



DIVE DEEPER //

Proactive Ad Fraud Detection

A new source of truth in
a web of lies.



Introduction

Here at DeepSee, we make tools for ad fraud & security researchers. Drawing from our backgrounds as ad agency analysts, web developers, data scientists, and white-hat fraud researchers, we're now building the tools we always wished we'd had to fight fraud.

Our technology at DeepSee catches complex systems of web traffic laundering in order to flag risky domains before you partner with them.

Fraud in Web Advertising

Ad fraud is endemic to the advertising industry, and shows no signs of slowing. At the same time, the on-page & creative-level trackers used to detect fraud are missing a large part of the story. It is estimated that in 2019, ad fraudsters stole between \$6 billion and \$23 billion from advertisers¹. The range is large because some ad fraud is tough to detect, and the technology to protect advertisers is immature.

Ad fraudsters are able to steal this amount of money with fake views, users, traffic, clicks and installations. These methods all have the same goal: make advertisers pay for events that never happen.

At DeepSee, we go beyond impression or click tracking to detect these fraudulent sites before you partner with them.

How Do Advertisers Protect Themselves?

Advertisers and agencies spend a significant amount of time and resources analyzing the performance of their campaigns. These campaigns are judged on a variety of criterias, including:

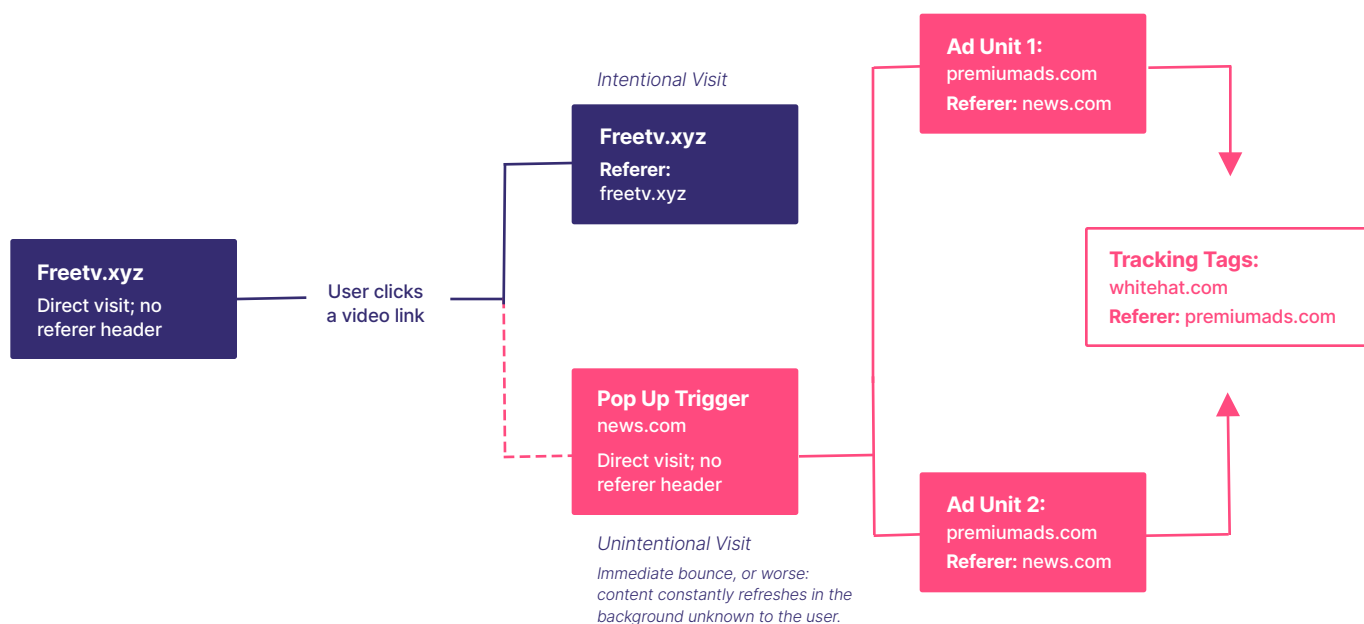
- Pay-per-click (CPC) campaigns, measured by click-through-rate
- Pay-per-visit (CPV) campaigns, measured by targeted user groups visiting a certain site/page.
- On-site views of internal calls-to-action like 'Buy Now' or 'Click Here' (CPA)
- Pay-per-Impression (CPM) campaigns, which pay publishers a certain price per-thousand times an ad appears

To better understand these events, advertisers attach tracking tags (3rd party scripts) to their advertisements, or to their landing pages. These scripts attempt to scrape all the publicly facing info that is exposed when a user visits a site. This information may include a user's IP, browser, screen size, plugins, cookies and more. When employed by white-hat fraud researchers, these trackers aim to gather sufficient data to make a probabilistic judgment about if a user is real.

[1] <https://www.emarketer.com/content/digital-ad-fraud-2019>

However, answering the question of whether or not an agent is a human or a bot is not sufficient to determine if fraud is being committed. Tracking tags only gets a shallow understanding of web traffic, and they leave analysts with a poor understanding of how a user arrives at a site. This is because impression trackers can only detect the URL of the container they are loaded in, or the website directly above them in the site tree.

Pop Up Obfuscation



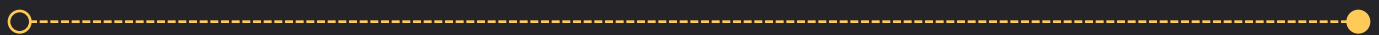
These trackers cannot detect sites loading as pop-ups, pages embedded deep within shady websites, or internal referrals that make site traffic appear to be direct visits. Ad impression/click tracking is the most popular method used to detect fraud, even though their shortcomings are well-known. Additionally, discrepancies between measurement vendors can cause significant headaches for the agency when it comes time to get paid. When this happens, it's the analysts the agency employs who are now scrambling to reconcile the judgements of these measurement vendors.

Having been down this road too many times, we recognize the need for a ground truth data-set which can reveal the types of connections that are out-of-bounds for tag based tracking. That's why we built it.

A New Tool in the Fight to Detect Threats

DeepSee is a SaaS system that augments an advertiser's security and intelligence technology stack by giving web analysts unparalleled insight into how websites and interconnected groups of websites behave. Our software employs machine learning models based on our robust historical data to always discover the next scheme. By targeting suspicious sites, we can reveal schemes that would otherwise go undetected, allowing users to make better data-driven decisions.

Using DeepSee, users can access in-depth reports on any domain, and pull large lists of domains using our advanced search feature. This insight allows users to make more informed decisions.



DeepSee can be used to:

- Build site lists for ad campaign targeting and avoidance
- Research inbound and outbound traffic for irregularities
- Research Competitors
- Investigate Security
- Probe the quality of publisher networks for SSPs looking to improve their onboarding practices

No Private Data is Used in the Making of our Products

By scraping a combination of request and response headers, cookies, and other information available in your browser, click and impression trackers collect a massive amount of information on individuals. This flippant collection of personally identifying information (PII) is a huge concern to citizens worldwide, and governing bodies are starting to notice. The European Union has already taken aggressive steps to protect its citizens' privacy, in the form the General Data Protection Regulation (GDPR), and that's just the beginning. Advertising solutions built on ingesting huge amounts of PII are simply not future proof, and face increasing risk to their business.

Unlike click or impression trackers, DeepSee's analytics are not predicated on collecting large amounts of data from web users. DeepSee looks at the nature of websites and rings of websites to discover places that harbor fraud. Where on-page trackers capture a point in time, DeepSee captures the journey, end-to-end, that a user might take.

Combining deep learning with domain expertise, we intelligently interrogate sites that produce exploitative behaviors. Using the data we acquire through this process, we help web analysts get the complete picture of website interactions and behaviors.

Augment Your Security Stack with DeepSee

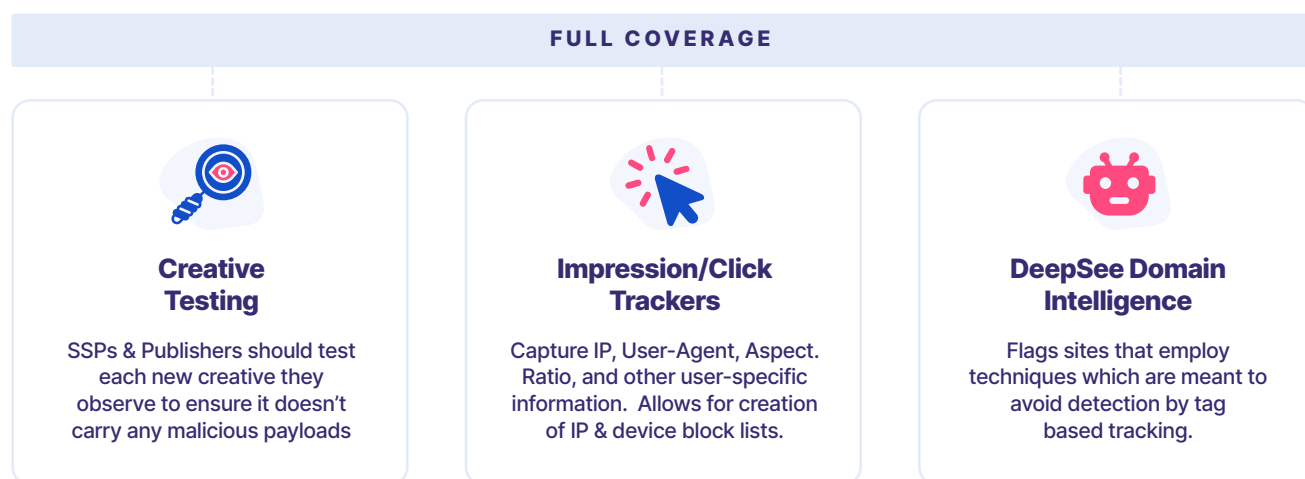
DeepSee fills in the much-exploited knowledge gaps left by tag based tracking technology, and traces the threads between risky domains and those who benefit from their bad behavior.

We see impression trackers can benefit from DeepSee's data because our knowledge alerts analysts to the ways that users arrive at a given website. By using this information in concert with data from tag based trackers, analysts and researchers can strengthen their probabilistic systems that assign risk to certain page visits and ad-views.

Even if you detect fraud with a tag based solution, the process of getting a refund can be difficult. If you know who is likely to commit fraud beforehand, you can avoid them entirely.

Best Practices for a Protected Advertising Supply Chain

Trifecta of Coverage for Programmatic Display/Video



This diagram shows an ideal situation wherein the entire supply chain is protected from fraud. SSPs protect buyers by ensuring their publishers meet our rigorous standard, the buyers verify ad-events don't come from known bots/datacenters, and the individual ads themselves are tested to ensure they don't hijack the user's browser.

DeepSee Complements Existing Ad Tracking Tech

DeepSee is complementary to existing tracker technologies. This is because DeepSee and ad trackers serve important, distinct functions.

Impression & page-level trackers get user information like cookie ids and IP addresses. This information can be beneficial; however, bad actors design schemes to manipulate traffic and con these tracking techniques.

DeepSee detects sites that are connected by fraudulent methods of laundering traffic. Impression trackers could only see this laundering if trackers were present on every page of the internet, including on malicious sites.

Since fraudsters won't be sharing their misuse of these technologies with partners, it is up to buyers to use DeepSee to investigate suspicious activity and avoid shady sellers.

5 Common Publisher Behaviors Plaguing Advertisers & Users

While the number of threats out there are countless, we have identified core behaviors of websites that extract value from the ecosystem. Each one relates to a specific way that ad measurement can be duped, or that a user's experience on a web page can be negatively affected. We expect these behaviors to be a concern for web users for years to come (as they have already been for years prior to this release).



POP UPS

Sites with a high Pop Up Risk produce suspicious pop-ups or load as pop-ups from shady sources. These domains pose a great risk to advertisers who place ads on the popped-up page, and they annoy users whose experience becomes suddenly interrupted.



REDIRECTORS

These sites force you to visit another page when you visit them. They sometimes force users through several URLs before placing them at their destination. This behavior can be triggered by an interaction with the page, or automatically as a "zero-click" redirect.



EMBEDDED PAGES

Sites with high embedded risk have untrustworthy referral values. These embedded pages load underneath the content of another page. This page or ad hiding behavior takes advantage of the limitations of on-page trackers and creative-level trackers, as visits to these sites may falsely appear as direct visits or impressions.



RESOURCE HOGS

Resource hogs have a high demand for your computer's resources for the duration of your visit. These sites run scripts that require lots of communication between the client and other web servers and have elements that refresh frequently. Often, sites with many different programmatic advertising partners are resource hogs. These sites may run slowly after a typical loading time. These sites may prompt users to block ads entirely.



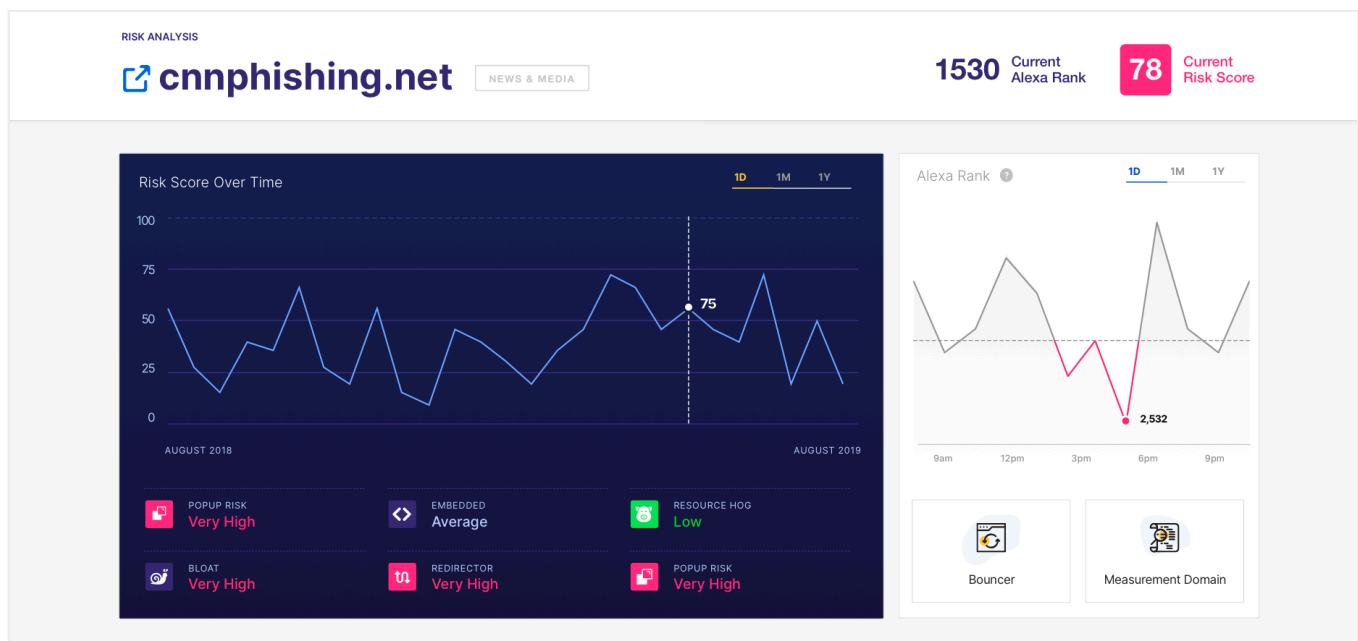
BLOAT

Bloated sites have a sizable initial loading time. This is due to the site containing many different images or containers. As opposed to Resource Hog, these sites may speed up after the initial page load. Bloat can cause trackers to fire in a way which causes inconsistencies when comparing measurement between multiple vendors.

What We Offer Our Customers

Deep Dive Into Any Site

Input a single domain of interest and get a multitude of information about that domain. Below we see a fraction of the information given in a single site view of cnnpishing.net.



The following information is included for every site:

A **0–100 Risk Score** as a 30-day time series: An easy to digest 0–100 score that reflects the risk of associating with any given domain, displayed for the last month.

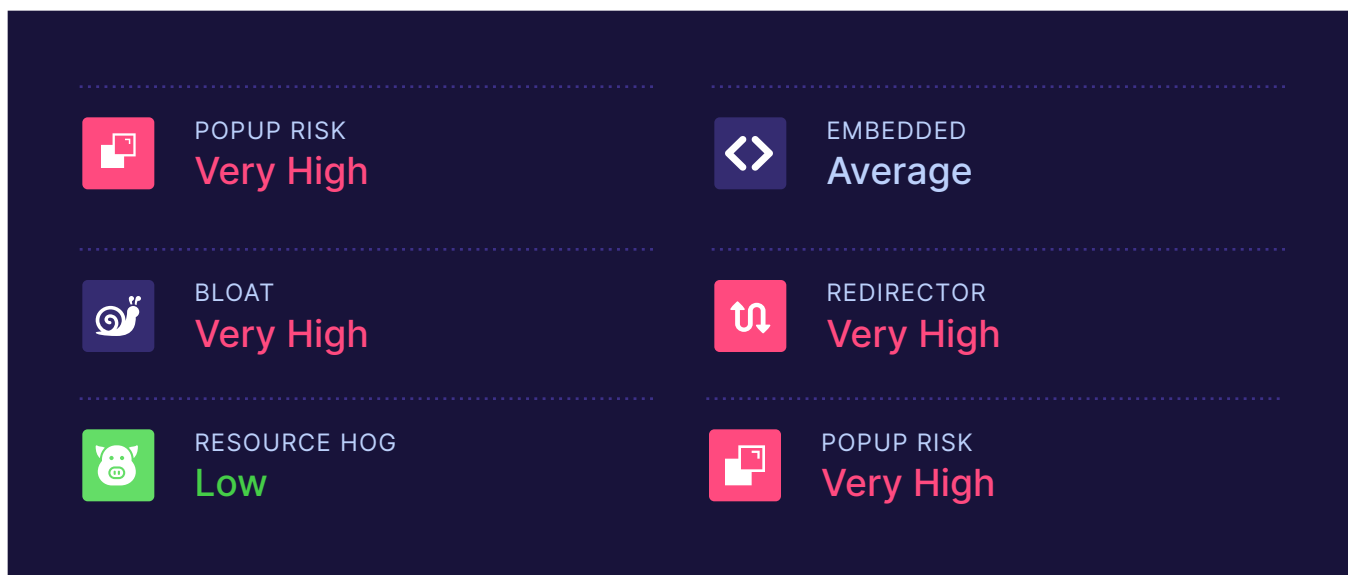
The **presence or absence** of an **ads.txt file**: Companies host these files on their web servers. The ads.txt lists the other companies authorized to sell their products or services. This list exists to allow online buyers to check the validity of the sellers from whom they buy. Ads.txt is a tool for internet fraud prevention.

Tranco Rank as a 30 day time series: Tranco unifies ranking information from four providers: Alexa, Cisco Umbrella, Majestic, and Quantcast. Tranco is not affiliated with any of these providers.

Tags: These are unique characteristics that are not necessarily related to risk. Currently, we label sites with the following tags:

- **Big Mover:** This domain's ranking has recently changed significantly.
- **Bouncer:** This domain performs an immediate redirect.
- **Advertising Domain:** This domain relates to an advertising platform.
- **Measurement Domain:** This domain collects information about visitors across many unique domains.
- **Piggybacked:** This domain appears as a query parameter in the URLs of other domains.
- **Widget:** This domain often appears as a widget.
- **Hub:** This domain has a large number of unique subdomains.
- **Top 1000:** This site regularly appears in the Tranco top 1000 global rankings.
- **Top 100:** This site regularly appears in the Tranco top 100 global rankings
- **Recently Registered:** This domain has been registered for a month or less.
- **High-Risk Advisory:** This domain has one or more risk types that are incredibly high.

Specific Risk Analysis: Offers behavior-specific risk likelihood.



Our Risk Types include:

Pop-up risk: Sites with a high Pop Up Risk produce suspicious pop-ups or load as pop-ups from shady sources.























Redirector Risk: These sites force you to visit another page when you visit them.

Embedded Risk: Sites with high embedded risk have untrustworthy referral values.

Resource Hog Risk: Resource hogs have a high demand for machine hardware resources for the duration of your visit.







Bloat Risk: Bloated sites have a sizable initial loading time.

Traffic Flow:

Sites Sending Traffic						
URL	TRANCO RANK	RISK	CHANGE (30D)	TOP FACTORS	TYPE	CATEGORY
 examplelongerdomain.com	9972	88	-10 ↗	  	Popup	Arts & Entertainment
 cats.com	9972	100	+5 ↘	  	Popup	Arts & Entertainment
 notabadsite.com	466	7	-11 ↗		Popup	Arts & Entertainment
 notabadsite.com	466	7	-11 ↗		Popup	Arts & Entertainment
 cats.com	9972	100	+5 ↘	  	Popup	Arts & Entertainment
 notabadsite.com	466	7	-11 ↗		Popup	Arts & Entertainment
 cats.com	9972	100	+5 ↘	  	Popup	Arts & Entertainment

The top 5 sites that send traffic to this site, the top 5 places that this site sends users, and some additional information on each. This feature looks beyond the most common connections to uncover more exciting trends in traffic.

Related Sites:

Possible Related Websites						
Check out other sites we feel are related.						
Related Sites						
URL	TRANCO RANK	RISK	CHANGE (30D)	TOP FACTORS	TYPE	CATEGORY
 notabadsite.com	466	7	-11 ↗		Popup	Arts & Entertainment
 cats.com	9972	100	+5 ↘	  	Popup	Arts & Entertainment

DeepSee is all about connections, the unseen relationships that dictate the flow of traffic on the web. This unique feature finds connections between sites that may seem unrelated. These sites don't necessarily load, or link to one another, but display one or more of the following related behaviors.

- **Same Parent:** When these domains appear, it's often in the same context.
- **Same Referrer:** When these domains appear, the value in the referer header is often similar. A referer request-header contains the address of the previous web page. This information allows servers to identify where people are visiting from and use that data for analytics.
- **Shared ads.txt:** These domains have remarkably similar ads.txt files.
- **Behaves Similarly:** These domains have similar activity profiles to each other.

For example, take the "shared ads.txt" similarity type. Ads.txt is a tool that ad buyers use to verify that person selling them space on a website has the right to do so. Looking at this file, you will see a list of ad systems and the publisher IDs which are allowed to be associated with that web page in a real-time-bidding situation.

If two domains have an identical ads.txt page, that means they get paid out through the same channel. This connection based on ad system & publisher ID confirms that two web sites are connected at a business level, even if they hide their web presence through anonymous registration.

All of this Single Site View information empowers users to facilitate more successful advertising campaigns for less cost and to make better decisions ensuring their ads are placed on verified reputable sites.

Filter Results

Tags

Example Tag x

Category

RECEIVES TRAFFIC FROM DOMAIN

☒ Link ☐ Referral ☐ Popup

www.text.com

SENDS TRAFFIC TO DOMAIN

☐ Link ☐ Referral ☐ Popup

Domain (example.com)



Domain has ads.txt file

POPUP RISK



EMBED RISK



REDIRECTOR RISK



BLOAT RISK



RESOURCE HOG RISK



Reset All Filters

Export Results To CSV

Advanced Search

The Advanced Search tool gives users the ability to pull information for a list of domains simultaneously. The information you are seeking can be customized with a variety of filters before being exported to a CSV. The filters available for pulling data include risk types mentioned above.

- **Category:** Arts and Entertainment, Education, etc
- **Ads.txt flag:** Yes or No
- **Risk tolerance:** A 0–100 score based on how risk-averse or risk-tolerant you'd like your site list to be.
- **Individual risk type tolerance:** Insight into specific risk types.
 - **Pop-up risk:** Sites with a high Pop Up Risk produce suspicious pop-ups or load as pop-ups from shady sources.
 - **Redirector Risk:** These sites force you to visit another page when you visit them.
 - **Embedded Risk:** Sites with high embedded risk have untrustworthy referral values.
 - **Resource Hog Risk:** Resource hogs have a high demand for machine hardware resources for the duration of your visit.
 - **Bloat Risk:** Bloated sites have a sizable initial loading time.
- **The tags described in the previous section:** Big Mover, Bouncer, Advertising Domain, Measurement Domain, Piggybacked, Widget, Hub, Top 1000, Top 100, Recently Registered, High-Risk Advisory
- **Sends traffic to domain:** Sites that send traffic to the user-specified domain
 - **Relationship Type:** This is how users arrive on a page: pop-up, referral, or link.
- **Receives traffic from domain:** Sites that receive traffic from the user-specified domain.
 - **Relationship Type:** This is how users arrive on a page: pop-up, referral, or link.
- **Site Rank:** The site's current Tranco Rank.

The Advanced Search also includes the following features:

- **CSV Upload:** Upload a custom list of domains, and get the following risk information for each with just one click. Then use this data to augment your list of sites with DeepSee risk information.
- **API:** All of the data from the advanced search feature can be accessed through an API to collect and interact with site information programmatically.

With the advanced search feature, you can create site lists on the fly based on your risk tolerance, category, rank, or even the relationships between domains. This data can then be used to enhance buying algorithms, unveil suspicious behaviors, reveal connections between domains, or give teams a clearer picture with which to make decisions.

Sample Use Cases

Programmatic Buyer or Supply Side Platform

This example will walk through how an analyst might collect data for a targeted ad purchase or search for prospective publishers to add to your network.

We are going to pull a list of sites with ads.txt and a risk score of less than 40.


Using the Advanced Search Function:

1. Set "Has ads.txt" to True.
2. Set the Risk Score Tolerance Slider so that the minimum risk is 0, and the maximum risk is 40.
3. Additionally, you have the option to select specific categories of sites that are relevant to you.
4. Hit search and download the result as a CSV.

Supply Side Platforms can use this feature for prospecting for leads and increasing impressions. Add DeepSee data into an ad-buying platform for domain targeting or avoidance to optimize your platform.

Filter Results

RISK SCORE



TRANCO RANK

MinMax

Tags

Category

Finance

RECEIVES TRAFFIC FROM DOMAIN

☒ Referral☒ Popup☐ Link

Domain (example.com)

DOMAIN	RISK SCORE	TRANCO RANK	ADS.TXT
forbes.com	19	76	Yes
cnbc.com	34	178	Yes
stackexchange.com	24	220	Yes
ft.com	7	329	Yes
investopedia.com	30	404	Yes
xe.com	5	938	Yes
fool.com	4	957	Yes
seekingalpha.com	37	1.1K	Yes
foxbusiness.com	19	1.4K	Yes
thebalance.com	27	1.5K	Yes
bankrate.com	6	1.7K	Yes
superuser.com	24	1.7K	Yes

Publisher

In this example, we will use the single-site view and the advanced search tool to understand how traffic flows to a particular site.

Prompted by a nonscheduled spike in traffic, a publisher might investigate his site with the single site view. This then leads to his use of advanced search to get a complete list of inbound connections and ultimately unveils a network of malicious actors.

1. *Using The single-site view, the publisher can check in on their property to see if any new and risky connections have appeared.*
2. *There are more than five inbound connections, perform an advanced search using a filter where the Outbound Domain = the user's domain.*
3. *We have a suspicion about pop-ups, so we add the filter "Outbound Connection Type" = pop-up*

Go deeper into the traffic from shady upstream sources and eliminate bad partners. Using this tool, analysts receive unique insights into how users arrive at their webpages. This saves time otherwise spent trying to figure out which partners send traffic by employing pop-up or forced redirection.

Security Researcher

Security researchers have to act fast to protect their customers from harmful browser exploits. DeepSee's system finds websites that exploit browsers, highlighting where these researchers should look.

In this example, we will fetch a list of sites that force visitors to go to any number of suspicious pages by hijacking their browser.

We will use 2 different advanced searches, each targeting a certain type of risky behavior.

1. *Set pop-up Risk Slider minimum to high and maximum to extremely high*
2. *After pressing SEARCH, the result will be downloadable as a CSV*



Domain has ads.txt file

POPUP RISK

EMBED RISK

6

bitefaucet.com

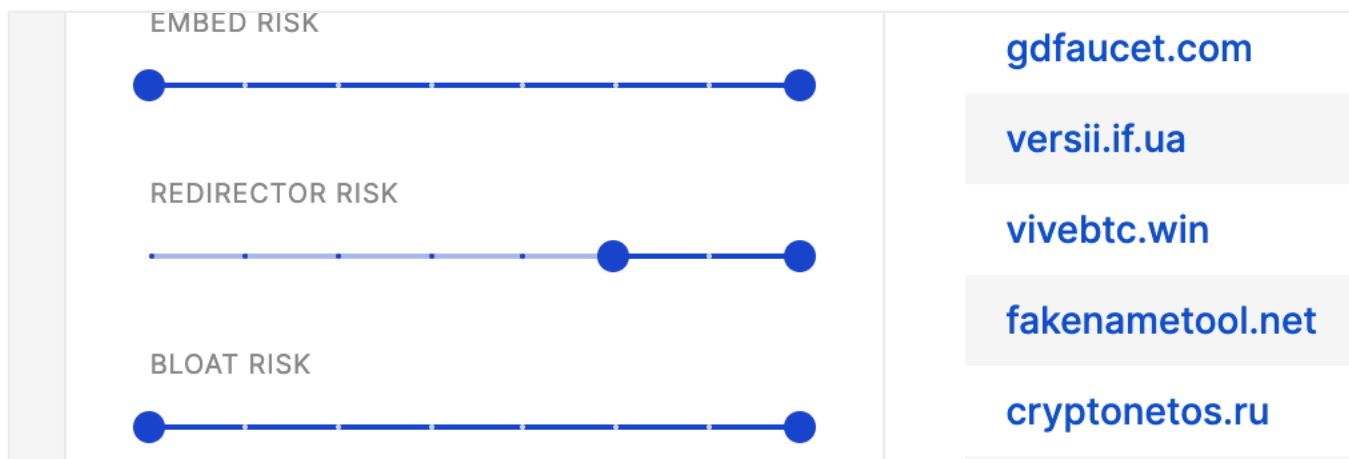
cryptomoneytechh.

infolaayoune.blogspot

makemoney1080.bl

gdfaucet.com

3. *Start a new search where the Redirector Risk Slider is set with min = high and max = extremely high*
4. *After pressing SEARCH, the result will be downloadable as a CSV*



EMBED RISK

REDIRECTOR RISK

BLOAT RISK

gdfaucet.com

versii.if.ua

vivebtc.win

fakenametool.net

cryptonetos.ru

This data is also available via an API for those looking for programmatic data solutions. For enterprise clients who are looking for real-time solutions, DeepSee can also send e-mails on your behalf as soon as a shady connection is observed with a publisher of interest.

Conclusion

More than four billion people use the internet today. Advertisers want their attention and spend billions of dollars annually on web ads to get it. Domains and publishers can boost the number of payable events recorded using fraudulent means like fake clicks and forced visits.

To gather data on how visitors arrive at a site, publishers use trackers embedded within their webpages to measure referral information like the previous destination and number of unique visitors. Similarly, advertisers record a variety of information about users each time an ad is loaded.

This type of data collection is not likely to go away. Experienced publishers, marketers, and advertisers know that common web behaviors like pop-ups and forced redirects can completely obfuscate ad tracking measurements. This results in falsely inflated invoices and causes misunderstanding about how visitors arrive on webpages.

As analysts ourselves, we consistently find situations where seemingly high-quality sites have spikes of suspicious traffic. In these cases, publishers must often reimburse clients or provide them with credit. Companies using DeepSee augment their intelligence stacks to enable better decision making and avoid dealing with the hassle of refunds.

Currently, DeepSee offers insights into browser-based traffic. In the future, we will expand this product to include iOS and Android mobile apps and connected TVs like Roku or Amazon Fire. With our information highlighting maleficent actors, ad buyers are empowered to make better decisions about who their business partners are.

As DeepSee grows and expands, we will continue to develop fraud detecting solutions as new ad markets emerge.

Smarter.
Explore
Further.
**Dive
Deeper.**