute.amazonaws.com	Portland	Oregon	1235	extrapdfdoc.com	US

Tier 1 domains are hosted globally
Tier 2 domains are mostly hosted on AWS
Tier 3 domains are hosted on Google, AWS



IP geo-spatial distribution, comparison



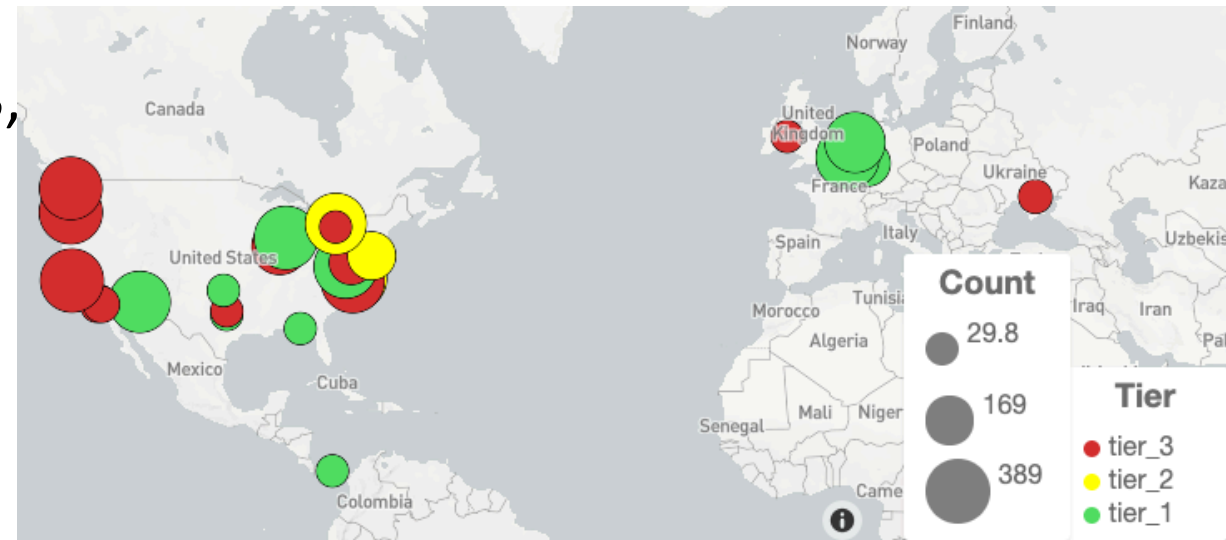
Campaign 1, all IPs on US



Campaign 2, most IPs in US and Australia

NOTE: Campaign 1 is relatively easy to take down, Campaign 2&3 are much harder, as IPs belong to multiple jurisdictions

Campaign 3, multiple IP locations across US and Europe



Roadmap

1. Method
2. Challenge and crawling architecture
3. Entry point (seed suspicious domains) analysis
4. Results: Discovered redirection campaigns
5. **Results: Fast flux evidence**
6. Results: Malicious chrome extension analysis
7. Conclusion

Evidence of fast flux

Tier 2 domain: jared-don.com

IP history source: <https://viewdns.info/iphistory/?domain=jared-don.com>

IP Address	Location	IP Address Owner	Last seen on this IP
54.84.174.180	Ashburn - United States	Amazon Technologies Inc.	3/26/20
52.71.209.190	Ashburn - United States	Amazon Technologies Inc.	3/26/20
52.4.32.92	Ashburn - United States	Amazon Technologies Inc.	3/26/20
52.207.32.96	Ashburn - United States	Amazon Technologies Inc.	3/26/20
52.202.53.245	Ashburn - United States	Amazon Technologies Inc.	3/26/20
35.169.74.130	Ashburn - United States	Amazon Technologies Inc.	3/26/20
35.168.147.213	Ashburn - United States	Amazon Technologies Inc.	3/26/20
3.225.81.82	Ashburn - United States	Amazon Data Services NoVa	3/26/20

Evidence of fast flux

Tier 1 domain: 16wmpo.com

IP history source:
<https://viewdns.info/iphistory/?domain=16wmpo.com>

The fast flux frequency is higher in reality: the DNS record keeps changing, and the IP changes accordingly

IP Address	Location	IP Address Owner	Last seen on this IP
64.32.8.68	Chicago	Sharktech	3/26/20
64.32.8.67	Chicago	Sharktech	3/25/20
46.166.182.110	Netherlands	Serverhosting	3/25/20
37.48.65.148	Netherlands	LEASEWEB	3/22/20
46.166.182.113	Netherlands	Serverhosting	3/21/20
	Manassas -		
207.244.67.215	United States	Leaseweb USA, Inc.	3/20/20
64.32.8.68	Chicago	Sharktech	3/19/20
	Manassas -		
207.244.67.216	United States	Leaseweb USA, Inc.	3/19/20
64.32.8.67	Chicago	Sharktech	3/18/20
46.166.182.114	Netherlands	Serverhosting	3/17/20
46.166.182.115	Netherlands	Serverhosting	3/16/20
64.32.8.69	Chicago	Sharktech	3/15/20
64.32.8.67	Chicago	Sharktech	3/14/20
46.166.182.111	Netherlands	Serverhosting	3/14/20
46.166.182.110	Netherlands	Serverhosting	3/14/20
37.48.65.136	Netherlands	LEASEWEB	3/14/20

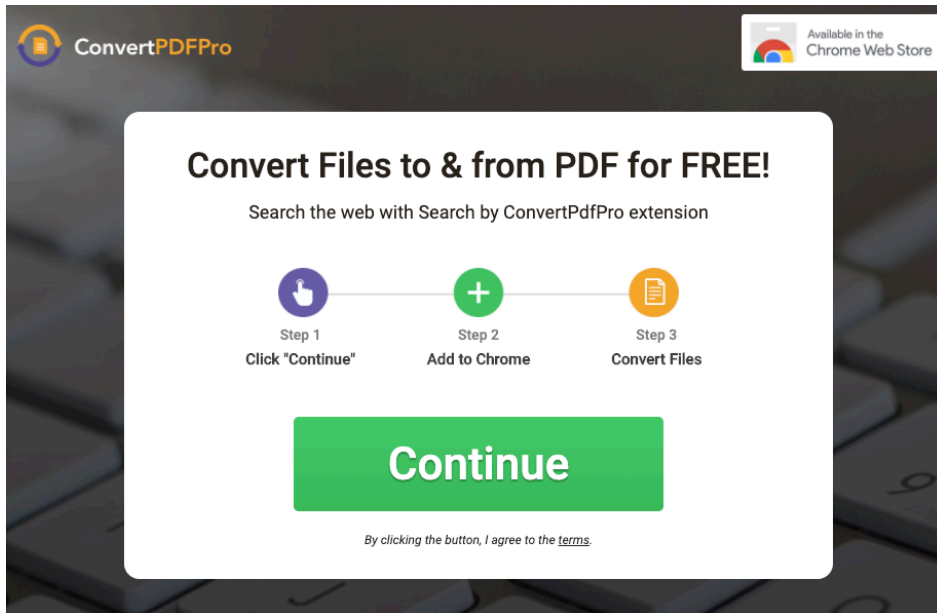
Roadmap

1. Method
2. Challenge and crawling architecture
3. Entry point (seed suspicious domains) analysis
4. Results: Discovered redirection campaigns
5. Results: Fast flux evidence
- 6. Results: Malicious chrome extension analysis**
7. Conclusion

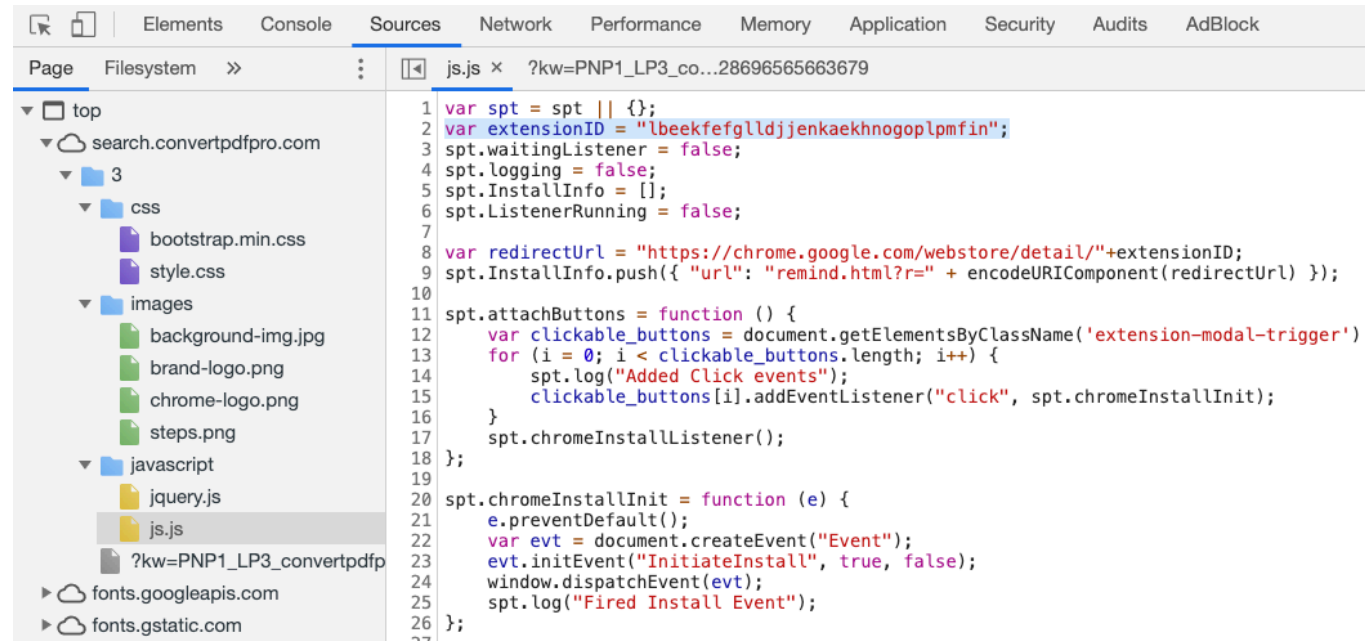
Evidence of malicious downloads and Chrome
extensions from final redirected URLs
-- Three detailed examples

Evidence of adware/malware, example 1

1. Entry point: <http://16wmpo.com/>
2. Redirected to:
https://search.convertpdfpro.com/3/?kw=PNP1_LP3_convertpdfprosearch&sid=11165151&said=16wmpocom&clickid=119277450507494203442931728696565663679



Screenshot of redirected page



Click "Continue" will trigger a javascript function that redirect users to Chrome Web Store

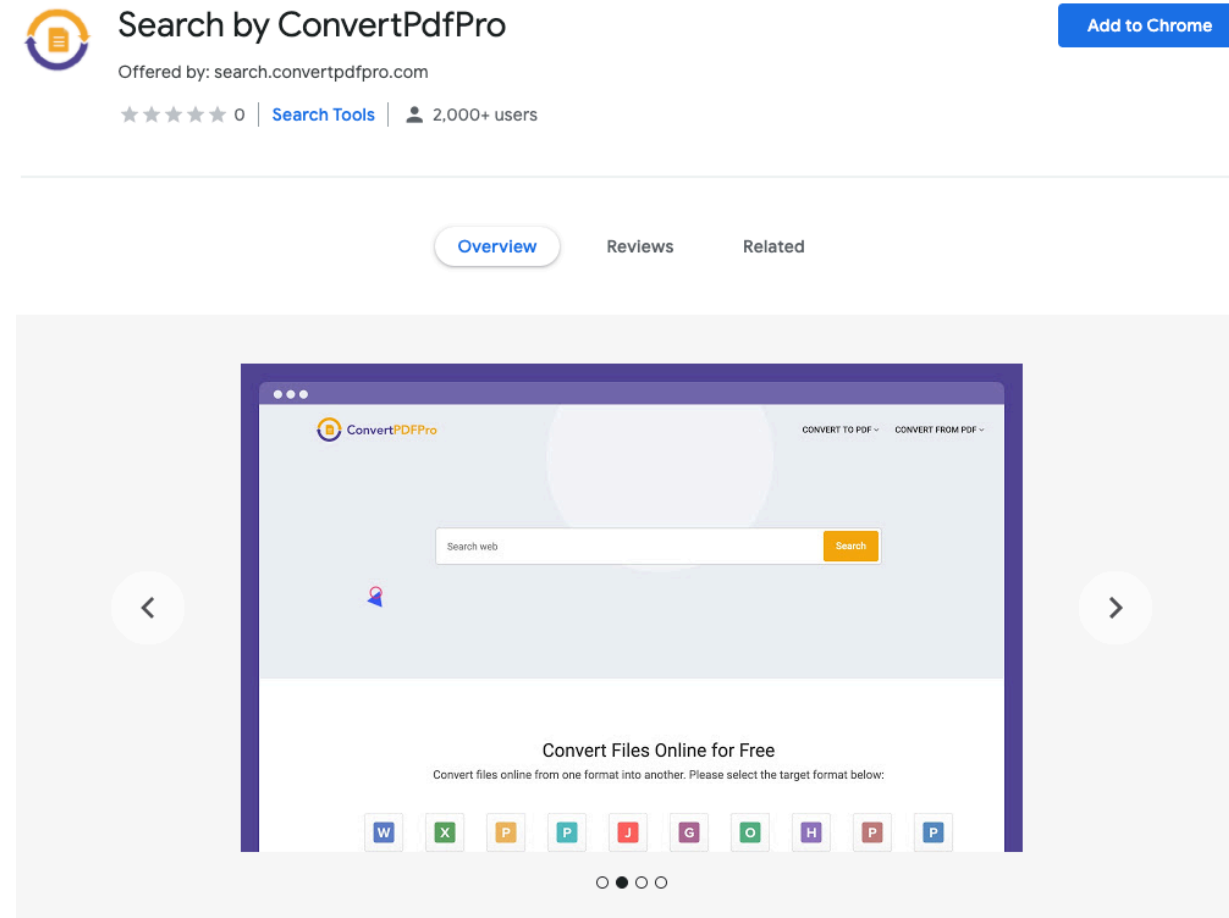
Evidence of adware/malware, example 1

3. Screenshot of redirected Chrome extension Screenshot



4. Download the extension source code

Red flag: extension overwrites default chrome search url



```
"permissions": ["tabs"],
"chrome_settings_overrides": {
  "search_provider": {
    "name": "Web",
    "keyword": "search",
    "search_url": "http://search.convertpdfpro.com/search.html?q={searchTerms}&s",
    "favicon_url": "http://search.convertpdfpro.com/assets/img/convertpdfpro.ico",
    "suggest_url": "http://api.convertpdfpro.com/api/search/autosuggestions?keyw",
    "encoding": "UTF-8",
    "is_default": true
  }
},
```

Evidence of adware/malware, example 2

1. Entry point: <http://realcodes.us/>
2. Redirected to: https://goosedsearch.com/lander?d=&utm_campaign=fe7a3c357ec38c8afce282eb17c010a75054a150
3. Extract Chrome extension URL from final landing page HTML

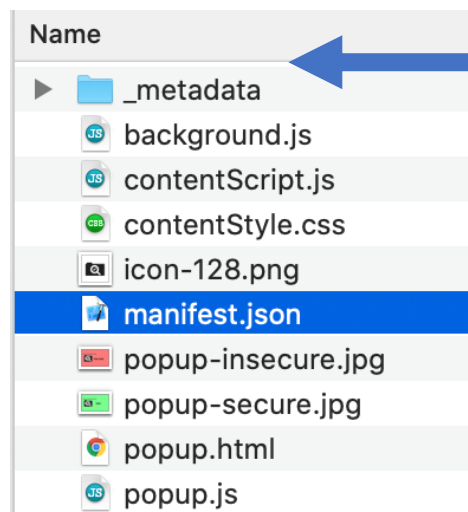
Security Check

ough a standard security check. You will also be redirected to the Chrome Store and directly to your destination. This extension will offer you a safer web search experier provider.

CONTINUE

```
<script>
  window.tid = '2';
  window.appId = '22';
  window.storeUrl = 'https://chrome.google.com/webstore/detail/
    safe-web-searching/hmmnhahdacolomjankkcljjocpaohkbj';
  window.fallbackUrl = 'https://getsecuredsearch.com/1501738800';
  window.fbm = '0';
</script>
<script src="/js/lander.js"></script>
```

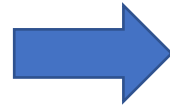
4. Download the extension source code



```
"chrome_settings_overrides": {
  "search_provider": {
    "name": "Safe Web Searching",
    "keyword": "sws",
    "search_url": "https:\\\\goosedsearch.com\\search-bing?q={searchTerms}&appId=22",
    "favicon_url": "https:\\\\goosedsearch.com\\favicon.png",
    "suggest_url": "https:\\\\goosedsearch.com\\suggest.php?q={searchTerms}",
    "encoding": "UTF-8",
    "is_default": true
  }
},
```

Evidence of adware/malware, example 3

1. Entry point: <http://rotthq.com/>
2. Redirected to: <http://fqgay.rubyinvest.icu/hyllkjit/n3w1p4csb/?n=1587470204>



Engine	Detection	Engine	Detection
Arcabit	Adware.MAC.Bundle.EGK	Avast	MacOS.Agent-FJ [Adw]
AVG	MacOS.Agent-FJ [Adw]	BitDefender	Adware.MAC.Bundle.EGK
Emsisoft	Adware.MAC.Bundle.EGK (B)	eScan	Adware.MAC.Bundle.EGK
ESET-NOD32	OSX/Adware.Bundle.DE	FireEye	Adware.MAC.Bundle.EGK
GData	Adware.MAC.Bundle.EGK	Kaspersky	Not-a-virus:HEUR:AdWare.OSX.Bnodler..

3. Download the installer

4. Scan for malicious code at Virustotal <https://www.virustotal.com/gui/file-analysis/YWYzOTI0Y2M2YjM3NmNmYTlmNzczOTMwZmMzOTg3OTQ6MTU4NjEwMDM0OQ==/detection>

Summary Table: Coordinated groups of Chrome Extensions, group one and two

Chrome Extension Name	Extension ID	Permissions	Overwrite Search URL
Group one			
Securify for Chrome™	pcfapghfanllmbdfii peiihpkojekckk	['<all_urls>', 'contextMenus', 'tabs', 'storage', 'cookies', 'webRequest', 'notifications', 'idle']	https://search.withsecurify.com/?dwfy&yh&q={searchTerms}
Securify for Chrome Desktop	dmakkciciccnjgmfjfl pbdfkdnmpfghp	['<all_urls>', 'contextMenus', 'tabs', 'storage', 'cookies', 'webRequest', 'notifications', 'idle']	https://search.withsecurify.com/?dwfy&yh&q={searchTerms}
Group two			
PDF Converter	pokhhkbhifimfkegn endnjkeickbckbf	["*://*.pdfsrch.com/*", "*://*.pdfswizard.com/*", "*://*.apiprxy.com/*", "cookies", "tabs", "webRequest", "webRequestBlocking", "contextMenus"]	https://pdfsrch.com/?q={searchTerms}
EasyConvert	ojmoedcpcgeminijl nogdmkelkcfafl	['*://*.srchbar.com/*', '*://*.doctopdftech.com/*', '*://*.apiprxy.com/*', 'cookies', 'tabs', 'webRequest', 'webRequestBlocking', 'contextMenus']	https://srchbar.com/?q={searchTerms}

Summary Table: Coordinated groups of Chrome Extensions, group three

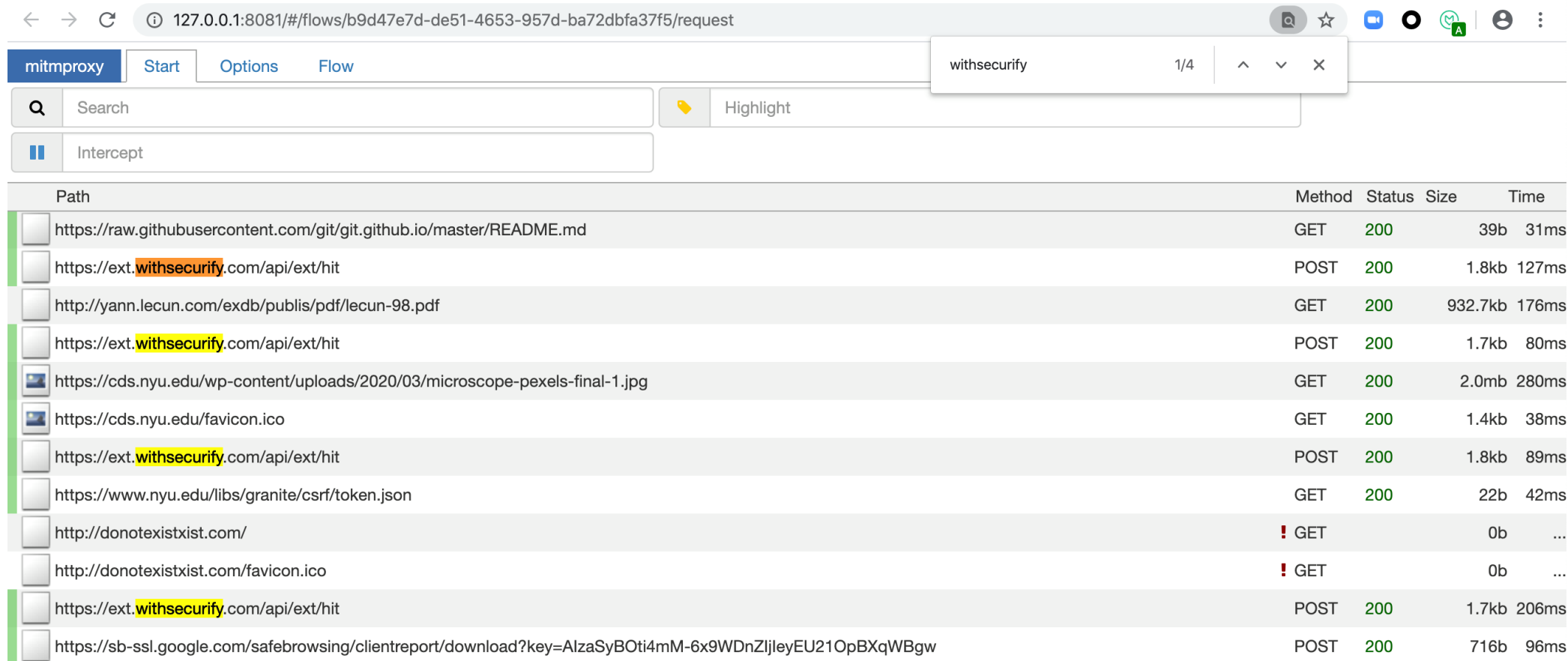
Chrome Extension Name	Extension ID	Permissions	Overwrite Search URL
Group three			
Secure Web Searching	dopmojabcdlfbnppmje aajclohofnbol	['tabs', 'webRequest', 'webRequestBlocking', 'http://*/', 'https://*/']	https://goldsearch.com/search- bing?q={searchTerms}&appId=27&src=bar
Safe Browsing Checker	hpcnoikpdmeemodpjii clgobemmbccmj		https://goodbyesearch.com/search- bing?q={searchTerms}&appId=20&src=bar
Secure Web Surfing	enmjojmecjhakabinfc bmkbcpdbgijh		https://protectedsearching.com/search- bing?q={searchTerms}&appId=25&src=bar
Web Searching Security	pkemkgkekbelcohkcbj pcepeogcagehl		https://goshsearch.com/search- bing?q={searchTerms}&appId=37&src=bar
Web Security Checker	enfgmdnkelcpecofafaa ingdocmknanl		https://browsingsecurityhub.com/search- bing?q={searchTerms}&appId=15&src=bar
Browse Safer	deiiiklocnibjflinkfmefp ofgcfhdga		https://searchprotectors.com/search- bing?q={searchTerms}&appId=31&src=bar
Safe Web Searching	hmmnhahdacolomjan kkcljjocpaohkbj		https://goosedsearch.com/search- bing?q={searchTerms}&appId=22&src=bar
Browsing Protector	npdfkclmbnoklkdebjfo dpendkepbjek		https://gobsearch.com/search- bing?q={searchTerms}&appId=33&src=bar
Browsing Safety Checker	dopkmmcoegcjggfanaj nindneiffpck		https://websitesecuritygroup.com/search- bing?q={searchTerms}&appId=16&src=bar

Steps to reproduce suspicious extension behavior

1. Environment: MacOS Catalina V10.15.2, Chrome Version 80.0.3987.149
2. Download extension **Securify for Chrome Desktop** <https://chrome.google.com/webstore/detail/securify-for-chrome-deskt/dmakkciciccnjgmfjflpbdfkdnmpfghp>, (version 1.5.43), enable extension
3. Use mitmproxy to intercept all http*, https* traffic
4. **Every time a user visits an URL, a POST request is sent to endpoint withsecurify.com, the payload is obfuscated**

Right: A suspicious POST request is sent after a GET

Other extension families will probably have different behaviors



Path	Method	Status	Size	Time
https://raw.githubusercontent.com/git/git.github.io/master/README.md	GET	200	39b	31ms
https://ext.withsecurify.com/api/ext/hit	POST	200	1.8kb	127ms
http://yann.lecun.com/exdb/publis/pdf/lecun-98.pdf	GET	200	932.7kb	176ms
https://ext.withsecurify.com/api/ext/hit	POST	200	1.7kb	80ms
https://cds.nyu.edu/wp-content/uploads/2020/03/microscope-pexels-final-1.jpg	GET	200	2.0mb	280ms
https://cds.nyu.edu/favicon.ico	GET	200	1.4kb	38ms
https://ext.withsecurify.com/api/ext/hit	POST	200	1.8kb	89ms
https://www.nyu.edu/lib/granite/csrf/token.json	GET	200	22b	42ms
http://donotexistxist.com/	! GET		0b	...
http://donotexistxist.com/favicon.ico	! GET		0b	...
https://ext.withsecurify.com/api/ext/hit	POST	200	1.7kb	206ms
https://sb-ssl.google.com/safebrowsing/clientreport/download?key=AlzaSyBOti4mM-6x9WDnZljleyEU21OpBXqWBgw	POST	200	716b	96ms

Steps to reproduce suspicious extension behavior, continued

5. The payload looks like: hN2Klpvd... loyU3YI=
6. After studying the source code, I reverse engineered the decoding protocol
7. From the plain text, we can see clearly that the extension tracks every web visit, including url, referral, tab number and visit time.

Right: The payload is in base64 format. To get plaintext, we first decode base64 to decimal, then we manipulate the decimal to get the right ascii character.

```
import base64
def decode(payload):
    # decode obfuscated code
    decoded_bytes = base64.urlsafe_b64decode(payload)
    decoded_decimal = [i for i in decoded_bytes]
    decoded_string = ''
    for num in decoded_decimal:
        decoded_string += chr(255 - int(num))
    return decoded_string
```

Right: Part of decoded payload.
The extension tracks every web visits (in the "extra" field)

```
{
  "active_tab_id": 138,
  "local": {
    "language": "en-US",
    "local_time": 1586545534.106,
    "local_timezone": 240
  },
  "extra": {
    "url": "http://beautifytools.com/csv-to-xml-json-converter.php",
    "tabId": 138,
    "ref": "http://beautifytools.com/base64-to-image-converter.php"
  },
  "hid": "68bb0385-d9f9-4429-abc5-37ee966bf547",
  "action": "risk"
}
```

Roadmap

1. Method
2. Challenge and crawling architecture
3. Entry point (seed suspicious domains) analysis
4. Results: Discovered redirection campaigns
5. Results: Fast flux evidence
6. Results: Malicious chrome extension analysis
7. **Conclusion**

Conclusion and next steps

1. There are large-scale, coordinated redirection campaigns to distribute adware/malware Chrome extensions. The entry point domains are not flagged by Safe Browsing
2. I'm currently tracing more malware campaigns, and comparing different temporal and geospatial patterns
3. Mitigation and intervention – need Industry partners' help and collaboration

References

- 1. Cloak of Visibility: Detecting When Machines Browse a Different Web**
(Invernizzi, et al., *2016 IEEE Symposium on Security and Privacy*)
 - From Google safe browsing team
 - Seed collection is top Google search results with suspicious keywords “LV, GUCCI”
- 2. What You See is NOT What You Get: Discovering and Tracking Social Engineering Attack Campaigns** (Vadrevu, et al., *2019 International Measurement Conference*)
 - Seed collection is a list of low-tier ad-publishers
 - Use reverse search to find all websites that embed javascript from those publishers
- 3. EVILSEED: A Guided Approach to Finding Malicious Web Pages** (Invernizzi, et al., *2012 IEEE Symposium on Security and Privacy*)
 - Introduced the concept of gadgets (which are basically similarity measurements)
 - The author used link, content, SEO, domain, and DNS trace similarity measurements