RAPID7

RESEARCH

Cloud Misconfigurations Report

BY TOD BEARDSLEY AND KWAN LIN

APRIL, 2022

Breaches are out there, but that doesn't mean you have to be a target.

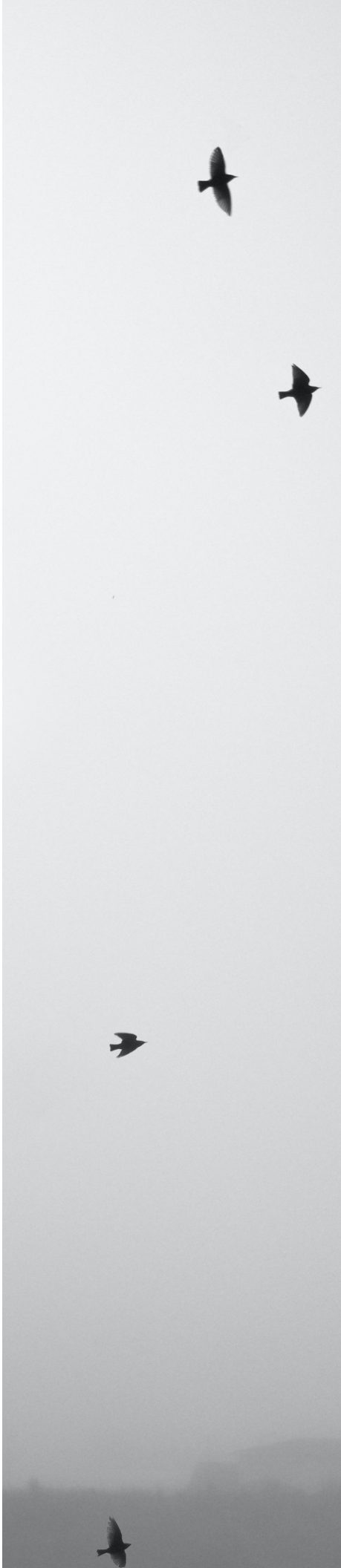
TABLE OF CONTENTS

Introduction

Welcome to the latest edition of our Cloud Misconfigurations Report, where we review data from publicly-disclosed breaches that occurred over the last year as well as data from our Internet scanners and honeypots.

The patterns we've extracted are quite revealing in terms of cloud-related breaches and persistent exposures.

The bad news is—surprise, surprise—breaches are still happening, and they're not likely to stop. The good news, though, is that we found many breaches to be caused by avoidable circumstances, such as using unsecured resources or users relaxing security permissions. Breaches are out there, but that doesn't mean you have to be a target, and keeping your organization safe may be simpler than you think.



In the News

We reviewed 68 different accounts of breaches from 2021, looking for details about the industries subject to breaches, types of data that were compromised, volume, and more.

Before we go further though, it's important for us to acknowledge that this sort of analysis is based on only a slice of all the attacks that happened in 2021. The set of accounts we're looking at are collected from primarily Western, English-speaking news sources. That means we didn't account for a large segment of the rest of the world and our data is acknowledgedly biased.

Additionally, we're looking at incidents that made the news. There certainly were more breaches that happened in 2021, but many of those breaches were probably never publicly disclosed. Without a standardized government mandate of some form requiring public disclosure of breaches, we'll likely never have a complete picture of the breaches that do occur.

The subset of breaches we're considering here provides a glimpse into some notable events, which may provide a sense of the types and severity of breaches that can occur. These accounts are useful warnings of what to be wary of.

What we do see in the sample we examined is a broad distribution of affected industries. Unsurprisingly, some of the most frequently breached industries include the information, healthcare, and public administration sectors[1].

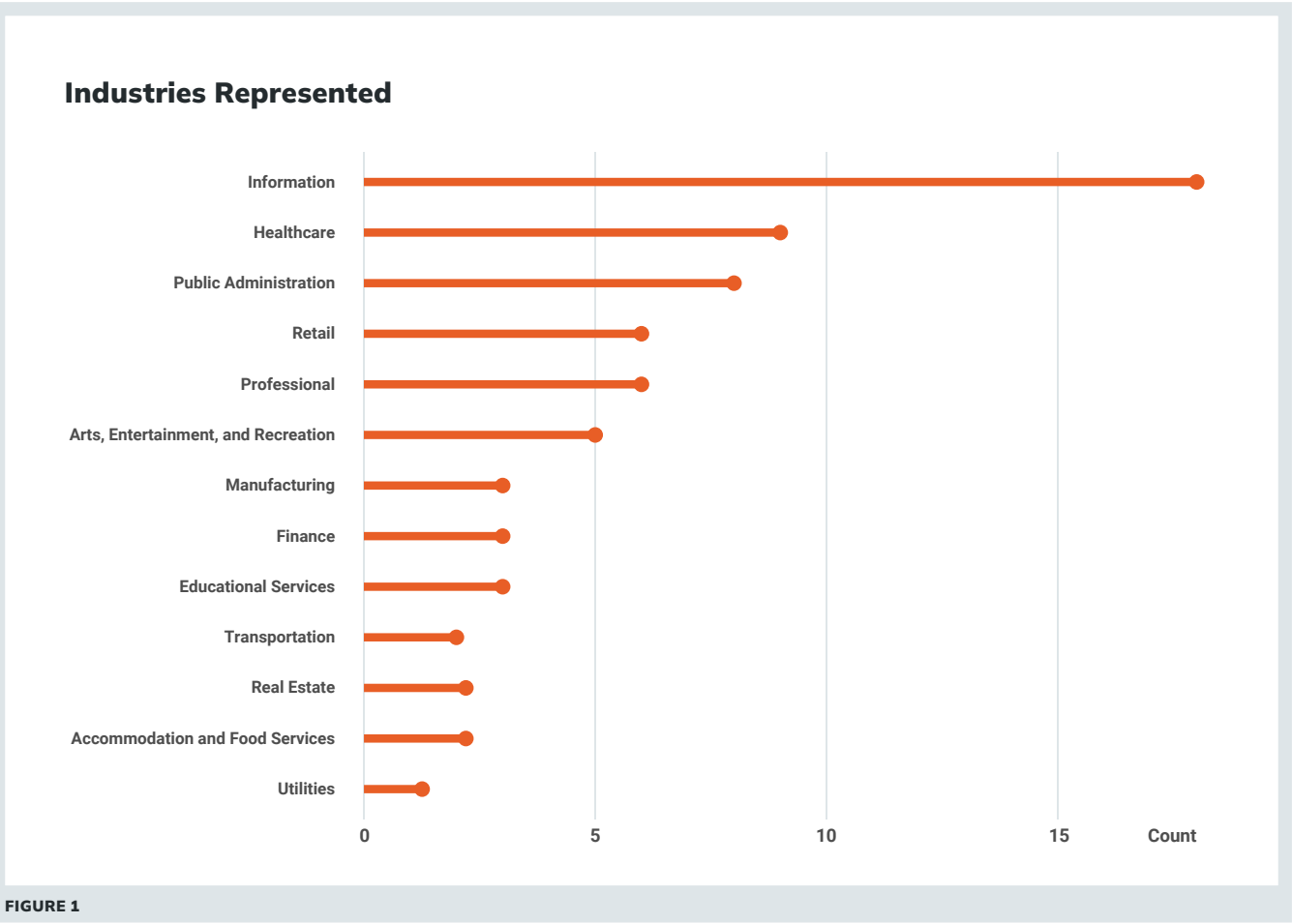


FIGURE 1

1 Industry naming convention adopted from the North American Industry Classification System (NAICS)

Some very prominent brands were represented among the organizations in the 2021 set of breaches, including some staples of the Fortune 500.

These are not startups operating on shoestring budgets; they're titans of industry with plenty of capital, resources, and personnel to deploy. Given their size, scale, and reach, we would assume that their IT teams and security functions are well-staffed and resourced.

If such presumably well-resourced organizations can be breached, the lesson we should take away from this is that anyone is susceptible.

The unfortunate reality is it really doesn't take much for a breach to occur. Breaches can be the result of something as simple as a misconfigured cloud storage instance or a slip-up with credential management.

Not all of these public accounts provide clear details about the specific resources that were compromised. Where compromised resources are specified, the resources in question are a mix of AWS, Elasticsearch, Azure, Google, and various Git-based services.

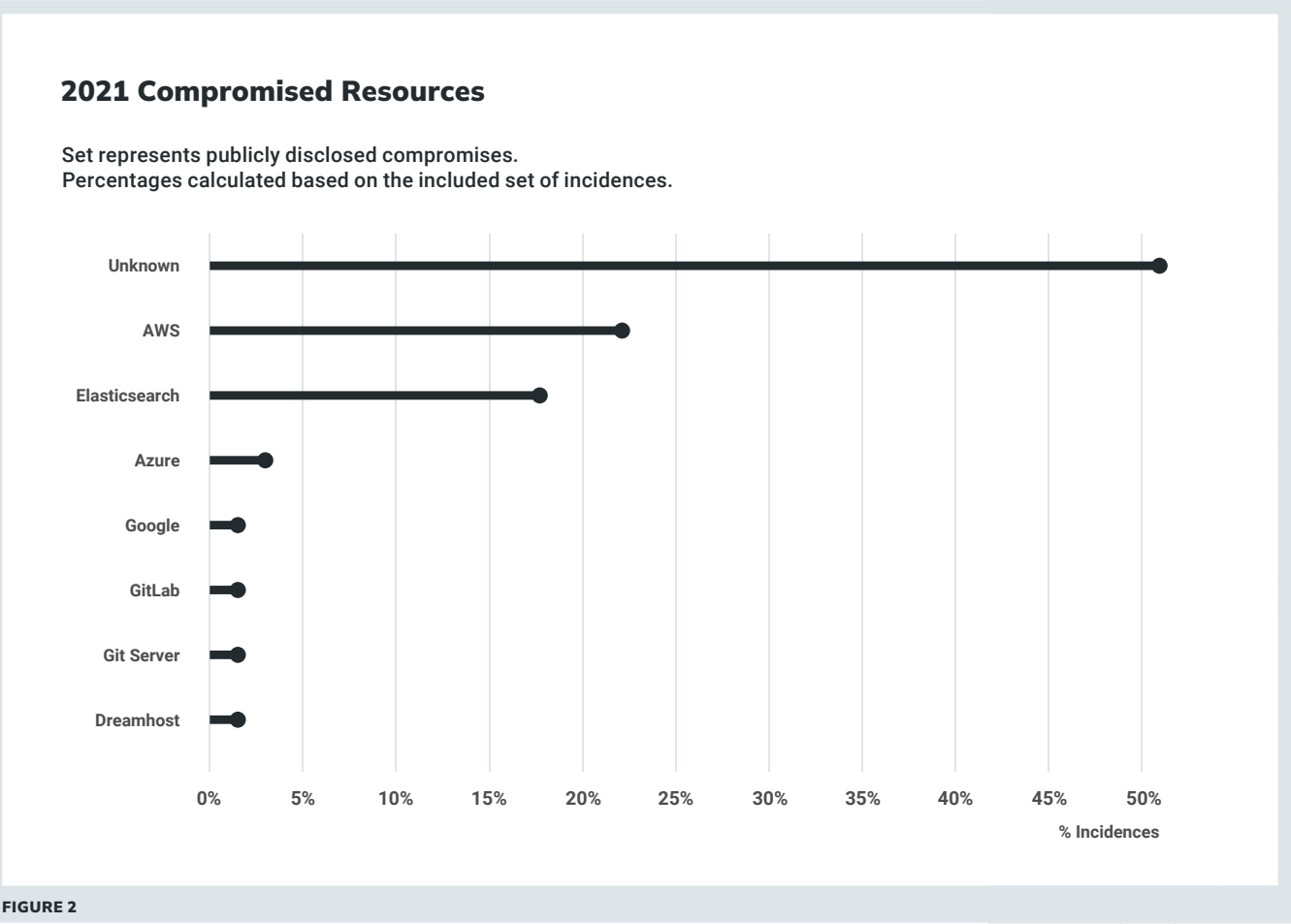
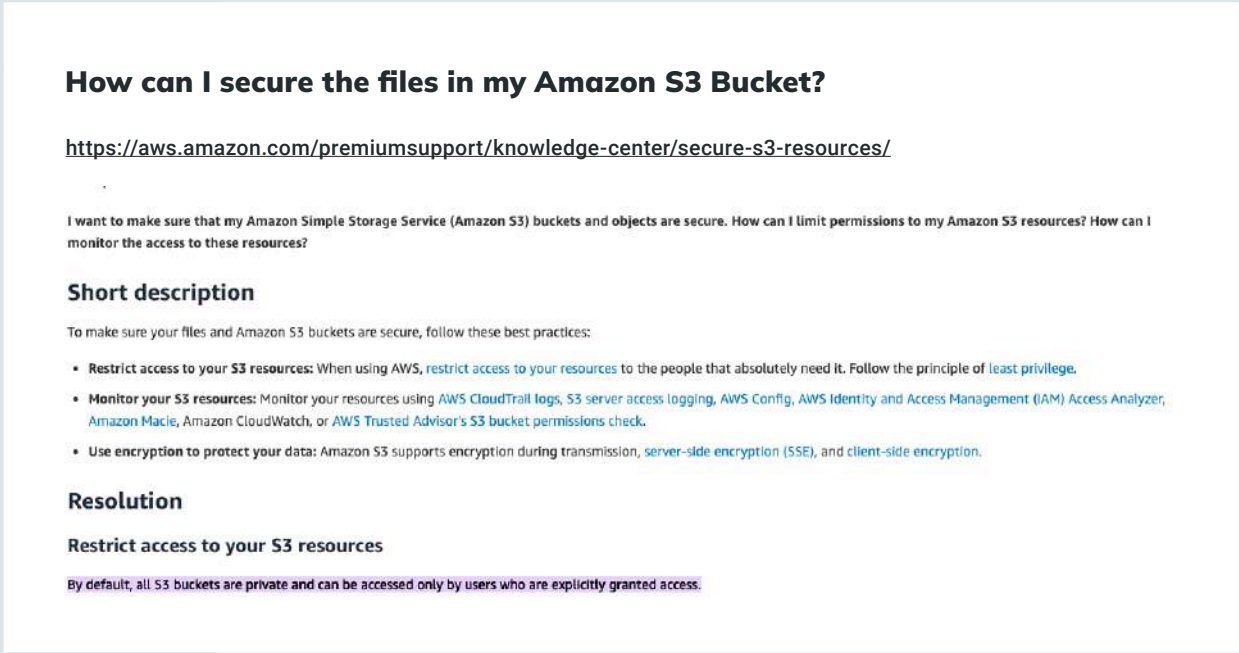


FIGURE 2



How can I secure the files in my Amazon S3 Bucket?

https://aws.amazon.com/premiumsupport/knowledge-center/secure-s3-resources/

I want to make sure that my Amazon Simple Storage Service (Amazon S3) buckets and objects are secure. How can I limit permissions to my Amazon S3 resources? How can I monitor the access to these resources?

Short description

To make sure your files and Amazon S3 buckets are secure, follow these best practices:

- **Restrict access to your S3 resources:** When using AWS, restrict access to your resources to the people that absolutely need it. Follow the principle of least privilege.
- **Monitor your S3 resources:** Monitor your resources using AWS CloudTrail logs, S3 server access logging, AWS Config, AWS Identity and Access Management (IAM) Access Analyzer, Amazon Macie, Amazon CloudWatch, or AWS Trusted Advisor's S3 bucket permissions check.
- **Use encryption to protect your data:** Amazon S3 supports encryption during transmission, server-side encryption (SSE), and client-side encryption.

Resolution

Restrict access to your S3 resources

By default, all S3 buckets are private and can be accessed only by users who are explicitly granted access.

FIGURE 3

What is revealing is that of the resources that were breached, many are secured and private by default. It takes intentional choice to make such cloud resources less secure and more susceptible to breach.

A service like AWS S3 for cloud storage, for instance, is by default set to private and limited in access to explicitly specified users.

For these resources to become accessible to unauthorized parties suggests that someone, somewhere intentionally made a change to reduce the security posture of the resource.

At the end of the day, this implies a lapse in security. Such lapses can be addressed with a combination of:

- Better user training
- Systems and controls to discourage the relaxing of security mechanisms
- Reviews of identified resources for appropriate configurations

We also pored through the various accounts under consideration for details of what was exposed or leaked.

In the following view, we've arranged the types of records in descending order from left to right, going from most frequently cited to least frequently cited. We've also arranged the industry sector in descending order based on the occurrence in the set where details of record loss type are provided.

Record Loss Type by Industry

Industries named based on the North American Industry Classification System (NAICS)

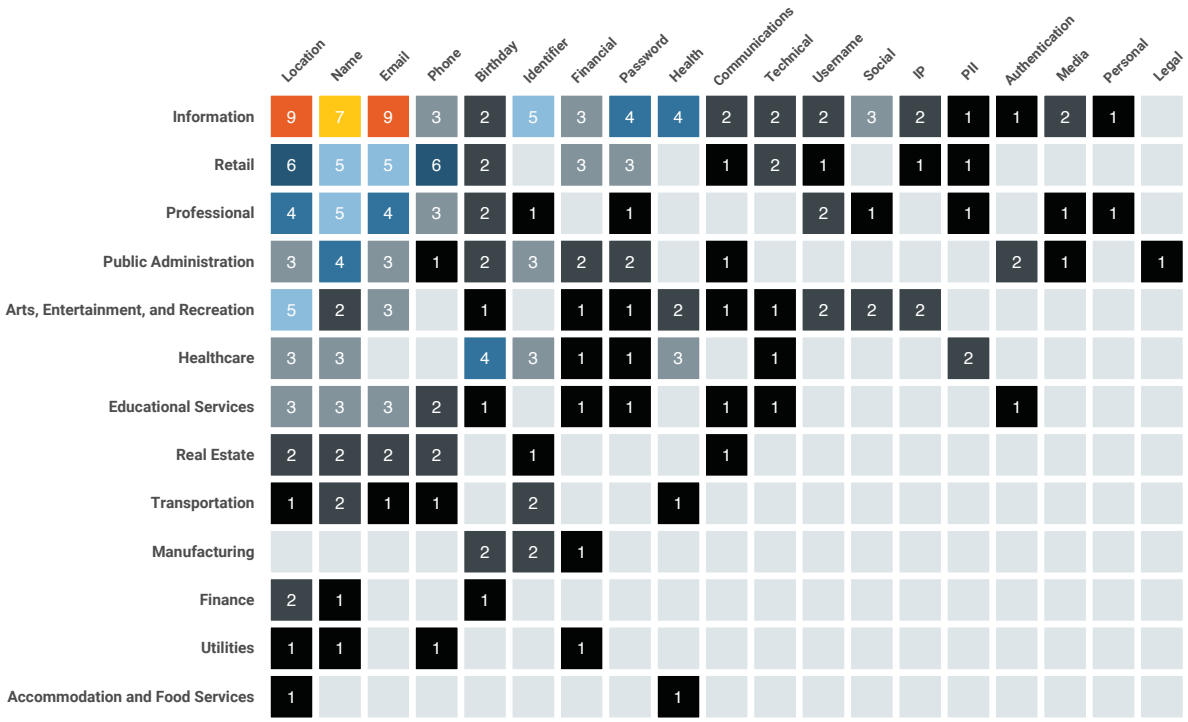


FIGURE 4

What we find is that details on physical location (such as addresses or latitude and longitude details), names, and email were by far the most commonly lost resources. The overall set of details lost represent a gold mine for opportunists and attackers, particularly of the social engineering variety.

In a number of the accounts we examined, we were provided with details about the time that a breach was discovered. Based on the date of the reporting, we also had a date of disclosure. Typically, disclosure happens after some sort of remediation has occurred. With that information, we're able to derive a sense of the duration that organizations remain compromised.

The good news is that the time between breach and disclosure typically falls under one month, which as far as security incident reporting goes is fairly brief. In such cases, there likely was an effective discovery process and remediation plan in place, or at least sufficient resources and people available to throw into remediating the problem in a timely manner.

The bad news is that long-duration breaches are still a thing. In some cases, the breaches lasted for years. Who knows what malicious actors could accomplish with such a broad window of access?

Time from Breach to Disclosure

Each dumbbell represents one distinct breach.
Records subset to only cases where breach date was provided.

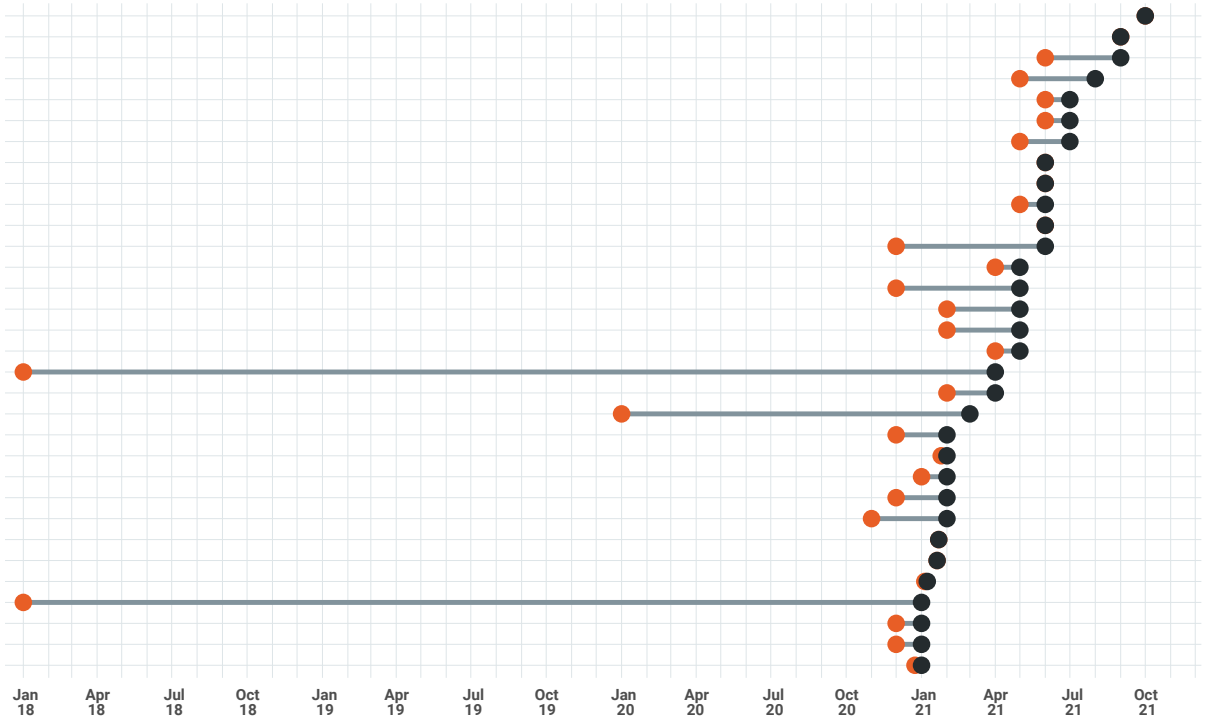


FIGURE 5

We did contemplate performing comparisons based on the size of data lost, but the accounts we found did not present the scale of losses in any consistent form or unit. Losses might be stated in terms of size of data lost (such as gigabytes or terabytes), in numbers of records lost, in terms of numbers of distinct users affected, and so on.

Frankly, size is probably not a very useful way to frame the severity of these breaches. After all, a terabyte of cute pet photos is probably less significant from a security standpoint than, say, individual governmental identifiers (like social security numbers) or confidential source code (which does occur—there were more than a few instances where Git repositories were exposed).

Systematic Data

When it comes to drawing generalizations about the cloud, it's preferable to work with more systematically collected, expansive data.

Fortunately, Rapid7 operates a number of data collection systems that can shed light on the matter of misconfigurations and exploitations in the cloudy wilderness.

One source of data is Project Sonar, an outbound scanner that looks across the internet, searching for public exposures across a range of ports and protocols.

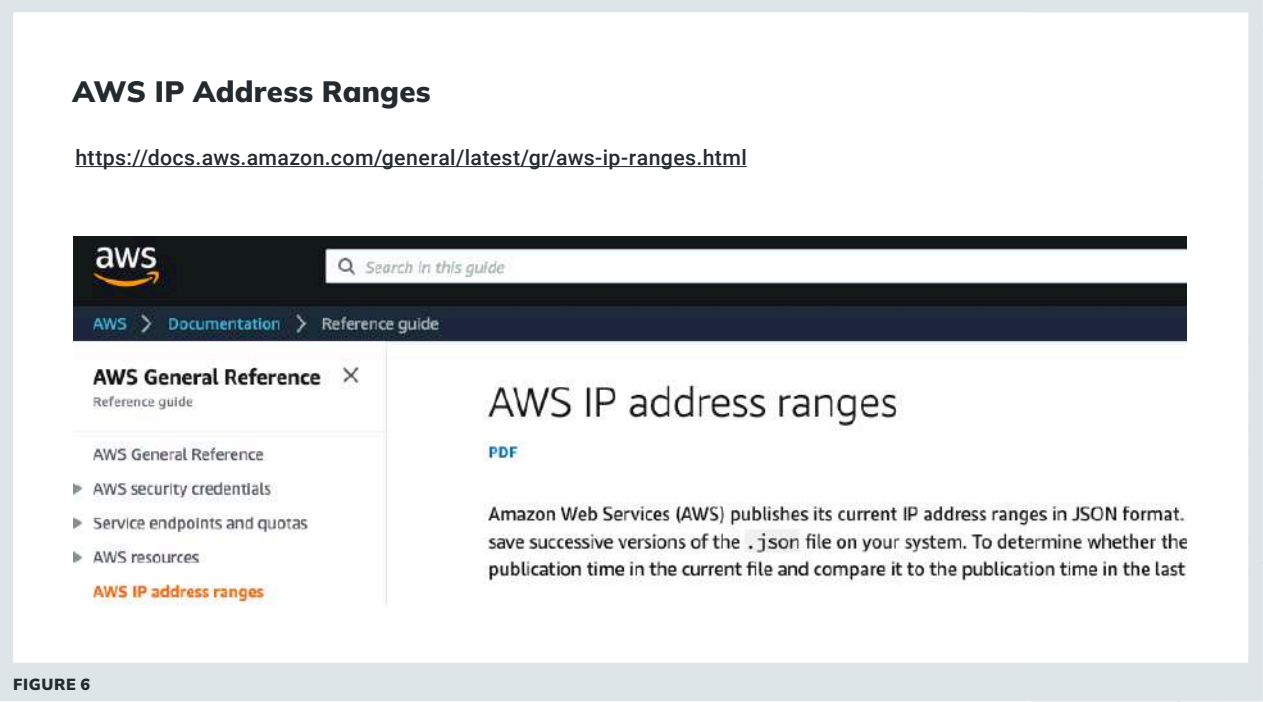


FIGURE 6

The different cloud vendors publicly disclose their cloud hosting service IP ranges, which makes it possible to filter our Sonar discoveries to just the exposures discovered on the cloud vendors in question.

With the public information about the cloud IP ranges, we can get a sense of their internet footprint. Based on the latest details from February 2022, AWS has a fairly large presence, followed by Microsoft Azure, then Google Cloud Platform.

IPs assigned to Cloud Vendors

February 2022

FIGURE 7

The opposite side of the coin is Project Heisenberg, a globally distributed honeypot network that Rapid7 maintains, with instances set up on different cloud vendors.

These honeypot instances sit idly, waiting for inbound connections. In a perfect world, the honeypots should receive no connections at all. If there are any inbound connections, it implies one of two things: there is a misconfiguration or there is some sort of malicious activity happening.

The misconfigurations can be the result of something as innocuous as someone having mistyped an IP address when attempting to connect to a networked resource.

We're able to filter out benign researchers conducting internet-wide research, so what's left is—we presume—malicious activity. These malicious connection attempts are effectively trying to scan the internet for exposed resources to exploit.

As the inbound connections come into Heisenberg, we collect all sorts of details, including timestamps for when the connections happen, what ports are being connected to, what protocols are being tapped, and any credentials that are being attempted, and more. In effect, we get a pretty good sense of what the bad actors are up to on a systematic basis.

It's worth pointing out that these honeypots are seeded across the internet in a fairly random manner. If we think of them as a random sampling of the internet, we can draw inferences about what's happening across the internet as a whole. So if we see unexpected connections of any form, we can assume that the same is happening across the internet.

What the Bad Guys Are Looking For

Below, we're taking a time series view of the distinct number of sources connecting to Heisenberg on any given day, across a range of ports and protocols. We get a sense of how many bad actors there are out there, and what it is they're trying to exploit.

The set of ports and protocols featured represent popular resources or protocols that we know have been targeted for exploitation in recent memory.

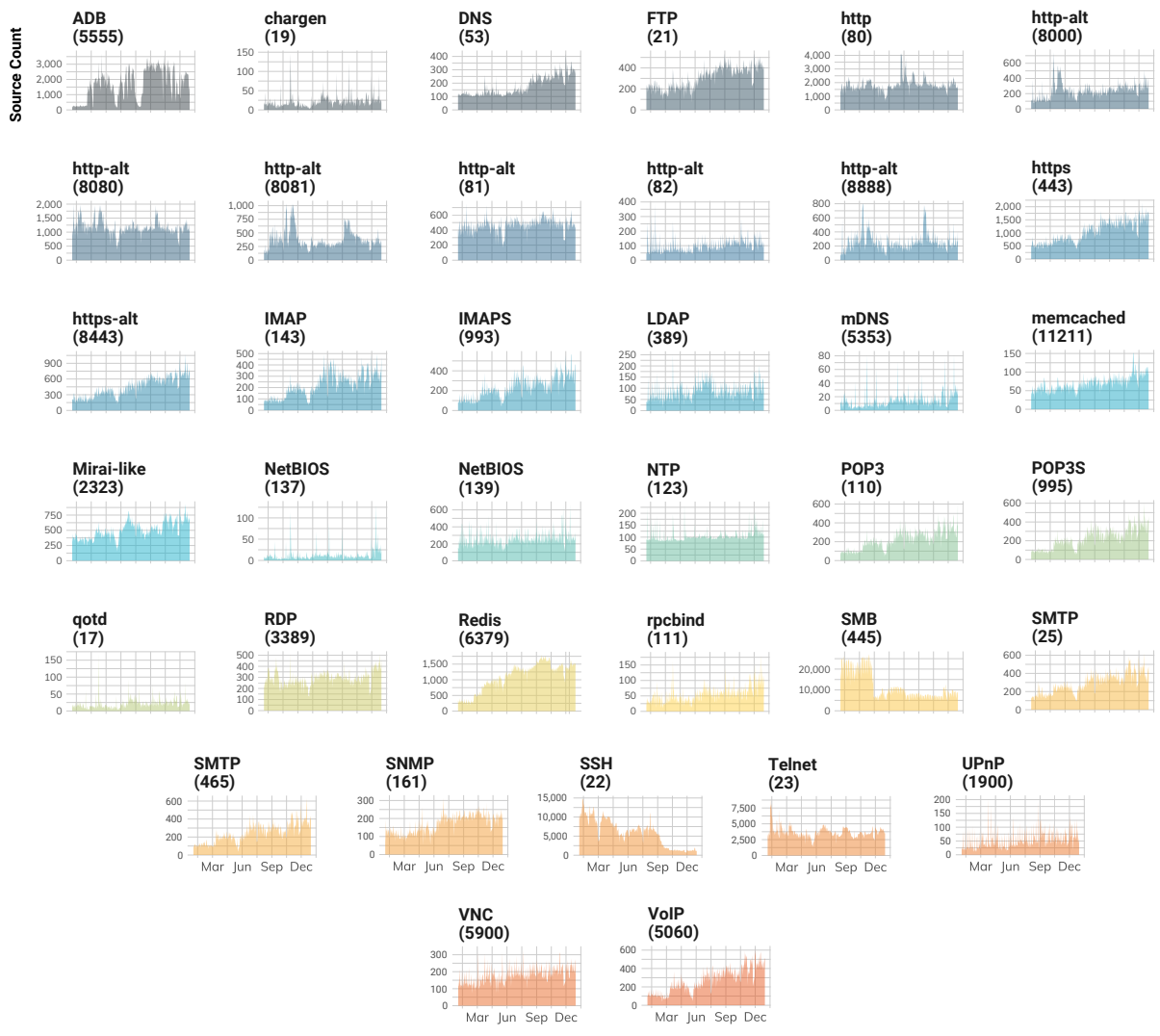


FIGURE 8

Protocols like Android Debug Bridge (ADB), Server Message Block (SMB), SSH, and Telnet were characterized by high levels of inbound connection sources throughout much of the year.

ADB might seem like an odd one initially, but it has been rampantly co-opted into botnets targeted at cryptocurrency mining in the past. Back in 2018, noted security researcher Kevin Beaumont **warned** "how thousands of internet connected Android devices now have no security, and are being exploited by criminals," ranging from "tankers in the US to DVRs in Hong Kong to mobile telephones in South Korea."

Targeting SMB is practically the norm these days: SMB is the vector through which the EternalBlue exploit operates, which has been the basis for a rash of ransomware campaigns in recent years, including WannaCry, Petya and NotPetya.

Telnet and SSH are widely used for remote access, so it's unsurprising that they would be targeted. It is unfortunate that we do see so many attempts directed at Telnet. If we are to assume that malicious actors are rational, in the sense that they do things that have positive payoffs, we can surmise that they consistently target Telnet because they've been able to successfully exploit Telnet before and expect to continue to do so, despite years of **warnings** from the security sector that Telnet is inappropriate for any internet-based usage today.

Deeper Dive on Telnet

We can now turn our attention to Telnet instances discovered on the cloud by Sonar, given its inherent weaknesses and its prominence from the Heisenberg side. Our findings are somewhat depressing, if not wholly unexpected. There is still a whole lot of Telnet out there, particularly on the cloud.

What that means is that there are people and organizations out there actively deploying Telnet instances on the cloud. Bear in mind that any communication through Telnet is inherently unencrypted, and therefore fully exposed and insecure. That warrants a severe look of disappointment.

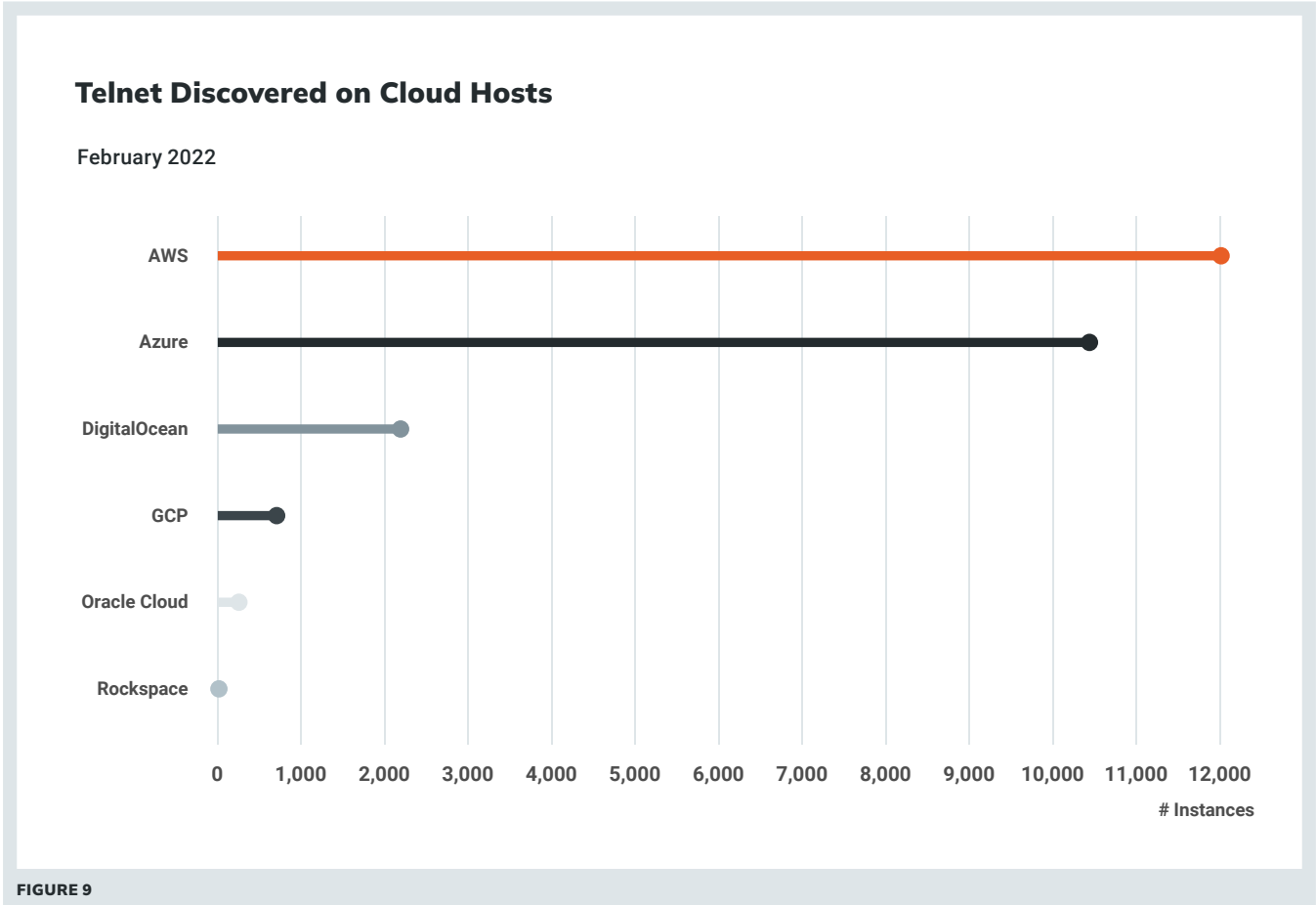


FIGURE 9

We're in a position to see what sorts of credentials are attempted on Telnet through our honeypots. We do find that particular protocol, username, and password combinations appear over and over. Credential stuffing is still very much a persistent concern, and is a common means for attackers to gain initial access into an organization.

Cloudy, With a Chance of Compromise

Previously, we mentioned that we have details about the IP ranges of cloud vendors, as well as details on the sources of connections to our honeypots. By combining those two sets of data, we can get a sense of what cloud sources are connecting to our honeypots.

Here we've provided a breakdown of the latest period of complete data—February 2022—in terms of inbound connections from different clouds to our honeypots.

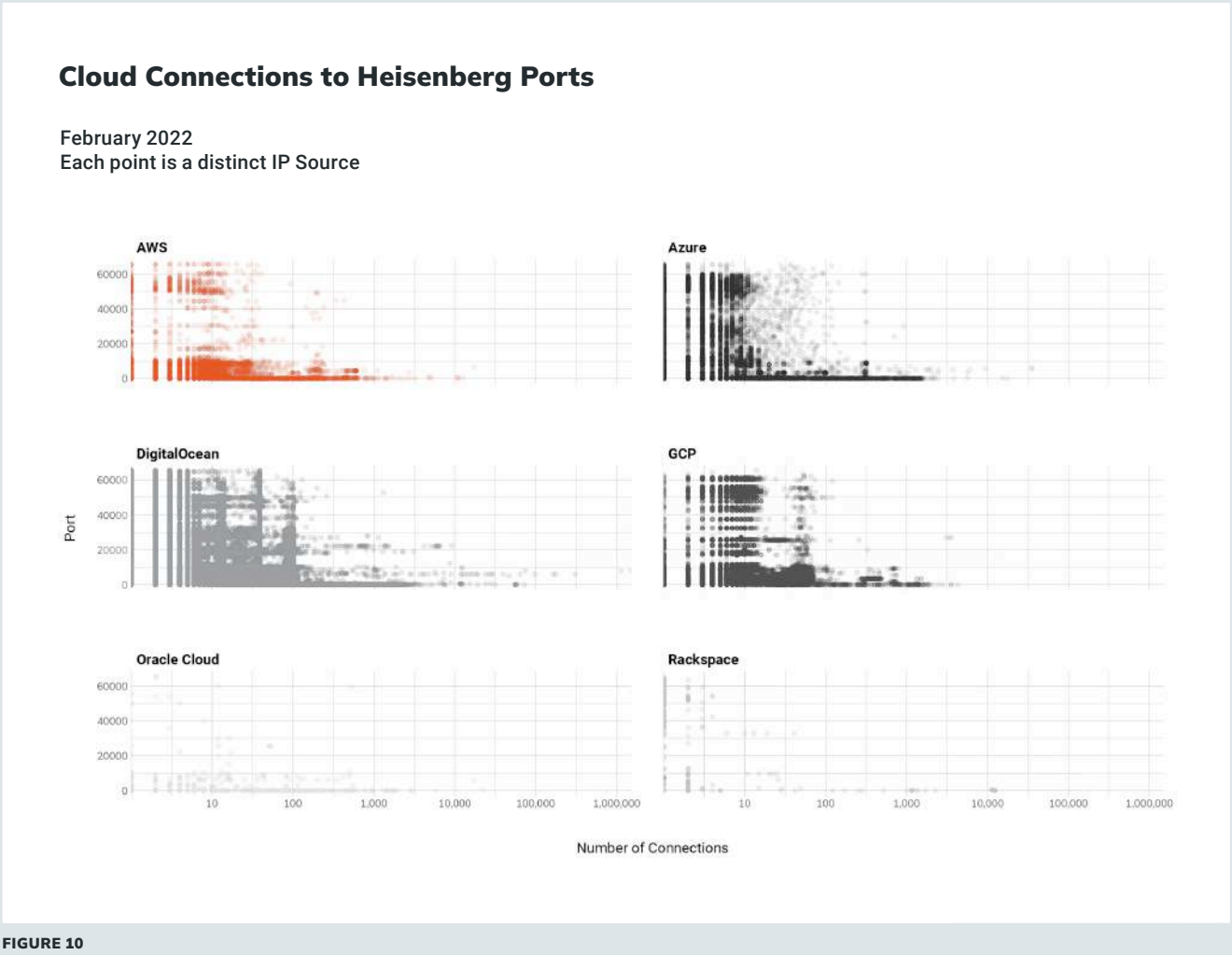


FIGURE 10

There's a lot going on. Each point represents a single source, and where it's positioned horizontally reflects how many connection attempts we saw from that particular source. Where the points are positioned vertically reflects which of the 65,536 possible ports inbound connections could hit. Looking at the port number in this manner is a bit nonsensical, but it does convey the spread of the types of things that are being targeted.

In general, we're seeing a lot of inbound connections from different sources originating from DigitalOcean. If we examine Digital Ocean's **Acceptable Use Policy**, we find that there's no specific language barring scanning, only abuse.

In contrast, at the other end of the scale, Oracle Cloud and Rackspace lag behind pretty significantly in terms of inbound connections to the Heisenberg honeypots. This could be attributed to their respective **Services Agreement** and **Acceptable Use Policy**, which contain language that discourage network discovery or monitoring.

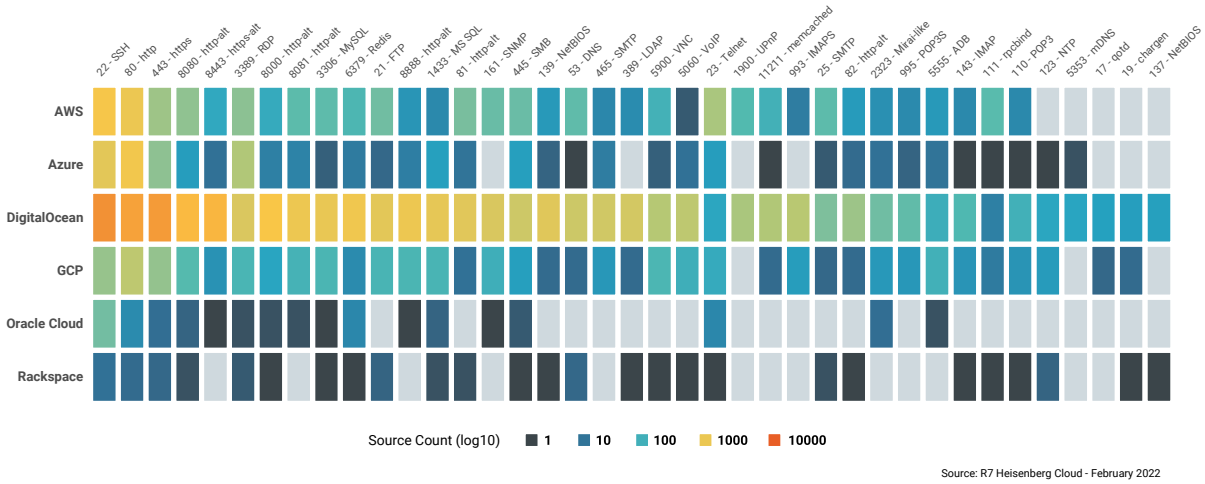
We can pare down the mass of data to focus on some prominent connection targets—the ones that have a recent history of exploitation. We can also separate that by cloud source, and present everything in a color-scaled manner to convey volume.

In this particular view, we're also sorting the ports and protocols in descending order from left to right. SSH, HTTP, and HTTPs were unsurprisingly the most commonly targeted protocols.

We also see a very brightly colored band across all of DigitalOcean, indicating that a lot of these undesirable connections originate from that particular cloud.

Distinct Sources Connecting to Heisenberg on Select Ports

Port/Protocol arranged in descending order based on count of distinct sources



Source: R7 Heisenberg Cloud - February 2022

FIGURE 11

Conclusion

So what does all this tell us?

Even with imperfect data—68 records is not a huge data set—a number of important trends were very clear.

We found that breaches can hit any organization, no matter their size and prestige (but organizations in high-risk industries like information, healthcare, and public administration should be especially cautious). And the data compromised isn't always the expected high-value nuggets, like credit card information or social security numbers. Simple data on names, locations, and email addresses can be powerful weapons in the hands of a skilled social engineer, so it's critical to keep these seemingly less important tidbits of information safe.

We also found that many breaches have causes that are easy to fix. Far too many breaches happen as a result of users manually relaxing security settings on cloud resources. Keeping some cloud resources safe can sometimes be as easy as leaving the default security settings intact. (Also, seriously, stop deploying unencrypted instances on the cloud.)

In a nutshell, better cloud security doesn't have to be hard. A few extra security checks here, and some user training there, and your risk of breach will plummet.

And if you'd like some helpful tools to make cloud security even easier, feel free to check out InsightCloudSec, Rapid7's own cloud-native security platform.

Our Philosophy

We believe that cybersecurity should be simpler and more accessible. Trusted by more than 9,300 customers worldwide, our best-in-class technology and strategic expertise draws on the insights of industry-leading researchers and contributions from the global security community to empower security professionals. The world has changed–Rapid7 is helping protectors be ready for what comes next.

Contact Us

North America:
+866.7.RAPID7
sales@rapid7.com

EMEA:
+44.1183.703500
emeasales@rapid7.com

APAC:
+65.3159.0080
apacsales@rapid7.com

@rapid7

rapid7.com

RAPID7