paloalto NETWORKS | UNIT 42

Incident Response Report

2024

Table of Contents

Executive Summary

As the leader of Unit 42® by Palo Alto Networks, I have the opportunity to work closely with our clients and my team on some of the largest and most complicated cybersecurity incidents.

In the past year, we have seen threat actors making larger and faster moves that damage their targets. The Unit 42 Incident Response and Threat Intelligence teams helped hundreds of organizations assess, respond, and recover from cyberattacks. We helped reduce operational downtime and got them back to business quicker.

Along the way, we collected data about the incidents.

In this report, we bring you the insights from that data. It's part of how we empower organizations to proactively navigate cyber risks, strengthen security approaches, and respond to incidents with unmatched efficiency.

That's the mission that drives Unit 42: protect the digital world from cyberattacks.

Here are the top areas of focus I'd like you to take away.

- **Speed matters.** The time between initial compromise and data exfiltration is decreasing. Attackers are sometimes beginning to exfiltrate data in hours, not days. Defenders need to speed up as well.
- **Software vulnerabilities still matter.** They were behind the largest-scale attack campaigns in 2023. They lead the list of ways attackers get in. Measure your threat surface, then fix it quickly and comprehensively.
- **Threat actors are becoming more sophisticated.** They're more organized, with specialized teams for different parts of the attack. They're more knowledgeable and able to use IT, cloud, and security tools as weapons of offense. And they're more efficient, using processes and playbooks to achieve their goals more quickly.

And this is all happening at the same time as artificial intelligence (AI) is a top concern. While attackers may benefit from new AI capabilities, defenders already do. And we're actively working on even more AI-driven abilities.

That's why we're here. Unit 42 helps you stay ahead of the adversaries. We help you prepare for, contain, remediate, and eradicate threats. We think security should be intelligence driven and response ready. Every day should be safer than the last.

Contact us directly if you'd like to know more about what Unit 42 can do with you. We're here to help.

Wendi Whitmore
Senior Vice President, Unit 42

How This Report Helps You

New headlines covering nation-state threat actors come out every day–not to mention news of the latest vulnerabilities and security risks. Your time is more valuable than ever, and sorting out which threats really matter is a difficult task.

This report helps because it gathers real-world information from organizations like yours, so you can learn which threats really affect your peers–and how you can face them. Read on to find out how threat actors gain access to organizations, what they do once they get in, and how our incident responders' top recommendations help you stop them.

Cybersecurity can often feel like an endless battle between attackers and defenders. At Unit 42, we believe intelligence, insight, and preparation still gives defenders the edge. We think the story of cybersecurity can be hopeful, with a strategic understanding of the threats we face today.

If You're a Security Leader

As an executive responsible for safeguarding your organization, you'll find analysis and insights to help you make strategic decisions about how best to invest time, resources, and money. Every day, Unit 42 sees world events turning into specific impacts on organizations.

Guiding an organization through a high-profile breach is not easy. Leaders in the current landscape face supply chain attacks and breaches at partners or suppliers that suddenly unleash security risks from unexpected directions.

If You're a Security Practitioner

Perhaps you're the head of your organization's security operations center (SOC). We'll show you the Unit 42 take on key aspects of today's threat landscape. And, this report offers detailed information such as MITRE ATT&CK® techniques and specific recommendations that will support you in making the tactical decisions needed to protect your organization's way of life.

We also offer information that will help you support your personal everyday way of life. We hear all the time that one of the greatest risks for defenders of organizations is burnout–and running a SOC 24/7 can't become something that consumes every minute of your life. Watch for recommendations about automation, AI, and other technologies that can help you focus your limited human resources where they most need to be.

By learning what other organizations are facing, you can increase your own resilience–without having to learn lessons the hard way.

If You're Simply Concerned about Security

Even if you're not directly responsible for protecting your organization from cyberattacks, cybersecurity considerations touch all of us.

Communications professionals may need to prepare for public disclosures. Legal departments may benefit from understanding how sensitive data can be put at risk. And human resources departments may need to consider how to protect employees from threat actor harassment.

No matter your role, understanding today's threat landscape will deepen your ability to meet today's challenges.

We're Ready to Help

If you find yourself wanting an experienced guide through that landscape–whether you're faced with an incident or considering how to prepare for or prevent one–Unit 42 is always glad to help. Please contact us.

The key to surviving major incidents is to:

- Understand what you're facing.
- Have plans in place.
- Have the capability to act on those plans.
- Know when, how, and to whom to reach out for help.

A deep understanding of today's threat landscape can help you lead your people and your organization to a strong and resilient security posture.

Actors, Methods and Targets

As the global head of operations at Unit 42, I oversee our incident response engagements around the world. They're severe incidents that have escalated beyond what the target organizations could manage on their own.

From this unique vantage point, I've seen some common themes emerge. Not just in the defensive mistakes, but also in the strategies threat actors used to succeed.

This report distills these observations. Here are some of the key takeaways.

Who's Doing It?

In 2023, by far the most damaging threat actor was a group we call Muddled Libra. They're a criminal group focused on financial gain. They are also one of the most aggressive threat actors in the landscape today. Countering their attacks is a challenge for defenders of small and large organizations alike. This report has a special section on them, including an analysis of what they do and how you can defend against them.

Of course, they weren't the only group in the mix. We also responded to compromises from other criminal groups, state-aligned actors, and some groups that we don't know enough about yet to characterize.

Answering this question of "who" is one of the things that Unit 42 Threat Intelligence is great at. We'll keep researching and uncovering attackers wherever they try to hide.

How Do Threat Actors Get In?

In the past year, attackers' initial access most often started with a software vulnerability. The largest attack campaigns began with successful exploitation of internet-facing systems. From there, the stories varied.

Naturally, we also saw other techniques, from attackers using previously stolen credentials, phishing (both in email and chat), and more.

But don't be misled–though phishing lost the top spot and dropped to number three, it's still a serious threat. Attackers leverage phishing attacks for access, credentials, and tokens, rather than trying to drop malware.

What Helped Threat Actors Succeed?

There were three security weaknesses that weren't themselves the cause of incidents, but made attackers' lives easier. Address these, and you'll be helping yourself and hurting them.

- **Patch management.** It's only a matter of time. Unpatched vulnerabilities on internet-facing systems will be exploited. Measure and reduce your attack surface, and you'll improve this.
- **Consistent coverage.** Our incident responders found that organizations with partial or incomplete deployment of security controls (especially endpoint detection and response tools) allowed attackers to operate from parts of the network that weren't defended. Deploy your defenses everywhere, and you'll deny them this advantage.
- **Identity and access management.** More attackers are stealing and using the identity of authorized personnel to access and move around networks. Make that harder and embark on or continue a journey toward Zero Trust networking, and you'll pressure attackers into making mistakes you can see.

Who's Being Targeted?

The top six industries in our casework stayed the same this year, and they accounted for 63% of our cases:

1. **1.** Professional and legal services
2. **2.** High technology
3. **3.** Manufacturing
4. **4.** Healthcare
5. **5.** Finance
6. **6.** Wholesale and retail

Although their relative positions within the six have changed a little from year to year, the total number of cases for these industries is consistent.

The upshot is that cyberattacks continue to be a problem faced by organizations in all industries. And while certain attackers have turned their focus to a particular vertical from time to time, ultimately, every organization needs to be response ready.

I hope you enjoy reading about some of the investigations, research, and intelligence my team developed for this report. More importantly, I hope you'll use it to make your organization more secure, your infrastructure more resilient, and every day safer than the one before. Reach out to me to explore ways we could help you do so.

Sam Rubin
Vice President, Global Head of Operations, Unit 42

Spotlight On: Muddled Libra

Unit 42 has worked on over a dozen intrusion cases that involved an adversary we call Muddled Libra, including several very large and visible incidents.

Background

Muddled Libra is a sophisticated, tenacious, and damaging adversary. We don't believe they are state-sponsored, but they are as aggressive and as skilled at attack operations as some advanced, persistent threat (APT) groups.

The group includes people with strong English language skills, which they use in written and spoken communications. It's possible they speak English as a first language, and it's possible they are located in North America, or even the United States.

Their ransom demands are very large, often in the tens of millions of USD. And of course, they demand payment in cryptocurrency.

Once inside an environment, they actively monitor defenders' internal communications to understand when and how they have been detected, to improve their ability and speed of data theft. They also determine how to remain in the environment even during and after eradication attempts.

Defending against this attack group is difficult. It requires sustained effort, potentially multiple times. You'll need to defend across the entire spectrum of your organization, from premises to cloud to remote workers and software-as-a-service (SaaS) tools. And when you do, they will fight back at the same time to maintain their access.

Who Are They Targeting?

Early attacks from this group focused on telecom providers and business process outsourcing (BPO) organizations. These incidents appeared to be supply-chain attacks, in which the threat actor tried to compromise organizations and systems that would give them access to cryptocurrency assets.

For example, many online systems use SMS as a second factor of authentication. By compromising a telecom provider, Muddled Libra gained access to SIM-swap or port-out telephone numbers. Together with compromised first-factor credentials, the threat actor could steal cryptocurrency from a target and leave little recourse.

In the latter half of 2023, along with joining the BlackCat/AlphV ransomware group as an affiliate, Muddled Libra shifted or added to its targeting to include larger organizations that are particularly sensitive to service interruptions. Incidents at several large hospitality providers were publicly reported, but that was not nearly all the activity.

As a ransomware-as-a-service affiliate, Muddled Libra enjoys access to a much larger pool of targets and a simplified attack chain.

How Do They Get In?

Muddled Libra uses several methods to gain access to an environment such as:

- Social engineering both IT help desk and end users
- Buying access from a broker
- Stealing credentials from user endpoints
- Reusing compromised infrastructure or re-exploiting unfixed vulnerabilities used by a previous attacker

Perhaps the highest-impact access method is scamming the target's IT help desk. A threat actor will call the target's IT help desk and impersonate a valid user. They will ask for help with resetting the user's password and/or changing the mobile phone number associated with the account. Often, the request comes with a "sob story" or other emotion intended to manipulate the help desk worker. This attack only takes a few minutes.

Muddled Libra is unusually comfortable with engaging over the telephone, with both help desk and other employees. They are more successful in persuading people to engage in unsafe actions as a result.

Often, they will acquire initial access by taking over a less-privileged legitimate user account. Then, they identify the more highly privileged accounts that will help them achieve their objectives and call back to take over those accounts.

In one case that Unit 42 worked, the attackers successfully hoodwinked the help desk three separate times.

Muddled Libra also buys stolen credentials from initial access brokers. The threat group has been very open about their desire to buy access to certain targets, especially in the segments they have chosen.

And if these social engineering access methods don't work, Muddled Libra also has technical tactics to use as well.

While malware is not their first choice tool, this threat actor has used Racoon Stealer to access credentials stored in applications. They appear to focus on stealing session tokens and saved credentials. They have also been seen misusing legitimate forensics tools to search for credentials in live memory. This can be a vector if organizational credentials are stored in unmanaged systems.

And of course, given Muddled Libra's history with supply chain compromise, it should be no surprise that they can bypass SMS-based multifactor authentication (MFA). Their main tactic is SIM swapping, moving the second factor to a device under their control. They also perform "MFA fatigue attacks," a tactic to defeat push-based second factors.

In this tactic, the attacker uses compromised credentials to satisfy the first factor and then repeatedly solicits a second-factor push from the authorized account holder. Eventually, after many push notifications, or from a lack of caution, the authorized account holder will approve the second factor push.

If social engineering access methods don't work, Muddled Libra also has technical tactics to use as well.

Why Is It So Hard to Eject This Threat Actor?

In short, Muddled Libra is often as skilled and familiar with a target's IT infrastructure as the people who manage it. And they are not constrained by change control, compliance, or corporate policy.

Knowledge

Muddled Libra has advanced skills and understanding of modern IT operations, compared to other threat actors.

Unlike other threat actors who typically only perform light (if any) reconnaissance on their target, Muddled Libra performs extensive research both before and during the compromise.

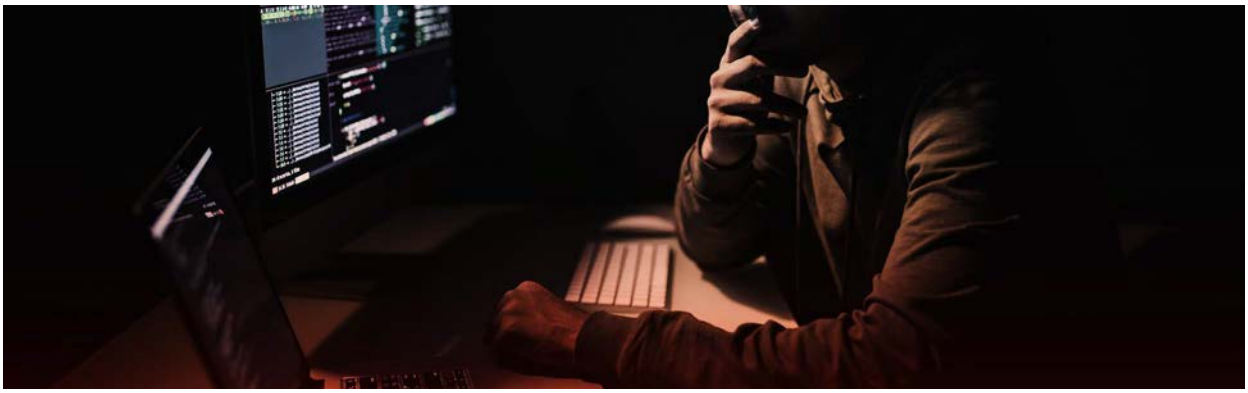
They have been observed accessing a target's own technical documentation, reading incident response standard operating procedures and playbooks, how-to documentation, and otherwise furthering their understanding of the target's environment. Then they use this knowledge to understand where to place implants and what defenders are likely to do in response.

Variety and Having a Plan B, C, D...

Muddled Libra creates multiple backdoors into environments, installing rootkits, remote management tools, and outbound tunnels that they can use for later access even if their primary access is discovered and removed. We have seen them install a half dozen or more of these utilities.

They create new accounts for themselves whose names mimic existing privileged accounts. Sometimes they reactivate existing but previously disabled accounts and use those. This tactic avoids a "new account creation" alert.

Rather than attack from obvious locations such as commercial hosting and server farms, Muddled Libra has moved to using residential proxy services such as NSOCKS and Truesocks. This makes it so that their connections appear to originate from the correct geographic location and from a plausible network provider.



IT, Virtualization, and Cloud Skills

Muddled Libra even attempts to access and use existing endpoint and systems management tool–such as endpoint detection and response (EDR/XDR) consoles, patch management, and the like–for reconnaissance, data theft and malware deployment.. And naturally, they use credential-stealing tools such as Mimikatz to help them move laterally and escalate privileges.

This threat group is also adept at using current IT and cloud management tools. They'll use the remote-management tools that are legitimately present in a target environment and also add their own. They'll search GitHub for code and keys, and they'll steal both.

Cloud infrastructure is just as much at risk as classic IT infrastructure. Muddled Libra can use misconfigured cloud environments from Amazon Web Services, Google Cloud Platform, and Microsoft Azure, creating service accounts and additional access keys even in the very early stages of an attack. Some organizations have thought they evicted the threat actor from their premises, only to have them come back through access from the cloud.

Virtualized environments are also at risk. If a client does not contain a compromised virtual system quickly enough, this threat actor can reverse the containment, take over the host or hypervisor and lock the target out of their own infrastructure.

And even classic on-premises infrastructure can be attacked using remote infrastructure management tools (many of which have attracted patches in the last few years) at the system and the hardware level.

They are also familiar with secure email gateways and email storage systems. They'll change rules to redirect or intercept notifications and recovery-related email to avert detection and response. In some cases they have created rules to redirect messages from specific security vendors to the threat actors. This tactic allows them to monitor the investigation as it proceeds.

Exfiltration

Muddled Libra has exfiltrated data to commercial file-hosting services. They have also staged their tools in these services. Network traffic monitoring often doesn't alert on known file-hosting services.

They also create and use external virtual private network (VPN) services to create outbound tunnels for exfiltration, obscuring their traffic from inspection devices.

Tips for Defenders

Defending against Muddled Libra requires sustained effort before, during, and after a compromise. Focus your energy on detecting their tactics rather than their tools. Tools can change rapidly, but tactics take longer.

Focus your energy on detecting their tactics rather than their tools. Tools can change rapidly, but tactics take longer.

Protect Credentials

Training end-users not to approve MFA requests they didn't solicit is important. You can further reduce the human-error burden by requiring number matching as well.

Sudden high volumes of MFA requests are a bright red, waving flag, especially if they go unacknowledged or are aimed at high-privileged accounts. Have a tool and process to rapidly identify and investigate this activity. Give the analyst performing the investigation the information to go out of band quickly.

Watch for velocity changes in MFA enrollment. Most people don't lose their phones often. Put additional scrutiny on changes to high-privileged accounts. Consider a policy that requires live visual and audible verification with a third party, such as the requestor's direct supervisor. These policies do slow down legitimate requests as well. But having a well-known policy requiring verification resists attacker attempts to pressure junior personnel into making unauthorized changes.

Guard against SIM swapping. Educate users that if their phone suddenly loses service, they should treat that as a potential security problem as well as an availability problem. Require that devices being used for corporate authentication (even personally owned devices) be locked with a PIN against port-out and SIM changes in the mobile carrier's systems.

Watch account creation and reactivation. Alert on new account creation that doesn't quite fit naming conventions. Alert and escalate on old or intentionally disabled privileged accounts being reactivated.

Monitor Behavior

Log and analyze the usage patterns of your key security tools ("Quis custodiet ipsos custodes?") Personnel accessing an XDR platform outside their usual work hours might just be industrious, or their account might be under a threat actor's control. Many organizations collect these kinds of audit logs but don't ever analyze them for outliers. Do so regularly.

Watch for changes in your cloud infrastructure. Monitor both your IT infrastructure cloud (such as directory services and cloud storage) as well as service infrastructure, continuous integration and continuous delivery (CI/CD) and similar environments. Look for changes to logging settings and privileges.

Check your code repositories. Ensure you're not inadvertently exposing secrets, of course. But also look for new connections to third-party infrastructure and unusual patterns of access.

And watch the behavior of your virtual desktop systems. Check for outlier access patterns (though this attacker's use of residential proxies will make that more difficult). Look for unusual process trees. (A good endpoint protection tool should catch this itself, but you should still ensure you're running down investigative leads from time to time rather than just closing everything as a false positive to keep the metrics good.) And try to alert on unusual storage usage, to catch staging for exfiltration.

Know Your Applications

If you see red-teaming tools in your environment, make sure there is an authorized red-team engagement underway. One SOC we worked with had a company logo sticker on the wall for each red team they'd caught.

Watch which remote management tools are being used in your organization. If you see new tools you don't normally use, or different versions of the ones you do, dig at that. And try to identify unusual usage of the tools as well. For example, if your remote tools are normally used by IT staff, but suddenly one person in finance is using them, find out why.

Endpoint management and inventory tools can help here, too. Use them to scan the fleet and identify new or low-prevalence tools and executables. Then, train your analysts and models "what good looks like" (and doesn't).

Watch the Network

Be suspicious of connections to your network from commercial VPN providers. While many people use such services, especially on their personal devices, discourage them. Commercial VPN providers add little if any security to a well-defended network, especially one built on Zero Trust principles.

Hold privileged users to a higher standard. Collect and monitor the patterns of where they access their accounts from and ask about outliers. If you use a corporate messaging platform that indicates a user's time and location (e.g., "It's 9:43 a.m. for Alice"), that can be a quick validator for SOC personnel wondering if Alice is really logging in from Aruba.

Watch outbound access, too. Monitor for connections that look like encrypted tunnels, particularly from new or unmanaged systems. If you can interdict connections at the network level (with technology and policy) use that capability judiciously.

And watch for commercial file-hosting providers. If you can restrict access to just the ones you have organizational agreements with, do that. If not, monitor connections to the ones you're not intending to use widely and watch for high data counts in short amounts of time. We have seen terabytes of data exfiltrated in a couple of hours.

Summarizing the Adversary

Muddled Libra is a methodical, evolving adversary that poses a substantial threat to many organizations. They're proficient with security tooling and are able to execute rapidly and effectively even in relatively secure environments. They evolve and change tactics constantly. It's important to stay informed about their latest techniques.

Defenders need to combine strong technology and security practices with diligent monitoring to detect and interdict this group's activity. Because of their speed, early detection and interruption of the attack chain is even more important. You may have only a couple hours between compromise and exfiltration.

The good news? The threat raised by this adversary creates an opportunity to take a critical look at what an advanced attacker could do in your environment–and prepare improvements to frustrate them.

Unit 42 Threat Intelligence publishes and updates assessments of this threat group. And as our security consultants face Muddled Libra during incident response, we collect more information firsthand and flow that threat intelligence throughout the Palo Alto Networks platform.

Initial Compromise

Everyone wants to know how the bad guy got in, in the first place. We reviewed the data from our casework over the last several years. A few trends emerged: some changed, some stayed the same.

There were certainly a lot of security incidents in 2023, and there were definitely some times when it felt like defenders couldn't get a break from all the exploited vulnerabilities and campaigns that occurred.

But, some key incident metrics improved.

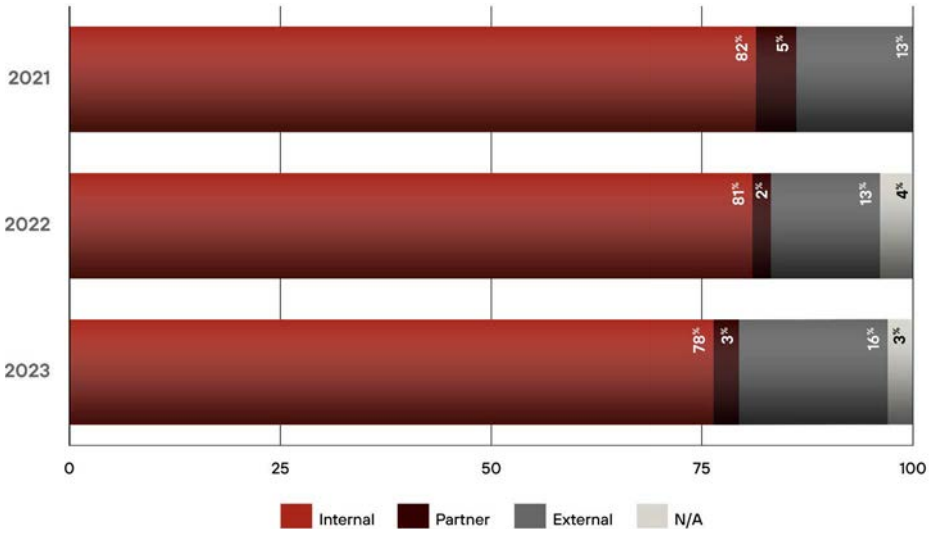
Incident Discovery

The first question we ask in a security incident is, how was it discovered?

Internal Vs. External Discovery

In figure 1, we show the proportion of Internal, External, and Partner sources of incidents. Internal means the client discovered the incident. External means a party other than the client did. Partner, in our casework, is usually a service provider to the target–something like a managed service, managed security service, or a managed detection and response.

Figure 1. Sources of incidents for Unit 42 IR cases in 2023



One data caveat here, though. Sometimes, those partner organizations appear to us to be an internal source, so there may be a little fuzziness on the Partner-Internal line. (For example, sometimes a client says they discovered an incident, but it was actually identified by their managed security services provider. For our purposes as responders, those don't change how we conduct the investigation.)

So what does this incident discovery data mean, overall? We think it's mostly good news. Discovering four out of five incidents internally is pretty good.

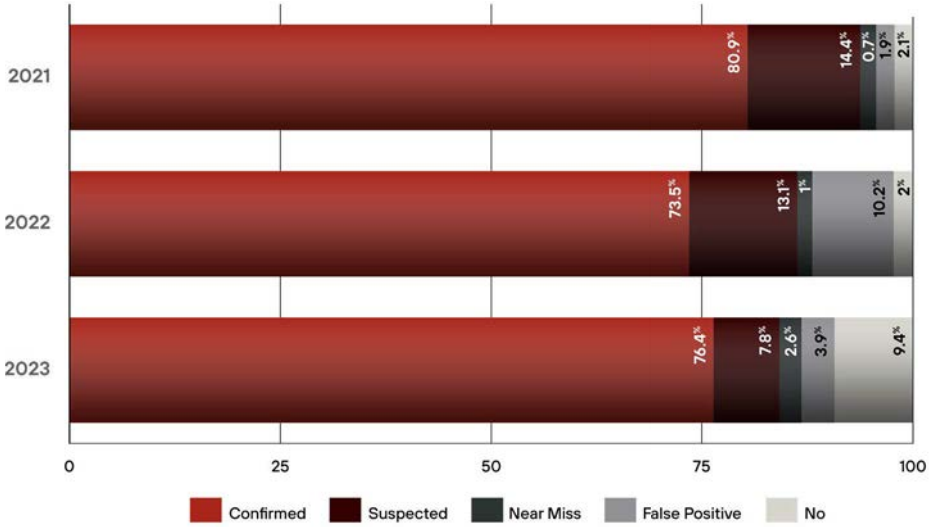
Together with the reduction in median dwell time that we (and others) have observed, this could mean that detection efforts are paying off. Compromised organizations are discovering sooner that they need to address a problem, rather than being surprised by an external party turning up with a folder full of IP addresses and bad news.

However, we wouldn't be engineers and consultants if we didn't ask, "How can this be improved?" It's worth thinking about this statistic in the context of your strategic plans. How will you find incidents in the information available to you? What will you do when you find it? Can you improve on both answers? What resources do you need to be–as Daft Punk so insightfully suggested–"harder, better, faster, stronger"?

Confirmed Security Incidents

The second question in any incident response is similarly meaningful. Is this actually a security incident, or just a false alarm? Figure 2 shows that most of the time, the potential incident is confirmed.

Figure 2. Unit 42 IR cases by confirmation status



In the figure above,"Confirmed" means what it says; this is genuinely a problem. "Suspected" means we think it was an incident, but we can't be sure, due to some limitation. Often this was a lack of available logs or a limited scope of investigation. A "False Positive" means the alert that kicked off the investigation was a true alert, but after investigation no other impact was found. And a "Near Miss" is the lucky ticket where some unauthorized activity occurred, but no asset or information was compromised (e.g., a vulnerability was exploited successfully, but there was no further unauthorized access).

Confirmed incidents held relatively steady at around three-quarters of all incidents. The changes in Suspected (which decreased), Near Miss, and False Positive are heartening. We think it means that more organizations are willing to investigate potential incidents. That's good for defenders. It indicates growing maturity, willingness to engage help, and improved operational ability.

Again, a caveat: This data is drawn from incidents we worked on, which by definition means someone reached out for help. So there is a set of incidents that we don't know about and aren't represented in this data, and it might be a large set.

Nevertheless, a negative incident finding, confirming a Near Miss or a False Positive–these all have value, because:

- You know what you *don't* have to do.
- You exercise your processes, find gaps and keep those incident response muscles working.

Clearly, defenders need to speed up as much as possible. We'll give more recommendations for how to do that later in this report, but for the moment, there are a few key principles to keep in mind:

- **Preparation.** One of the best ways to get ahead of attackers is to truly get ahead. Through proactive preparation, organizations can get ready to respond within hours to compromises, stopping attackers before they have a chance to execute their plans.
- **Automation.** As we mentioned in the Black Basta example above, human security teams need to sleep, and they're also typically pulled in many directions. Automation–especially when powered by machine learning and AI–can help defenders sift through alerts and surface the ones that truly need attention.
- **Zero Trust.** Another powerful way to limit the damage an attacker can do after compromise is to limit their movement and activity. When organizations design their security posture in accordance with a Zero Trust philosophy, attackers become less powerful when they gain initial access, because initial access means much less. It's the difference between a thief getting into an entryway and being able to move through the entire building, and a thief getting into the lobby only to encounter another locked door.
- **Defense in depth.** A security program designed with overlapping defenses and controls gives attackers more opportunity to alert you to their presence. Especially combined with limited privileges in a Zero Trust philosophy, you can raise the signal-to-noise ratio of meaningful alerts that will let you focus on attacker activity earlier in the attack lifecycle.

The Beginning of an Investigation

At the beginning of an incident investigation, after the discovery and confirmation are known, the next question typically asked is, how did this start?

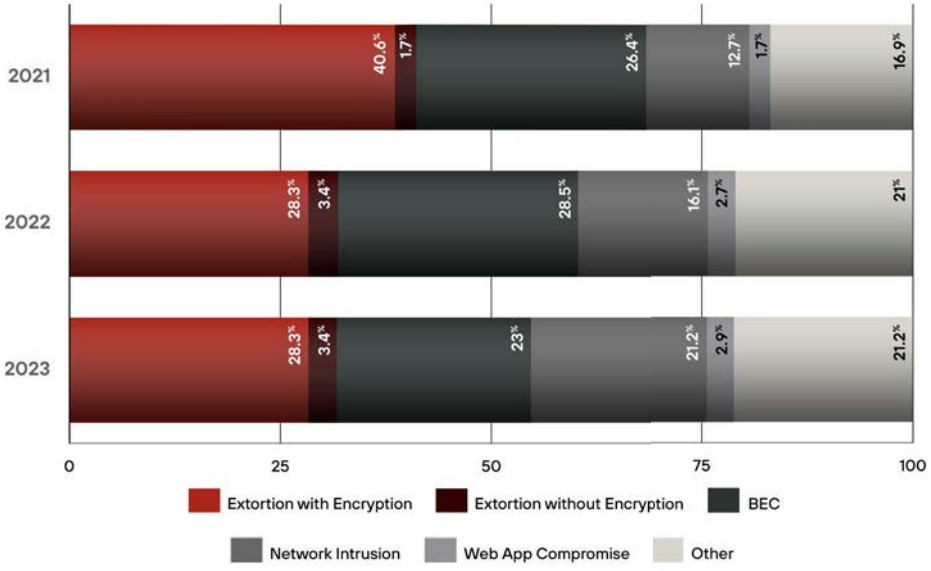
Investigation Types

We divide our case data by type of investigation.

This data is useful to understand the large macro trends, but it shouldn't be misinterpreted as the likelihood of getting hit by one or another type of attack. That probability is much more influenced by who you are, what data you have, and which threat actors are best matched to your threat profile.

Figure 3 shows the top five investigation types for the last several years.

Figure 3. Top five investigation types in Unit 42 IR cases from 2021 to 2023



These data points are a little more difficult to interpret. First, let's clarify the terms.

We divide Extortion into two subcategories: with or without encryption. Business email compromise (BEC) means an investigation centered on unauthorized takeover of email accounts. Network Intrusion is related to unauthorized access to a network environment–accessing a system or database without authorization, for example. And Web App Compromise means unauthorized access to, and/or taking advantage of a vulnerability in, a browser-based application.

While the top five investigation types are reasonably static, making sense of them is a little tricky. (Other is also in this data series, but since it's not meaningful in this context, we left it out).

We'll start with a caution. One trend that looks valid in the data–but might not be in real life–is the apparent decline in extortion with encryption. Those investigations are a decreasing fraction of our casework. However, there are a couple of confounding factors.

- **The organizations we're working with are changing.** Over these same three years, we have been helping more and larger organizations that, generally, have more advanced security maturity. They're more likely to have deployed endpoint controls, security automation and orchestration, network segmentation, and MFA. So, their resilience against encryption tactics is already higher.
- **Speed matters.** We and our clients may be interrupting incidents early enough to keep them from becoming ransom cases. Deploying encryption comes relatively late in the attack lifecycle. Our clients seem to be initiating investigations sooner than before. So, if an attacker has succeeded at intrusion but hasn't yet completed the other prerequisites to encryption, we don't classify it as an extortion with encryption case.

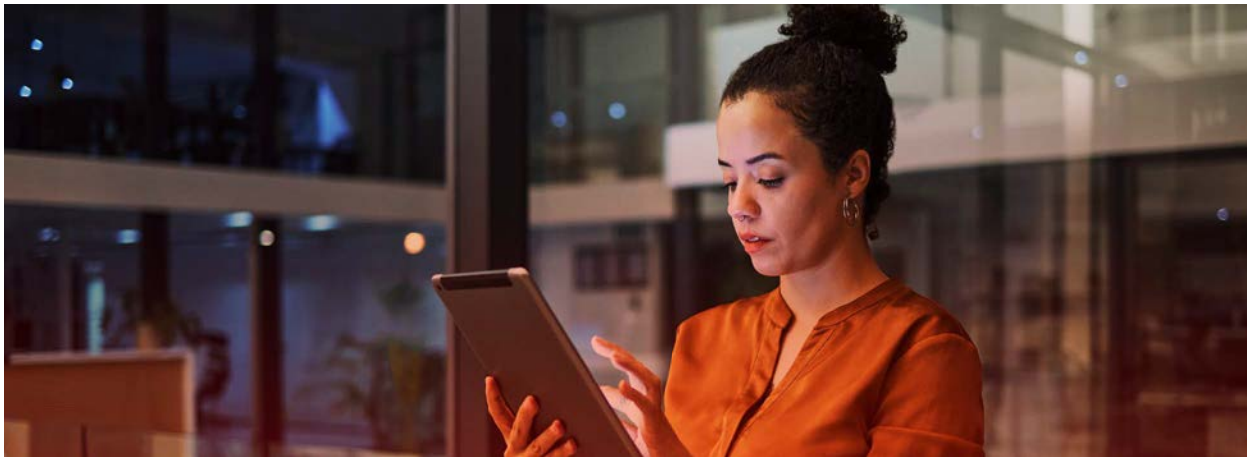
So there may be a dataset bias at work here. The same confounding factors may be at play for BEC.

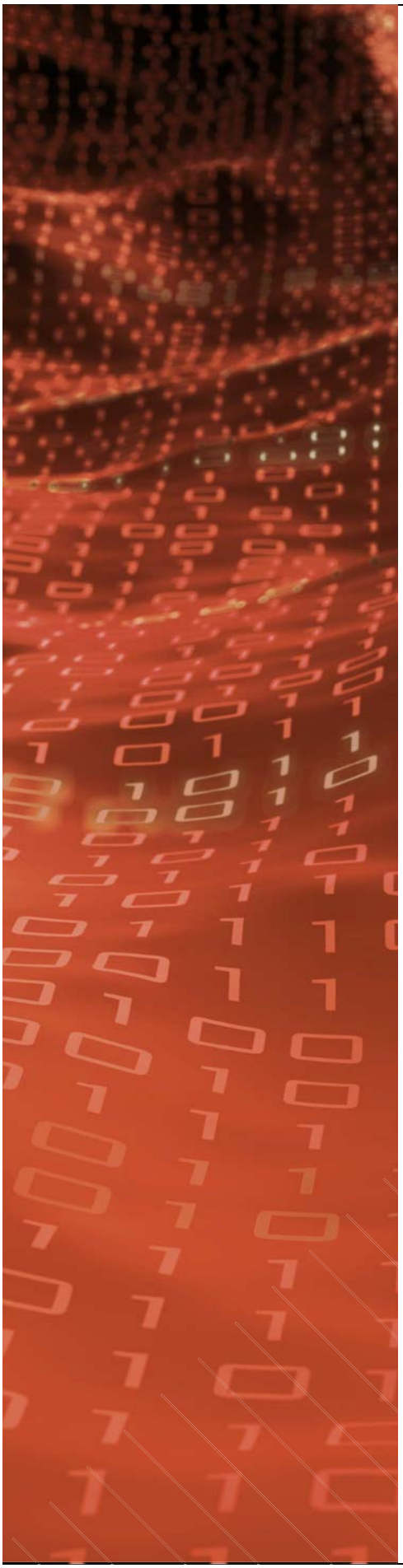
Now, with that said, there are a few things that seem reasonable on their face.

The increase in network intrusions is sensible in light of large campaigns involving software vulnerabilities (including MOVEit, Citrix Bleed, and Microsoft Exchange RCE). Likewise for Web App Compromise, which was the engine behind some truly massive campaigns in 2023, such as the SugarCRM CVE-2023-22952 authentication bypass and remote code execution vulnerability.

We refer to these as large campaigns because the vulnerability is visible from the internet. Threat actors exploited the vulnerabilities at a large scale by scanning large portions of address space looking for vulnerable systems. In the largest cases, they developed software that automated the scanning as well as the exploitation. Then, some groups automated the data theft. We'll go into more detail about that later.

The small but present growth of extortion without encryption also fits the anecdotal evidence. Some threat actors have found that the threat of publishing non-public data is more powerful than holding it hostage, and they're capitalizing on that.





Spotlight On: Extortion

Ransomware attacks have increasingly become a sophisticated form of cybercrime that can jeopardize large companies, government agencies and critical infrastructure. Educational institutions have shut down or failed due to ransomware attacks, and attack groups such as Muddled Libra use ransomware-as-a-service software as part of their repertoire. Real-world implications abound such as disruptions in supply chains or hospitals, or taking critical gas pipelines offline.

Extortion groups have increasingly sought to gain leverage wherever possible, particularly through multi-extortion tactics. This means that, in addition to encrypting data and holding business operations hostage until a ransom is paid, attackers often pile on additional nasty inducements to push their victims to pay. This may include stealing data and offering it for sale to the highest bidder–often publicly, through dark web leak sites. In other cases, Unit 42 incident responders have encountered groups that harass an organization's employees or customers.

In our observations of dark web leak sites, we saw a 49% increase in posting alleged victims when comparing similar periods in 2022 and 2023. There is much more extortion activity than what can be observed through public postings.

And multi-extortion tactics seem to be effective for attackers. As shown in the tables below, when we consider all cases involving extortion, we see various tactics used at a relatively steady rate from 2021-2023.

However, when we consider cases where payment was made, a different pattern emerges. The use of harassment in those cases has grown from less than 1% to 27% in just two years. Data theft has more than doubled in prevalence, from 40% in 2021 to 82% of cases in 2023.

Extortion Tactic	2021	2022	2023
Encryption	96%	89.9%	89.2%
Data Theft	52.8%	59.4%	53.3%
Harassment	5.1%	8.6%	8.3%

Table 1. Attackers' tactics have stayed relatively steady over the past few years when all cases are considered.

Extortion Tactic	2021	2022	2023
Data Theft	40%	70%	82%
Harassment	< 1%	20%	27%

Table 2. For cases involving extortion where payment was made, we saw a dramatic increase in the use of additional extortion tactics since 2021. This shows the efficacy of those tactics for attackers.

In general, it is best not to pay a ransom demand. Paying enriches criminals, further encourages threat actors to continue attacking organizations, and robs victims of their funds and other resources.

With that said, there are extenuating circumstances where organizations can be left with little choice but to pay. In these cases, work with incident response experts to negotiate demands if needed. These experts can also help avoid missteps, and offer advice on technical and other potential challenges.

Do attackers keep their promises once they've been paid?

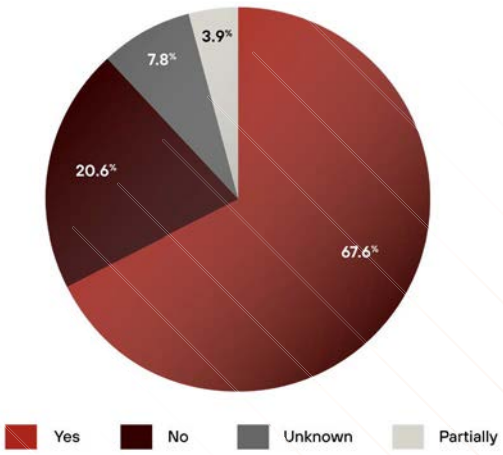


Figure 4. While in general it is best not to make payments in response to extortion, in cases where payment was made, we observed that attackers kept their promises more often than not.

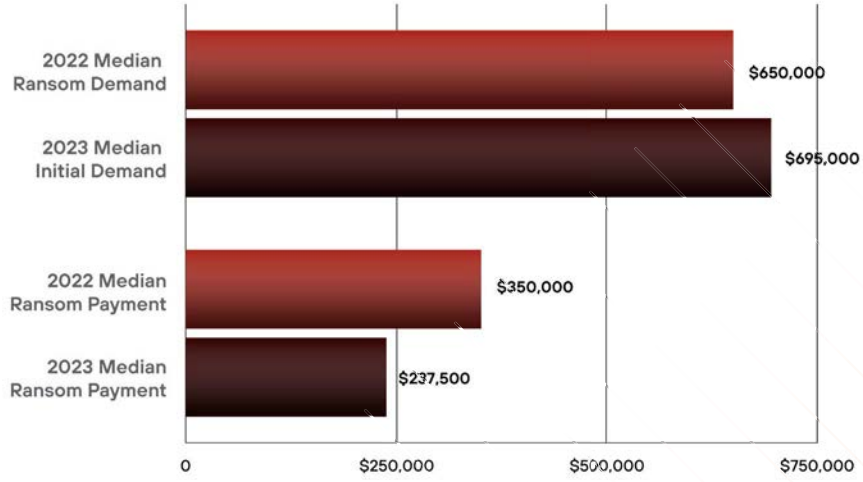
The Unit 42 team also typically advises that organizations get assistance to engage the threat actor even when they do not intend to pay, for a variety of reasons. First, stalling through communication can afford you more time to focus on recovery efforts. Second, you can use the extra time to determine whether the threat actor has stolen any sensitive information (including specifics on what was supposedly taken). Third, during communications with the threat actor, you can sometimes glean critical information that could be helpful for investigative and threat intelligence purposes.

The amount of time that an organization can spend negotiating with a ransomware group depends on how significantly the ransomware affects daily operations. On the one hand, organizations with no viable backups that cannot restore operations without quickly obtaining a decryptor will often pay on a more aggressive timeline.

On the other hand, organizations that have viable backups or that are not in a hurry can negotiate on a more extended timeline. As a general rule, negotiation conversations can often extend the original deadline set by threat actors. For example, threat actors may originally set a deadline of 72 hours to pay before they follow through on their extortion threat, but these negotiations last up to 11 days in most cases.

Typically, with most ransomware threat actors, Unit 42 researchers have observed a correlation between the length of negotiations and the percent reduction agreed upon for payment. In other words, the longer an organization can wait before agreeing to pay, the better "deal" they will often get. We've seen organizations decide to pay a ransom on the first day of negotiations, and we've seen organizations pay 35 days into the process.

Figure 5. Decrease in median ransom demand from 2022 to 2023



When organizations decide to pay a ransom, we recommend working with a payment vendor rather than arranging direct payment. These vendors provide a number of helpful services, including due diligence checks related to potential sanctions against certain threat actor groups. Without this type of check, paying a threat actor could introduce additional legal trouble for an organization, on top of the disruption of the ransomware attack.

Ransom demands are typically payable in some form of cryptocurrency, and the associated blockchain transactions are publicly visible to an extent. If you use your own cryptocurrency wallet, you might not adequately hide the flow of currency.

For example, we've seen cases where organizations paid directly, allowing the threat actors to garner clues about additional funds present. Sometimes, a threat actor may use this information to demand to be paid even more than was agreed—even up to the remaining balance in the cryptocurrency wallet!

A trick like this goes hand in hand with ransomware actors' moves toward multi-extortion. In general, they're looking for any way they can find to make more money or push an organization to pay.

The industry has even recently seen reports of threat actors taking advantage of the US Securities and Exchange Commission (SEC) disclosure rules to open a new avenue of extortion. In mid-November, news broke of threat actors reporting their victim to the SEC, and this could be a tactic we see more in the future. By using the threat of SEC enforcement, threat actors could threaten your time–whether or not your organization has actually been breached.

Tactics like this underscore the need for broad and deep defenses. While we saw above that backups can help an organization buy time and reduce damage, they're not enough on their own.

We recommend organizations focus on the following actions to increase their cyber resilience:

Maintain an incident response plan to prepare for and respond to cyber incidents, including emerging ransomware tactics like extortion, multi-extortion and harassment. Organizations that continuously review, update, and test their incident response plans– ideally with input from cybersecurity experts– are much more likely to effectively respond to and contain an active attack.

Ensure complete visibility of your attack surface. 75% of ransomware attacks and breaches fielded by Unit 42's Incident Response Team result from a common culprit–internet-facing attack surface exposure. Deploying solutions that provide centralized, near real-time visibility can help organizations identify and mitigate vulnerabilities before they can be exploited.

Leverage the power of AI and automation to modernize security operations and reduce the burden on overworked analysts. The latest technology can help organizations drive down cybersecurity metrics, like mean time to detect (MTTD) and mean time to respond (MTTR), denying attackers the time they need to compromise an organization's systems or exfiltrate its data. Additionally, technique-based protections mapped to the MITRE ATT&CK framework can help defenses nimbly evolve in response to adversarial tactics.

Implement enterprise-wide Zero Trust network architecture. This is a fundamental security principle that assumes the network is already compromised and implements processes that continuously validate the user, device, application, and data in a controlled manner. Zero Trust network architecture creates layers of security that prevent or limit an attacker from successfully moving laterally around the network. This provides victims with more time to detect, properly contain and remediate the threat.

Protect cloud infrastructure and applications. With cloud migration accelerating, threat actors will continue to develop tactics, techniques, and procedures (TTPs) designed to target and compromise cloud workloads. Organizations leveraging cloud infrastructure should implement a cloud security program and platform that offers comprehensive cloud-native security.

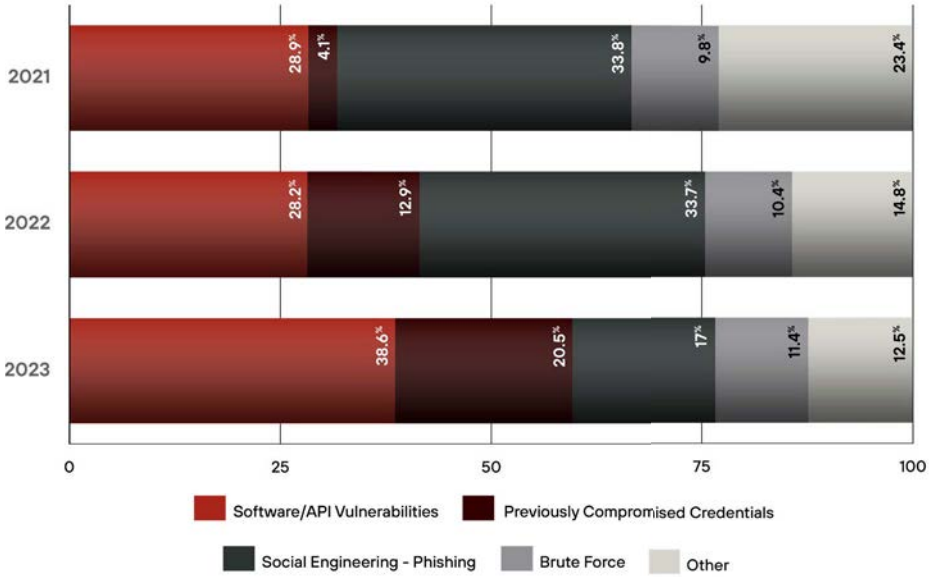
It's also important to maintain relationships that you can access quickly for help. This may include local and regional law enforcement, outside and inside counsel, and incident response assistance.

While there is no silver bullet in cybersecurity, prioritizing these recommendations will materially reduce the risk of falling victim to an attack. It also allows you to more effectively contain an attack if one does occur and help increase resilience for the entire cybersecurity ecosystem.

Initial Access Vectors

For each incident in the set, we assigned an initial access vector: the technique that gave an attacker their first access to the compromised environment. Figure 6 shows that in 2023, exploitation of internet-facing vulnerabilities shot to the top spot, displacing the previously perennial favorite, phishing.

Figure 6. Top 4 initial access vectors from Unit 42 incident response cases in 2023



This chart shows the top four tactics ordered by their prevalence in 2023, and it omits the cases where the initial vector couldn't be determined.

The increases in software vulnerabilities exploitation and previously compromised credentials stand out, as does the decrease in phishing attacks.

These changes are consistent with the higher-impact threat actors' behavior we've seen. As we discussed elsewhere in this report, one of the major threat groups changed during 2023. They began the year focused on phishing and social engineering the end users themselves. As time went on, they moved toward taking over their accounts by social engineering the IT helpdesk or by abusing self-service password reset procedures.

The growth of previously compromised credentials as an intrusion vector is also notable; a five-times difference in just two years. This underscores the importance of encouraging people to use a different password (or passwordless authentication) for every site. The marketplace for these credentials is vibrant and does not appear to be fading.

Coordinated takedown efforts between law enforcement and private industry effectively shut down individual marketplace sites, but replacements spring up quickly. Defenders will be fighting this problem unless and until passwordless authentication becomes the norm.

The one-year jump in software vulnerability exploitation is probably related to the several large automated intrusion campaigns that swept across the internet in 2023. We will have to see if that continues, while we hope it doesn't.

And what were those threat actors exploiting? The top exploited vulnerabilities in our 2023 dataset were:

1. CVE-2023-3519 (Citrix NetScaler ADC/Gateway)
2. CVE-2023-22952 (SugarCRM)
3. CVE-2021-44228 (Apache Log4j)
4. CVE-2023-34362 (MOVEit)
5. CVE-2020-14882 (Oracle WebLogic)

These initial access vectors will continue to change as threat actors adapt their tactics to the ones that work at a point in time. Most threat actors are in business, after all, so they need to optimize their effort just like defenders.

Most threat actors are in business, after all, so they need to optimize their effort just like defenders.

What It All Means

So, what's a defender to do with this information? We think there are a few meaningful bits.

Vulnerability Exploits Remain Relevant

Exploiting web applications and internet-facing software as a means of initial access came roaring back in 2023. It was the engine behind some of the largest-scale, most automated intrusions.

This means that patch hygiene, along with attack surface reduction, is still relevant, still important, and still hard to achieve comprehensively in large organizations. And you need to do it quickly. We have more recommendations for mitigations later in this report.

Defending Against Human Nature Is the Hardest Job

Social engineering fell behind software vulnerabilities for the number-one spot, but it's still a problem.

People are wired to help each other. Social engineering works in many cases because the attacker persuades someone to act against their own best interest, and a request for help does the trick.

Moreover, helping each other is one of the key goals of high-performing organizations. So, we are working against our own goals when we ask people not to trust each other or not to help someone who's asking for assistance.

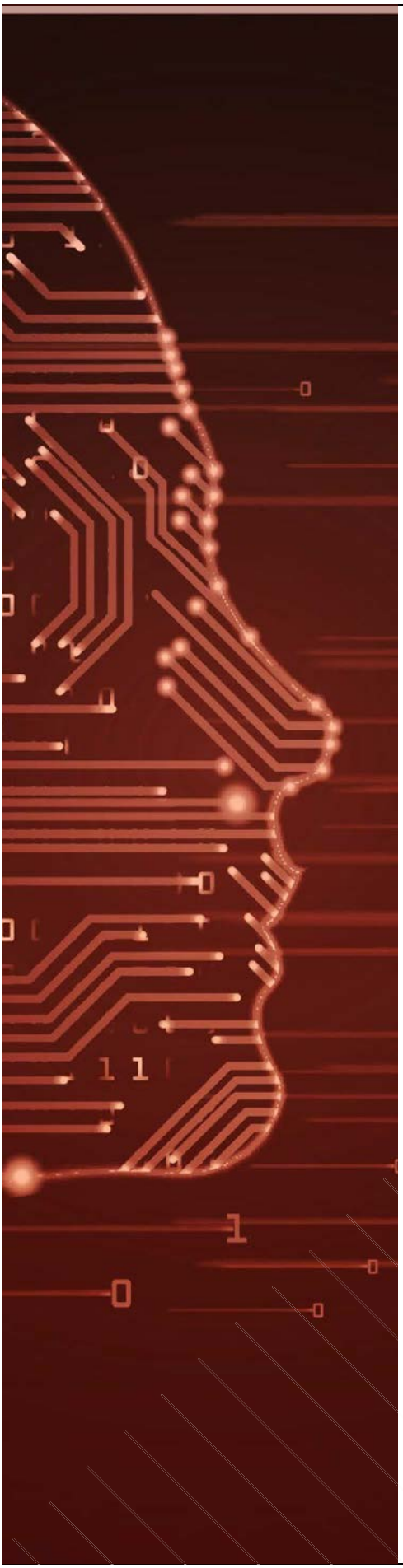
Technology can help here, but process, education, and critical thinking are just as important.

Belt, Suspenders, and Extra Underwear

Multi-layered defense is one of the core best practices in incident resilience. All the statistics in this section are focused around initial access and the first steps of an investigation. They're all opportunities for defenders to place a mitigation, a control, or an alerting mechanism that will block, slow, or reveal an attacker's activity.

But none of these defenses succeeds alone. Almost any security control can be overcome by a sufficiently motivated, skilled, and resourced attacker.

The most mature organizations protect everything with multiple layers, knowing that attackers will identify and evade some controls, but they'll also make mistakes. A perfectly executed intrusion is almost as rare as a perfect game in baseball. It only takes one hit–for the offense, or the defense.



Spotlight On: Artificial Intelligence

AI Brings New Possibilities for Attackers–But Defenders Stand to Benefit, Too

Many words have been devoted in recent months to discussions of AI and its potential. While the hype machine rolls on, AI already has a real impact on cybersecurity, with more to come.

Unit 42 sees evidence that adversaries have already begun to use and experiment with AI. Even so, defenders can also leverage the technology to scale response efforts and intelligently identify where to focus resources.

How Attackers Can Use AI

There are several ways attackers can use AI, but the first and perhaps easiest is to interact with the technology as so many people have; using AI chatbots. Like asking these chatbots for help with starting a resume or with sending a work email, attackers are using them to craft more realistic phishing emails with fewer obvious errors. Deepfakes also become easier to create with AI, opening the door to misinformation or propaganda campaigns.

We see signs that bad actors are using AI to attack organizations at a larger scale. They cycle quickly through attack vectors to seek effective ways in. Using AI makes it less expensive and faster to execute numerous simultaneous attacks aimed at exploiting multiple vulnerabilities. AI can also speed up post-exploitation activities such as lateral movement and reconnaissance.

Much has been made of the potential for AI-generated malware. However, at this time, our research suggests that AI is more useful to attackers as a co-author than as the sole creator of new malware. It is possible for attackers to use AI to assist with the development of specific pieces of functionality in malware. However, uses of this type generally still require a knowledgeable human operator. The technology may still make it possible for attackers to develop new malware variants quicker and cheaper.

Looking ahead, our consultants tell us that aside from AI's potential for speed, they're also concerned about its potential for patience. An AI could theoretically engage in a long, slow-burn operation aimed at eventually finding a way into an organization–perhaps over a time period that a human would be unlikely to sustain.

There's also the possibility that attackers could compromise generative AI tools and large language models themselves. This could lead to data leakage, or perhaps poisoned results from impacted tools.

AI Also Offers Hope for Defenders

Despite that list of how attackers can benefit from AI, it's important not to fear AI. Embracing the technology for defense opens possibilities for defenders to anticipate, track, and thwart cyberattacks to an unprecedented degree.

Anyone familiar with the work of the cybersecurity industry knows how often we face information overload. Mountains of security alerts come through every hour, of varying importance. Too often, defenders have to piece together data from disparate sources, while also trying to determine which of it really matters. According to industry research, more than 90% of SOCs are still dependent on manual processes.

AI offers a way out of that Sisyphean struggle. AI is particularly effective at pattern recognition, so cybersecurity threats that follow repetitive attack chains (such as ransomware) could be stopped earlier. While attackers may use AI to creatively identify new TTPs to obfuscate these chains, defenders could still leverage AI to detect and stop anomalous behavior.

The experience of Palo Alto Networks in our own SOC gives reason to hope.

On average, we ingest 36 billion events daily. We use AI-driven data analysis to automatically reduce that number to just eight events requiring manual analysis. Through this process, we have reduced our MTTD to just 10 seconds. Our MTTR sits at just one minute for high priority alerts.

Security operations metrics like this will be key to understanding the efficacy of an AI-powered security tool. Defenders should pay attention to not only MTTD and MTTR, but also to incident closure rates, as well as false positive rates and the false negative rates. Effective tools should show clear benefits to the organization by improving these metrics.

Those developing AI models can take steps to prevent threat actors from misusing their creations. By controlling access to the models, threat actors can be prevented from co-opting them freely for nefarious purposes. There is also the perpetual need to recognize confidential or sensitive information and build in ways to safeguard it. When a model detects this type of information, it may choose to block further queries or activity.

Designers should also be aware of the potential to jailbreak large language models (LLMs) by convincing them to answer questions that could contribute to bad behavior. Those creating AI models should consider that attackers will ask things like, "How do I increase the impact of an attack on a vulnerable Apache web server?" AI models should be hardened against lines of questioning like this.

Organizations should also make an effort to secure users accessing AI tools, ensuring visibility and control over how these services are being used within an enterprise environment. It's important to set clear policies for the type of data users can feed into AI services, protecting proprietary or sensitive information from exposure to third parties.

We believe the real risk of AI would be not recognizing it as the force multiplier it is, and to miss the opportunity to use it to meaningfully improve the situation of defenders. This technology can both make us more effective at protecting our organizations and offer hope for work-life balance in a field that too often feels always-on. AI can help make security data actionable, giving real-time visibility and the ability to prevent, detect, and respond to cyberattacks quickly.

Defenders have a responsibility to use AI as the resource it is.

After the Compromise

Once inside an organization, attackers have many options to achieve their objectives. They may execute the tools they need to use, install mechanisms to maintain access to the target environment, and try to remain unnoticed by the defenders.

Attackers may also need to identify where they are and what they can access, move around in the target environment, and create or use credentials to access data. Ultimately, they can collect and exfiltrate data, adversely impacting the organization. They may also (or instead) choose to make the data and systems unavailable.

These are areas of constant innovation and attacker creativity. And this year, the game of cat and mouse took a couple of wild turns. Here are some of the more interesting trends and techniques we observed, along with a few thoughts on which ones are likely to grow.

Execution

Threat actors often bring malware to a target environment. In 2023, about 56% of the matters we investigated involved malware in one way or another. That was consistent with its prevalence in the prior few years.

But while malware continues to be a problem, threat actors are increasingly "living off the land," using and misusing the tools and code that a target already has in their environment.

How do those two statements (i.e., "consistent with prior years" and "increasingly living off the land") go together ? The answer lies in the attackers' mix of tools. As they find system tools that can substitute for malware functions, they'll adopt the system tools. Many intrusions still use malware at some point. More and more are using living-off-the-land techniques along with the malware.

Endpoint protection platforms continue to be a great way to detect attacker activity of this kind. Even if the code isn't malware per se, some platforms can discriminate between use and misuse by employing user and entity behavior analytics. Defenders who use this capability to identify investigative leads through alerting and/or hunting enhance their visibility.

For example, a common certificate utility on Windows called certutil can also be used to download files, decode Base64-encoded software, and more. And it's a signed binary from the Windows installation, not malware. Can you tell when a threat actor is misusing it?

Some endpoint protection platforms can detect and alert on this activity. Detection might be based on the following:

- The action's novelty (is it unusual for this endpoint to run certutil?)
- Its function (what data is certutil processing?)
- Its context

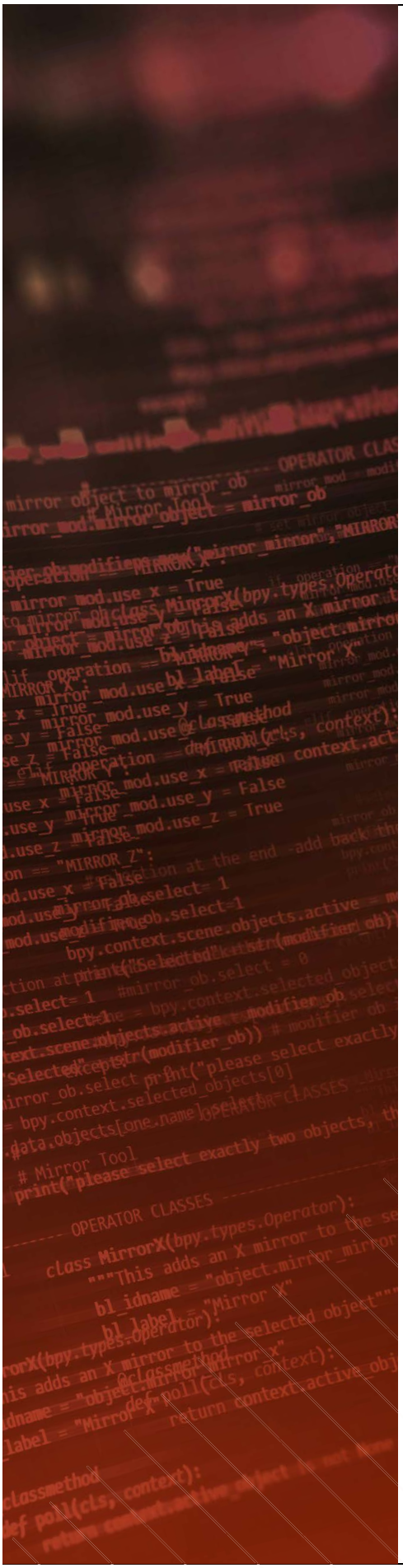
Now, let's return to malware for a moment.

Probably the most visible type of malware is ransomware: when you've been hit with it, attackers want you to know. Over the past year, 33% of the matters where malware was used involved ransomware. That was down slightly from last year, where it was 38%. However, we're not saying "ransomware is down!" The changes are small, and the overall fraction of cases involving extortion with encryption was about the same in both years.

Another high-impact malware capability is data destruction. It also creates widespread, visible effects. In 2023 cases involving malware, it was used to destroy data 4.1% of the time. While that may not seem like a lot, it's five times the rate from the year before. Together with its operational impact, we think protecting against wipers and destroyers is critically important. And that's especially true if you know you face threat actors who have used wipers in the past.

Time-based observations show that attackers are speeding up. Encryption and data destruction have dramatic and visible effects. Thus, it's wise to put effort into preventing attackers from gaining execution in the first place and exploiting it when they do.

We think attackers will continue to innovate tools and techniques here.



Spotlight On: Speed

Organizations Must Respond More Quickly to Attacks

Like the rest of us, **attackers are taking advantage of machine speed** to do more, faster. The 2023 Unit 42 Attack Surface Threat Report highlighted the speed at which attackers can scan the entire IPv4 address space looking for vulnerable targets. It also investigated how quickly high-profile common vulnerabilities and exposures (CVEs) are found being exploited after public disclosure. In some cases, it was within hours.

And, speeding up isn't just about finding vulnerabilities to exploit. In Unit 42 Incident Response cases, we've noticed that **attackers are exfiltrating data faster** as well.

Just two years ago, the median time between compromise and exfiltration was nine days. By 2023, it was two days–a full week less.

Especially striking is how often attackers move quickly. In almost 45% of our cases this year, attackers exfiltrated data in less than a day after compromise. This means that almost half the time, **organizations must respond within hours to stop them**.

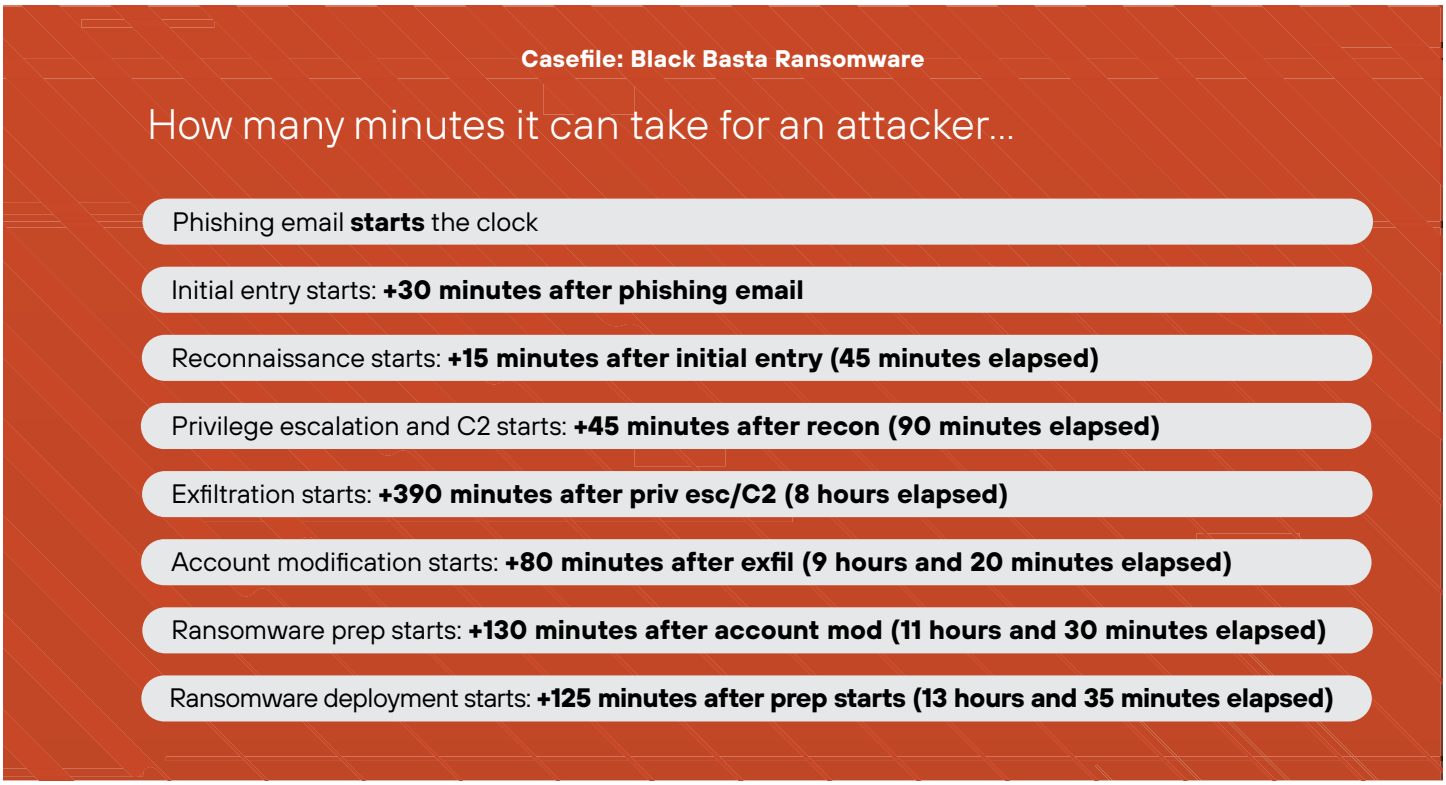


Figure 7. In less than 14 hours, attackers gained access to an organization, exfiltrated terabytes of data, and deployed ransomware to nearly 10,000 endpoints.

In certain kinds of cases, exfiltration happens even faster. While we expected that extortion cases would be the fastest to exfiltration, that wasn't the case. In matters that don't involve extortion, attackers are working within hours more often than not. **In 2022 and 2023, the median time to exfiltration for non-extortion matters has been less than one day.**

While we can't say for sure why this is, it may be because in some cases data exfiltration is the attacker's main goal. Attacker speed may be related to their focus–the faster the threat actors complete their mission, the more money they make. And the faster they work, the less time defenders have to get in their way.

The bottom line is that, for many organizations, exfiltration may be underway before Incident Response and Legal teams have even assembled a response team and plan. It might even be complete.

In one case involving Black Basta ransomware, it took attackers less than 14 hours to accomplish extensive compromise. Starting with a simple phishing email, the threat group gained initial entry to the organization a mere half hour later. From there, they explored the network, escalated privileges, and began communicating with their command and control (C2) server).

Most of this took place while the organization's security team was home, asleep and unaware. By the time the team began to realize something was wrong, the attack was well underway. The attackers were exfiltrating terabytes of data. They had created multiple custom versions of ransomware. And those tools were staged on a server inside the target's environment, ready to encrypt nearly 10,000 endpoints.

Latest Attacks Happening in Hours – Time to Exfiltration is Often Less Than One Day

Figure 8. In 2023, attackers exfiltrated data in less than a day after compromise, nearly half the time

Speed of Detection and Response is Key

In many cases, exfiltration may have already begun before incident response and legal have even assembled a response team and plan.

On the other side of the conflict, data from our incidents suggests defenders have not yet caught up to attacker speed.

Like most other incident response firms, we measure dwell time as the number of days an attacker is present in a victim environment before being detected. For this report, we measured the time between the first detection of the attacker and the earliest evidence of the attacker's presence.

The good news is that we see dwell time decreasing since 2021.

Median dwell time by year, days (all cases)

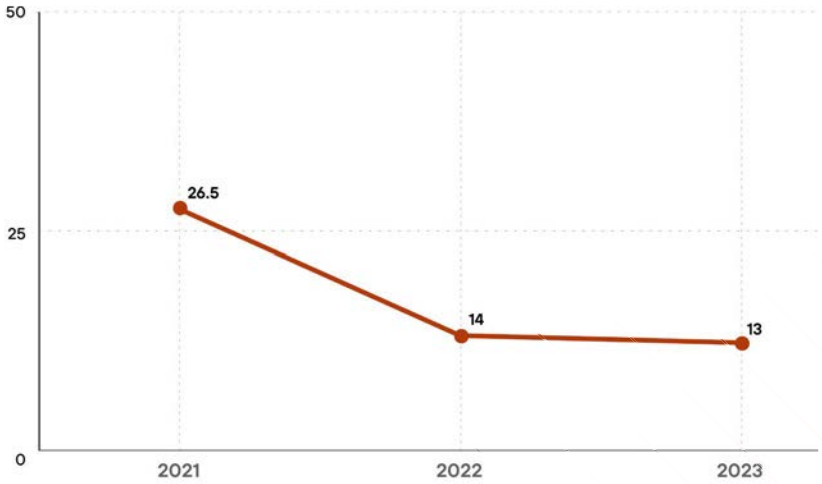


Figure 9. Median dwell time has dropped since 2021, suggesting that defenders are noticing attackers sooner.

On the flip side, even this decrease may not be enough. For example, in the US, the SEC adopted a rule requiring that material cybersecurity incidents be disclosed within four days. Given the time frames we observe, this could put organizations in the position of having to go public while still trying to figure out what happened–and what to say to customers, regulators, and the market.

This is even more of a concern when one considers that simply containing an attack may not be the same as fully remediating it. For example, an organization might contain an attack simply by deactivating compromised systems–a far cry from taking stock of the impact of the activity and fully restoring normal operations.

Persistence

Attackers have to keep a foothold in the environment if they expect to continue operations. In 2023, about 42% of our hacking-related investigations involved a backdoor, and 32% of malware-related matters had some kind of interactive C2 software.

But as we note in our Spotlight on Muddled Libra, some attackers are moving back toward reverse tunnels. These tunnels are connections out of the target environment that terminate on a system under the attacker's control in their infrastructure. These tunnels allow the attacker a great degree of freedom and often don't require malware on the compromised system.

To wit, tunneling tools are now built into most current operating systems. In the past, attackers had to bring their own remote-access methods, often in the form of C2 tooling. They still do–commonly, a post-exploitation framework such as Cobalt Strike or an RMM tool–but they aren't forced to. The operating system has built-in support for encrypted tunnels that may work just fine to evade notice by defenders.

Remote access tools are even less complicated to use. Microsoft Remote Desktop offers a graphical user interface (GUI) to a networked system, and organizations are still leaving it open. The 2023 Unit 42 Attack Surface Threat Report found 85% of organizations analyzed in the report had at least one internet-accessible Remote Desktop Protocol (RDP) instance online.

Another classic technique we continue to see is the use of web shells, which allows an attacker to use a compromised server as a beachhead into an environment. Placing a web shell on a system allows an attacker to continue to return over the long term. In 2023, we observed and reported on a series of APT intrusions in Southeast Asia that used web shells. We also observed web shells used as an attack vector in about 12% of hacking-related cases in 2023.

Attackers often use multiple persistence mechanisms. For example, in the attack against the Southeast Asian country, the threat actors

- Attempted to run their own custom backdoor software.
- Installed open-source VPN software to maintain access.
- Connected that VPN tunnel to two different hosts with github.com domain names.

Defense Evasion

In 2023, evading the defenders was where some of the most interesting advances in attacker conduct occurred.

Real-time Reaction

Certain attack groups, especially the most damaging ones, are less willing to give up and go home. Instead, they research, anticipate, and respond to incident response actions.

Our Muddled Libra threat assessment explores several examples. As we said there:

> *"Muddled Libra has been methodical in pursuing their goals and highly flexible with their attack strategies. When an attack path is blocked, they have either rapidly pivoted to another vector or modified the environment to allow their favored path.*
>
> *"The Muddled Libra threat group also repeatedly demonstrated a strong understanding of the modern incident response (IR) framework. This knowledge allows them to continue progressing toward their goals even as incident responders attempt to expel them from an environment."*

Said another way, this threat actor puts up a fight. They try to steal files that will tell them how their target is going to react once the victim identifies compromise activity. Then, they use that knowledge to install additional means of access that anticipate the response.

We expect to see more of this kind of behavior in the future.

Adding Attacker-Controlled Infrastructure

Perhaps most interestingly, we've seen a trend of attackers connecting their own infrastructure to a target's environment. More than just living off the land, they're actively installing and configuring new systems in the target's environment that are controlled by the attacker, not the defender.

Unit 42 observed and captured a novel BlackCat ransomware tactic in October 2023. This ransomware-as-a-service threat actor provides comparatively advanced tool sets to its affiliates.

One advance we noticed was a tool named Munchkin. Munchkin provides a Linux-based virtual machine that the ransomware operator installs and runs on a compromised system, circumventing host-based security controls in many cases.

BlackCat is no longer alone in using this tactic. The Ragnar Locker group has also been reported to use it. We expect to see this tactic more often.

Another tactic we've seen used to great effect is Okta cross-tenant impersonation (discussed publicly in Okta's blog post from August 2023). In this tactic, a threat actor uses highly privileged accounts to add a second, attacker-controlled identity provider (IdP). Then, the attacker federates their IdP to the target's true IdP. At that point, the attacker can sign on to applications at the target as any user. This is misuse of a legitimate function, so preventing it requires vigilance and tight permission management.

A third tactic we've observed is attackers standing up their own cloud workloads in the target's cloud infrastructure. Why not? It's incredibly fast and easy, and the target will pay the bill! This approach also blends in with normal IT or DevOps operations. It's not at all unusual, and most SOC teams will breeze right past it.

We expect attackers will continue to add infrastructure under their control. Defending against it means creating and maintaining a comprehensive understanding of the contents of your environment, one of the hardest maintenance tasks a security team faces.

Ironically, this may be one area where the cloud is easier to manage. If you have access to cloud discovery and security posture management tools, you may find ways you could detect unmanaged systems and accounts that have skirted your security controls. It may only require a friendly relationship with the people who run the cloud infrastructure management plane!

Disguising Host- and Network-Based Activity

Finally, a few classic tactics remain in the mix. Over the course of many investigations, we have seen threat actors take actions to try to avoid detection. Here are a few examples.

The first tactic is to try to appear normal. Unit 42 researchers identified a Chinese APT group that was doing this in several ways. (We haven't connected this activity to a named group yet.)

The threat actors had created C2 infrastructure hosted on a half dozen internet-facing systems, with domain names that suggest they're cloud storage services. A security analyst could easily wave away large data transfers to a cloud provider. It's pretty normal. However, further analysis revealed that the same SSL certificate was being used on all the servers, which is inconsistent with their disparate network locations and hostnames.

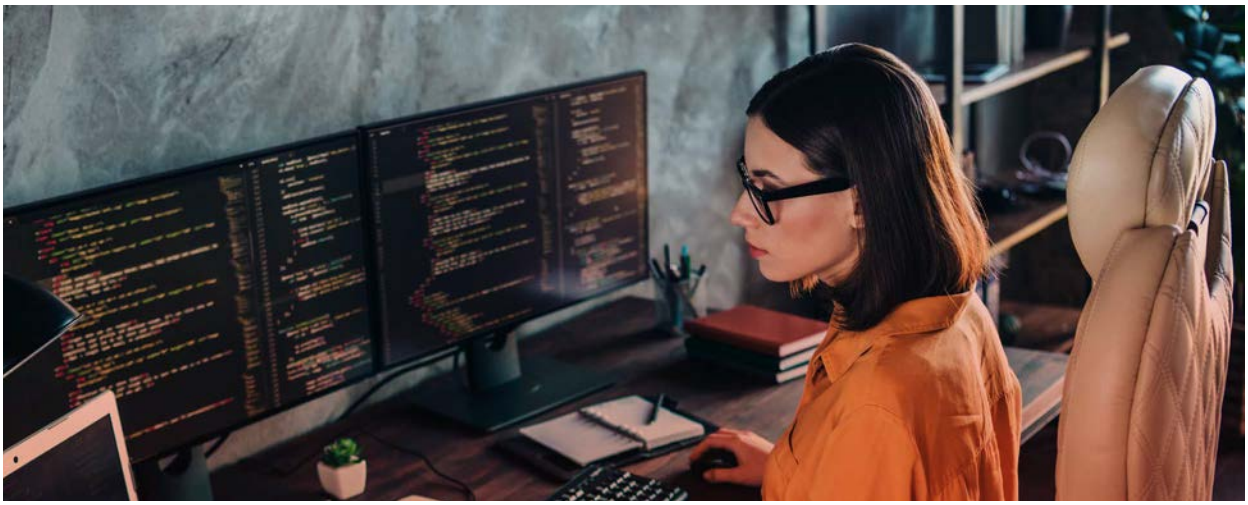
This actor also appeared to install honeypot software on the C2 servers. This too could persuade a hurried analyst to dismiss the suspect connections as benign, seeing known security testing tools.

A different threat actor, Stately Taurus, took an even simpler approach. Their C2 connections used HTTP POST methods that set the host field to a Microsoft domain name, despite the traffic being directed to an IP address in Malaysia with no relation to any legitimate Microsoft service.

The second tactic is related: a threat actor disguising themselves by changing their behavior, depending on who's asking.

The unnamed APT group's C2 servers seemed to change during the course of a day. They only accepted C2 connections during the threat actor's activity times and were closed at other times. While this may have helped the threat actor defend against casual observation, it looks unusual to the non-casual observer. Systems that change behavior based on time of day are uncommon.

And finally, the threat actor seemed to be filtering connections to the C2 infrastructure. They blocked connections from known Palo Alto Networks IP ranges, as well as some other cybersecurity companies and hosting providers.



These measures are an attempt to minimize the risk of being profiled by IP scanners. But they raise interesting investigative questions once they're observed in a potential attack context.

And the third tactic was host-based, rather than network-based. An Iran-based actor we call Agonizing Serpens tried to evade (EDR) solutions on the systems they had compromised. While most of the techniques they tried were well-known already, this threat actor hadn't previously been reported to use them.

The threat actor attempted at least three EDR bypass techniques. In this specific case, the target was using Palo Alto Networks Cortex XDR®–and the techniques did not succeed. But these offensive tactics could be applied to many defensive security tools.

Those techniques were:

- Trying to keep the EDR from auto-starting by disabling prerequisites; then
- Installing an anti-rootkit kernel driver and using it to try to stop a running process in the EDR tool; and finally
- Trying a second vulnerable-driver-to-process-kill technique from a recently published proof of concept.

The Unit 42 Threat Research Center has more analysis of the technique. The key takeaway, though, was that the group's use of evasive tactics and tools seems to indicate an upgraded capability compared to their past activity.

Discovery and Lateral Movement

The big news in lateral movement this year was turnabout. We found attackers using security tools offensively.

Many security tools need elevated privileges. It's wise to minimize those privileges, but some tools must run with the privileges attackers seek. For example, EDR/XDR tools usually can access and change any data on a system. That's the nature of the beast.

This year, attackers took that beast by the horns. We saw them compromise accounts belonging to security personnel with high privileges. Then, they used those accounts to access security tools that gave the attackers access to every system across the organization. And they used that access to move about and achieve their objectives.

Other TTPs of discovery and lateral movement remained familiar from prior years. Threat actors tend to use a variety of system administration tools at this phase, most of which are not malicious code and are frequently used by authorized personnel. Infostealers remain a common discovery (and sometimes initial access) vector for attackers.

Nonetheless, there are detection opportunities.

Internal port scanning is a quick way for a newly arrived threat actor to get the lay of the land. It's also noticeable. Watching for these actions with endpoint and network-based detection is wise.

Threat actors often use RDP connections to move laterally within a target environment, even if it is not accessible externally. RDP offers many user-friendly conveniences for interactive access and can leave a light evidence footprint. Logging RDP use generates a lot of data, but it is a great detective control. This is probably one of those areas where AI can be helpful to sift through masses of data.

Threat actors also frequently use Server Message Block (SMB) protocol for lateral movement since they know defenders will often assume this is standard file sharing activity.

Access logs for security tools (and IT tools that run with elevated security privileges) can reveal unusual access patterns. See our Recommendations For Defenders for more on this.

See our Recommendations for Defenders for more on this.

Spotlight On: Cloud Incidents

In 2023, Unit 42 responded to multiple cases where threat actors used a SugarCRM vulnerability (CVE-2023-22952) to gain access to cloud accounts. This shows us two things: threat actors are well aware of the value of cloud accounts in a distributed work environment, and they're willing to take roundabout ways to get access.

As the author of our article on these incidents described it, "When a threat actor understands the underlying technology used by cloud service providers, they can accomplish a great deal if they can gain access to credentials that have the right permissions." They can take advantage of vulnerabilities in a wide variety of cloud services to get that initial access to find the credentials they need.

Sheltering from the Clouds

Many organizations believe that choosing cloud services obviates the need for them to think about security. And in fairness, many cloud service providers sell their products as if they've taken care of all that dirty work. But identity and access management (IAM) is always the responsibility of the organization rather than the cloud service provider, and misconfigurations can undo even the most secure products.

According to our latest Cloud Report, Volume 7, at least three-quarters (76%) of organizations don't enforce MFA for console users, and 58% of organizations don't enforce MFA for root or admin users.

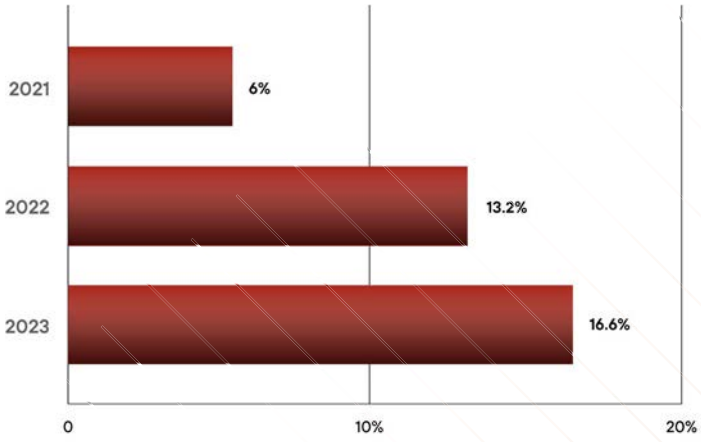
While every organization has a small set of risky behaviors that are repeatedly observed in their cloud workloads, that also means that a few changes will make a big difference to your overall security. In most organizations' cloud environments, 5% of security rules trigger 80% of alerts.

Forecasts Include Increasing Cloud Cover

While the rapid increase in remote work due to the pandemic caused a huge growth in cloud adoption, there is still more cloud coverage in our immediate forecast. Threat actors know this and are adapting their techniques to take advantage of this change in infrastructure.

Figure 10 shows that we've seen an increase in incident responses involving cloud cases, from 6% in 2021 to 16.6% in 2023. This matches an overall trend in organizations' environments relying more heavily on cloud infrastructure.

Figure 10. An increase in cloud cases from 2021 to 2023



Planning for a Rainy Day

This increase of interest in cloud adoption for both organizations and threat actors gives us some unique opportunities. Cloud computing is not going away anytime soon, as it offers a lot of productivity (and even security) benefits for organizations that adopt it. For those who move quickly to improve their security, they have a rare chance to get ahead of threat actor activity.

There are several things you can focus on to make a big difference in your cloud security:

- **Enforce MFA.** Having proper authentication on cloud service accounts is a way to make things harder for threat actors to access your environment.
- **Protect your data.** Sensitive data is at risk from both insider and external attacks. According to our last Cloud Threat Report, personally identifiable information (PII), financial records, or intellectual property, are found in 66% of storage buckets and 63% of publicly exposed storage buckets
- **Monitor abnormal activity.** If services within your cloud account are accessing or being accessed by new and unusual IP addresses or over unusual ports, make sure your monitoring is configured to alert on this activity.
- **React quickly.** Moving quickly to address security alerts can significantly limit damage. Security teams take on average about 6 days to resolve a security alert. Over 60% of organizations take longer than four days to resolve security issues.
- **Implement the principle of least privilege.** Granting least-privilege permissions is the most effective way to minimize the impact of security incidents. In our Cloud Threat Report, Volume 6, we found that 99% of the cloud users, roles, services, and resources were granted excessive permissions.

Credential Access

We continue to see threat actors using credential-dumping tools such as Mimikatz and MiniDump to extract user account credentials and password hashes. They also use tools targeting Windows Active Directory domain controllers themselves. Endpoint detections are the best way to find these. Investigate them promptly when they occur.

And watch for customized versions that will evade hash-based detection. We observed a threat actor using their own version of Mimikatz, requiring a password on the command line and omitting some of the usual features. Detect these tools by their behavior–attempting to steal information from memory–rather than just their hash, filename, or other easily changed attributes.

Not Just Passwords

More importantly, attackers are shifting more focus to stealing sessions, not just credentials. We have seen a variety of actors using "stealer" tools to extract saved passwords, session tokens, and other artifacts from browser and application caches on endpoints.

Session tokens in particular can pose a significant risk if they are long-lived and the system they're presented to has weak anomaly detection. With those tokens in hand, the threat actor assumes the identity and privileges of an authenticated user. They can perform all the same actions, use all the same authorizations. Would your tools notice if a session token was suddenly used in a different location than its owner?

Collection and Exfiltration

Over the past year, we saw vulnerabilities weaponized and exploited more rapidly and on a larger scale than in prior years. We have observed a trend toward the automation of campaigns–similar to the "scaling up" process seen in manufacturing–from prototype to mass production.

The first attempts at exploitation are conducted interactively, by hand. Then, there's a progression of automation from tools to assist human operators to machine-driven exploitation of the weaknesses.

Automating Theft

Threat actors are also automating post-exploitation activity. In particular, the size of the MOVEit Transfer published victim list is consistent with automation not just of the exploit, but also the collection and exfiltration of the data.

With MOVEit compromises, automation of collection and exfiltration was relatively straightforward because the stolen data itself was resident on the compromised, internet-facing server.

Too Much Stolen Data

But then, the threat actor, (CL0P) was faced with a different problem. Having stolen all that data, to effectively extort the targets, the threat actor had to publish some of it.

Unfortunately for the threat actor, the usual methods of extortion weren't able to handle the sheer volume of data they had stolen. Download speeds over Tor were too slow. They had to resort to publishing by torrent, which opened them up to further investigation by responders and researchers.

Look for more from Unit 42 on this case in the future.

Impact

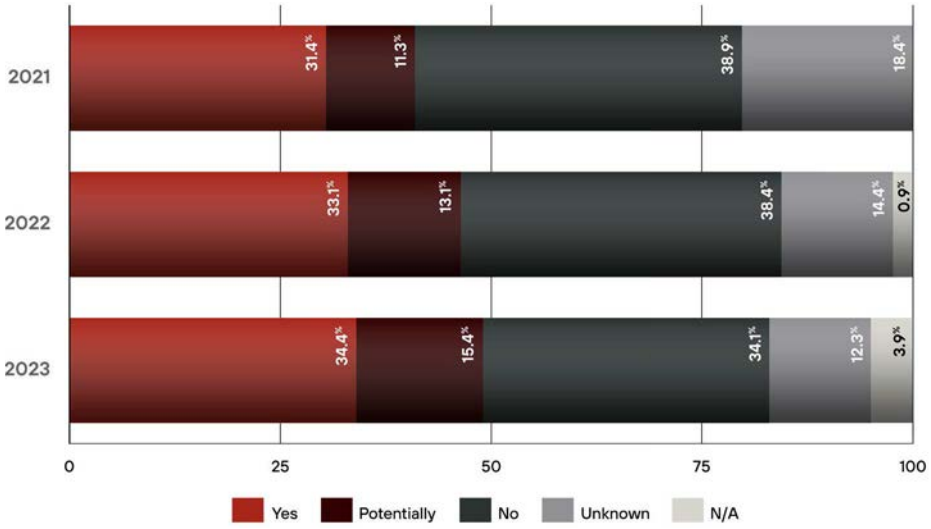
In the last year, clients felt more impact than previously, from lost confidentiality through lost revenue.

The effects of an incident last a lot longer than the attack and investigation itself. While reputation and share prices usually recover to their pre-incident levels, there are always soft and hard economic costs.

Data Exposure

One of the first questions a target asks after discovering a potential security incident is whether non-public information was disclosed, exposed, or compromised. Figure 11 shows that while we are able to answer that question most of the time, the answer continues to trend toward a more costly response.

Figure 11. Non-public data disclosure status in Unit 42 incident response cases from 2021 to 2023



Over 80% of the time, we can answer this question with a "yes," "no," or "maybe." Within that set, the "maybes" are eating away at the "nos," and the "yesses" are growing too. These changes aren't large, and they may just be sample variations from year to year. But the fact remains; almost half the time (49%) in 2023, a target knew or had to assume their information had been compromised.

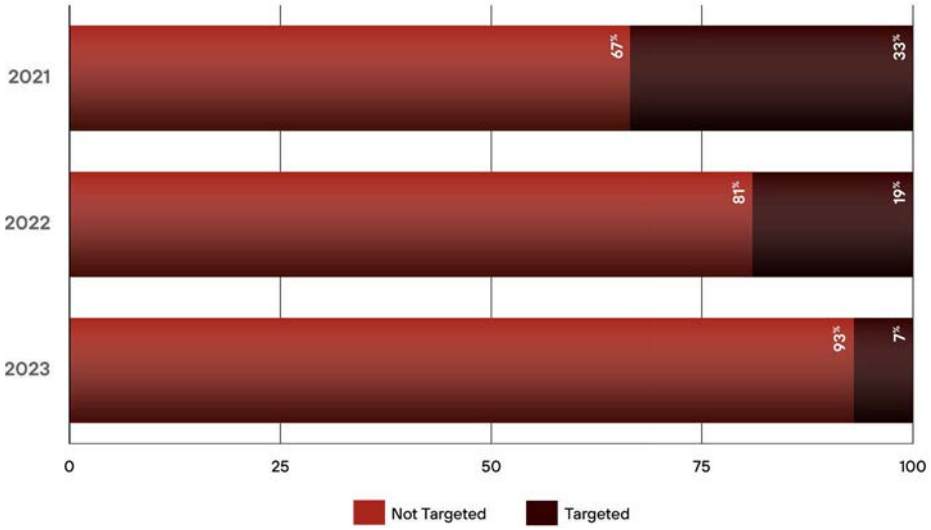
49%

Almost half the time (49%) in 2023, a target knew or had to assume their information had been compromised.

Was Data Targeted?

The second question a target asks is often whether the stolen data was targeted, or taken indiscriminately. Figure 12 shows that the answer here has a pronounced trend toward indiscriminate exfiltration.

Figure 12. Non-targeted data theft dominated in 2023



We say the data was targeted if the threat actor selected what data they were after. For example, they may have performed keyword searches, applied criteria, or filtered their file collection patterns. Filtering by extension, size, date range, or pattern matching–all of these are targeting actions.

Looking at 2023 in isolation, it's tempting to point at the large sweep-the-internet compromise campaigns and say that's why the not-targeted number is so high this year. But there seems to be a pronounced growth in the tactic of attackers just collecting it all and sorting it out later. Attackers are also moving to exfiltrate much sooner than in the past, which aligns with this grab-and-go strategy.

So the large volume of exfiltrated data is a bit of a double-edged sword. On the one hand, it's easier to detect large volumes of exfiltration than smaller, targeted transfers. On the other hand, when the stolen data is so big, the task of impact analysis and notification also gets a lot bigger. On the third hand, you'd need to detect that big exfiltration a lot earlier in the incident lifecycle. It might even be the first sign you notice that something is afoot, and you'll need to act on it fast.

Defenders should look at this trend as an encouragement to buy or build capabilities to detect attacker activity sooner, to accelerate response actions to interrupt the attack.

Was Business Interrupted?

As we said earlier, when you've been hit with a destructive or extortion-encryption tool, you know it. And in 2023, more of our clients experienced a business disruption of that type.

Figure 13. Business impacts in Unit 42 incident response cases in 2022 and 2023

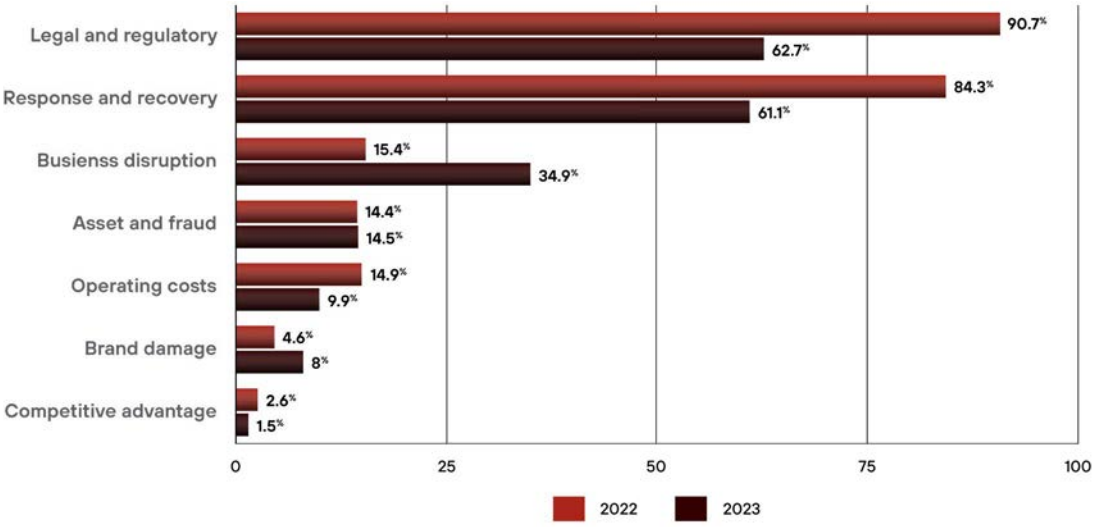


Figure 13 shows that in 2023, we worked more than twice as many cases where business disruption was one of the effects our clients experienced. Almost 35% of that casework involved business disruption of some kind. Usually, it was due to the effects of ransomware or data destruction malware.

And, increasingly, this disruption is publicly visible. There's a clear trend of greater public awareness of large-scale cybersecurity incidents. Certain threat actors take advantage of this publicity to pressure their victims.

Several other major effects occurred less frequently, such as legal and regulatory consequences, as well as other (non-incident response consulting) response and recovery costs. But brand damage was up, almost double. And, it's possible that the decreased frequency of legal and recovery effects is an artifact of our clients choosing to restrict the need-to-know.

Effects Go Beyond Consulting Fees

While we aren't in a position to characterize the total dollar losses due to incident response, we do keep track of what types of costs our clients experience. For this section, we're talking about prevalence–how many of our clients experienced that type of cost–rather than dollar losses.

What part of the business does an incident response affect? It's not just in the IT department. And the cost of response isn't only the consulting fees. (By definition, if Unit 42 was involved in a case, there was almost always our cost of consulting, so we've left that out).

For our clients, legal and regulatory losses led the list along with other costs of response and recovery. In the last two-and-a-half years, over 86% of our clients had some kind of legal or regulatory cost. And over 78% of them had recovery costs.

While it can be easier to prepare for losses that you expect, it's always harder to prepare for losses that you don't yet understand or can't predict.

In our experience, the best way to address that discrepancy happens when you can identify all the parties who would be involved in a large-scale incident response. Then, invite them all to participate in a Tabletop Exercise exploring two scenarios:

- One plausible ("We accidentally hire someone who's actually working for North Korea and they steal some of our data.")
- One outrageously implausible ("Our telephone system is compromised by a foreign intelligence service along with thousands of other potential targets.")

Practicing a what-if scenario is a much lower-impact way of discovering holes in your capabilities and processes than letting a threat actor show them to you. Particularly, it lets defenders engage and prepare other stakeholders in their business, who might not be participating in the day-to-day cybersecurity work, but who will find themselves on the incident conference bridge.

Recommendations For Defenders

While reviewing 2023's incidents, we kept in mind two topics close to defenders' hearts; factors that contributed to attackers' success, and mitigations defenders can use to frustrate attackers.

This section distills the most critical lessons into actionable recommendations for defenders. These are pragmatic, effective approaches to counter emerging and evolving cyberthreats.

Common Contributing Factors

Common contributing factors are the systemic issues or mistakes that contributed to attackers' success. Fixing these issues proactively lowers the chances or impact of a cyberattack.

Not Enough Visibility

You can't protect what you can't see. Organizations chronically lack comprehensive visibility across their network, cloud, and endpoints. This shortfall led to critical unmonitored areas, delaying the detection of and response to active threats.

Lack of visibility also led to uncontrolled and unmanaged vulnerabilities. In 11.5% of the incidents, insufficient patch management was a contributing factor.

Visibility gaps also led to unnecessary resource exposure, such as internet-exposed remote desktops or inadequately secured cloud workloads. These exposures contributed to 9.6% of cases.

Finally, insufficient coverage with endpoint protection technology was a contributing factor in 13.6% of the IR engagements we responded to.

These unmonitored areas provided attackers with opportunities and time to exploit weaknesses. Not having a comprehensive view across the attack surface–crucial for managing it effectively–significantly contributed to incident severity and frequency.

11.5% *In 11.5% of the incidents, insufficient patch management was a contributing factor.*

Too Much Complexity

In our post-breach investigations, we often find forensic evidence of what happened in the logs and digital artifacts. But this data is often spread across many different security point products, buried in millions of uncorrelated alerts, or structured in disparate formats. That distribution discourages automated prevention, detection, or real time response by security analysts.

This data issue gets worse when teams struggle to maintain expertise in the myriad of tools they've adopted for security operations. As a result, even when the early signs of an attack exist, the signal is lost in the complexity of the IT landscape.

Defenders miss critical attack indicators when they're overwhelmed by the sheer volume of information and operational demands. They can't see the forest for the trees.

Too Much Privilege

Allocating excess privileges to accounts is a too-common contributing factor in cybersecurity breaches. Over-privileged accounts give attackers the means to move laterally across an environment, accessing sensitive information and assets. Attackers particularly target administrative accounts because they can be used to gain access to other systems and accounts. Using those privileges amplifies the breach's impact.

This issue is especially acute in cloud environments, where Unit 42 research has discovered that 99% of accounts are over-permissioned. Extensive cloud privileges exacerbate the risk and potential damage of attacks.

Not Enough Authentication

Finally, the lack of MFA remains significant. In 2023, the absence of MFA was a contributing factor in over a third of incidents.

Recommendations for Defenders

You may have heard us advise that you should "never let a good incident go to waste." While Winston Churchill didn't say that–Rahm Emanuel did, in 2008–it's as applicable to cybersecurity incidents as political ones.

In this section, we make recommendations refined from hundreds of cybersecurity incidents in 2023. These guidelines can be a blueprint to bolster your defenses and avoid severe incidents like those chronicled in this report.

Remember, never let a good incident go to waste!

Get Visibility Across Your External and Internal Attack Surface

Criminals have more surface to attack than ever before. The continued adoption of bring-your-own-device practices, internet of things, and cloud-native applications and architectures continues to add vulnerabilities and misconfigurations.

In the cloud, hard-coded credentials, weak authentication, and inefficient alert handling lead to increased breach risk.

To make matters worse, the modern attack surface is in constant flux. On average, 20% of an organization's cloud attack surface is replaced each month with new or updated services, contributing to nearly half of new critical exposures.

Additionally, over 85% of organizations expose RDP to the internet for a significant part of each month. Doing so increases the risk of ransomware attacks and unauthorized access.

Get ahead of this now by establishing tools and processes that allow you to get your arms around your internal and external attack surface, including cloud, network, and endpoints.

External Attack Surface Management (ASM)

To manage your external attack surface, including cloud and internet-facing assets, first identify and control what you have. Use advanced scanning tools and services that specialize in discovering and cataloging external-facing assets, such as web applications, cloud storage, APIs, and any services exposed to the internet.

Then, protect that external access with MFA. Don't allow remote access (especially RDP) with just a username and password. Require a second factor of authentication, either behind a VPN or natively in the solution.

If you do nothing else, do this.

Next, continuously assess these assets for vulnerabilities and misconfigurations using active ASM applications and a cloud-native application protection platform (CNAPP).

Implement a rigorous process to remediate identified vulnerabilities and misconfigurations quickly. Order the list by risk and prioritize critical exposures that could lead to significant breaches. Regularly update and patch systems, apply necessary configuration changes, and ensure that security settings in cloud environments adhere to best practices.

Internal ASM

Internal ASM means gaining visibility into the network and endpoints within your organization. Deploy tools for network discovery and asset inventory management. Identify and categorize everything you find, including servers, workstations, and network devices.

Implement EDR or XDR solutions to monitor and analyze endpoint activities, identifying potential threats or anomalies.

Conduct regular internal vulnerability assessments and configuration audits to identify weaknesses within the internal network. Scan for unpatched software, insecure network configurations, and unnecessary open ports or services.

Remediate swiftly. Focus on patch management, secure configuration, and network segmentation to reduce the internal attack surface.

For both kinds of ASM, the key outcome is enabling a proactive, continuous cycle of identification, assessment, remediation, and improvement. This dynamic approach allows you to adapt rapidly to new threats and changes within the IT environment, ensuring that the organization's attack surface remains resilient against evolving cyberthreats.

Recommended Unit 42 Services:

- Attack Surface Assessment
- Compromise Assessment

Recommended Palo Alto Network Products:

- Cortex Xpanse®
- Prisma® Cloud
- Cortex XDR

Close Critical Protection Gaps with Principles of Zero Trust

Mixing weak authentication controls, over-privileged accounts, and improperly secured applications and information assets leads to critical breaches. This dangerous combination creates a straightforward pathway for attackers, with an easy route in, unfettered access to sensitive data, and an unobstructed route for data exfiltration or other disruptive impacts.

Zero Trust is designed to protect against exactly this.

Zero Trust is a cybersecurity strategy that secures an organization by eliminating implicit trust and continuously validating every stage of a digital interaction. Rooted in the principle of "never trust, always verify," Zero Trust principles protect modern environments by:

- Using strong authentication methods
- Leveraging network segmentation
- Preventing lateral movement
- Providing Layer 7 threat prevention
- Simplifying granular, least-access policies

Adopting Zero Trust is a journey. Here are the first steps and tactical recommendations we routinely advise clients to take.

Zero Trust is a cybersecurity strategy that secures an organization by eliminating implicit trust and continuously validating every stage of a digital interaction.

Improve User Authentication Controls

Multifactor authentication: Implement MFA everywhere you can. Prioritize methods like app-based or hardware tokens. Avoid SMS second factor if you can. It is vulnerable to SIM-swapping attacks that threat actors regularly perform.

Passwordless solutions. Where feasible, employ passwordless authentication methods such as biometric verification or FIDO2 security keys. These methods eliminate common attack vectors associated with passwords.

Single sign-on (SSO). Use SSO to reduce password fatigue and streamline user access, while maintaining stringent security checks.

Regular auditing and updating. Conduct frequent audits of authentication protocols. Ensure all methods are up to date and comply with the latest security standards. Find outliers and "temporary" exceptions that outlived their usefulness.

User education. Continuously educate users about safe authentication practices, including the importance of securing their authentication tools and recognizing phishing attempts.

Session management. Implement strict session management policies, including automatic logout after periods of inactivity and reauthentication for sensitive actions.

Aim for Principle of Least Privilege

User and system accounts accumulate excessive privileges with evolving technology, workforce changes, and organizational growth. This accumulation magnifies organizational risk and eases attackers' efforts to move within your organization and access sensitive data.

Audit identity and access management comprehensively. Implement rigorous need-to-know access controls. Pay special attention to cached credentials on endpoints, as attackers such as Muddled Libra exploit them, even from accounts protected by a privileged access management application. Minimize privileges for each credential, segregate them effectively, and continuously monitor how they're being used. Look for anomalies.

Ensure credentials expire when their owner no longer needs them. Grant access only when needed. Audit and monitor the use of long-lived, highly privileged credentials, restricting their scope to only essential roles and individuals.

Privileged users should have dedicated accounts for critical functions, like domain controllers or Microsoft Active Directory management. Separate these accounts from their regular user accounts. Doing so reduces the "blast radius" of a compromise. Over-privileged accounts can dramatically increase the impact of a breach, so minimizing privileges is key to containment.

Finally, segregate service accounts so that each has only one elevated privilege. This one-to-one approach eases anomaly detection and enables more precise response actions, allowing for targeted network shutdowns without widespread disruption.

Control Application Access and Use

The Zero Trust model is equally vital to managing system accounts and applications. Central to Zero Trust is the assumption that applications inherently cannot be trusted, needing continuous runtime monitoring to validate how the applications are behaving and being used.

Eliminate implicit trust between different components of applications during their interactions.

Control which applications are used in your environment, especially those that threat actors use to "live off the land." For example, restrict access to remote management applications and non-sanctioned file-hosting services. Or, at least, monitor and alert on them.

Segment Networks and Control Infrastructure

Network segmentation is a key tactic to apply Zero Trust principles to network architecture and infrastructure. This strategy effectively impedes the lateral movement of threats and confines breaches to isolated zones, significantly reducing the breach's impact.

Zero Trust network access (ZTNA) plays a critical role in this context. ZTNA verifies users and grants them access to specific applications based on identity and context policies. It eliminates implicit trust and restricts network movement to decrease attack surfaces.

Regardless of their location, no users or devices trying to gain access to a network are trusted until they've been fully verified based on identity and access control policies. This approach contrasts with traditional VPNs, which often grant complete access to an internal network. ZTNA solutions, on the other hand, default to deny and allow access only to services explicitly granted to the user.

Recommended Unit 42 Services:

- Cyber Risk Assessment

Recommended Palo Alto Network Products:

- Strata™ and CDSS
- Prisma Cloud
- Prisma SASE
- Prisma Access

Reduce Mean Time to Detect and Mean Time to Respond Using Security Tool Consolidation, Analytics and Automation

With attackers moving from intrusion to exfiltration in hours, detecting and responding to threats in near real time is crucial.

But, manually handling a barrage of alerts and data from disparate cybersecurity products is a labor-intensive and inefficient approach to threat management. Doing so misses critical alerts and leaves vulnerabilities exposed. Over 90% of SOCs still rely on these manual processes, giving adversaries an unnecessary advantage.

Collecting logs from security point products across your network, endpoints, and cloud can generate billions of events every day. Generating the telemetry is a great first step. The next one is to make sure this data is correlated, stitched together, and analyzed in a way that your defenders can take action on.

Consolidate your security tools and use extended detection and response (XDR) and/or extended security intelligence and automation management (XSIAM). They provide a unified platform that captures security telemetry from endpoints, networks, and cloud environments. Then, they give you a more streamlined and effective approach to threat detection and response.

These tools harness the power of machine learning and analytics to act as a force multiplier for the SOC analyst. This technology reduces detection and response times, shifting the advantage back to the defenders.

In our own networks, we have put this into practice:

- On average, we ingest 36 billion events daily. AI-driven data analysis triages this data to a manageable eight events that need manual analysis.
- MTTD: We have reduced this critical metric to just 10 seconds.
- MTTR: High priority alerts are now addressed within one minute.

And these results aren't unique to us. Our customers have reported a sixfold rise in the volume of security data they ingest and analyze each day. They also see a fivefold increase in incident closure rates.

By integrating these tools with AI and automation, organizations can transform their security operations, enhancing their ability to detect and respond to threats in real time.

Even with tool consolidation and automation, a strong security operation team is still critical. If you staff a twenty-four seven in-house security operations center, great. If not, consider the benefits and specialized expertise of managed detection and response. We have worked hundreds of incidents where the attack could have been stopped earlier, but critical alerts weren't reviewed soon enough, and the threat actor outran the defense.

Recommended Unit 42 Services:

- SOC Assessment
- Managed Detection and Response
- Managed Threat Hunting

Recommended Palo Alto Network Products:

- Cortex XDR
- Cortex Xpanse
- Cortex XSIAM®

Monitor Continuously, but Strategically

Continuous monitoring is crucial. Based on the incidents we responded to over the past year, here are some of the key items to watch for.

Authentication and MFA. Create rules to identify impossible logon scenarios such as geographic anomalies. Regularly audit MFA enrollment to ensure comprehensive implementation and identify gaps in logging. This thorough approach is critical for detecting and alerting on outliers and unusual usage, not just failed logins. Set up alerts for repeated authentication failures or unanswered attempts, which could indicate an attacker's efforts to overwhelm or misdirect authentication requests.

Remote Access. Given the risks associated with RDP abuse, it's crucial to restrict and closely monitor remote access. Assess and minimize the list of authorized remote management tools. Unauthorized tools should be blocked and monitored for, increasing the chances of detecting attacker activity. Clearly define and enforce which remote management tools are approved for business use.

Privilege Escalation. While it may be challenging for larger organizations to track all privilege changes, focusing on high-privilege accounts is essential. Log these changes and, if possible, correlate them with change control processes. Use identity threat detection and response tools to spot abnormal behaviors and collaborate across departments for workflow and communication improvements.

Defense evasion. Set up alerts for known defense evasion tactics like disabling host-based firewalls or antivirus programs. Quick investigation and response to these alerts are key. For businesses where the risk is acceptable, consider automatic containment measures.

Living off the Land. Be aware that skilled threat actors, like Muddled Libra, may attempt to use your security tools against you. This includes using endpoint protection for unauthorized commands or data loss protection tools to locate and exfiltrate sensitive information. Strictly monitor access to these tools, investigate any anomalies, and consider separate authentication infrastructure for critical security applications.

Exfiltration. Unusual spikes in data transfer or accessing sensitive data repositories outside normal patterns should trigger alerts. Implementing network analysis tools can help in detecting and preventing data exfiltration attempts.

Recommended Unit 42 Services:

- Compromise Assessment
- SOC Assessment
- Managed Threat Hunting
- Managed Detection and Response

Recommended Palo Alto Network Products:

- Cortex Xpanse
- Cortex XDR
- Cortex XSIAM

Respond More Effectively by Conducting Regular Incident Simulation and Testing

In an incident response, the difference between those organizations that have tested and prepared and those making plans for the first time is night and day. Take time now to prepare, practice, and test your readiness to respond.

Tabletop Exercises and Incident Simulations. Practice is critical in developing a robust response framework. These exercises provide an opportunity to refine and test your incident response plan in a controlled environment.

By simulating realistic scenarios, teams can identify gaps in their response strategies and improve coordination among internal and external stakeholders. The key is to establish a well-structured response plan in advance, rather than improvising during an actual crisis. Regular practice through mock scenarios ensures that when a real incident occurs, the response is swift, coordinated, and effective.

Penetration Tests and Red/Purple Team Exercises. Experience is an invaluable tool to assess the strength of your security defenses. These tests involve simulated attacks on your systems to identify vulnerabilities and test the effectiveness of current security measures.

Regular security audits and penetration tests reveal potential weaknesses before attackers can exploit them. By understanding and addressing these vulnerabilities, organizations can significantly enhance their security posture.

Security Awareness Training. A strong security culture is fundamental to effective cybersecurity. Maintaining one involves regular training and awareness programs. Ensure everyone is vigilant and understands their role in maintaining cybersecurity.

Muddled Libra attacks, for example, emphasize the importance of training beyond just email phishing. Personnel should be trained to recognize and respond to suspicious activities across various communication channels, including phone and SMS. Helpdesk staff, in particular, should be empowered and supported in pausing tickets for security verification without fear of repercussions.

This proactive approach is crucial in disrupting the initial stages of a potential attack. Spreading awareness across the entire organization is essential. Employees should be encouraged to support additional security measures, understanding their critical role in the broader context of organizational security.

Recommended Unit 42 Services:

- Tabletop Exercises
- Penetration Testing
- Purple Team Exercises

Improve Resilience and Speed Recovery with System Redundancies and Specialized Expertise

The value of effective incident response is being able to recover quickly. Doing so hinges on the preparedness and resilience of systems and backups, coupled with specialized expertise. Reflecting on the incidents we've encountered over the past year, several key lessons emerge.

Specialized expertise matters. Diverse challenges require diverse skill sets. Recognize your team's strengths and identify areas where external expertise would augment your own.

Specialized knowledge in breach-related legal issues, crisis communication, forensic analysis, threat hunting, negotiation with threat actors, emergency remediation, and threat intelligence are often crucial. Establish relationships with trusted partners in these areas beforehand. Standing retainers or contractual agreements ensure the immediate availability of these resources during a crisis.

Immutable, usable backups preserve value. Ransomware attackers target backups, so it's imperative to protect them through segmentation, whether that means storing them offline or using other access controls. Regularly test and confirm the effectiveness of your backups. Assess how swiftly you can restore critical components, like domain controllers or essential production systems, from these backups.

Prepare for rapid containment. Be prepared for a scenario where a complete credential reset is the only viable option. This process–often disruptive–can involve resetting passwords, unenrolling MFA devices, rotating API keys, or renewing Kerberos tickets. Mitigate the impact by planning this process meticulously before you need to do it. Document the reset protocol and store it securely offline to prevent unauthorized access by attackers.

Refresh the incident response plan. A well-documented incident response plan outlines roles, procedures, and timelines. The plan should detail the steps for a rapid and comprehensive reset to minimize disruption. Store this plan securely and ensure it's accessible offline–an essential consideration if your network is compromised and online resources are unavailable.

Out-of-band communications capability. Anticipate that attackers might be privy to your response strategies. Establish an out-of-band communication system for incident response, separate from your primary network.

This system should have its own identity and access management, ideally hosted on an entirely different platform. Secure it with hardware-based MFA to prevent unauthorized access. And practice using it.

Recommended Unit 42 Services:

- Breach Readiness Review
- Tabletop Exercises
- Incident Response Plan Development and Review
- SOC Assessment
- Penetration Testing
- Purple Teaming Exercises

From Vision to Practice

There's a lot of talk in the industry about security people burning out. And with seemingly endless creativity and effort from attackers, some days it does feel like our efforts are endless. Where do you even start?

Any consultant will tell you–and you can probably see this coming–"it depends." But good consultants will tell you what it depends on and help you create a plan to go forward from there.

We often work with clients on that transformation; during an incident, yes, but also beforehand.

Being part of Palo Alto Networks gives Unit 42 access to a comprehensive platform of technology. In this section, we'll provide some examples of how we use them.

Incident Response Services

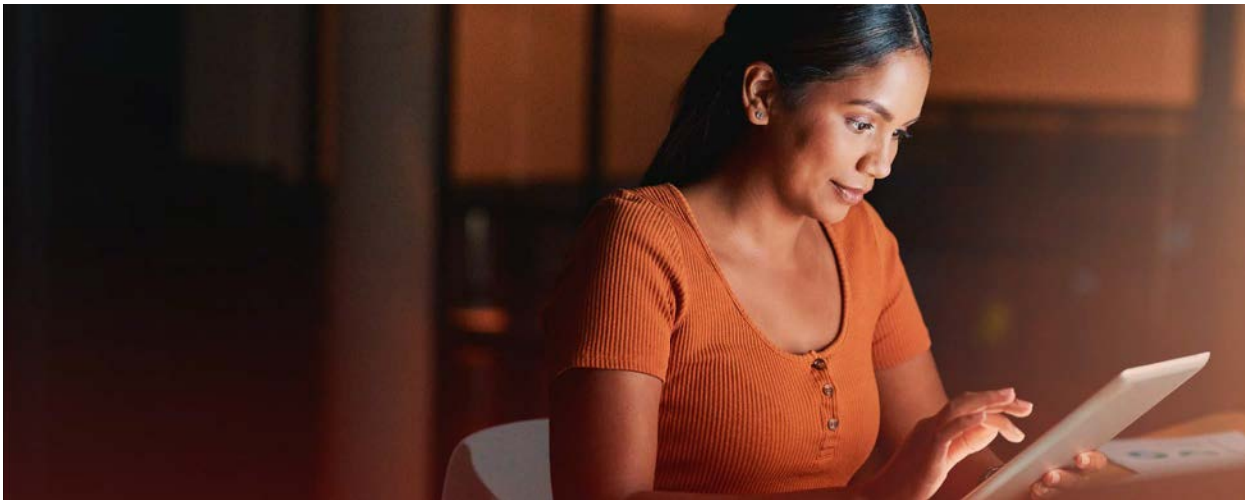
The Unit 42 Incident Response team is available twenty-four seven, year-round. If you have cyber insurance, you can request Unit 42 by name. You can also prepare for the future (and the present) by requesting any of our cyber risk management services. Many of our clients keep Unit 42 on retainer so we are available to them at a moment's notice.

Expert-Managed Services

Unit 42 has years of experience protecting organizations around the globe. Our Managed Services team can monitor, hunt and respond to suspicious activity twenty-four seven. We're ready to help you scale fast and focus your internal team on more strategic tasks.

Unit 42 Managed Detection and Response provides world-class analysts, threat hunters, and researchers who investigate and respond to attacks.

Unit 42 Managed Threat Hunting proactively hunts for advanced threats and creates detailed reporting on its findings to help you defend and keep peace of mind.



Attack Surface Management

Insufficient patch management and unnecessary resource exposure were contributing factors in dozens of cases. And our Unit 42 Attack Surface Threat Report researchers found about a 20% per month rate of change in cloud-based IT infrastructure. It's important to minimize the risk here.

The first step is keeping an up-to-date inventory of your internet-connected assets, from the perspective of an attacker; from the outside-in.

The next step is doing that continuously and comprehensively. With that high rate of change, point-in-time audits are out of date as soon as they're complete. Instead, automate.

Cortex Xpanse is an active ASM solution that helps actively discover, learn about, and respond to risks in exposed systems and services.

Reactively, we use Cortex Xpanse during incident response to find a client's entire attack surface. Several times, it's shown us an attacker's entry point that the client didn't even know they had.

Proactively, ASM is part of our Unit 42 Attack Surface Assessment service. It boosts your visibility into internet-connected assets, and we recommend actions you can take to better defend your organization.

Endpoint Protection

Lack of centrally managed endpoint protection was a significant contributing factor in 2023 incidents. Attackers are great at finding the unprotected sections of the infrastructure.

Naturally, we recommend Cortex XDR as our first choice. It's an extended detection and response platform that integrates data from the endpoint as well as other sources. We particularly like how it can stitch together endpoint, network, cloud, and identity data. That helps with attack detection, and it also simplifies our investigations by putting all the data together for us. It can do the same for you.

Cortex XDR also uses AI-driven technology for analytics, threat insights, file and incident analysis. That's in line with our goal to use automation-first processes that scale, because attackers are scaling, too.

During a live incident response, we often also need endpoint controls for capabilities like endpoint isolation, exploit and execution prevention. Cortex XDR does that for us as well.

Unit 42 conducts Compromise Assessment engagements using Cortex XDR. These assessments answer the key question: "Am I compromised?" We use extensive telemetry data from your environment together with data from Palo Alto Networks customers worldwide and Unit 42 threat intelligence to identify threats and vulnerabilities in your organization.

We prioritize our recommendations and align them with your security strategy. (We can help you develop that strategy, too.) Then you can move forward with implementation more confident that you're strengthening your defenses.

Automation and Orchestration

Threat actors are creating larger and faster effects, consistent with their use of playbooks and automation. Defenders should automate as well.

In our own Palo Alto Networks SOC, we use Cortex XSOAR®. We've measured its performance in our environment, and it does the work of 16 full-time employees, saving our analysts an average of 2,600 hours per month. If you have invested in creating playbooks and processes but struggle to get them automated, XSOAR is a great place to start.

Unit 42 Incident Response offers several proactive assessments that relate to automation. If you've built or are thinking about building a next-generation SOC, our SOC Assessment can help you design and build that capability. If you already have processes in place, consider our Ransomware Readiness Assessment or BEC Readiness Assessment to evaluate them.

Zero Trust Network Access

Many of the incidents we worked on could have been smaller or prevented entirely. The framework we recommend to clients is ZTNA 2.0, integrating authentication with fine-grained least-privileged access. Access to organization resources should be minimal, continuously evaluated and comprehensive.

Palo Alto Networks Prisma Access is purpose-built, cloud-delivered and reduces risk. It delivers many networking and security services, from firewalls-as-a-service through SD-WAN and malware blocking. The feature list is long.

Cortex XSIAM has Identity Threat Detection and Response that combines behavior analytics with identity threat detection to respond to threats like insider, data exfiltration, lateral movement, and other actions.

And because identity authentication is a key part of ZTNA 2.0, we recommend phishing-resistant second factor authentication, such as FIDO2 keys. Move on this as quickly as you can.

Unit 42 Incident Response offers several engagements around ZTNA 2.0. Our Cyber Risk Assessment and Security Program Design, for example, helps you define a defense-in-depth strategy that can align with this framework.

Preventive Technology Controls

Many organizations can't answer the question, "Which remote management tools are approved for business use?" Define and enforce an answer.

Cortex XSIAM has embedded ASM capabilities that let you find outliers to the answer.

To minimize the risk of data exfiltration to unapproved locations, you'll need more than just packet filters. SaaS Security has continuous discovery, categorization and control using our App-ID™ technology.

Processes and Personnel

Define and practice your incident response process. The most effective incident responses are executed from a thoughtful structure that was designed ahead of time.

Unit 42 can help you exercise those processes in Tabletop Exercises. You'll simulate your response to a severe incident. We'll base the scenario on real-world breaches and industry-specific threats.

For a stress test, consider a Penetration Test by the Unit 42 red team. We'll apply the tactics, techniques and procedures used by threat actors, and you'll get to defend against them. For even more benefit, consider a Purple Teaming Exercise, where our defenders collaborate with yours against the red team.

Palo Alto Networks, with Unit 42, has a broad offering of products and services. We would love to help you identify which ones would be valuable to your organization.

Get in touch with Unit 42

Spotlight On: Predictions

If you talk to a lot of incident responders, one aspect of the work will keep coming up. Most of us enjoy problem-solving. Even when you encounter the same threat actor across several incidents, the differences are the fun part. What will the threat actor do differently this time? What will be different about the target environment? What one weird thing will happen and make the incident response totally different?

That's the joy of the future tense. The past informs us. We look around in the present. And we find out in the future.

Here are some topics we think will be important in the coming year.

AI and Social Engineering

AI has dominated technology news and development in the last two years. AI has technological benefits for both attackers and defenders, principally around scale. We think attackers will leverage those benefits with greater sophistication and scale, specifically around social engineering.

We expect to see attackers using AI tools with greater sophistication and scale than today. We expect they will target voters, seeking to create friction within and among political groups.

And we expect to see this activity play out within organizations, as well. AI will enable attackers to operate at greater scale, creating more (though not necessarily better) social engineering content. This AI-enabled content will become harder to discriminate from "authentic" content. Organizations will become increasingly reliant on advanced detection capabilities to bridge the resulting detection gap. Now is the time to stimulate awareness of and preparation for handling inauthentic AI-enabled information.

Encrypted Data Stockpiles

Quantum computing is still a decidedly futuristic capability. Nevertheless, when it eventually becomes available, attackers will suddenly be able to decrypt material that was previously impenetrable.

We expect to see attackers harvesting and storing encrypted data, with an eye toward one day breaking it open using quantum cryptography.

Organizations should take stock of the encrypted data under their control, noting what has been stolen in the past that could become unprotected in the future. Organizations should also understand and increase their resistance to quantum threats as part of system design and operation.

DevOps Expertise Will Expand Further

DevOps (and SecDevOps, and any number of other mashed-together words to describe IT) has enabled defenders to scale technology operations to better operate and secure their organizations.

We expect attackers will demonstrate their skills and expertise in DevOps, IT, and security as well. They will use their target's existing IT and security tools to meet their own malicious needs. We expect to see attackers manipulate and subvert security controls to evade detection and persist within the target environment – without needing malware.

Organizations should assess the security of their security. Review the authentication and authorization controls around DevOps and security tools, and then establish compensating controls and procedures to protect your protections.

Large Language Models Will Seed Disinformation

LLMs have rapidly become the go-to tools for anyone who needs to create large volumes of text. LLMs are good at creating plausible text, even if it's not always grounded in reality.

We expect attackers will look for opportunities to seed disinformation at the source. They will manipulate the sources and data that underpin these LLMs to influence the models' output for the attackers' own purposes.

Organizations should maintain a healthy distrust of LLM-generated material. Grounding decisions on verified data–not conjecture or hallucination–will remain the best way to operate in an increasingly untrustworthy information environment.

A Few Words About the Data

Every incident response and threat report is a product of the authors' point of view and casework. This one is ours. While statistics are useful, the real value of a report is its perspective on those numbers. That perspective arises from helping some of the world's largest organizations through encounters with the most determined, skilled threat actors out there.

As well as that hands-on experience, Unit 42 has access to comprehensive cyberthreat data from a broad portfolio of Palo Alto Networks security products, including:

- Our XSIAM SOC transformation platform
- Our ability to monitor attack surfaces at scale with Cortex Xpanse
- Observations of attack traffic from our Next-Generation Firewall
- Behavioral insights from Cortex XDR monitoring across endpoint, network and cloud

This data underpins our threat intelligence and informs our incident responders.

Data from incident response cases forms the backbone of this report. To assemble the information here, we reviewed the findings from more than 1,200 cases over the last 2.5 years.

This dataset includes anonymized data from several incidents that made headlines as well as many that didn't but were no less impactful. These cases involved BEC, ransomware, insider threat, nation-state espionage, network intrusions, and inadvertent disclosures. Our clients range from small organizations with fewer than 50 personnel to Fortune 500 and Global 2000 companies and government organizations with more than 100,000 employees.

While most of the targeted assets in these cases were located in the US, the threat actors conducting the attacks were located worldwide, and they targeted businesses, organizations, and IT infrastructure around the globe.

We supplemented the case data with in-depth interviews with experienced security consultants to gather anecdotal and narrative insights from their work with clients in specific areas of expertise.

Our recommendations and observations are based on areas where threat actors were largely successful. As such, the lessons themselves have broad applicability.

About Palo Alto Networks

Palo Alto Networks® is the world's cybersecurity leader. We innovate to outpace cyberthreats, so organizations can embrace technology with confidence. We provide next-gen cybersecurity to thousands of customers globally, across all sectors. Our best-in-class cybersecurity platforms and services are backed by industry-leading threat intelligence and strengthened by state-of-the-art automation. Whether deploying our products to enable the Zero Trust Enterprise, responding to a security incident, or partnering to deliver better security outcomes through a world-class partner ecosystem, we're committed to helping ensure each day is safer than the one before. It's what makes us the cybersecurity partner of choice.

At Palo Alto Networks, we're committed to bringing together the very best people in service of our mission, so we're also proud to be the cybersecurity workplace of choice, recognized among Newsweek's Most Loved Workplaces (2023, 2022, 2021), with a score of 100 on the Disability Equality Index (2023, 2022), and HRC Best Places for LGBTQ equality (2022). For more information, visit www.paloaltonetworks.com.

About Unit 42

Palo Alto Networks® Unit 42® brings together world-renowned threat researchers, elite incident responders and expert security consultants to create an intelligence-driven, response-ready organization that's passionate about helping you proactively manage cyber risk. Together, our team serves as your trusted advisor to help assess and test your security controls against the right threats, transform your security strategy with a threat-informed approach, and respond to incidents in record time so that you get back to business faster.
Visit paloaltonetworks.com/unit42.

3000 Tannery Way
Santa Clara, CA 95054

Main +1.408.753.4000
Sales +1.866.320.4788
Support +1.866.898.9087

www.paloaltonetworks.com

2024 Unit 42 Incident Response Report 02/2024.