Investigating Site 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

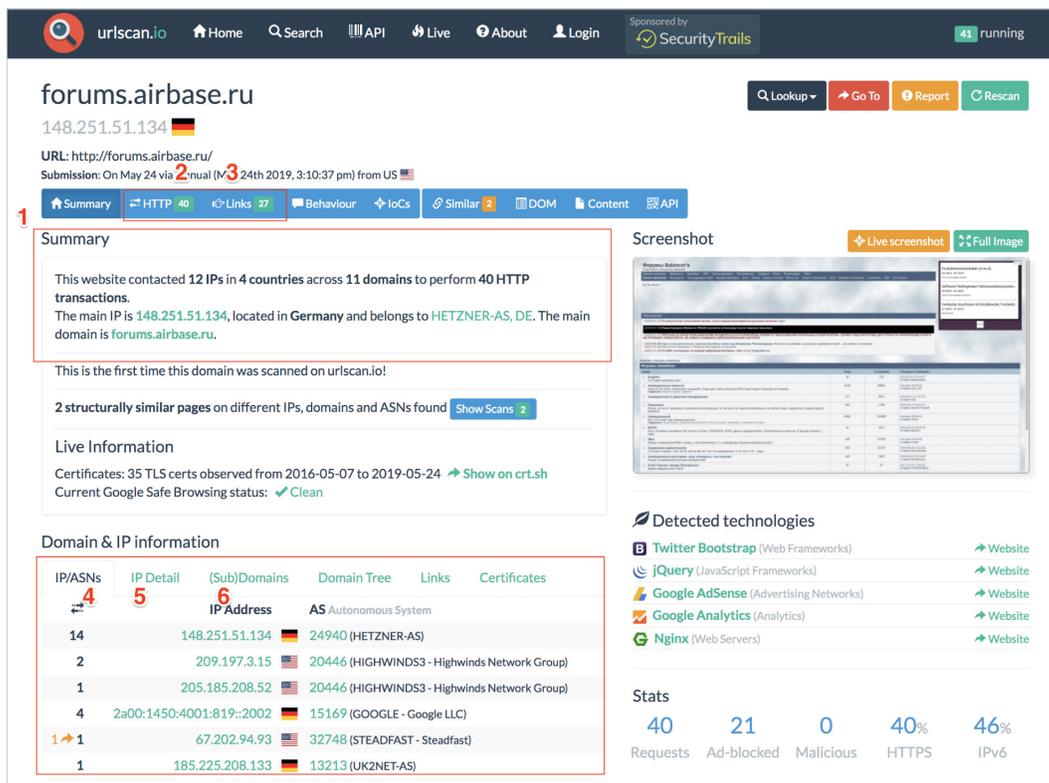Analysts can leverage the following sites and services:

- **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
  - DomainIQ: https://www.domainiq.com
- 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://carbondate.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.

Breakdown of URLscan.io result panels:

1. "Summary" provides a top level summary of what country the site is hosted in.
2. "HTTP" details how many HTTP connections are made during initial load.
3. "Links" details what other sites are linked to on the main page.
4. "IP/ASN" details the IPs of everything used upon initial load and the geographic location as well as ASN.
5. "IP Detail" contains the exact city/state/country an IP address is assigned to, and redirects.
6. "(Sub)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.

WHOIS Record Analysis: Hosting Research

Hosting Research provides the end user with historical information on the servers hosting the site. This can be useful as servers often host multiple sites from the same webmaster or have valuable information like the owner information available.

Breakdown of Hosting Research result panels:

1. Hosting Server History contains the historical IPs which hosted the site of interest, and details what other domains were on that server and the server's IP subnet.
2. "Domains on this IP" is opened when clicking on an IP. This details what other sites have WHOIS information that point to this IP.

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.

silo
By Authentic8

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.

Breakdown of Hosting Research result panels:

1. "Estimated creation date" pulls the earliest date from the result set.

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

3. The web.archive.org result is the earliest result set; with a URL you can follow to view the earliest iteration of the site.

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.

Carbon Dating The Web

Predict the Birthday of a Webpage!

forums.airbase.ru

Carbon Date!

1 Estimated creation date: 2003-10-28T00:36:45

2

```
{
  "self": "http://carbondate.cs.odu.edu/cd/forums.airbase.ru",
  "uri": "http://forums.airbase.ru",
  "estimated-creation-date": "2003-10-28T00:36:45",
  "earliest-sources": [
    "web.archive.org"
  ],
  "sources": {                                3
    "web.archive.org": {
      "uri-m": "http://web.archive.org/web/20031028003645/http://
      "memento-datetime": "2003-10-28T00:36:45",
      "memento-pubdate": "",
      "earliest": "2003-10-28T00:36:45"
    },
    "archive.md": {
      "uri-m": "http://archive.md/20140825015428/http://forums.ai
      "memento-datetime": "2014-08-25T01:54:28",
      "memento-pubdate": "2014-08-25T01:54:28",
      "earliest": "2014-08-25T01:54:28"
    },
    "bing.com": {
      "earliest": ""
    },
    "bitly.com": {
      "earliest": "2014-03-21T04:07:24"
    },
    "google.com": {
      "earliest": ""
    },
    "last-modified": {
      "earliest": "2019-05-24T15:11:09"
    },
    "pubdate": {
      "earliest": ""
    },
    "twitter.com": {
      "earliest": ""
    }
  }
}
```

The most commonly used Google Dorks are:

Intitle	This identifies any mention of search text in the web page title.
Allintitle	This will only identify pages with all of the search text in the web page title.
Inurl	This identifies any mention of search text in the web page URL.
Allinurl	This will only identify pages with all of the search text in the web page URL.
Intext	This will search for any mentions of search text.
Site	This will limit your results to only those within the site specified.
Filetype	This will limit your results to only the specified file type.
Cache	This will show the most recent cache of a site specified.
Around(X)	This will search for two different words within X words of one another.

silo
By Authentic8

The most commonly used Boolean logic search operators are:

AND	This will search for content mentioning two phrases anywhere.
OR	This is used in multi part search, and will search for content mentioning any combination of the first search term and two unique second search variables.
*	This will act as a wildcard and search for any word or phrase.
–	This will exclude any specific word or phrase (if using brackets or quotes). *Note: this is a dash sign.*
( )	This will 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
By Authentic8

CONNECT WITH US

+1 877-659-6535
www.Authentic8.com

in twitter G2

PROTECT THE THINGS YOU CARE ABOUT FROM THE THINGS YOU CANNOT TRUST

Authentic8 enables anyone, anywhere, on any device to experience the web without risk. The Silo Web Isolation Platform by Authentic8 separates the things you care about like apps, data and devices, from the things you cannot trust like public websites, external users and unmanaged devices. Silo executes all web code in a secure, isolated environment that is managed by policy, to provide protection and oversight.

Today, the world's most at-risk organizations rely on Silo to deliver trust where it otherwise cannot be guaranteed.