censys

WHERE THE WEIRD THINGS ARE

How to Investigate Unusual Internet Artifacts with Censys Search Data

The Internet is a Big, Weird Place.

And it's no wonder; there's *a lot* going on here. Think of all the things we can access at any given moment – the products we can shop, the apps we can launch, the IoT devices we can engage...the list is endless. Almost every part of how we live our lives and do business is in some way tied to the internet. That equals a lot of touchpoints across an increasing number of servers, networks, and cloud-connected devices. In addition to accessing what's already on the internet, we're also contributing to its vastness and variety: by uploading our own content, spinning up our own sites, housing our personal information, or logging into remote work sites from multiple devices.

All of this activity across so many access points can pose a challenge for security teams, who are tasked with discerning which activity is "normal" and which might be a threat to their organizations. The malicious tactics that bad actors deploy when looking for a way in can be hard to discern at-a-glance. A security team might discover software that claims to be one version but is actually another, or control panels with no authentication. Are these cybersecurity threats, or harmless anomalies from someone who wasn't quite sure what they were doing?

In these instances, it's critical that threat hunting security teams have a tool they can use to quickly gather more intel and make proactive security decisions. Doing so could mean the difference between being able to carry on with business as usual – and the catastrophic loss of data, money, and brand reputation.

That's why in this ebook, we'll be covering how these security teams can gather the intel they need using Censys Search.

Let's get started.

But First - Who is Censys?

If you're new to Censys, allow us to introduce ourselves. We like to say that we're the one place to find everything on the internet. Meaning, through our best-in-class internet scanning capabilities, we enable security teams and threat hunters to identify attacker command and control infrastructure, locate vulnerable and compromised hosts, uncover trends in activity that occur across the globe, and improve their defensive operations.

In fact, Censys maintains the **most comprehensive view of the public internet in the world by**:

- Multi-perspective global scanning and daily coverage of the largest number of popular ports in the industry to provide on-demand, accurate, and contextual data
- Enabling complex and custom searches to find advances threats, IOCs, and trend analysis
- Providing visibility into all things connected on the internet so you can see each connection and pivot to different assets to further your investigations
- A massive collection of historical data that looks back two years

These capabilities fuel our Universal Internet Dataset, which is the engine behind our powerful Censys Search tool. Censys Search is used by security teams of all sizes in both commercial and government sectors to conduct threat intelligence and proactively protect the organization.

Censys Internet Scanning Intro

Check out this in-depth guide to learn more about how Censys scans the internet.

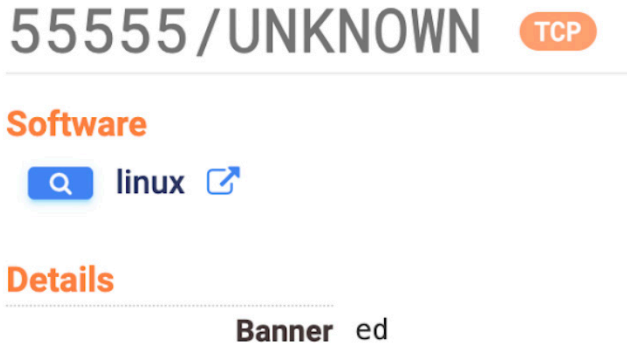
Read Now

For the purposes of this ebook, we'll be focusing on the Hosts dataset – which provides accurate, up-to-date records on the public IPv4 and IPv6 hosts and virtual hosts – and we'll walk you through a step-by-step guide to how we evaluate anything "weird" we might find online.

Read on to let the sleuthing begin!

Step One: Evaluating the "Weirdness" of an Observation.

Our researchers at Censys recently found something weird: a cluster of hosts was running an unrecognized service – all on port 55555, all on one autonomous system, and all with the same cryptic two-character service banner. This is unusual – particularly because of this unique banner message – but not surprising.



But how do we know if this strange cluster of hosts is benign or something of concern? This is where we begin our investigation.

Like any good security practitioner, we need a map of sorts to get started. Here's a rudimentary list of questions that might lead to understanding what this weird thing could be:

1. How widespread is this observation? How many hosts display these characteristics?
2. What autonomous systems are these hosts distributed across?
3. Which geographic regions are these hosts located in?
4. What other services are these hosts running?
5. Was there a spike in the number of active hosts displaying these characteristics on a certain day?

While the first four questions can be answered using Censys Search, the fifth question requires us to expand our toolset. We'll need to access snapshots of our Universal Internet Dataset in Google BigQuery. Let's tackle these questions in order.

Step Two:
Assessing Your Scope.

Censys Search makes it easy to quickly determine the public-facing footprint of an internet phenomenon.

Each particular host in Censys Search is populated with detailed information about its IPv4 address, autonomous system, open ports, services running on those ports, and much more. These attributes are also searchable entities.

For example, we could search for all detected hosts located in Ireland that are running SSH. (See documentation of all searchable fields that are populated for each host.)

When faced with loads of data, we want to hone in on the attributes that make our observed host "weird" so that we can narrow our focus. Here are a few that tend to be useful:

1. service.port
2. services.service_name
3. services.banner
4. Autonomous_system.name

Now let's rewind back to our first research question:

Q1: How many hosts display these characteristics?

We can tackle this by writing a Censys Search query that will grab all the hosts that match those characteristics. Along with the matching hosts, Search will also return the total number of results, the time it took to grab them, and a breakdown of the results by some basic filters.

In the case of our odd 55555 port, we want to filter for hosts with one specific combination of autonomous system, port, service, and banner message. Our Censys Search query looks something like this:

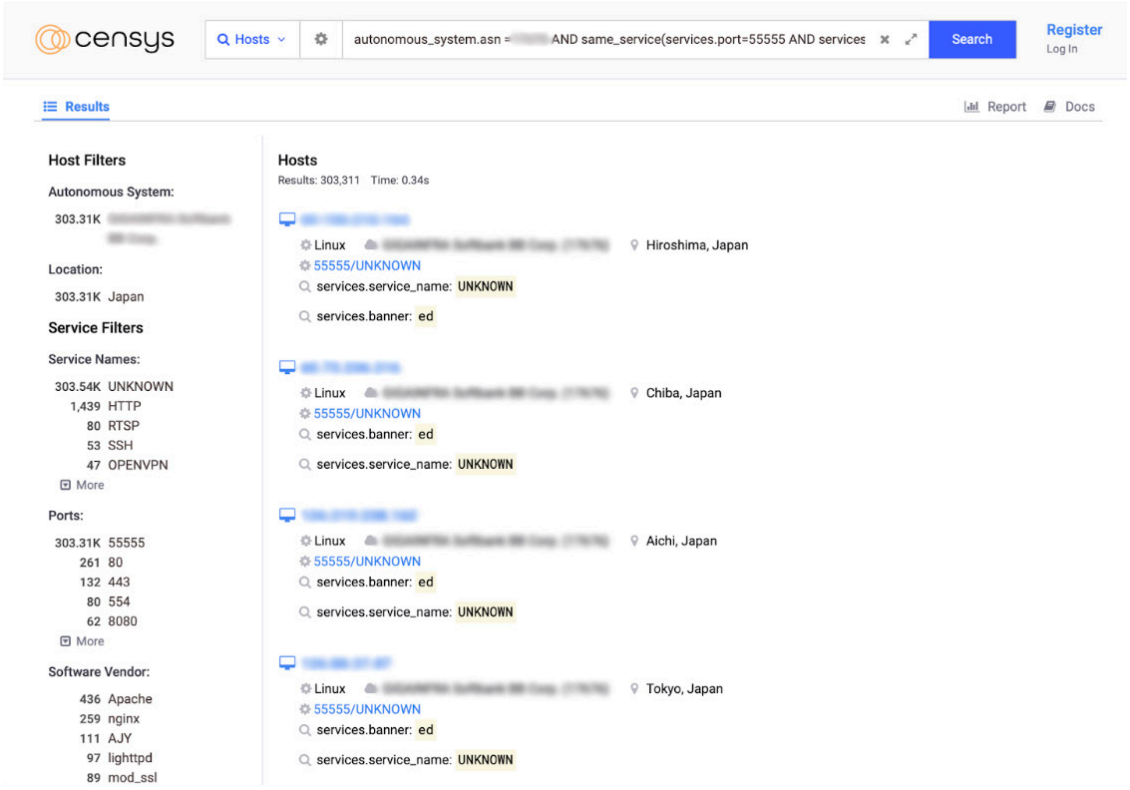
```
autonomous_system.asn=ASNx
AND same_service(service.port=PORTx
AND services.service_name='UNKNOWN'
AND services.banner='BANNERx'
AND services.truncated=`false`)
```

(line breaks added for clarity)

Adding that handy services.truncated=false at the end of the query will exclude any hosts that are running more than 100 services, which can often be a marker of honeypots or pseudo services. You can refer to our Search FAQ to learn more about the truncated field.

Using ":" instead of "=" is the syntax for running a "fuzzy" search that doesn't require an exact match. This is best for cases where you only have a keyword or snippet to go on. To learn more about how to write well-formed Censys Search queries, check out our Search documentation and these example Host queries.

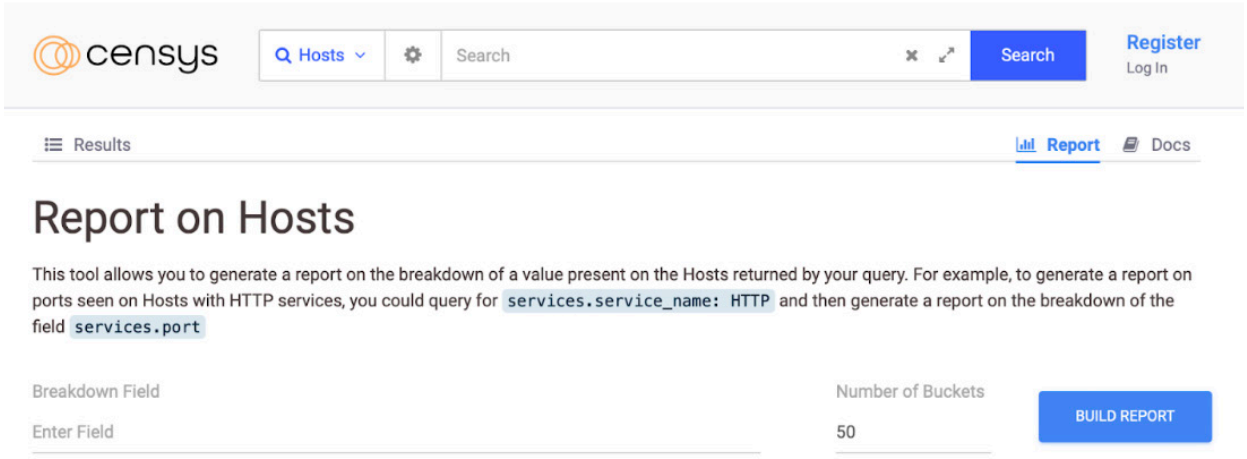
Running the above query took 0.34 seconds and returned 303,311 results:



In the grand scheme of things, that's not an enormous number of hosts…but it's also not a small number. Let's continue down our list of questions.

Step Three: Characterizing the IP Space.

Now that we have an easily indexable list of all our “weird” hosts, we can dig deeper into their other characteristics using the Reports feature of the Censys Search interface.



This feature provides an easy breakdown of search results, allowing us to see how search results compare with each other across a specific attribute. To access it, simply click on the Report tab in the upper right hand corner of the search results page.

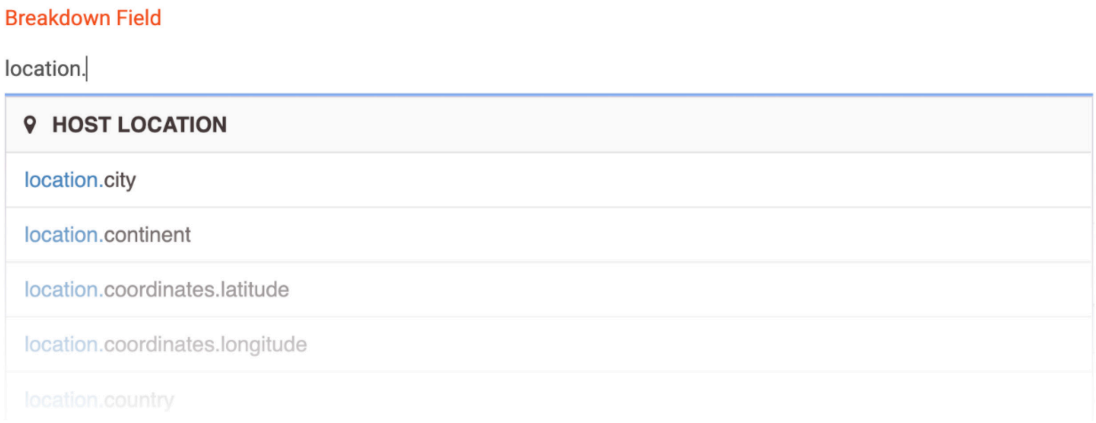
Continuing on with our questions:

Q2: What autonomous systems are these hosts distributed across?

We can easily get a breakdown of our “weird” hosts by autonomous system by generating a report and specifying the attribute autonomous_system.name or autonomous_system.asn as the Breakdown Field.

Q3: Which geographic regions are these hosts located in?

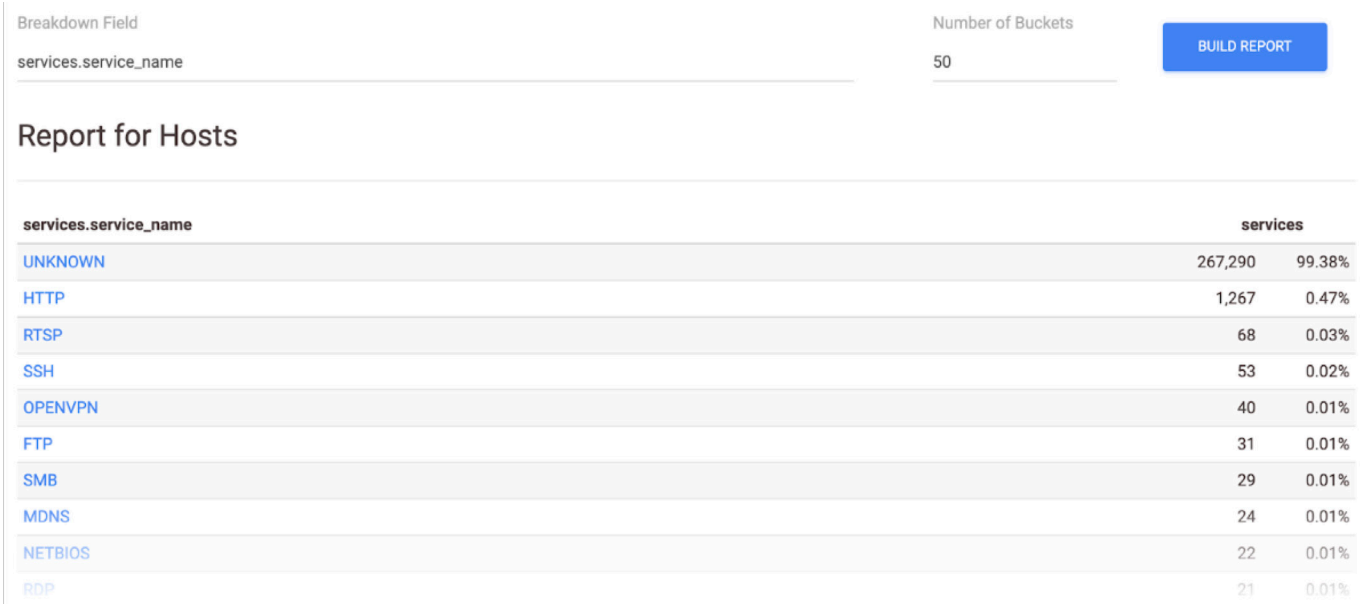
Let’s generate the same report as the above, except now we can set our Breakdown Field to any one of the attributes under the location field, depending on whether we want to investigate at the scale of cities, countries, continents, etc:



Q4: What other services are these hosts running?

Each service running on a host is captured by the services.service_name attribute. Generate a report with this attribute set as the Breakdown Field to get further insight, noting that there can be multiple services running on one host. Keep your eyes peeled for any services that are particularly known to be frequently exploited by threat actors, such as SMB, RDP, and FTP.

Step Four:
Examining Your Report.



services.service_name	services	
UNKNOWN	267,290	99.38%
HTTP	1,267	0.47%
RTSP	68	0.03%
SSH	53	0.02%
OPENVPN	40	0.01%
FTP	31	0.01%
SMB	29	0.01%
MDNS	24	0.01%
NETBIOS	22	0.01%
RDP	21	0.01%

At the very top of our report, we see that a whopping ~99% of our weird hosts are running some service labeled "UNKNOWN." What does that mean?

Censys can detect 105 Layer 7 protocols (a.k.a. services). While these automatic protocol detection techniques are pretty sophisticated, sometimes they come up inconclusive – either because the service doesn't adhere to the protocol in some way, or because we don't have a protocol-specific scanner written for it.

When the scanner is not able to recognize the service running on a particular port, it categorizes it as UNKNOWN. It could be interesting to dive deeper into what these UNKNOWNs might be, but the presence of them typically isn't an indicator of anything suspicious on its own.

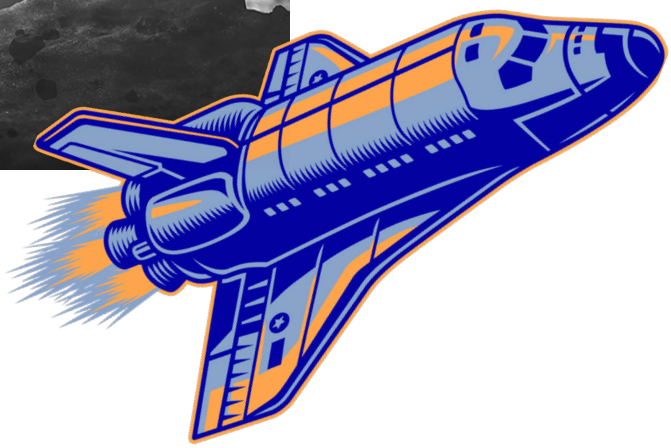
Going down the report, HTTP is the second most common service represented. That checks out, since HTTP makes up most of the services we see in Censys's scan data.

Scanning the list in its entirety, it looks like there's a smattering of hosts running FTP, SMB, and RDP. That warrants a little more snooping.

Click on any of the cells in the report to go to the Censys search results for that subset of hosts.

Examine several of the results, looking at their HTTP response bodies, banner messages, TLS certs, and software versions for anything that seems amiss. Are there any fishy JavaScript elements? Expired or self-signed certs? Software associated with known CVEs? This is not an exhaustive analysis, but it suits our goal of getting a quick feel for what we're dealing with.

Overall, we didn't find anything blatantly concerning here. Onward to the next question!



Step Five: Analyzing Historical Trends in the Data.

Visualizing the fluctuations in the number of active hosts that match these characteristics over time is another key piece of information. A significant increase or decrease in hosts on a certain day could indicate a number of suspicious events, for example the potential start of a threat incident.

Q5: Was there a spike in the number of active hosts displaying these characteristics on a certain day?

To answer this question, we'll need to step beyond the Censys Search Web UI and API. Searches executed on these platforms are always run against hosts in the most recent snapshot of Censys's Universal Internet Dataset. In order to get historical data, we need to run our query over snapshots from multiple days.

On any host page in the Search interface, navigating to the History tab will show you a historical chronology of events, but searches using history are not currently supported. Historical data can be pulled by running SQL queries through Google's BigQuery interface. Enterprise customers who download or access daily snapshots in BigQuery can search the internet as it was observed by Censys at a historical point in time.

The SQL query below counts every unique IPv4 address with services that match our search criteria between the range of two dates, grouped by each day. The `LAG()` function here calculates the difference in number of hosts from one day to the next.

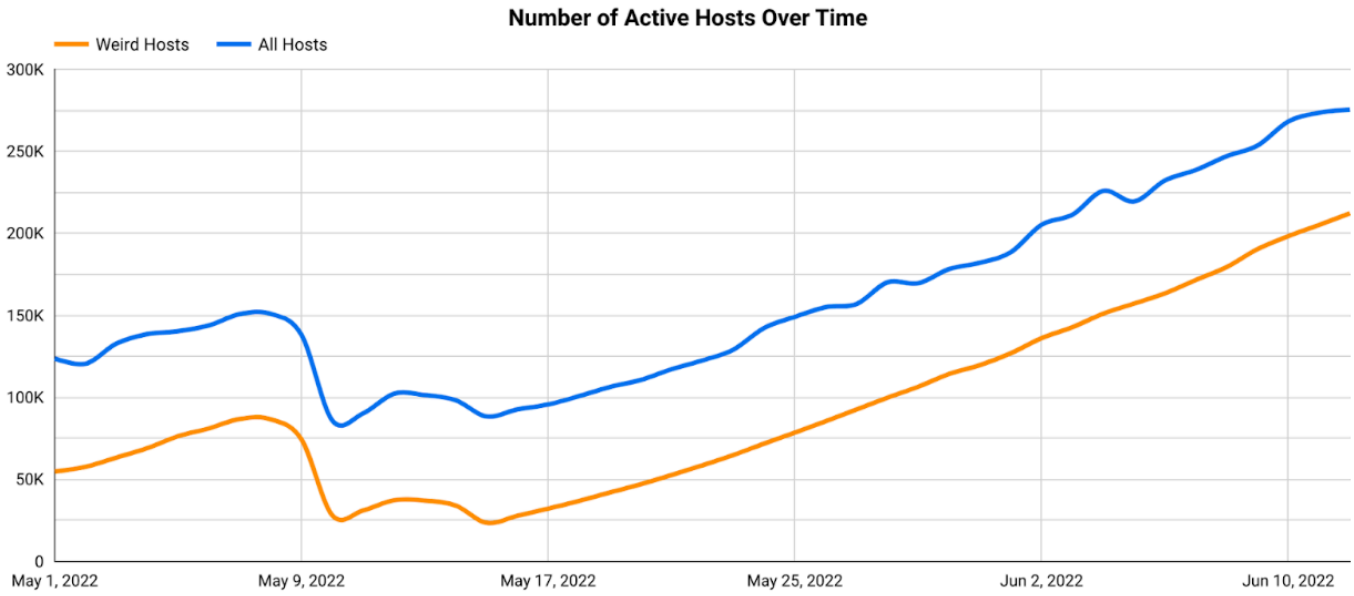
The criteria under the `WHERE` clause in the SQL query will have similar syntax to our earlier Censys Search query.

```
WITH
active_hosts AS (
SELECT
DATE(snapshot_date) AS day,
APPROX_COUNT_DISTINCT(host_identifier.ipv4) AS ipcount
FROM
`censys-io.universal_internet_dataset.universal_internet_dataset`,
UNNEST(services) AS svc
WHERE
autonomous_system.asn=ASNx
AND svc.port=PORTx
AND svc.service_name = 'UNKNOWN'
AND SAFE_CONVERT_BYTES_TO_STRING(svc.unknown.banner)='BANNERx'
AND svc.truncated=`false`
AND DATE(snapshot_date) BETWEEN '2022-05-01' AND '2022-06-14'
GROUP BY
day
ORDER BY
day
)
SELECT
day,
ipcount,
ipcount - LAG(ipcount) OVER (ORDER BY day ASC) AS delta
FROM
active_hosts
GROUP BY
day,
ipcount
ORDER BY
day
```

Step Six: Comparing Trends Across Hosts.

It can also be helpful to compare the trends in "weird" hosts side-by-side with trends in a larger subset of hosts, such as the broader country or industry. In the case of the strange 55555 port we came across, we want to analyze trends in all of the active hosts under this particular autonomous system. **To broaden the group of hosts you pull historical data for, adjust the filters under the WHERE clause in the SQL query from the previous page:**

```
WHERE
autonomous_system.asn=ASNx
AND DATE(snapshot_date) BETWEEN '2022-05-01' AND '2022-06-14'
```



This visualization isn't very useful to us until we know how to attach meaning to the patterns we see.

When looking at fluctuations in internet hosts, there are a few things that should raise some red flags for us. Any sort of large, sudden change in the data is suspicious – such as a large spike or trough on a particular day. The former could indicate the scaling of something malicious, whereas the latter could represent connectivity problems such as an outage, scheduled maintenance, or a security incident. In addition, pay attention to any patterns in one trendline that aren't reflected in the others. These anomalies might hint that something unintended is happening on those machines.

Let's make some initial observations of our graph:

- The trends in our "weird" hosts track pretty closely with those of all hosts in this Autonomous System.
- There is a hump in the data toward the beginning of May, but it's relatively small (~50K increase) and gradual, taking place over longer than a week.
- There are two troughs following that, but again they're both pretty minor and take place over the course of multiple days.
- Notice the continuous increase in both trend lines starting from around mid-May and continuing into June.

What might we hypothesize from this?

Step Seven: Drawing Initial Conclusions.

From our brief analysis, it appears more likely that the “weird” hosts are not anything explicitly malicious. The similarity between the two trendlines and relatively slow rate of the changes in our historical data visualization suggests that these hosts may be part of this network’s intended infrastructure scaling. There is still the possibility that this is something malicious – but now we have a solid hypothesis to report!

Remember that the internet is a BIG, weird, and constantly changing place. **With just a few tools, we were able to understand a small chunk of this cyberspace. If you know which questions to ask, Censys Search data can help build a bridge to the insights that answer them.** Now we have a valuable springboard for a deeper exploration of this phenomenon.

Now It's Your Turn.

Begin your own internet investigation with Censys Search. Security practitioners and their teams at companies of all sizes use Censys Search to proactively defend against advanced threat actors and protect the organization. Our continuous scanning of 101 protocols across the top 3,500+ ports on the full IPv4 address space, and the top 138 IPv4 ports daily, produces a high-resolution map of the public internet that gives best-in-class visibility to threat hunters and other security professionals.

With Censys Search, you can:

- Identify attacker command & control infrastructure
- Research new and emerging threats
- Understand state prior to compromise
- Enrich network indicators

Users can choose a Censys Search package that best matches their goals.

Community	Pro	Pro Plus	Enterprise
Free and recommended for testing different use cases	Ideal for small threat intelligence or incident response teams	Ideal for scaling threat intelligence or incident response teams	Ideal for larger scale threat intelligence or cyber operations

Interested in learning more? Check out censys.io/data-and-search.