



Dorking the Internet

Unlocking Secrets in Open Directories

search.censys.io

Hosts ▾ services: (service_name=HTTP and http.response.html_title: "Index of /")

Hosts

Results: 382,257 Time: 1.17s

000.000.000.000 (super-secret-things.com)

Linux UNIFIEDLAYER-AS-1 (46606) Utah, United States

file-sharing remote-access email database login-page

21/FTP	>_ 22/SSH	25/SMTP	26/SMTP
80/HTTP	110/POP3	143/IMAP	443/HTTP
587/SMTP	993/IMAP	995/POP3	2077/HTTP
2082/HTTP	2083/HTTP	2086/HTTP	2087/HTTP
2096/HTTP	>_ 2222/SSH	3306/MYSQL	5432/POST

000.000.000.001 (extra-secret-things.com)

Red Hat Enterprise Linux 7 NAMECHEAP-NET (22612) Arizona, United States

Executive Summary

Having data unintentionally exposed to the internet could have severe consequences if attackers are able to find, identify and exploit the information. In this report, we used very simple techniques to identify potential threats related to misconfigured web servers using old, tried, and tested methods.

- Censys indexed the entirety of 313,750 hosts that had open directory listings enabled and contained one or more files. Through this process, we indexed and categorized 477,330,039 files, which would take up 2,000 terabytes of space if stored on a single disk.
- **We saw 9,897 hosts exposing over 6,000 gigabytes worth of database data.** After filtering out small-sized files, which could just be database schemas, we found thousands of files across hundreds of hosts containing database-related files over 100 megabytes. **Database file sizes of this nature should be a red flag, as this is indicative of a database backup that has been exposed to the internet.**
- We observed thousands of hosts serving millions of files with common spreadsheet file extensions, and looking closer at the file names, **we found over 9,000 had an indication of being related to financial data. These types of files have a high probability of containing very sensitive information about businesses and their customers.** We also looked for files that are known to explicitly contain authentication and credential data, netting over 3,000 hosts and exposing over 4,800 files that contain this type of sensitive information.
- **File exposure through open directories is in no way a new issue.** However, when examining the last-modification timestamp of files we observed, **we discovered that most of the exposed data was created or modified in 2023**, illustrating that this old problem is still going strong even as organizations become more security-conscious.
- By classifying the autonomous systems (ASes) for these hosts into different service offerings, we noted a predominant number of them were web hosting providers. These types of service providers frequently cater to small businesses and individuals, which often face limitations in terms of IT and security resources. **Compared to large organizations, where identifying a point of contact for security issues is straightforward (a security.txt file or a profile on LinkedIn), hosting provider customers are not as easily located, putting them at a disadvantage if their data is inadvertently exposed.**

INTRODUCTION

- Even when web-server developers go out of their way to make directory listings harder to enable accidentally, we still see it happening all the time. **This is likely the result of hasty decision-making when attempting to create and download backups or the lack of proper education when it comes to server maintenance.**

Introduction

In our [2023 State of the Internet Report](#), we discussed the topic of data leaks on the internet. We identified numerous hosts serving (potentially) sensitive data from HTTP servers with directory listings enabled. This report intends to expand on our findings by pushing our scanner beyond the landing page of an open directory listing and delving deeper into these exposures.

This type of exposure is not new and has long been a driving force behind the infamous Google search method known as “[Dorking](#).” One of the most popular forms of Dorking is searching around for open directories. The concept itself is straightforward yet remarkably effective. By searching for web servers with HTML titles containing phrases like “index of” or “directory listing for” and perhaps incorporating additional terms to target specific file types, one can initiate a search and manually sift through the results to identify any noteworthy findings.

However, with the help of Censys, you can execute [similar search queries](#) but with a broader scope. Rather than only uncovering standard HTTP hosts that were stumbled upon by the Google crawler, Censys allows you to explore results for every host and nearly every port across the entire internet. This encompasses not just bare IP addresses but also virtual-hosted websites.

Different Web Servers Render Slightly Different Outputs

Name	Last modified	Size	Description
1GB.dat	2023-05-02 08:28	1.0G	
1MB.dat	2023-05-02 08:26	1.0M	
50kb.dat	2023-05-02 08:26	50K	
500kb.dat	2023-05-02 08:26	500K	
bak_index.html	2023-01-11 09:59	27	
cacti/	2021-10-27 14:08	-	
index.nginx-debian.html	2022-08-23 07:11	612	
odir-walker	2023-05-02 08:15	7.0M	
stuff/	2023-05-02 12:14	-	
test/	2023-05-03 09:13	-	
things and stuff	2023-05-02 13:26	0	

*Apache-based directory listing
(file sizes are prettified)*

- [1GB.dat](#)
- [1MB.dat](#)
- [500kb.dat](#)
- [50kb.dat](#)
- [bak_index.html](#)
- [cacti/](#)
- [index.nginx-debian.html](#)
- [odir-walker](#)
- [stuff/](#)
- [test/](#)
- [things and stuff](#)

*Python-based directory listing
(note: no last-modification
timestamp)*

Index of /			
../			
cacti/	11-Jan-2023 05:00		-
stuff/	02-May-2023 19:14		-
test/	03-May-2023 16:22		-
1GB.dat	02-May-2023 15:28	1048576000	
1MB.dat	02-May-2023 15:26	1048576	
500kb.dat	02-May-2023 15:26	512000	
50kb.dat	02-May-2023 15:26	51200	
bak.index.html	11-Jan-2023 17:59		27
index.nginx-debian.html	23-Aug-2022 14:11		612
odir-walker	02-May-2023 15:15	7297078	
things_and_stuff	02-May-2023 20:26		0

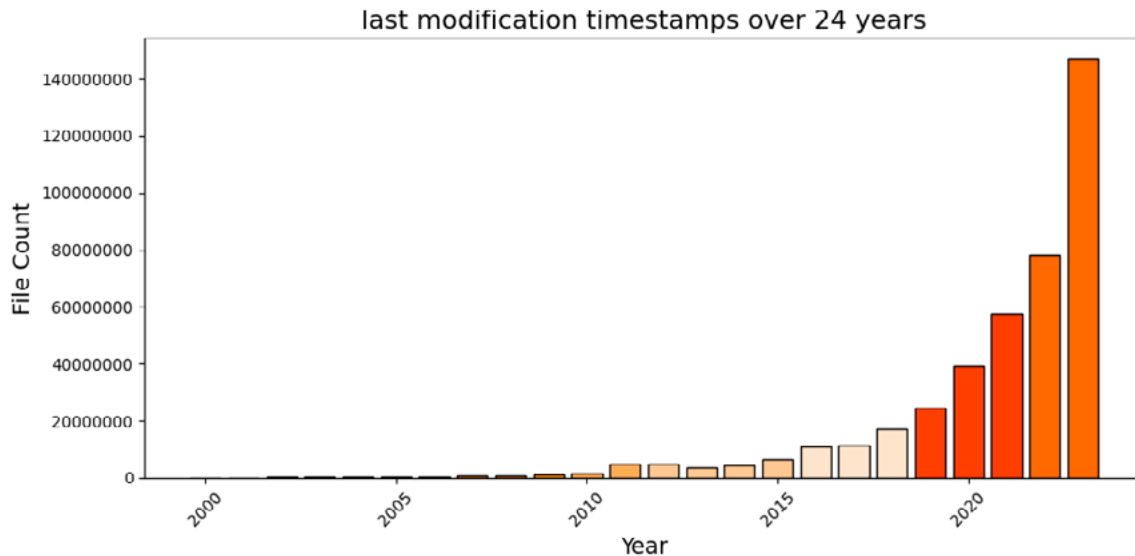
*Nginx-based directory listing
(file sizes are in bytes)*

But what are open directories, and why do they exist? Open directories are the result of web servers like Nginx or Apache that have been configured to render a directory listing when an index.html file isn't found. These configurations are (usually) not default, and most major web servers require an administrator to enable a module and configuration directive for them to even work. So in many cases, when you come across an open directory, someone somewhere at some time explicitly set it up. But if you're not an administrator and only a user of the web services, then you may not even know that your files would be exposed like this.

Since Censys does not crawl the entire tree of an HTTP endpoint and only indexes the contents of the root directory, we can still use this data as a starting point to feed into a custom scanner that walks the entirety of a directory listing. Along the way, our scanner will make a note of the file names, paths, file sizes, and last-modification timestamps. The result of this is one of the most comprehensive databases of all open directories on the internet.

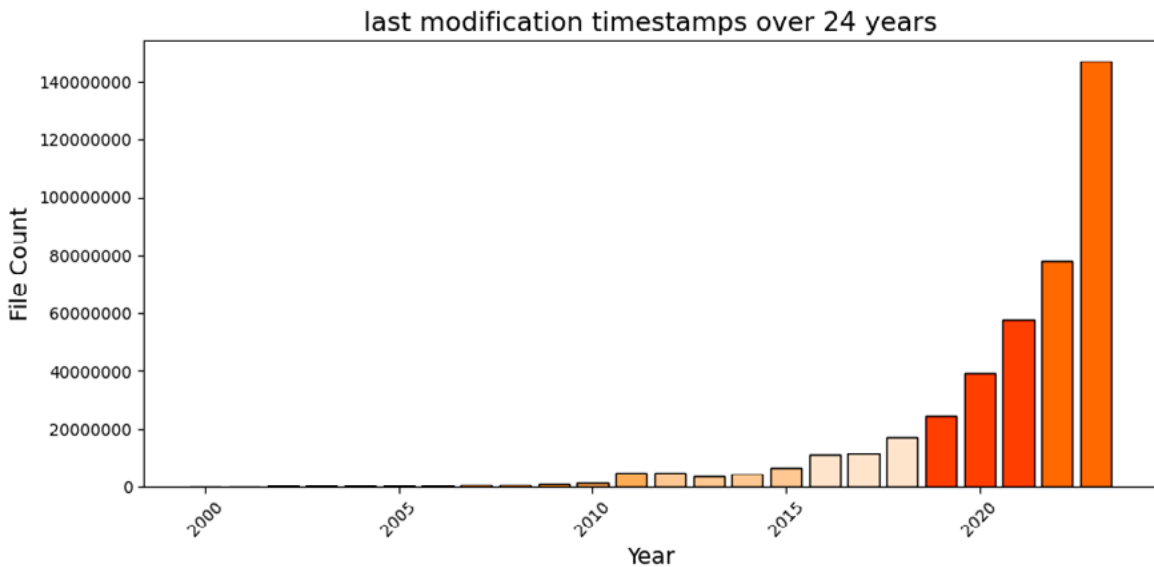
Overall

This investigation looked at 313,750 distinct hosts from a single day snapshot with open directory listings (with at least one file) and indexed the entirety of each. Contained within these hosts were 477,330,039 files, amassing a global storage size of over 2,000 terabytes of data (and that's only counting the hosts that present the file size in their HTTP responses, not all do).



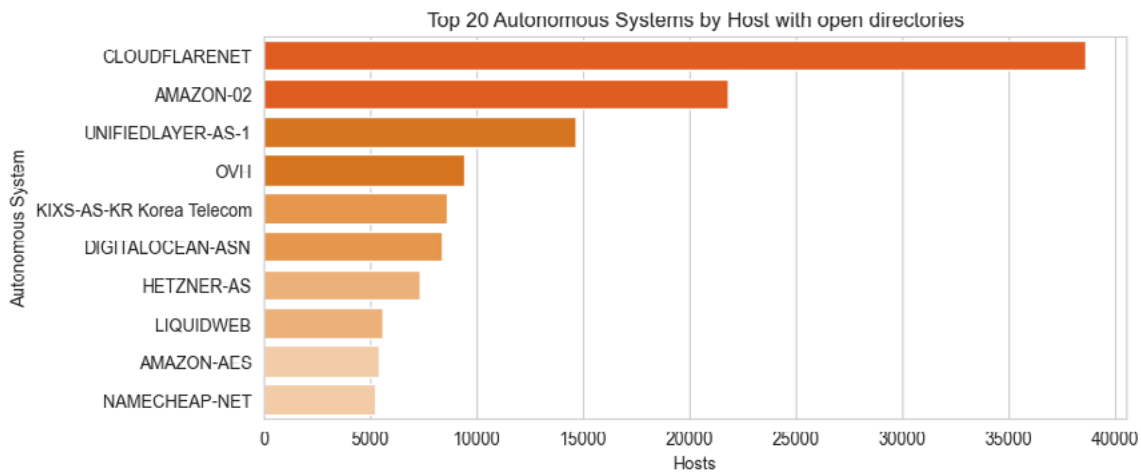
In the graph above, which shows file count over the last 24 years, we see the vast majority of the files we came across were created or modified this year (note: only for the files we had a timestamp for), with 146,882,511 files in 2023, while only 78,156,370 files in 2022, and each year before we see fewer and fewer files being exposed. This does not mean the files have been online for this long; it just means the files that are exposed are of this age.

OVERALL



This graph shows the number of hosts that have one or more files within each year for the past 24 years; much like the file count graph, the majority of hosts have files that have been created or updated in 2023 (59,050 hosts), which is over 7,500 more hosts than the 51,360 hosts hosting files created or modified in 2022.

In 2023, the problem of open directories is still going strong. Even as organizations become more security conscious, we're still seeing issues like this crop up to this day.



As for where these hosts can be found, we can look at the top autonomous systems (AS) hosting open directories. In the graph above, we see that the majority of hosts are sitting behind the Cloudflare CDN with 38,614 different web servers and AMAZON-02 with 21,805 hosts.

OVERALL

On its own, this information may not offer significant insights. However, by examining the top 100 autonomous systems (by host count) and classifying the (types of) services these ASs provide, we can gain a more nuanced understanding of who configures these servers. We created four distinct categories that each of the top 100 autonomous systems falls under:

Hosting

Hosting providers are companies that provide basic managed and unmanaged hosting services like virtual hosting, shared hosting, virtual private servers (VPS), and dedicated servers (rent-a-server) for individuals, and small to medium-sized businesses.

Cloud

Cloud providers are internet services like Amazon AWS, Google Cloud Platform, and Microsoft Azure. These are different from hosting providers in that they offer many different ways to store and access data.

CDN

These are hosts that fall within a Content Delivery Network like Cloudflare or Akamai.

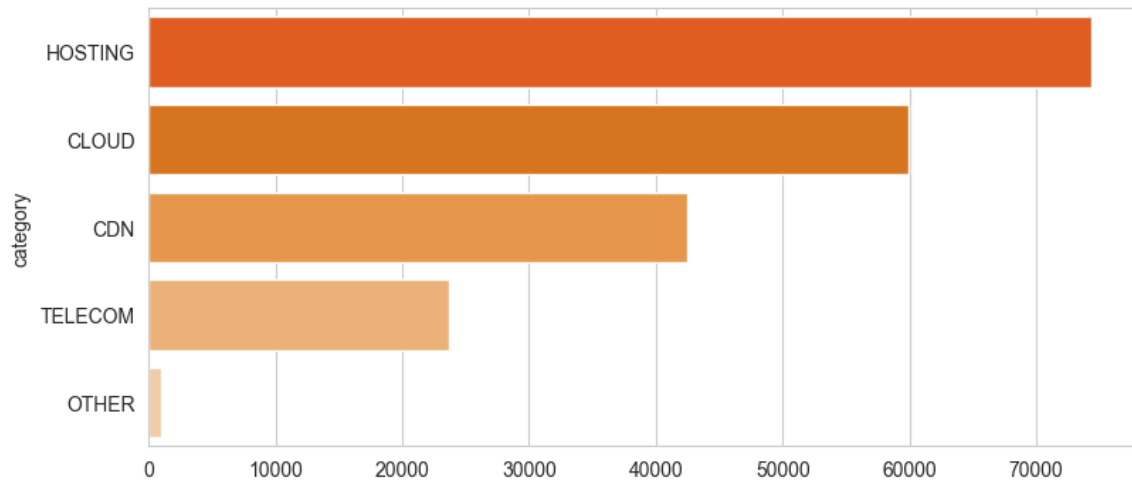
Telecom

Telecoms are companies that provide various communication technologies such as telephone, internet, and wireless. These would include companies like Comcast, AT&T, or Deutsche Telekom.

Other

Systems that don't fall under any of the above categories.

Top 100 Autonomous Systems - Categorized Host Count

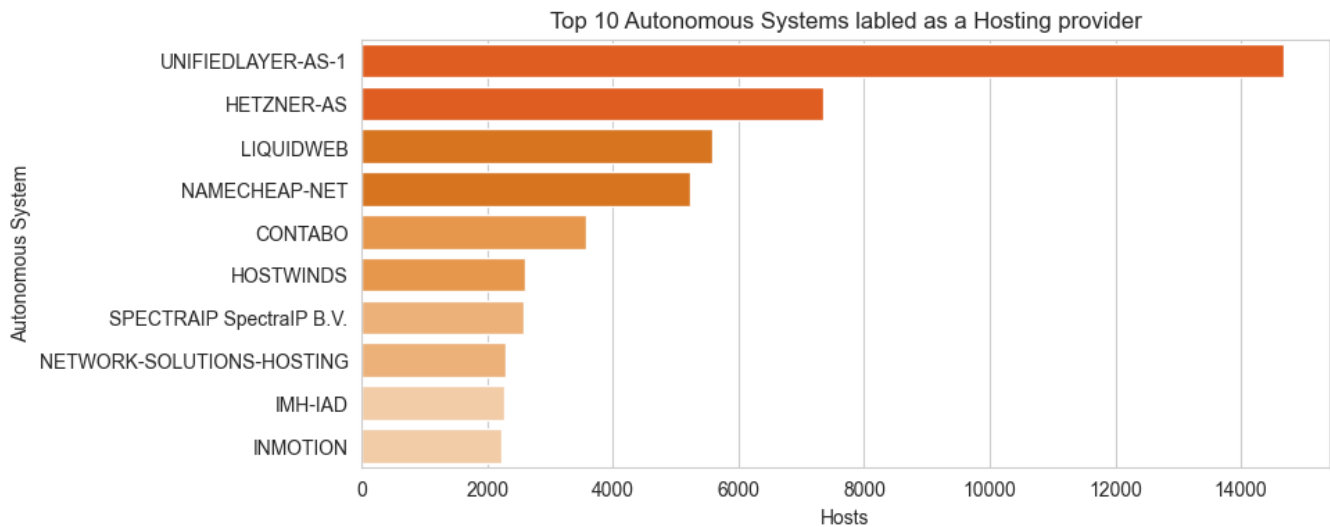


Category	Host Count
Hosting	74,304
Cloud	59,969
CDN	42,434
Telecom	23,690
Other	1,019

This provides an alternative perspective to the AS statistics by showing the actual user base that manages these ill-configured servers, with a total count of 74,304 hosts identified as operating within an autonomous system specifically designated as a hosting provider. These hosts and the people running these hosts represent a diverse range of server management expertise, spanning from individuals running personal blogs to web hosting resellers and businesses with limited or no internal IT support.

These types of organizations often seek more robust capabilities than what a standard web hosting service can offer, nor do they want to deal with any of the issues that come with moving to the cloud, but they also don't have the want or need to maintain their own internet infrastructure.

Top 10 Hosting Provider Autonomous Systems



Autonomous System	Host Count
UNIFIEDLAYER-AS-1	74,304
HETZNER-AS	7,342
LIQUIDWEB	5,581
NAMECHEAP-NET	5,222
CONTABO	3,563
HOSTWINDS	2,597
SPECTRAIP SpectralIP B.V.	2,576
NETWORK-SOLUTIONS-HOSTING	2,285
IMH-IAD	2,270
INMOTION	2,231

While we cannot definitively explain the disparity in the number of hosting providers versus hosts on other types of ASes, we can theorize about why this might happen.

It is likely that these ISPs cater to individuals and small businesses, many of which possess minimal or non-existent IT departments. Consider [Namecheap](#), a one-

OVERALL

stop-shop platform offering domain name purchasing and hosting services for those domains. Their focus is not on targeting large corporations with an extensive online presence but instead on providing a user-friendly service tailored to a diverse customer base. And these types of customers probably don't even know they misconfigured one of their servers since they don't have the expertise or insight to find them.

This points out another more significant issue. When dealing with misconfigurations found within large organizations and companies, finding a point of contact is (usually) straightforward; these organizations often have a proper [security.txt](#) file on their hosts or a security contact can easily be found on sites like LinkedIn.

But it's much more challenging to find contact information for these small businesses and individuals that run their sites and applications from a hosting provider like the ones above. Some providers (like HOSTWINDS) will have a fully functioning [Referral Whois](#) - a setup that can give some insight into the owners of a specific IP address, but more often than not, either RWhois is not available, or it is opt-in only (the customer has to enroll).

They are difficult to attribute, and they are hard to contact. So the only real option is to hope these site owners find their mistakes before someone else does.

The predominant portion of the files we observed pertained to websites, encompassing various elements such as images and HTML content, and in all probability, were meant to be online (maybe not exposed via an open directory). But a substantial number of hosts were observed sharing content that the host owner may not even be aware of.

OVERALL

Below is a table illustrating the assortment of file types discovered throughout our investigation, along with the respective number of hosts on which these file types were found and the cumulative size of data across all of them. The file extension to category mappings was mainly based on the work found in [dyne/file-extension-list](#).

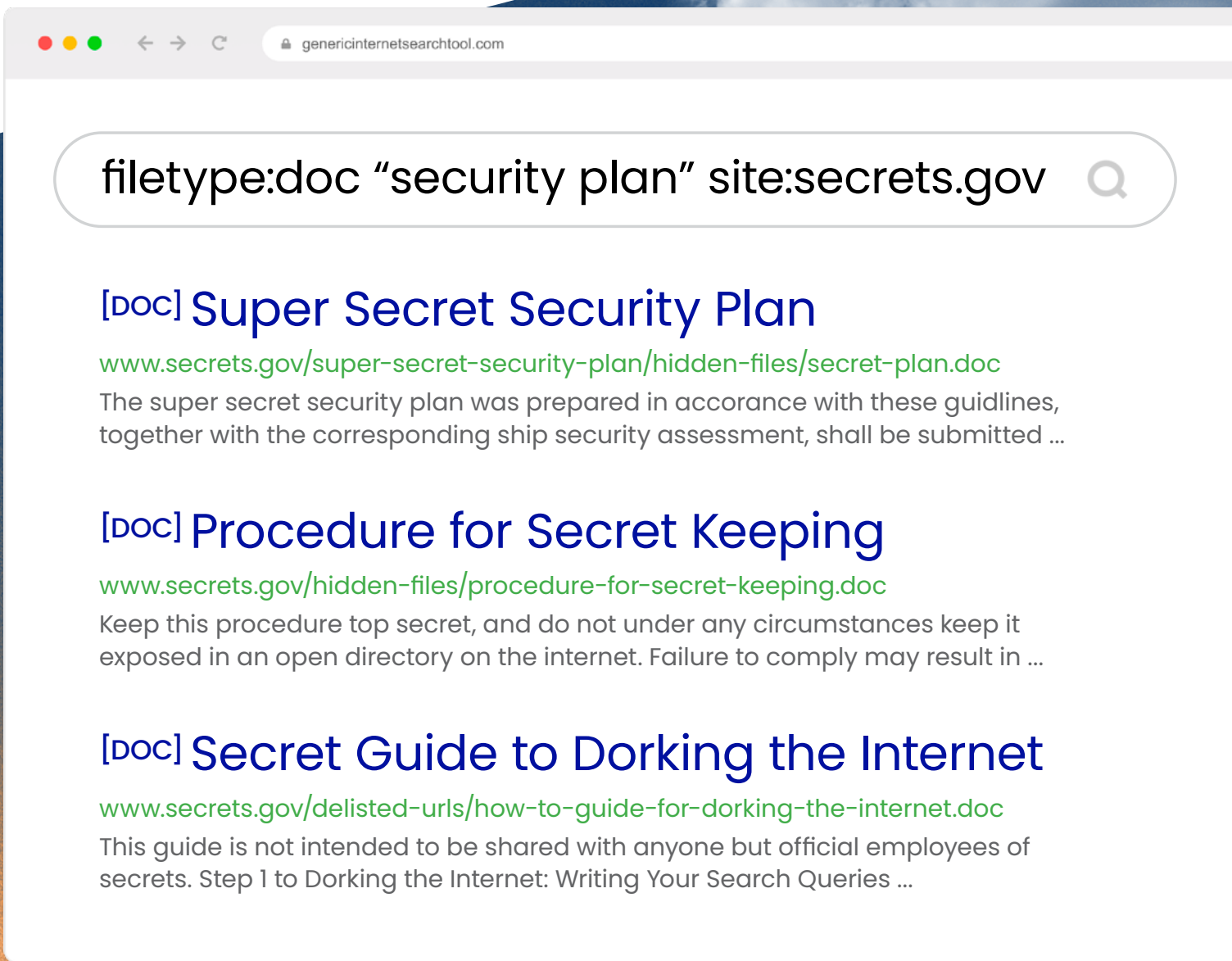
Note: these are based on the file extension, so if a file is compressed (i.e., .sql.bz2), the file will be categorized using the final file extension.

File Type	File Count	Host Count	Total Size
Images files containing image data (.png, .jpg, etc..)	208,595,961	51,741	50,300.20 GB
Archives files and directories that are combined into a single file (.zip, tar, gz, .bz2, .dmg, etc..)	51,448,372	41,654	548,134.81 GB
Documents document-related files (.doc, .rst, .pdf, etc..)	37,880,385	60,409	44,959.71 GB
Web files related to websites (.html, .php)	24,449,096	61,920	3,357.19 GB
Audio audio-related files (.mp3, .wav)	18,840,364	9,029	41,358.46 GB
Code files related to programming (.go, .jsp)	15,514,895	59,838	28,893.66GB
Data Exchange human-readable data structures (.json, .yaml)	13,835,492	39,459	1,997.45 GB
Business files containing business-related data (.xls, .tsv, .vcf, .ics)	11,894,178	8,467	5,814.00 GB
Video files containing video data (.mov, .mkv)	9,643,559	9,965	683,209.06 GB
Log Files app and server log files	7,458,453	13,705	15,236.98 GB

OVERALL

File Type	File Count	Host Count	Total Size
Executable files containing executable code (.exe, .dll)	3,654,925	18,441	60,476.24 GB
Cryptography crypto-related files (.sig, .pem, crt)	2,060,704	9,297	212.93GB
Database files containing database information (.sql, .mdb)	523,158	11,913	3,663.45 GB
Configuration files with configuration data (.conf, .cfg)	328,201	13,966	3.42 GB
Shell Files shell-related files (.bashrc, .profile)	252,109	15,536	195.41 GB
Fonts files containing font data (.ttf, .woff2)	226,461	10,954	33.01 GB
Books files used for e-books (.mobi, .epub)	140,812	372	548.69 GB
Presentation files used for presentations (.ppt, .pptx)	23,095	1,629	149.17 GB
Financial files used in financial applications (.mny, .qfx)	33	4	0.14 MB

Sensitive Data Exposures



The image shows a browser window with the address bar containing "genericinternetsearchtool.com". The search bar contains the query "filetype:doc 'security plan' site:secrets.gov". The search results are displayed in a list format, with each result including a document type indicator, a title, a URL, and a snippet of text.

filetype:doc "security plan" site:secrets.gov 🔍

[DOC] Super Secret Security Plan
www.secrets.gov/super-secret-security-plan/hidden-files/secret-plan.doc
The super secret security plan was prepared in accordance with these guidelines, together with the corresponding ship security assessment, shall be submitted ...

[DOC] Procedure for Secret Keeping
www.secrets.gov/hidden-files/procedure-for-secret-keeping.doc
Keep this procedure top secret, and do not under any circumstances keep it exposed in an open directory on the internet. Failure to comply may result in ...

[DOC] Secret Guide to Dorking the Internet
www.secrets.gov/delisted-urls/how-to-guide-for-dorking-the-internet.doc
This guide is not intended to be shared with anyone but official employees of secrets. Step 1 to Dorking the Internet: Writing Your Search Queries ...

Sensitive Data Exposures

Our analysis process strictly involved indexing publicly accessible data, specifically focusing on filenames, directories, and other metadata. As a result, we did not download or access the actual content within the files. Consequently, it is not feasible for us to definitively state whether sensitive or private information exists within these files. However, by applying pattern matching to file names, extensions, and paths, combined with an examination of modification timestamps and file sizes, we can form a general assessment regarding the potential presence of sensitive information.

LOG FILES

We found that 17,657 hosts were exposing 10,199,077 log files, totaling over 28,000 gigabytes (27 terabytes) of stored data.

Application logs are likely benign from a security perspective, provided that the logs don't divulge too much information about the goings-on of a runtime environment. Application debug logs, on the other hand, could contain a wealth of information about the runtime environment. HTTP access logs could contain information about the IP addresses that are being used to access the service and thus could provide insights into the relationship between the application and its users, giving an attacker the ability to launch more targeted attacks.

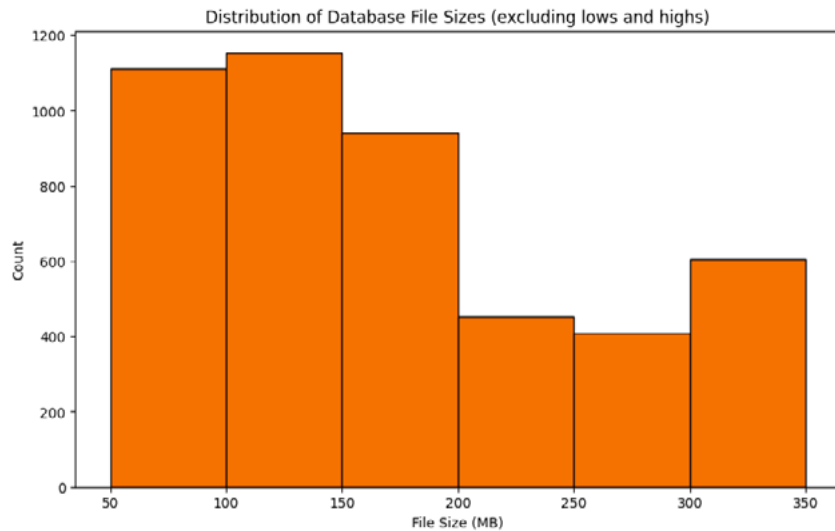
DATABASES

For this subset of the data, we have filtered out some of the rows that have a "database-like" file extension but were not actually "true" database files, for example:

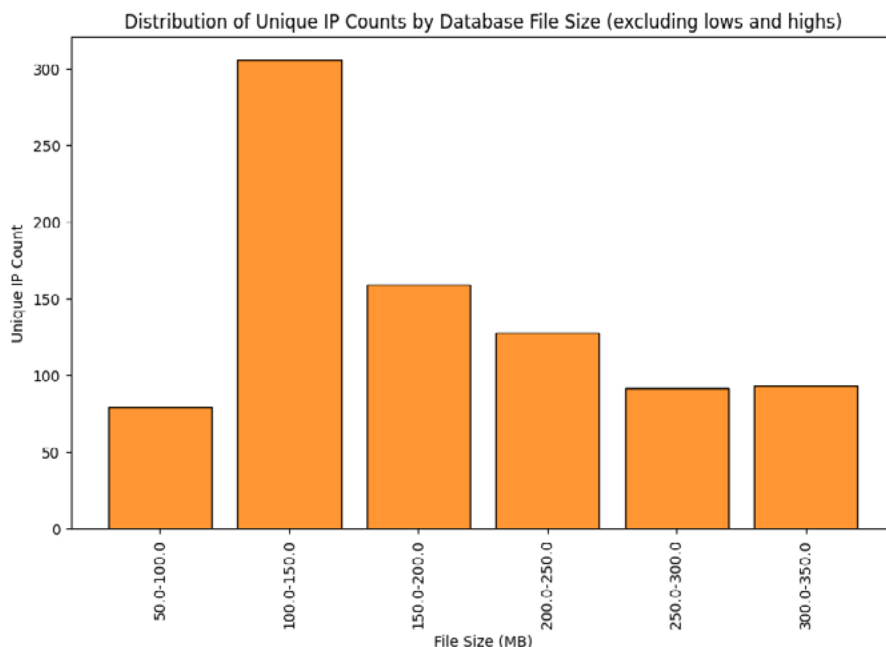
- **Thumbs.db** - are a cache of thumbnails that Microsoft operating systems use to display images in directories.
- **MYI** files - these are MySQL MyISAM Index files that do not contain anything that could be considered questionable.

We discovered a total of 9,897 internet hosts revealing over 6,039 gigabytes distributed across 573,966 individual files we considered a database. While this quantity may initially appear alarming, it is crucial to analyze the file sizes and other metadata of these files to ascertain the potential presence of complete database datasets or backups (which could indicate a data leak).

Studying file sizes, a significant portion of the examined database files fell within the range of zero to fifty megabytes, encompassing 353,169 files. Conversely, we encountered a mere four db files (across four distinct hosts) exceeding 120 gigabytes of data (indicative of a large database backup). By excluding these extreme cases and only including the 99th percentile of file sizes, we gain a clearer perspective on hosts housing sizable db files.



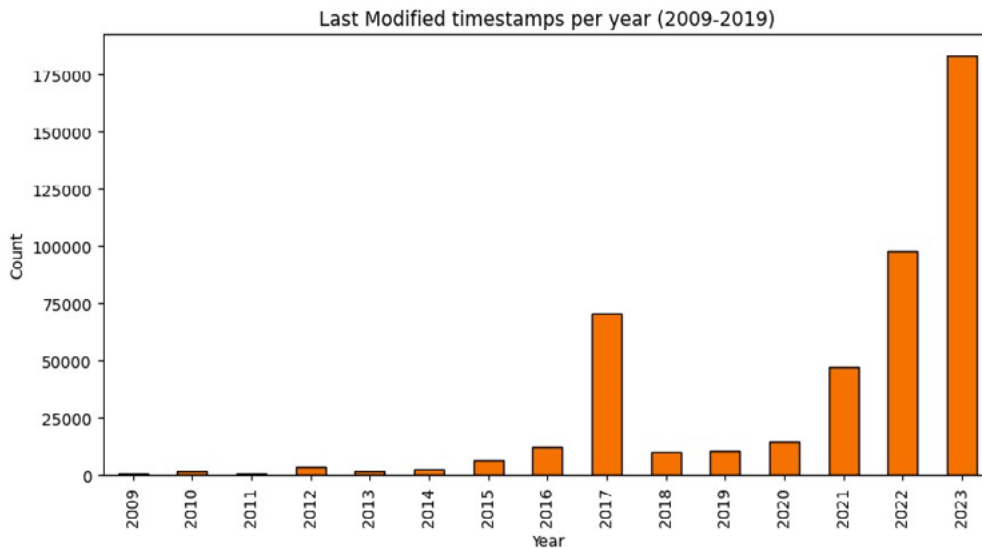
The accompanying graph illustrates the prevalence of database files within the size range of 100-150MB, constituting a total of 1,154 files. Another notable bump we saw here was the 300-350MB range, comprising 605 files. However, as we will explore in the subsequent graph, those larger sizes are confined to a limited number of hosts.



SENSITIVE DATA EXPOSURES

The data shows a similar pattern with the distribution of unique host counts in the 100–200 range, mirroring the trend observed in the file count distribution. Notably, the highest concentration of unique hosts is found within the 100–150MB file size range, with a total of 306 distinct hosts. Additionally, there are 159 hosts falling within the 150–100MB range. However, we do not observe a similar trend in the 300–350MB ranges, indicating that these larger database files are limited to a select number of hosts.

We can also analyze the last-modified timestamps of these files to gain insights into the temporal information. For files we have timestamps for, we see that a significant portion of the database files, precisely 183,313, were either created or modified this year, while 97,708 files were created or modified in 2022 and 47,025 in 2021.



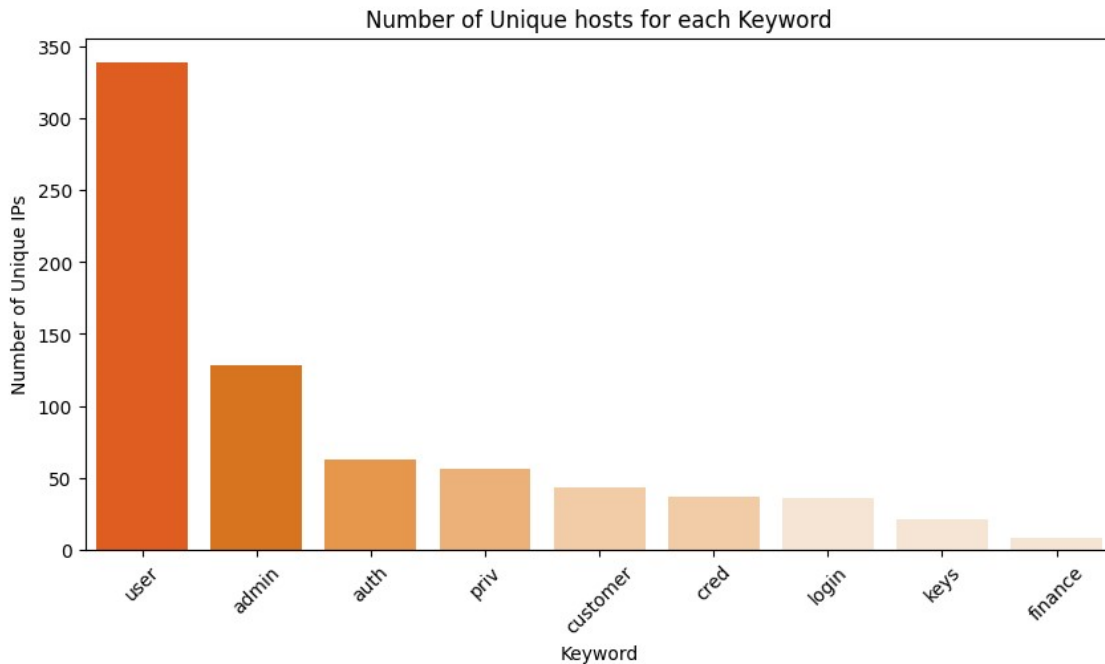
Some interesting tokens in the above pictures indicate these are files that potentially contain sensitive information, for example:

- **backup** - 713 occurrences, which indicates that these files are part of a database backup
- **dump** - 334 occurrences, a term often used to describe a full copy of a database (often generated by `mysqldump`, a database backup tool)
- **com** - 292 occurrences, found as part of a domain name in a path, which may indicate this is a full database dump for an entire website.

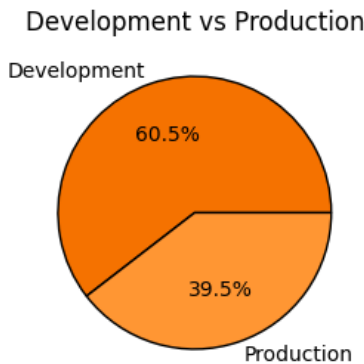
Alternatively, we observe some tokens that could be interpreted as not so dubious:

- **schema** - 308 occurrences - while the exposure of a database schema is not intended for public viewing, it is a relatively less severe issue than a complete database backup.
- **aarch64 / ppc64le** - 269 / 127 occurrences, respectively - these are directories containing software packages like SRPMs, which often are stored in directories based on the CPU architecture they are compiled for. In other words, these are probably database files that are distributed with open-source software.
- **EPEL** - Similar to the last bullet point, the occurrences of EPEL files can be attributed to open-source packages distributed under the label of "Extra Packages for Enterprise Linux" (EPEL)
- **references** - 299 occurrences - this likely indicates reference data, which makes it more likely to be test-like data and less likely to be sensitive data

We can also use the full file path to look for specific keywords that may indicate the database file in question may contain sensitive information. In this example, we used the following terms: "priv," "keys," "admin," "login," "cred," "auth," "user," "finance," and "customer." The following graph shows the number of unique hosts that housed a file with one of these terms somewhere in the full file path.



In the same vein, we can also attempt to determine if any of the files indicate they are either part of a development or production-related project by grouping several keywords together. In this example, any path or filename that contains the terms "dev," "test," or "staging" are labeled as "Development," while files containing the terms "prod," "live," or "prd" is labeled as "Production."



Here we see that we observed 43,533 (63.5%) database files containing a "Development" label and 25,427 (39.5%) database files with a "Production" label.

From our perspective, this data indicates that there is a potential goldmine of database-related information exposed on the internet that could be used by malicious parties to exploit weaknesses, compromise sensitive information, and launch targeted attacks.

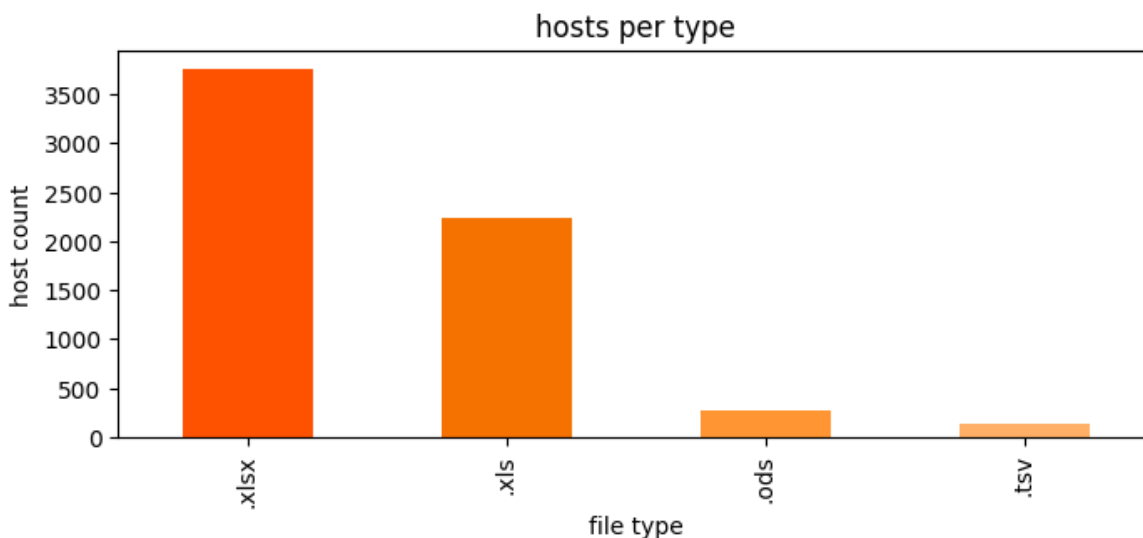
NETWORK CAPTURES

Network captures refer to files containing unprocessed network data obtained using tools such as tcpdump and Wireshark. These files aid in debugging network or security issues by enabling administrators to gain visibility into the system's inbound and outbound network traffic. The contents of these files can encompass a broad spectrum, ranging from unencrypted communications between clients and servers to valuable insights pertaining to an internal network. Having these readily available for just anyone to download is not ideal.

We observed 1,024,803 files (totaling over 203 gigabytes in size) with characteristics of a network packet capture, but a small percentage of hosts are responsible for a significant proportion of the total number of network capture files (1.7% of hosts account for 90% of the total number of these files). We observed only 172 distinct hosts which actually housed those files.

SPREADSHEETS

Everyone's favorite file format! The spreadsheet is most commonly used for the analysis and manipulation of any sort of data in a tabular structure, but one of its key areas of use is for handling financial data, catering to the needs of businesses, accountants, and financial analysts.



Here we will look at four of the most commonly used file types that include spreadsheet-like data:

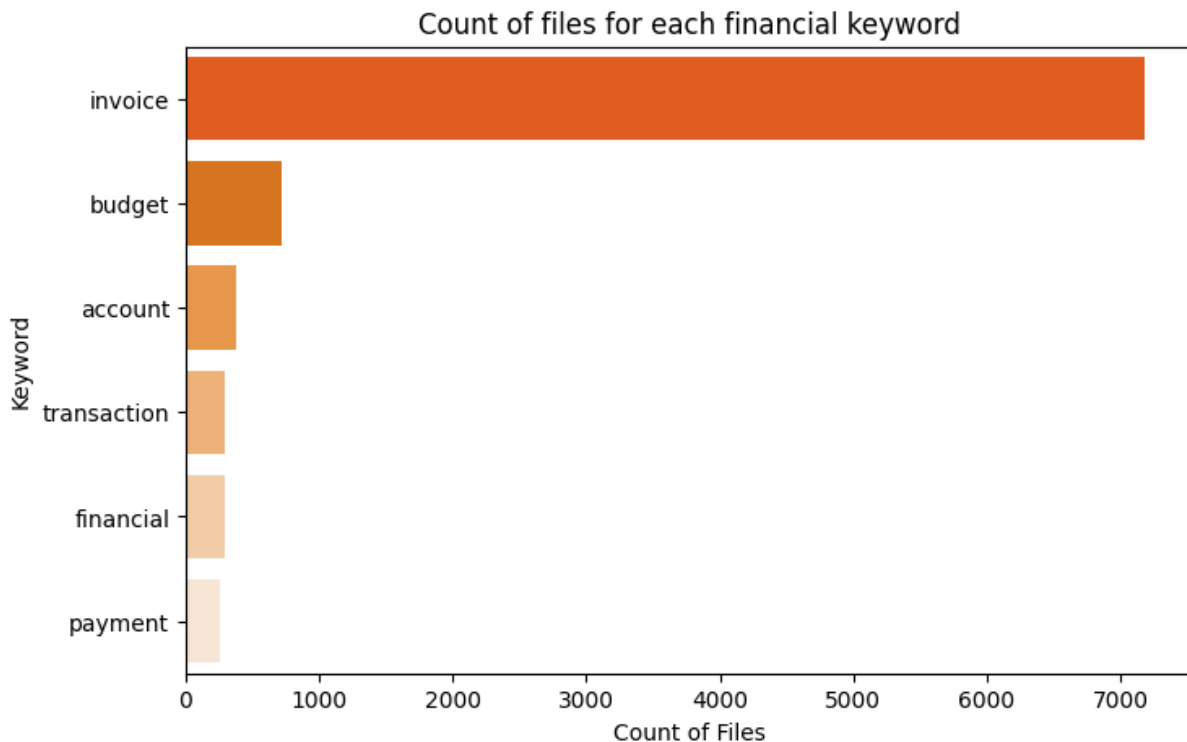
- XLSX - XML-encoded spreadsheet files that are generated via Microsoft Excel
 - ◇ 3,760 hosts had one or more XLSX files

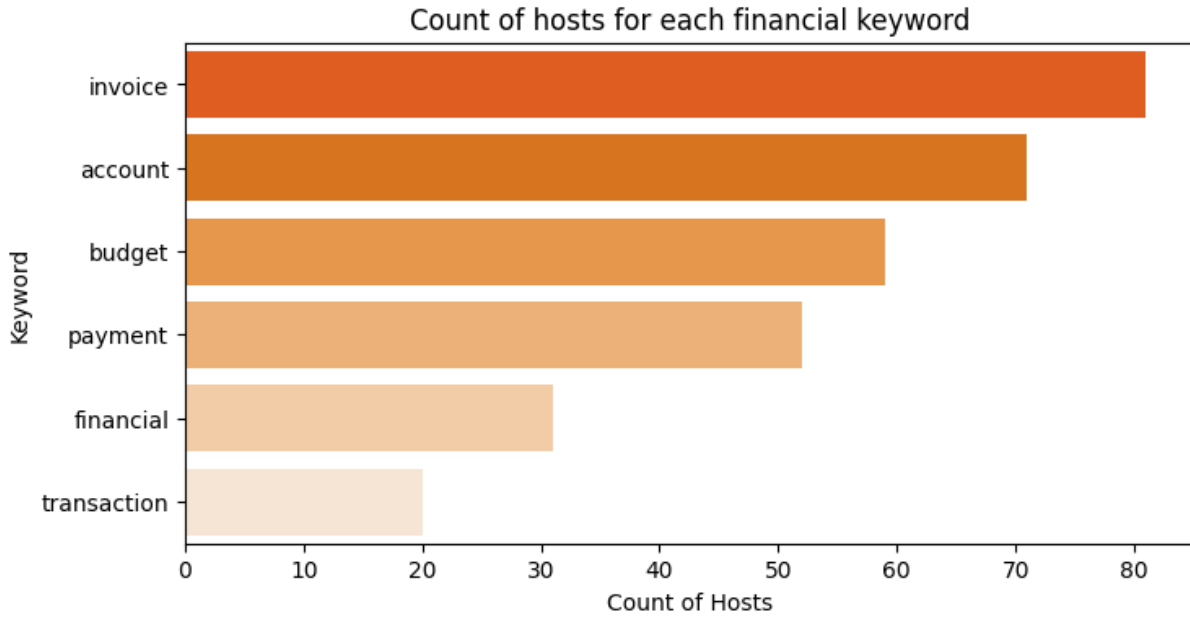
SENSITIVE DATA EXPOSURES

- XLS – Proprietary file format used by older versions of Microsoft Excel
 - ◊ 2,242 hosts had one or more XLS files
- ODS – (Open Document Spreadsheet) is an open standard format used by open-source office applications like LibreOffice
 - ◊ 264 hosts had one or more ODS files
- TSV – a plaintext file format used for storing tabular data.
 - ◊ 138 hosts had one or more TSV files

In total, we observed 5,133 unique hosts serving 1,118,599 spreadsheet files with one of these four file extensions, totaling over 370 gigabytes worth of data.

Given that these types of files are most often used for business and financial data, we looked for filenames that contained the tokens “invoice,” “payment,” “account,” “financial,” “transaction,” or “budget,” which were then translated multiple languages such as Spanish, French, German, Italian, and Dutch to determine if any of these files may reference financial data that shouldn’t be publicly available. The results were... interesting.





Keyword	Host Count	File Count
invoice	81	7,186
account	71	382
budget	59	717
payment	52	256
financial	31	292
transaction	20	256

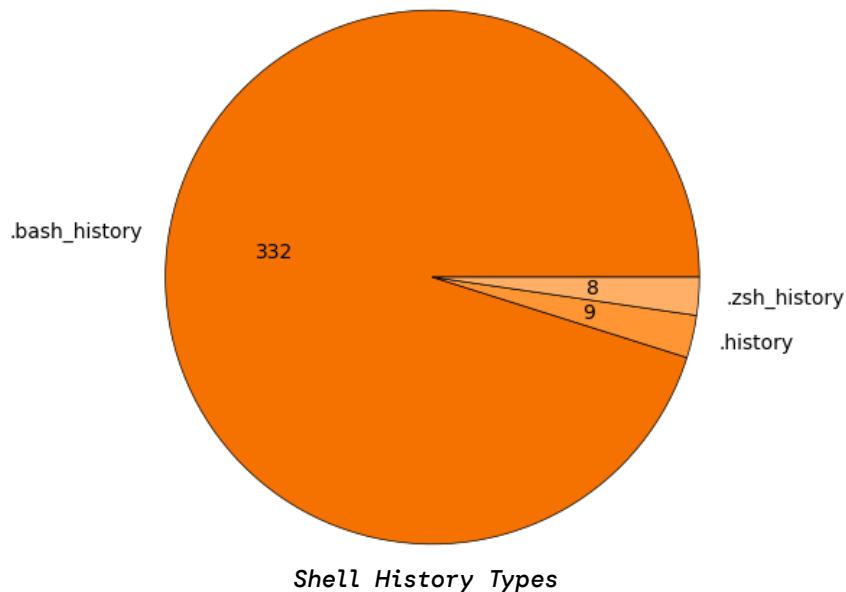
An anomaly here is the “invoice” keyword. There were 7,186 files that contained this term, but only 81 hosts that served them. What we can derive from this is that a few hosts accidentally exposed a large number of their electronic invoices within an open directory.

used by many popular Mail Transfer Agents (MTA) like Postfix, qmail, and Sendmail to store received and sent emails. But upon closer inspection, it seems that all of those files are related to the [Podesta Wikileaks emails](#); they are already public information.

While there are many email files exposed to the internet, the majority appear to be internal, machine-generated communications that likely don't contain sensitive information.

SHELL HISTORIES

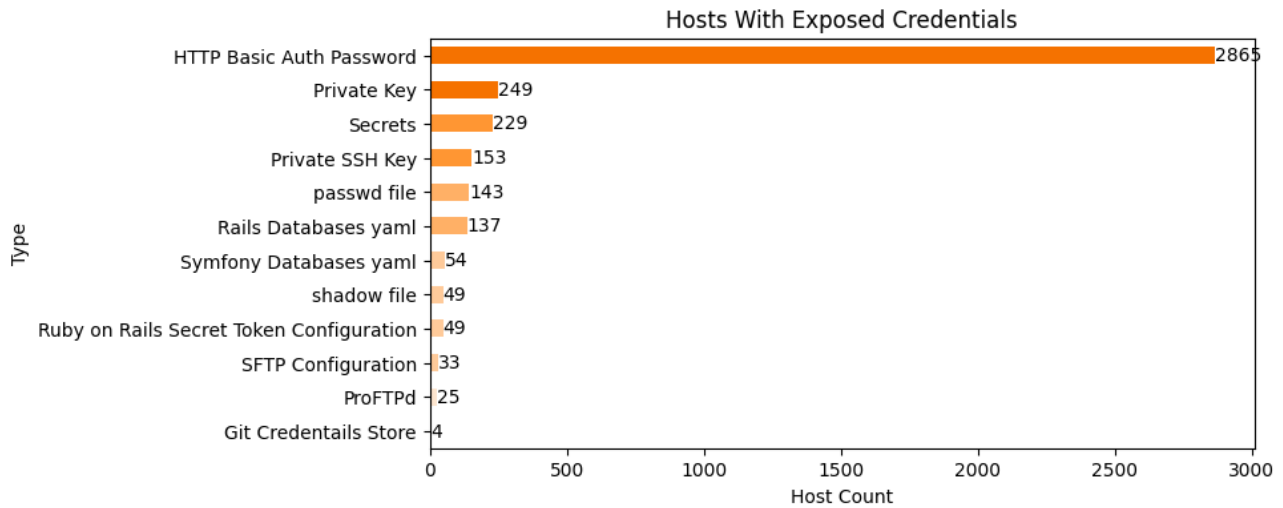
Shell history files contain the commands that were executed by a user on a shell, like Bash or ZSH. These files can potentially contain sensitive information like environment variables and even passwords (who hasn't accidentally pasted a password or key into our shell?). Luckily most of us have the luxury of knowing that nobody will see these files other than ourselves. Unfortunately, 349 shell history files are spread across 317 distinct IP addresses.



By far, the most common shell history file we observed was ".bash_history" with 332 instances, and trailing by quite substantially is 9 CSH/TCSH ".history" files and only 8 ZSH history files.

EXPOSED CREDENTIALS

This section describes several files we encountered, based on patterns in the file path and file name, that often contain credential-related information, like usernames and passwords, along with keys for various encrypted services.



Hosts - 2,865, Files - 2,869

HTTP Basic Auth Password (.htpasswd)

- An htpasswd file is a text-based configuration file deployed within a specific host directory, serving as a directive to the web server for initiating a basic authentication exchange with the client. This mechanism restricts access to a designated section of a website exclusively to individuals with the correct credentials.
- The passwords within these files are not typically stored in plaintext, but hashed, which still renders the credentials susceptible to being cracked through brute-force techniques.

Hosts - 249, Files - 498

Private Server Keys

- These files contain some combination of "server.key", a typical filename pattern for storing a server's private SSL key.

Hosts - 229, Files - 451

Secrets

- These files contain filenames containing “credentials” or “secrets,” accompanied by one of the following file extensions: .yaml, .json, .conf, .ini. This naming convention is commonly employed to store application credentials.

Hosts - 153, Files - 236

Private SSH Keys

- These are files generated by the SSH client which contain the private SSH key of a user. While passwords can be set to unlock the key, this is an optional feature. It is never recommended to share SSH keys, let alone make them available for download.
- OpenSSH will generate recognizable names for its different keys, such as id_dsa, id_rsa, id_ed25519, or id_ecdsa (and their public key counterparts suffixed with “.pub,” not included here). Alternatively, windows-based SSH clients such as Putty will write the key to a file ending in .ppk (Putty Private Key).

Ruby on Rails

- `database.yml`

Hosts - 137, Files - 217

- ◊ Since most Ruby-on-Rails-based applications utilize a database of some sort, the `database.yml` file is used to set up all the information needed to connect to the database, including the host, username, and password.

- `secret_token.rb`

Hosts - 49, Files - 65

- ◊ This file contains a secret token used to sign cookies set by the server application. If an attacker has access to this token, they can gain the ability to impersonate any user within the application.

Hosts - 54, Files - 81

Symfony databases.yml

- The Symfony PHP framework is a set of reusable PHP components that are commonly used in microservices, content management systems, reporting applications, and eCommerce platforms—a moderately popular framework, sitting at 28.5k stars on GitHub.
- The databases.yml file contains just about everything an attacker would need to know to access an application's database, including the database type, hostname, port, username, and password.

*NIX Password Files

- **etc/passwd**

Hosts - 143, Files - 220

- ◇ The etc/passwd file, commonly found in Unix-like operating systems, have a plaintext format and serve as system databases storing crucial user account information.
- ◇ However, it should be noted that these files do not contain the actual hashed credentials. Although, if you come across the presence of such a file on a public host, it typically indicates a potential issue or misconfiguration.

- **etc/shadow**

Hosts - 49, Files - 92

- ◇ etc/shadow files are similar to etc/passwd files but contain usernames and the actual hashed passwords.
- ◇ These files can easily be cracked with tools like John the Ripper
- ◇ The existence of such a file indicates a critical misconfiguration.

Hosts - 25, Files - 28

proftpdpasswd

- These files are generated and managed by the [cPanel](#) web hosting administration system containing usernames and passwords for FTP users
- These files can contain multitudes of different usernames and passwords, and should never be exposed to the public.

BONUS:

git-credential-store (.git-credentials)

- These are files that have been written via the git helper tool [git-credential-store](#), which will store unencrypted passwords to disk.
- While there are not many hosts exposing such data, the fact there are even a few is concerning, as this may give attackers access to entire development infrastructures.

Hosts - 4, Files - 4

WHY DOES THIS HAPPEN?

Most major web servers do not enable directory listings by default; some even make enabling them a little tricky. So why is it we're seeing so many instances of this issue out on the internet? We can't say for sure, but we make a few hypotheses based on our observations:

- It appears that some of these servers have been hastily configured, indicating situations where administrators sought quick access to files on old servers. So, directory listings were enabled, and old data was downloaded, but cleanup was neglected.
- We have observed the utilization of the Python built-in HTTP server (`python -m http.server 8080`) and its directory listing functionality on numerous hosts. Executing this on the command line exposes the current directory to the public, and if they're never stopped, they will continue to serve.
- A significant number of the directory listings resemble those of hosting resellers who have implemented only minimal measures to safeguard their customers' data. Many resellers manage their servers solely in administration UIs like [cPanel](#) or [Plesk](#), and anything outside of those UIs will be neglected.

WHAT CAN BE DONE?

Open directories are a very old problem that has persisted for decades. For some, finding these hosts and sharing them with communities and friends is a hobby. And while it is interesting to see what's out there, it can also be a massive problem if the wrong things are exposed, and the data makes its way into the wrong hands.

[You can use simple Censys search queries to find heaps of exposed directory listings online.](#) These queries can be used as a starting point to find any assets that you may own. And if you have identified one, then fixing that server is one [Google search away.](#)



About Censys

Censys is the leading Internet Intelligence Platform for Threat Hunting and Exposure Management. We provide governments, enterprises, and researchers with the most comprehensive, accurate, and up-to-date map of the internet to defend attack surfaces and hunt for threats.

Founded by the creators of ZMap, trusted by the U.S. Government and over 50% of the Fortune 500, Censys' mission is to be the one place to understand everything on the Internet.

For more information, visit www.censys.com.