



2021 HANDBOOK

Tools, tips and tricks for cyberthreat intelligence analysts

Table of contents

Top tools to collect and analyze cyberattack data	3
What is Shodan?	16
What is exif data?	19
Investigating site ownership and history	21
Silo for Research	25

Validate, engage and enrich without risk

[Silo for Research](#) is an online investigation solution with secure, anonymous and centralized access to the surface, deep and dark web. Built on Authentic8's patented, cloud-based Silo Web Isolation Platform, it provides 100-percent protection from all web-borne threats and complete oversight of all research activity.

Threat intelligence and SOC analysts can accomplish their goals without introducing risk to the organization or revealing intent. And all web activity is logged and encrypted to ensure and monitor compliance.

See how Silo for Research can give analysts the access they need without risk — [visit our Experience Center now](#).

Full isolation:

All web code is executed on Silo servers, not end-user devices

Cloud-based:

Turn-key, cloud-hosted solution that creates a clean instance every time

Managed attribution:

Configure the browser fingerprint and egress location

Access surface, deep or dark web:

One-click access to any destination without tainting your environment

Workflow enhancements:

Integrated tools for content capture, analysis and storage

Complete audit oversight:

Encrypted audit logs of all web activity are captured in one place and easily exported



Top tools to collect and analyze cyberattack data

Cyber threat intelligence (CTI) and security operations center (SOC) analysts collect, process and interpret threat data to come up with actionable insights. To do their jobs effectively, they need a set of tools, strategically chosen for their specific features and capabilities. Authentic8 engineers offer a curated collection of tools to help analysts along every step: from collecting threat indicators, to identifying threats and analyzing potential risks to the enterprise and its assets.

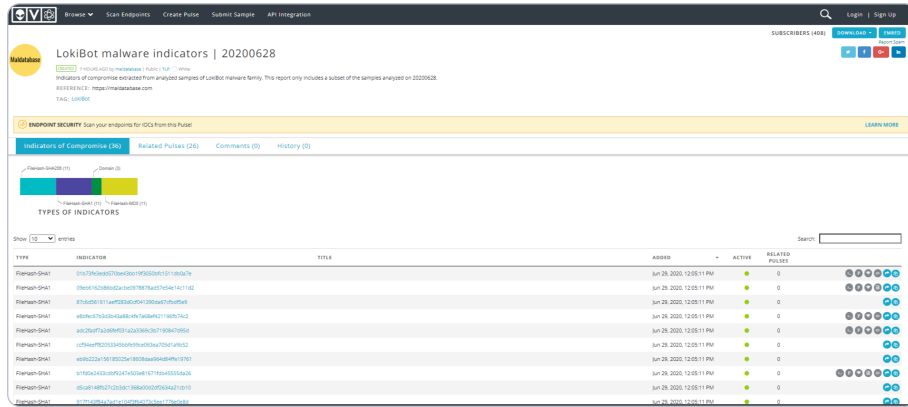
This suggested tool list will make your research work more successful and productive.

Threat intelligence stages

ASSET TYPE	INDICATOR COLLECTION	THREAT IDENTIFICATION	RISK ANALYSIS
IP/Domain	AlienVault OTX CISCPC Exploit Database FireHOL GreyNoise HoneyDB VirusTotal	DNSdumpster DomainTools Exploit Database HackerTarget Nmap Online Port Scanner HackerTarget OpenVAS Vulnerability Scan HackerTarget Reverse DNS Lookup Shodan Talos	DomainTools Exploit Database Maltego MITRE ATT&CK VirusTotal
URL	AlienVault OTX CISCPC OpenPhish VirusTotal	HackerTarget HTTP Header Check OpenPhish	Maltego MITRE ATT&CK VirusTotal
Files/hash	AlienVault OTX CISCPC VirusTotal		ANY.RUN Hex-Rays IDA Pro Maltego MITRE ATT&CK VirusTotal
Email	AlienVault OTX OpenPhish	Have i been pwnd OpenPhish	Maltego MITRE ATT&CK

AlienVault OTX

<https://otx.alienvault.com/>



THREAT INTEL STAGES

Indicator collection

ASSET TYPE

- IP/domain
- URL
- Files/hash
- Email

WHAT IT IS

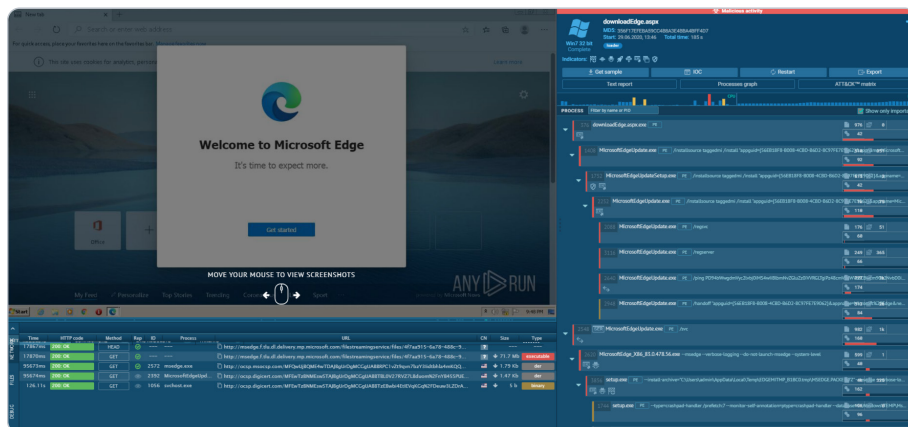
AlienVault Open Threat Exchange (OTX) is a global “neighborhood watch” of the intelligence community. It enables private companies, independent security researchers and government agencies to openly collaborate and share the latest information about emerging threats, attack methods and malicious actors, promoting greater security across the entire community.

USE CASE

In AlienVault OTX, anyone in the security community can contribute, discuss, research, validate and share threat data. Companies can integrate community-generated OTX threat data directly into their security products, making sure their threat detection defenses are always up to date with the latest threat intelligence.

ANY.RUN

<https://any.run>



THREAT INTEL STAGES

Risk analysis

ASSET TYPE

Files/hash

SIMILAR TOOLS

Valkyrie

<https://valkyrie.comodo.com/>

WHAT IT IS

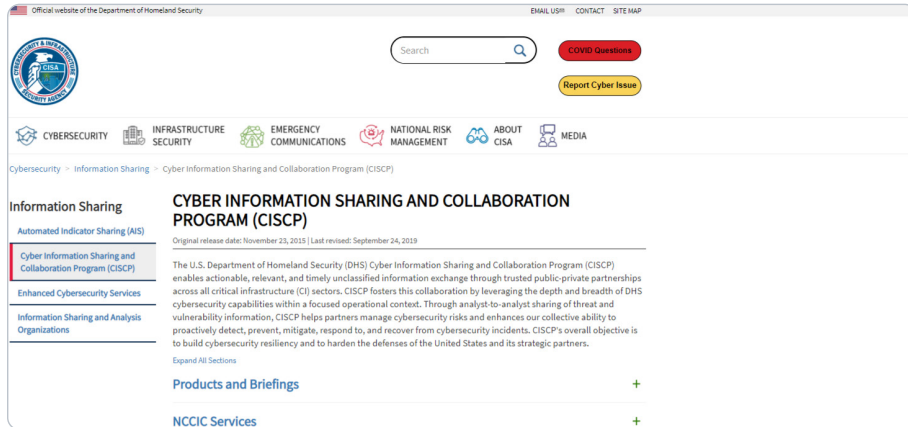
ANY.RUN is an interactive malware analysis sandbox for cyber defense.

USE CASE

ANY.RUN lets you analyze malicious files and share them with the community; train analysts to dynamically analyze malware; or privately hunt for threats using advanced features.

CISCP

<https://www.cisa.gov/ciscp>



THREAT INTEL STAGES

Indicator collection

ASSET TYPE

IP/domain
URL
Files/hash

SIMILAR TOOLS

AIS
<https://us-cert.cisa.gov/ais>

WHAT IT IS

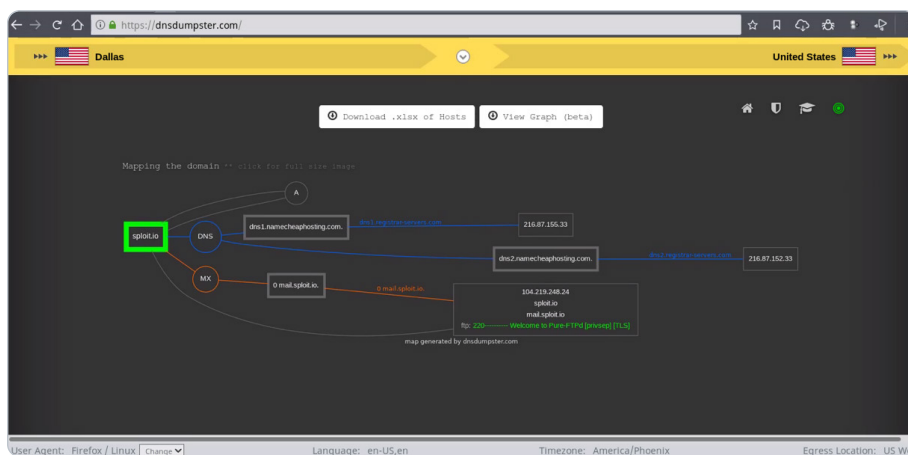
The U.S. Department of Homeland Security (DHS) Cyber Information Sharing and Collaboration Program (CISCIP) enables actionable, relevant and timely unclassified information exchange through trusted public-private partnerships across all critical infrastructure (CI) sectors.

USE CASE

CISPR leverages the depth and breadth of DHS cybersecurity capabilities to help its partners manage cybersecurity risks. The CISPR threat intelligence feed helps enhance the collective ability to proactively detect, prevent, mitigate, respond to and recover from cybersecurity incidents.

DNSdumpster

<https://dnsdumpster.com/>



THREAT INTEL STAGES

Threat identification

ASSET TYPE

IP/domain

WHAT IT IS

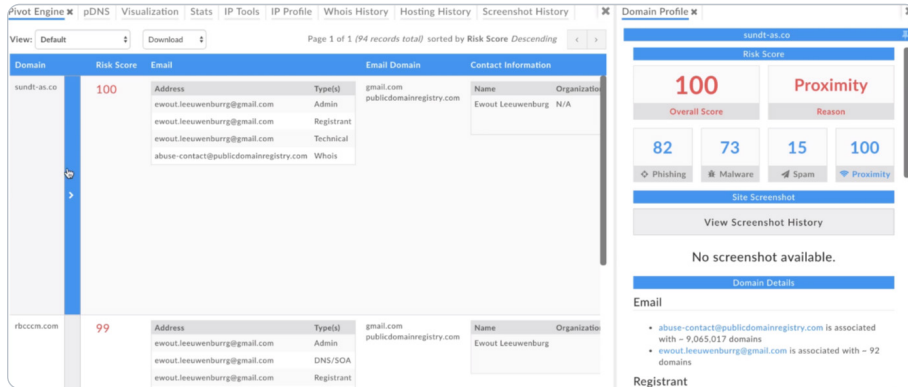
DNSdumpster is a free domain research tool that can discover hosts related to a domain. Finding visible hosts from the attackers' perspective is an important part of the security assessment process.

USE CASE

After a user enters a domain name, DNSDumpster identifies and displays all associated subdomains, helping map an organization's entire attack surface based on DNS records.

DomainTools

<https://www.domaintools.com/>



THREAT INTEL STAGES

Threat identification
Risk analysis

ASSET TYPE

IP/domain

SIMILAR TOOLS

SecurityTrails

<https://securitytrails.com/>

WHAT IT IS

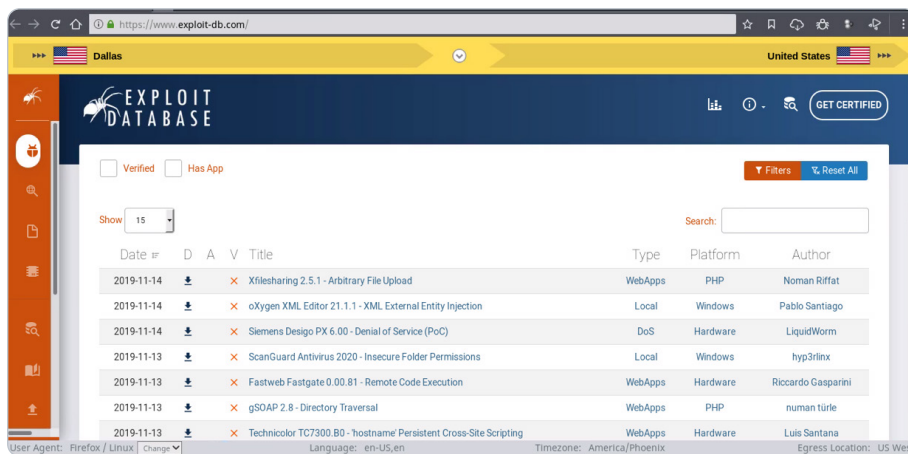
DomainTools helps security analysts turn threat data into threat intelligence. DomainTools collects OSINT data from many sources, along with historical records, in a central database. They index and analyze the data based on various connection algorithms to deliver actionable intelligence, including domain scoring and forensic mapping.

USE CASE

Use data from DomainTools to make informed risk assessments, help profile attackers, guide online fraud investigations, and map cyber activity to attacker infrastructure.

Exploit Database

<https://www.exploit-db.com/>



THREAT INTEL STAGES

Indicator collection
Threat identification
Risk analysis

ASSET TYPE

IP/domain

SIMILAR TOOLS

National Vulnerability Database

<https://nvd.nist.gov/>

MITRE Common Vulnerabilities and Exposure (CVE)

<https://cve.mitre.org/>

WHAT IT IS

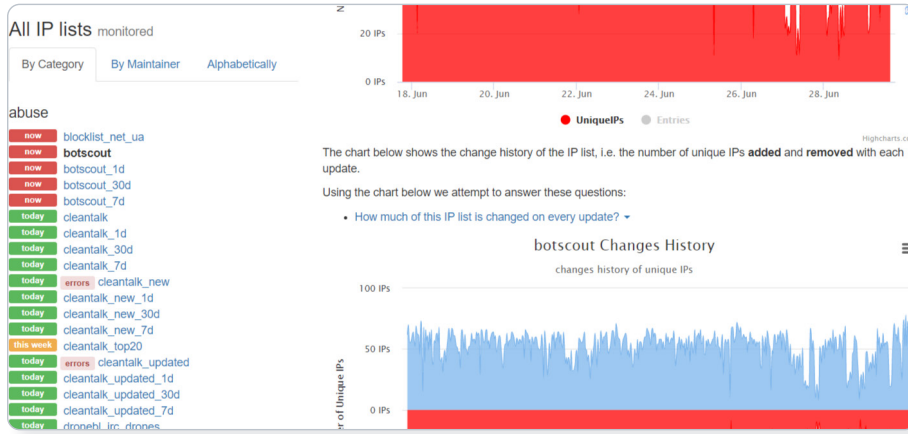
The Exploit Database is an archive of public exploits and corresponding vulnerable software, developed for use by penetration testers and vulnerability researchers. Exploits are collected throughout the internet and through user submissions, and archived for community use.

USE CASE

The Exploit Database is a repository for publicly available exploits, making it a valuable resource for those who need actionable data at their fingertips.

FireHOL

<http://iplists.firehol.org/>



THREAT INTEL STAGES

Indicator collection

ASSET TYPE

IP/domain

WHAT IT IS

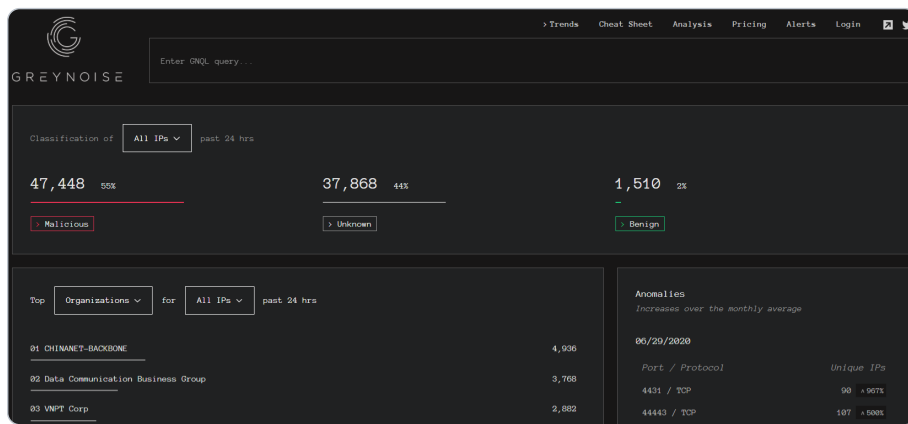
FireHOL is an IP threat intelligence feed aggregator. Its objective is to create a blacklist that is safe enough to be used on all systems with a firewall, to block access entirely, to and from its listed IPs.

USE CASE

Beyond the normal blacklist use case, FireHOL IP Aggregator can be used to determine the risk of an IP, track Tor nodes and proxies, and monitor a variety of different attacks, spam and malware.

GreyNoise

<https://greynoise.io/>



THREAT INTEL STAGES

Indicator collection

ASSET TYPE

IP/domain

WHAT IT IS

GreyNoise collects and analyzes untargeted, widespread and opportunistic scan and attack activity that reaches every server directly connected to the internet. Mass scanners, search engines, bots, worms and crawlers generate logs and events omnidirectionally on every IP address in the IPv4 space. GreyNoise gives you the ability to filter the noise out.

USE CASE

GreyNoise helps you distinguish between targeted and opportunistic attacks, filter out the noise from your network monitoring tools and identify compromised devices and emerging threats.

HackerTarget HTTP Header Check

<https://hackertarget.com/http-header-check/>

```

GET THE HTTP HEADERS

HTTP/1.1 301 Moved Permanently
Server: Varnish
Retry-After: 0
Content-Length: 0
Cache-Control: public, max-age=600
Location: http://www.cnn.com/
Accept-Ranges: bytes
Date: Mon, 29 Jun 2020 22:40:52 GMT
Via: 1.1 varnish
Connection: close
Set-Cookie: countryCode=US; Domain=.cnn.com; Path=/; SameSite=Lax; geoData=ashburn[VA]20147[US][NA]-400[broadband]39.030|-77.490; Domain=.cnn.com; Path=/; SameSite=Lax
X-Served-By: cache-bwi5040-BWI
X-Cache: HIT
X-Cache-Hits: 0

HTTP/1.1 301 Moved Permanently
Server: Varnish
Retry-After: 0
Content-Length: 0
    
```

THREAT INTEL STAGES

Threat identification

ASSET TYPE

URL

WHAT IT IS

A lot of information can be gathered in a check of the HTTP headers from a web server. Server-side software can be identified often down to the exact version running. Cookie strings, web application technologies and other data can also be gathered from the HTTP header.

USE CASE

HTTP header information is used when troubleshooting. It can also be a key resource when trying to prevent an attack against the web server.

HackerTarget Nmap Online Port Scanner

<https://hackertarget.com/nmap-online-port-scanner/>

Free Port Scan to check **any IP address** and test **10 common TCP ports** with Nmap version detection (-sV) enabled. Once you see how easy it is grab a membership and get immediate full access.

Ports Checked in Free Scan
 21 File Transfer (FTP)
 22 Secure Shell (SSH)
 23 Telnet
 25 Mail (SMTP)
 80 Web (HTTP)

110 Mail (POP3)
 143 Mail (IMAP)
 443 SSL/TLS (HTTPS)
 445 Microsoft (SMB)
 3389 Remote (RDP)

QUICK NMAP SCAN

```

Starting Nmap 7.70 ( https://nmap.org ) at 2020-06-29 23:02 UTC
Nmap scan report for www.cnn.com (157.166.226.25)
Host is up.

PORT      STATE SERVICE
21/tcp    filtered ftp
22/tcp    filtered ssh
23/tcp    filtered telnet
80/tcp    filtered http
110/tcp   filtered pop3
    
```

THREAT INTEL STAGES

Threat identification

ASSET TYPE

IP/domain

WHAT IT IS

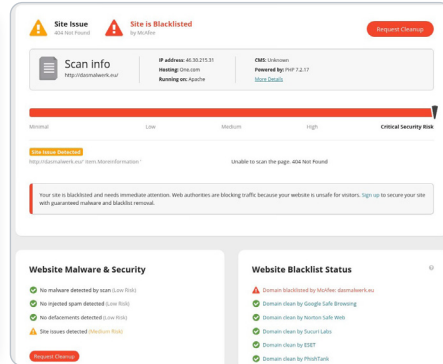
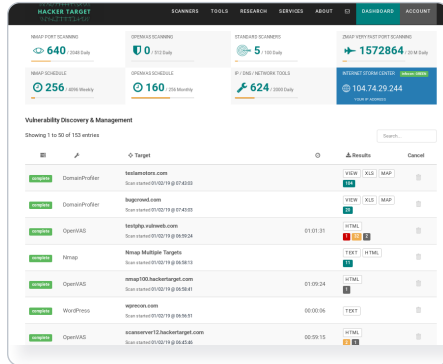
Nmap provides an accurate port status of a system's Internet footprint.

USE CASE

Nmap is used to test servers, firewalls and network perimeters to determine status of host and network-based firewalls, find open ports on cloud-based virtual servers, detect unauthorized firewall changes and troubleshoot network services.

HackerTarget OpenVAS Vulnerability Scan

<https://hackertarget.com/openvas-scan/>



THREAT INTEL STAGES

Threat identification

ASSET TYPE

IP/domain

WHAT IT IS

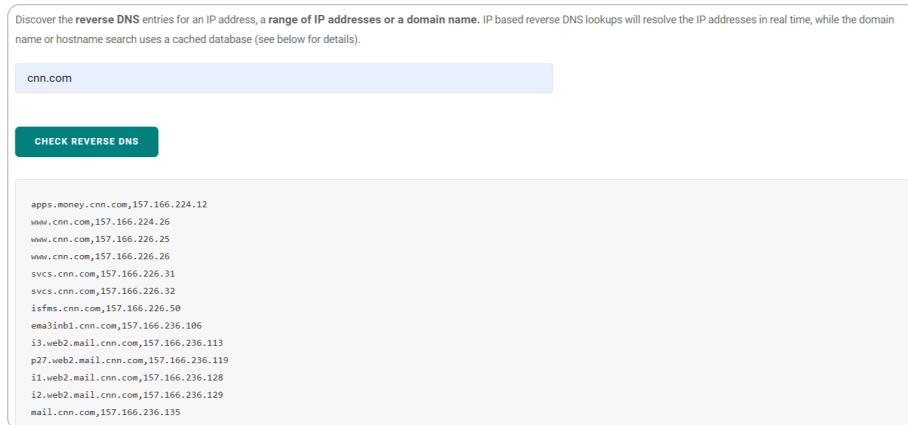
The OpenVAS scanner is a comprehensive vulnerability assessment system that can detect security issues in all types of servers and network devices. Use this hosted version of the OpenVAS software to effortlessly test your internet infrastructure. Results are delivered via email for analysis, allowing you to start remediating any risks your systems face from external threats.

USE CASE

The primary use for this scan type is to perform comprehensive security testing of an IP address. Once listening services are discovered, they are tested for known vulnerabilities and misconfiguration using a large database. The results are compiled into a report, including detailed information regarding each vulnerability and notable issues discovered.

HackerTarget Reverse DNS Lookup

<https://hackertarget.com/reverse-dns-lookup/>



THREAT INTEL STAGES

Threat identification

ASSET TYPE

IP/domain

WHAT IT IS

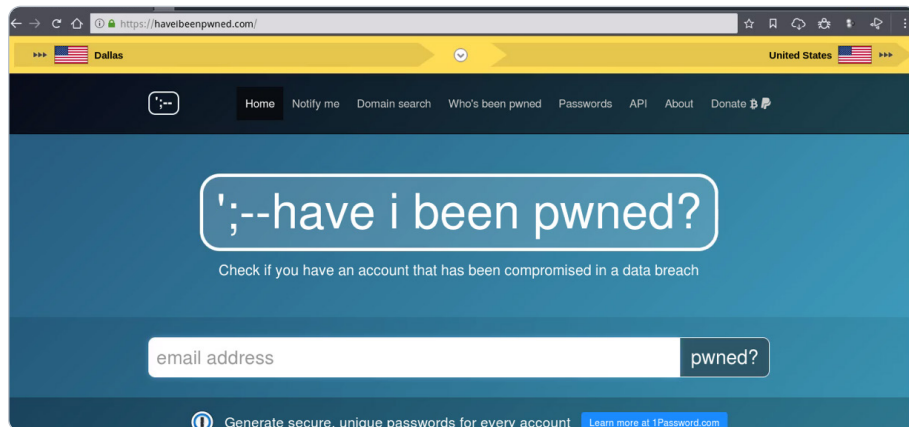
Reverse DNS Lookup allows you to discover reverse DNS entries for an IP address, a range of IP addresses or a domain name. IP-based reverse DNS lookups will resolve the IP addresses in real time, while the domain name or hostname search uses a cached database.

USE CASE

When an attacker assesses an organization, they will commonly attempt to map the footprint of the organization in order to find the weak points. With this reverse DNS tool, you are able to resolve single IP addresses or a range of IP addresses, or search for all reverse DNS containing a domain name.

Have i been pwnd

<https://haveibeenpwned.com/>



THREAT INTEL STAGES

Threat identification

ASSET TYPE

Email

WHAT IT IS

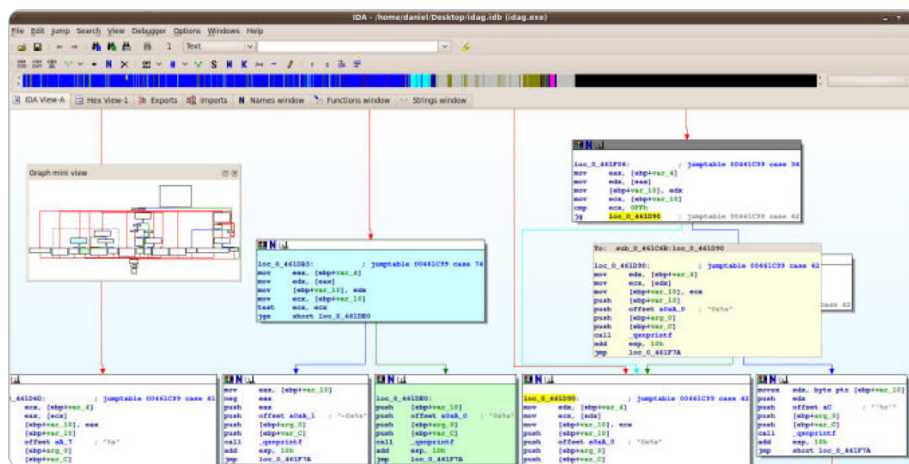
The service exposes the severity of the risks of online attacks, while helping victims of data breaches learn about compromises of their accounts. Users can subscribe to receive breach notifications, and search for pwned accounts and passwords across domains.

USE CASE

Users can securely enter email addresses and passwords to find out if they have been hacked. The site returns a complete list of breaches where specific accounts have been exposed, and what types of data (email addresses, names, passwords, locations, etc.) have been stolen.

Hex-Rays IDA Pro

<https://www.hex-rays.com/products/ida/>



THREAT INTEL STAGES

Risk analysis

ASSET TYPE

Files/hash

SIMILAR TOOLS

XORI

<https://github.com/endgameinc/xori>

Ghidra

<https://ghidra-sre.org/>

WHAT IT IS

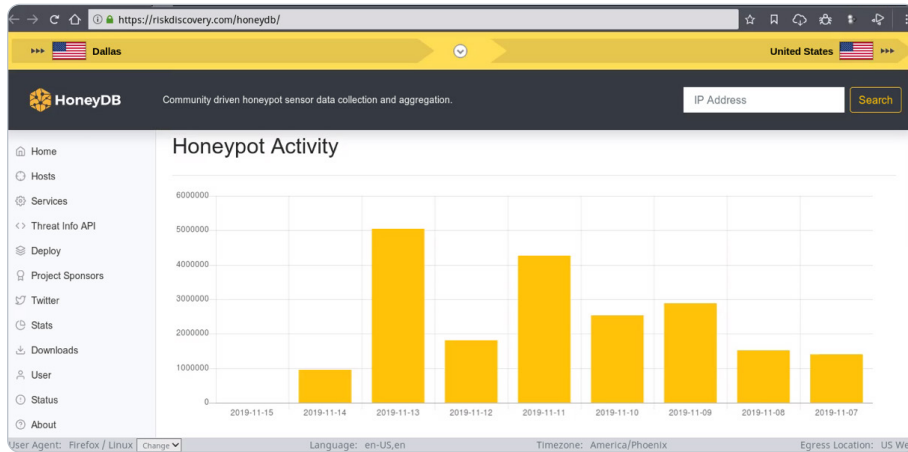
The source code of the software isn't always available. A disassembler like IDA Pro translates machine-executable code into readable assembly language source code, enabling research specialists to analyze programs that are suspected to contain malware or spyware.

USE CASE

An incident response team loads a malicious artifact found on a breached server into IDA Pro to further analyze and understand its behavior, potential damage and method of traversal. IDA Pro can also be used as a debugger to aid analysts in reading and examining the hostile code.

HoneyDB

<https://honeydb.io/>



THREAT INTEL STAGES

Indicator collection

ASSET TYPE

IP/domain

WHAT IT IS

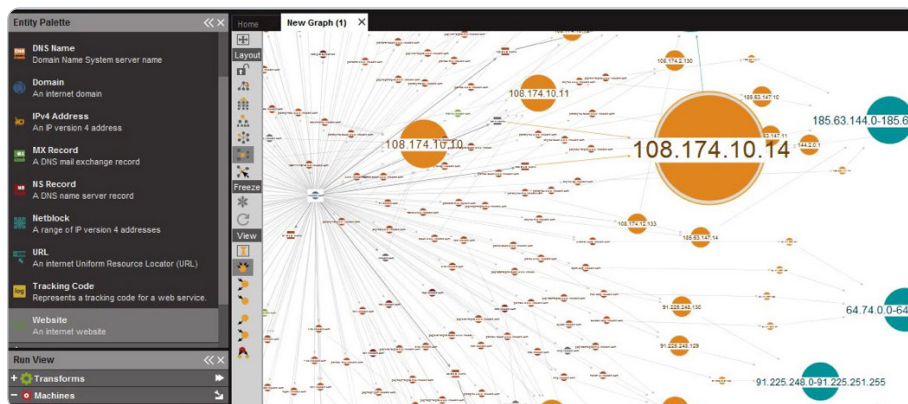
HoneyDB has multiple honeypots throughout the internet waiting to be attacked. The service logs complete details of an attack (including IP address) and the binary that was used to execute it, and it lists these items in the database. HoneyDB enables users to run a reverse search on IOCs and correlates it back to campaigns that are happening on its honeypots.

USE CASE

A campaign that uses a unique exploit to commit a wide-spread attack on every system possible, would most likely infect one or more of the honeypots. A user then accesses detailed information on the attack to gather information about its intentions and perpetrators.

Maltego

<https://www.maltego.com/>



THREAT INTEL STAGES

Risk analysis

ASSET TYPE

IP/domain

URL

Files/hash

Email

WHAT IT IS

Integrate data from public sources, commercial vendors and internal sources via the Maltego Transform Hub. All data comes pre-packaged as Transforms, ready to be used in investigations. Maltego takes one artifact and finds more.

USE CASE

A user feeds Maltego domain names, IP addresses, domain records, URLs or emails. The service finds connections and relationships within the data and allows users to create graphs in an intuitive point-and-click logic.

MITRE ATT&CK

<https://attack.mitre.org/>

THREAT INTEL STAGES

Risk analysis

ASSET TYPE

IP/domain

URL

Files/hash

Email

WHAT IT IS

MITRE ATT&CK® is a globally-accessible knowledge base of adversary tactics and techniques based on real-world observations. The ATT&CK knowledge base is used as a foundation for the development of specific threat models and methodologies in the private sector, government and the cybersecurity product and service community.

USE CASE

ATT&CK can be used in a variety of ways to help security operations, threat intelligence and security architecture. Common use cases include: detection and analytics; adversary emulation and red teaming; assessment and engineering; and more.

OpenPhish

<https://openphish.com/>

Top 10 Targeted Brands		Top 10 Sectors		Top 10 ASNs	
ASB Bank Limited	27.0%	Financial	44.2%	AS46606 Unifield Layer	29.6%
RuneScape	10.0%	Gaming	11.3%	AS60503 FNIX Techno...	8.3%
Apple Inc.	6.2%	Online Services	8.8%	AS32344 Liquid Web...	6.5%
Outlook	4.9%	Retail/Service	7.9%	AS15169 Google LLC	5.8%
PayPal Inc.	3.7%	Email Provider	7.8%	AS22612 Namecheap...	4.7%

THREAT INTEL STAGES

Indicator collection

Threat identification

ASSET TYPE

URL - both stages

Email - both stages

WHAT IT IS

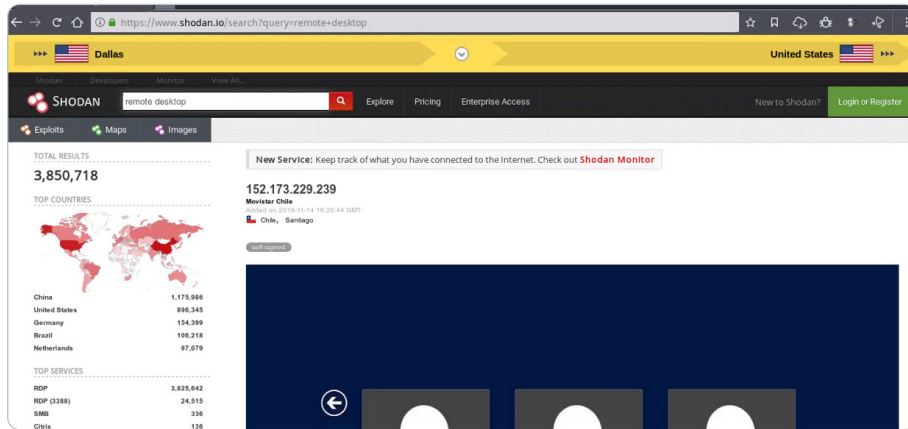
OpenPhish is a fully automated, self-contained platform for phishing intelligence. It identifies phishing sites and performs intelligence analysis in real time without human intervention and without using any external resources, such as blacklists.

USE CASE

OpenPhish detects and tracks new and live phishing sites downloadable for free (updated every 12 hours).

Shodan

<https://www.shodan.io/>



THREAT INTEL STAGES

Threat identification

ASSET TYPE

IP/domain

SIMILAR TOOLS

Censys

<https://censys.io/>

WHAT IT IS

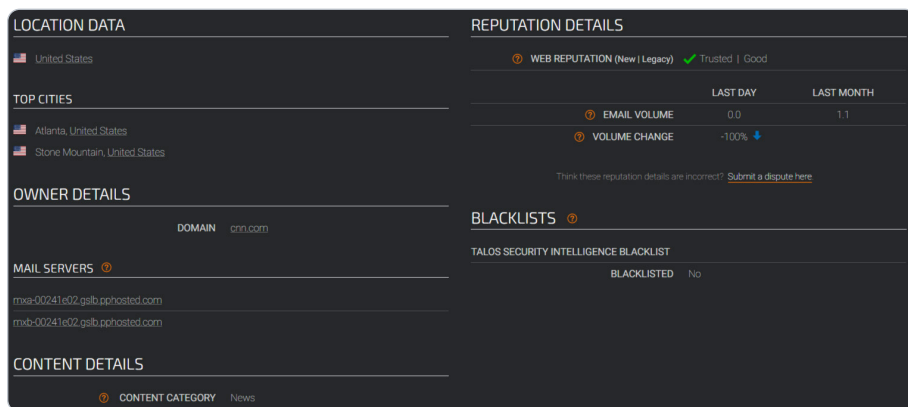
Websites are just one part of the internet. Shodan allows analysts to discover which of their devices are connected to the internet, where they are located and who is using them.

USE CASE

Shodan helps researchers monitor all devices within their network that are directly accessible from the Internet and therefore vulnerable to attacks.

Talos

<https://talosintelligence.com/>



THREAT INTEL STAGES

Threat identification

ASSET TYPE

IP/domain

WHAT IT IS

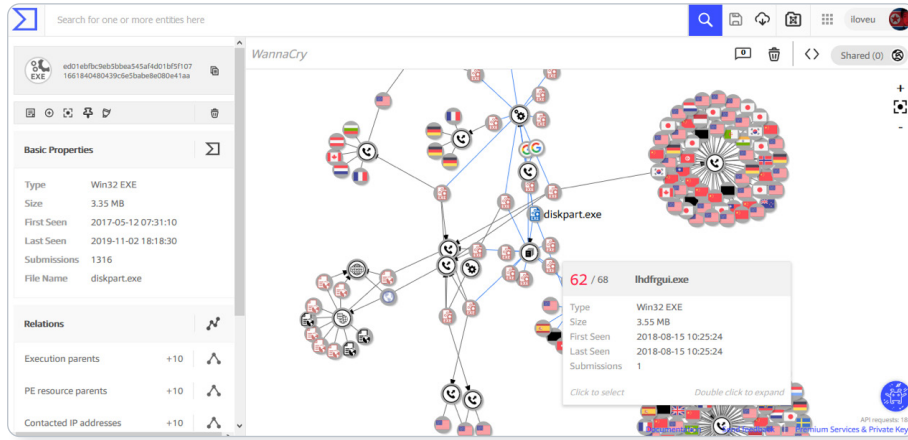
Cisco Talos Intelligence Group is one of the largest commercial threat intelligence teams in the world. Talos defends Cisco customers against known and emerging threats, discovers new vulnerabilities in common software and interdicts threats in the wild before they can harm the internet at large.

USE CASE

A variety of free software, services, resources and data from Talos are available to the public.

VirusTotal

<https://www.virustotal.com/>



THREAT INTEL STAGES

Indicator collection

Risk analysis

ASSET TYPE

IP/domain - all stages

URL - all stages

Files/hash - all stages

WHAT IT IS

VirusTotal inspects items with over 70 antivirus scanners and URL/domain blacklisting services. Scanning reports produced by VirusTotal are shared with the public to raise the global IT security level and awareness about potentially harmful content.

USE CASE

Users can select a file from their computer using their browser and send it to VirusTotal. Results are shared with the submitter, and also between the examining partners, who use this data to improve their own systems.

What is Shodan?

According to [GitHub](#), Shodan is “the world’s largest search engine for internet-connected devices”. But what exactly does this mean?

Most search engines are text indexes, meaning they allow search for content based on keywords. However, the task of scanning, indexing the ports and services running, and then searching for internet-connected devices at the scope and scale of the internet has been largely impossible to do.

With Shodan, it is now possible to identify nearly any internet-connected device, such as industrial control systems running specific software, internet-of-things devices like smart TVs, FTP servers with sensitive information and even very small aperture terminals (VSATs) on naval vessels.

How Shodan works

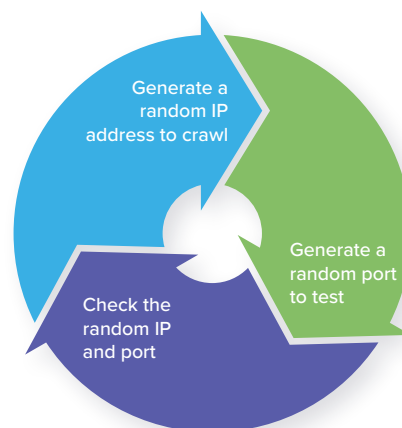
Shodan maintains servers across the globe that scan the internet-connected devices and harvest the banner of whatever is running on the server.

The diagram at right shows how these servers crawl.

These internet-connected devices return different banners depending on the different service running on it.

Example search returns

Two examples are below, one for an IP camera and one for an FTP server (FTP runs on port 21):



Document Error: Unauthorized

62.112.117.205

OA0 MGTS

Added on 2019-05-07 10:56:51 GMT

Russian Federation, Odintsovo

Technologies: IIS|confidence:50

HTTP/1.1 401 Unauthorized

Server: Cam-Webs

Date: Tue May 7 13:20:55 2019

WWW-Authenticate: Basic realm="Megapixel_IP_Camera"

Pragma: no-cache

Cache-Control: no-cache

Content-Type: text/html

188.225.26.71

vds-olga@irsova.timeweb.ru

hosting & vds

Added on 2019-05-28 17:02:17 GMT

Russian Federation

220 (vsFTPD 3.0.2)

230 Login successful.

214-The following commands are recognized.

```
ABOR ACCT ALLO APPE CDUP CWD  DELE EPRT EPSV FEAT HELP LIST MDTM MKD
MODE NLST NOOP OPTS PASS PASV PORT PWD  QUIT REIN REST RETR RMD  RNFR
RNT0 SITE SIZE SMNT STAT STOR STOU STRU SYST TYPE USER XCUP XCWD XMKD...
```

Basic Shodan searches and filters

Shodan allows for advanced search using filters. Filters are entered in a simple format: a filter, a colon and the search value, with no spaces between these three components.

Filter format	<code>filtername:value</code>
Filter example	<code>City:Moscow</code>

If searching a value that includes a space, double quotes must be used.

Filter example	<code>City:"Saint Petersburg"</code>
----------------	--------------------------------------

Examples of Shodan's most useful geographic filters

Country using two-letter geocode	<code>country:XX</code>
City using city name	<code>city:cityname</code>
Geographic coordinates in a bounding box	<code>geo:top-left-lat,top-left-long,top-right-lat,top-right-long</code>
Region	<code>region:region-name-or-state</code>

These filters are useful when attempting to identify something of interest in a specific AOR.

For example, a search for webcam `City:Incirlik` would find webcams, with some hopefully located near Incirlik Air Base.

Examples of software-focused filters

Firewall port	<code>port:XX</code>
Product name	<code>product:XX</code>
Product version	<code>version:XX</code>
Product vulnerability CVE	<code>vuln:XX</code>

These filters are useful when searching for a particular technology, like a database, a file server or vulnerable software.

For example, a search for `port:21 country:"RU" "login successful"` would find file transfer protocol (FTP) servers in Russia that do not require logins. This could yield valuable unsecured information if found in a location of interest, or can be used as a non-attributable temporary data transfer point.

Examples of organization-focused filters

Device hostname	<code>hostnames:XX</code>
Organization assignment	<code>org:XX</code>
Network CIDR range	<code>net:XX</code>

Examples of Shodan’s temporal filters

Results before a given date	<code>before:00/00/0000</code>
Results after a given date	<code>after:00/00/0000</code>

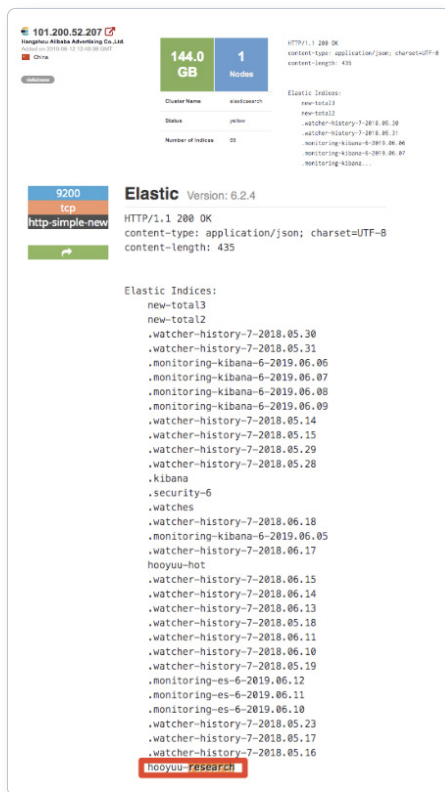
Finding open databases

A few databases openly list their indices: MongoDB, ElasticSearch and CouchDB.

Below are the baseline searches that allow you to quickly identify open databases with potentially valuable information sources.

Example database searches

Elasticsearch databases	<code>product:elastic port:9200</code>
MongoDB databases	<code>product:MongoDB</code>
CouchDB databases	<code>product:couchdb</code>
Kibana visualization of Elasticsearch	<code>kibana content-length: 217</code>
Gitlab software repos	<code>http.favicon.hash:1278323681</code>
Rsync utilities	<code>product:rsyncd</code>
Jenkins software automation	<code>jenkins 200 ok</code>



Combining these search filters and other key phrases allows analysts to identify high value and unsecured information.

Example search for Elasticsearch databases in China mentioning “research”:

`product:elastic port:9200 country:cn research`

This results in identifying an IP address hosting an open elasticsearch index with mentions of research. In this case, the research is about “Hooyuu,” a Chinese social media site.

The others range from what looks like security research, notifications and some form of alerting.

For more information please contact osint@authentic8.com.

What is exif data?

When a digital image is captured, metadata specific to that image is stored. This information is called exchangeable image file format — “exif” for short — data. Some examples of exif data are date, time and file size. This information can be extremely useful when conducting image analysis. Analysts can exploit exif data to find the location of the image, camera make and model, and other information that is valuable to the intelligence production cycle.

Incorporating exif data

To find exif data, an analyst can use a number of different tools. [FotoForensics](#) is the service used for the workflow described here. In the example in this report, we’ve taken an image of a cargo ship from a [ship-spotting forum](#) (see figure 1) and uploaded it to FotoForensics to analyze the exif data.

User-uploaded images in forums will likely have their exif data intact. However, if the analyst tries to pull exif data from an image on social media, there will likely be little to no data present. Social media platforms have begun to strip exif data off of user images to protect user privacy.

Once on FotoForensics, the analyst will have two options for image analysis. The analyst can paste an image URL or upload a file for analysis (see figure 2).

For this workflow, the analyst can save the above image of the cargo ship, and then upload the .jpg file into FotoForensics.

When the upload is complete the analyst should select the metadata field from the “Analysis” dropdown list (see figure 3). The analyst can then scroll down and begin to review information pertinent to the investigation.



Figure 1 | Image from [shipspotting.com](#)

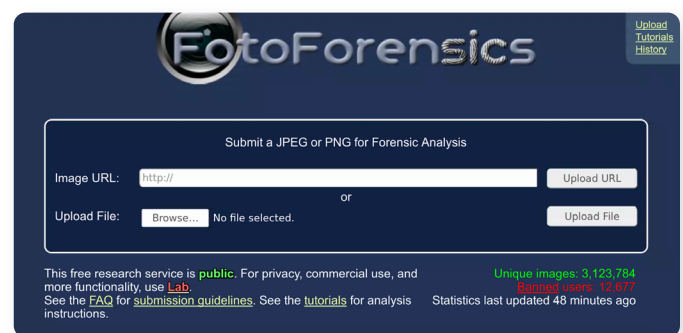


Figure 2 | FotoForensics user interface



Figure 3 | FotoForensics post image upload with metadata analysis selected

After reviewing the exif data collected by FotoForensics, a few pieces of information stand out. The analyst can glean what type of device was used to capture the image (see figure 4). This information can be useful when investigating a party of interest that may have a standard issue camera for reconnaissance.

FotoForensics also provides the analyst with an approximate latitude and longitude coordinate (see figure 5). This coordinate can be further incorporated into a targeting packet or reconnaissance mission.

Overall, the information captured from exif data can greatly enhance a unit's analytic ability. The exploitation of images, whether of an adversarial object or person or of a location, can help the analyst to further understand their battlespace or objective.

Conclusion

This workflow covers how to extract and incorporate exif data into the intelligence product. The analyst found a .jpg file of a cargo ship and leveraged FotoForensics to conduct exif data analysis. Results from the analysis included key identifiers such as equipment used and location data that can then be incorporated further into a finished intelligence product.

For more information please contact osint@authentic8.com.

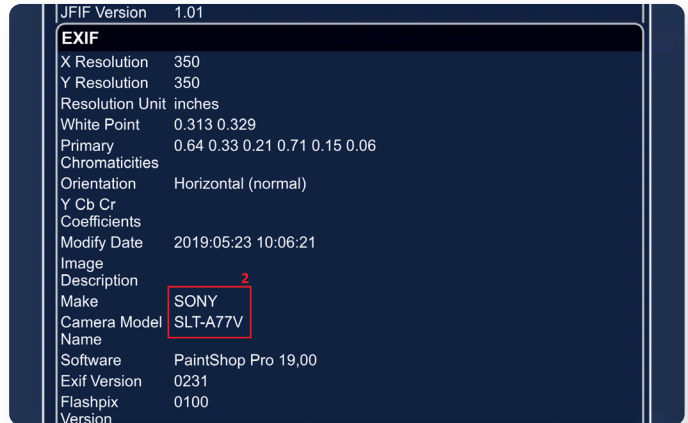


Figure 4 | FotoForensics exif data results including camera model

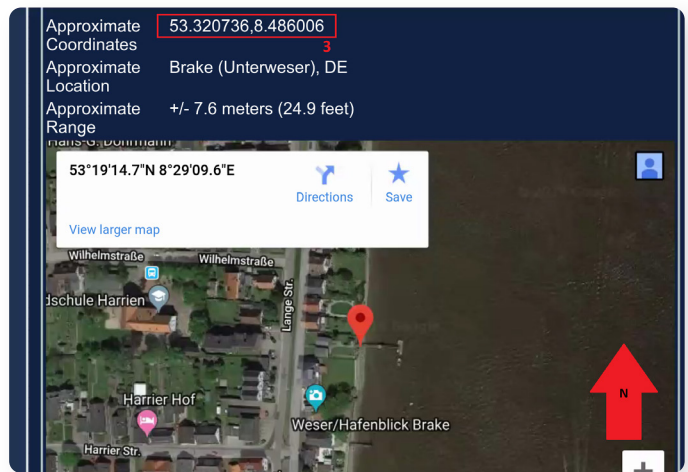


Figure 5 | FotoForensics exif data results including geographic coordinates.

Investigating surface websites' ownership and history

Analysts collecting publicly available information (PAI) encounter various sites and services with valuable information. While this information is of intelligence value, there are biases, agendas, and different reasons for the dissemination of such information.

To identify these reasons, analysts have to find information on the individuals/organizations behind the site/service which hosted, maintained, and funded them.

This information is commonly obfuscated, but accessible with proper research tools and tradecraft.

Resources used for site ownership research

Analysts can leverage the following sites and services:

- **WHOIS records:** WHOIS records provide top level domain (e.g., russianmilitaryblog.com) information such as exact dates of registration, addresses, names, and phone numbers associated with the domain. In addition, it provides web host information.
 - URL Scan: <https://urlscan.io>
- **Advanced search engine use:** Using advanced search engines and search engine parameters on uniquely identifying information found on the site or WHOIS records (i.e., emails, names, mail servers, other IP addresses, etc.) can provide additional information on the site or service administrator/s.
 - Carbon Date: <http://carbodate.cs.odu.edu>
 - Google Dorking: <https://www.google.com>

On the following pages we describe how to use these tools and give examples of information that can be gleaned from them.

For more information please contact osint@authentic8.com.

WHOIS record analysis: URLscan.io

URLscan.io conducts analysis of a domain, providing the end user with information on all HTTP connections made during the site’s retrieval, outbound links from the page, as well as detailed IP address information.

The screenshot shows the URLscan.io interface for the domain **forums.airbase.ru**. The main IP address is **148.251.51.134**, located in Germany. The analysis shows 82 HTTP transactions, 27 links, and 14 frames. The site is hosted on a Hetzner-AS server. Detected technologies include Bootstrap, AppNexus, DoubleClick Ad Exchange (AdX), DoubleClick Campaign Manager (DCM), and Google AdSense.

Callout 1: “Summary” provides a top level summary of what country the site is hosted in.

Callout 2: “HTTP” details how many HTTP connections are made during initial load.

Callout 3: “Links” details what other sites are linked to on the main page.

Callout 4: “IP/ASN” details the IPs of everything used upon initial load and the geographic location as well as ASN.

Callout 5: “IP Detail” contains the exact city/state/country an IP address is assigned to, and redirects.

Callout 6: “Domains” identifies how many subdomains a top level domain contains.

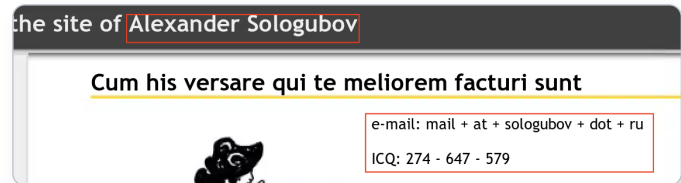
Example analysis of result panels

Forums.airbase.ru, a russian military forum, uses hosting primarily in Germany, which is likely due to Germany’s strict data privacy laws. From the HTTP panel, the site uses Google Analytics for user tracking and also uses Yandex.ru for email. From the Links panel, a live “Telegram” chat is also available for users.

Example analysis of the result panel

Only one other IP aside from the current German IP has been used for hosting forums.airbase.ru.

This IP is 95.31.43.16, which also is used by a range of other domains — one of which, sologubov.ru has personal information on the individual behind forums.airbase.ru. This reveals the web host's full name, email, and ICQ number for further targeting.

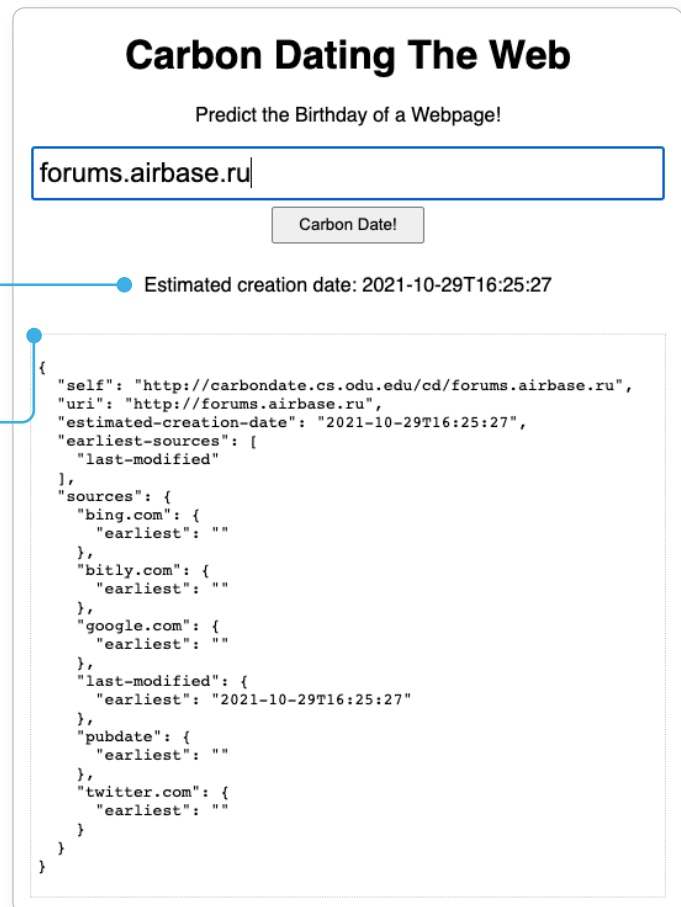


Advanced search engine: Carbon Date

This advanced search engine automates advanced searches against web.archive.org, archive.md, Bing, bit.ly, Google, and Twitter to identify the earliest scrape/index or mention of a website on the web.

“Estimated creation date” pulls the earliest date from the result set.

The result set shows the results from each source searched, and when available, a URL to the direct source itself.



Carbon Dating The Web
Predict the Birthday of a Webpage!

forums.airbase.ru

Carbon Date!

Estimated creation date: 2021-10-29T16:25:27

```
{
  "self": "http://carbondate.cs.odu.edu/cd/forums.airbase.ru",
  "uri": "http://forums.airbase.ru",
  "estimated-creation-date": "2021-10-29T16:25:27",
  "earliest-sources": [
    "last-modified"
  ],
  "sources": {
    "bing.com": {
      "earliest": ""
    },
    "bitly.com": {
      "earliest": ""
    },
    "google.com": {
      "earliest": ""
    },
    "last-modified": {
      "earliest": "2021-10-29T16:25:27"
    },
    "pubdate": {
      "earliest": ""
    },
    "twitter.com": {
      "earliest": ""
    }
  }
}
```

Example analysis of the results panel

The earliest mention of forums.airbase.ru was in October of 2003. To view the first ever scrape of this site by web.archive.org, use the URL in the “uri-m” field.

Advanced search engine: Google Dorking

Advanced Google search parameters and features are used in a technique called “Google Dorking.”

Users must combine various search parameters to effectively search and filter down results of interest to them.

The most commonly used Google Dorks are:

- **Intitle:** identifies any mention of search text in the web page title
- **Allintitle:** only identifies pages with all of the search text in the web page title
- **Inurl:** identifies any mention of search text in the web page URL
- **Intext:** only identifies pages with all of the search text in the web page URL
- **Site:** limits results to the specified file type
- **Filetype:** limits results to only the specified file type
- **Cache:** shows the most recent cache of a site specified
- **Around (X):** searches for two different words within X words of one another

The most commonly used Boolean logic search operators are:

- **AND:** searches for content mentioning two phrases anywhere
- **OR:** used in multi-part search, searches for content mentioning any combination of the first search term and two unique second search variables
- *****: the asterisk acts as a wildcard and searches for any word or phrase
- **-**: the dash excludes any specific word or phrase (if using brackets or quotation marks)
- **()**: the parenthesis group specific terms or search operators together

Example analysis using advanced Google Search parameters

```
site:sologubov.ru ICQ OR email
```

This search will find mentions of ICQ or email on a site of interest, resulting in an ICQ number and email previously unknown to an analyst.

```
site:forums.airbase.ru contact OR admin OR mod OR moderator OR donation
```

This search will find uniquely identifying information that can be linked to a person, such as mentions of a moderator, a contact page, or a donation page (such as Paypal, Bitcoin, etc), resulting in multiple pages with mentions of the moderator and a donation page for their health bills.

```
"95.31.43.16"
```

This search will find exact mentions of forums.airbase.ru, resulting in mentions on another forum of Russian censorship of the servers IP address.

Conclusion

This workflow covers how to investigate the ownership and hosting information related to a site/service of interest. Results from the analysis include key identifiers such as server IPs, other related domains, and the webhost's email address/name/ICQ number that can then be incorporated further into a finished intelligence product.

For more information please contact osint@authentic8.com.

Silo for Research

Safe and anonymous access to all areas of the web

Silo for Research embeds security, identity and data policies directly into the browser, eliminating the risk of the web, and protecting your applications and data from exploits and misuse.

Silo for Research is a purpose-built solution for conducting online research without exposing analysts' digital fingerprint. Safely pursue investigations across the surface, deep or dark web through an isolated, cloud-based browsing interface while controlling how you appear online.

Protect your identity and your investigation

Adversaries exploit tracking mechanisms in traditional browsers to uncover analysts' identity and intent — and spoil the investigation or retaliate against them. Silo for Research manages the details they see, so analysts don't arouse suspicion.

Manage attribution

Blend in with the crowd while conducting sensitive online investigations. Silo for Research equips investigators with dozens of options to spoof their geolocation, utilizing Authentic8's global network of internet egress nodes.

But building a complete "location narrative" requires more than just changing egress. Investigators using Silo for Research can control a range of details including:

- **Browser fingerprint:** time zone, language, keyboard, operating system, device type, web browser
- **Network address:** physical location, internet provider, subscriber information
- **Data transfer and protection:** isolated browsing session, one-time-use browser (no persistent tracking), policy control to restrict upload/download, copy/paste, etc.

Isolate browsing

Ensure 100% segregation between your device — including the apps and data it holds — and all that's encountered during online investigations — like trackers, malware and more — across the surface, deep and [dark web](#).

HOW THE BROWSER BETRAYS YOU

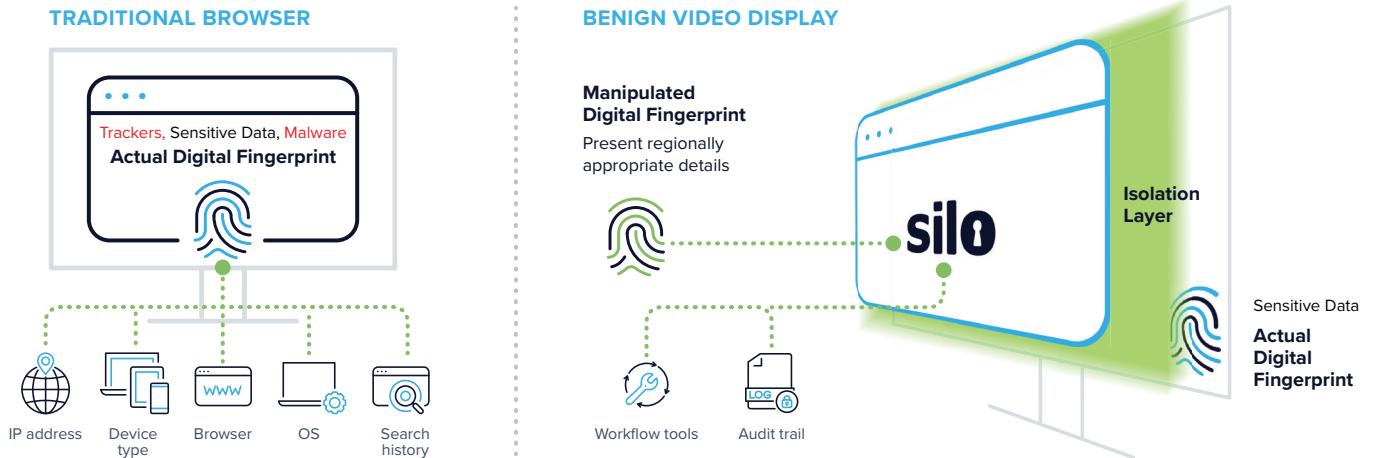
Traditional browsers disclose a range of information about you to the websites you visit.

- Passed by your browser: device type, OS, software/plugins installed, time zone, audio/video devices
- Stored in your browser by websites: cookies, HTML5 local storage
- Derived from content displayed: HTML5 canvas fingerprinting, audio

By combining these details, the subjects of your investigation can get a highly unique picture of who you are. Once they realize they're under investigation, they could hide, feed you disinformation or retaliate — online or in real life.

Silo for Research is built on Authentic8’s patented, cloud-based Silo Web Isolation Platform, which executes all web code in a secure, isolated environment that’s managed by policy. All web activity is logged and encrypted so compliance teams can be sure that the tools are being used appropriately.

And, each session is launched as a one-time-use browser, ensuring cookies and supercookies don’t follow investigators, even between sessions.



Improve efficiency

Purpose-built tools and third-party integrations give investigators the workflow tools they need to move through their caseload effectively. Built-in features for translation, capture and annotation simplify the data collection and analysis process. Authentic8 Secure Storage also makes it easy to save and collaborate safely on information, while adhering to policy.

[Additional features](#) are available to automate analysts’ tasks, including for collection and multi-search workflows, while adhering to tradecraft best practices.

More than 500 of the world’s most at-risk enterprises and government agencies rely on Silo for Research to conduct secure and anonymous online investigations, including for:

- Trust and safety
- Intelligence and evidence gathering
- Security intelligence
- Fraud and brand misuse
- Corporate research and protection
- Financial crime and compliance

To learn more about Silo for Research, [request a demo](#) or [contact a sales representative](#).