Mandiant

Issue 6

Cyber Snapshot Report

Google Cloud

The Defender's Advantage Cyber Snapshot report provides insights into cyber defense topics of growing importance based on Mandiant frontline observations and real-world experiences.

Contents

When AI Becomes a Crown Jewel

Mandiant works with clients across a mix of industries and geographies with varying challenges and targeted outcomes; however, a growing commonality across all of them is how to surface the risks and implement mitigations to safeguard their consumption and use of Artificial Intelligence (AI). When it comes to technology and cyber risk, we believe that newer technologies are not dissimilar to other operational risks within an organization, and that businesses need a prioritized approach to understanding where those AI technologies reside, their dependence on them, and their impact on the business if misappropriated.

The spectrum of AI

AI technologies are progressing at an immense pace. The opportunities for organizations to leverage developing technologies are almost limitless and bound only by an organization's own ingenuity and creativity. Companies are rapidly exploring how they can use these innovations to extract additional value from within their organizations, provide unique and rewarding experiences to their customers, and differentiate themselves in an ever-growing competitive landscape. An outcome of this is a spectrum that now presents itself where organizations are finding themselves to be: consumers of AI, creators enabled by AI, or somewhere firmly planted in the middle.

No matter where an organization finds themselves across this AI spectrum, there is a growing need by the global industry to ensure the appropriate safeguards are incorporated by 'default', and not only within the technologies leveraged, but also the very processes that are used to create them. In 2023, Google released its Secure AI Framework (SAIF) designed to promote a safe and collaborative way to adopt AI for the betterment of organizations while maintaining community safety. The tenets of the SAIF framework are built around the following principles:

1. Expand strong security foundations to the AI ecosystem
2. Extend detection and response to bring AI into an organization's threat universe
3. Automate defenses to keep pace with existing and new threats
4. Harmonize platform-level controls to ensure consistent security across the organization
5. Adapt controls to adjust mitigations and create faster feedback loops for AI deployment
6. Contextualize AI system risks in surrounding business processes

Understanding and identifying specific AI-related risks to your business

To truly understand what forms of risk could materialize from AI, it is important to first understand the business context (similar to guidance from SAIF), how AI will integrate into the organization from end to end, and what business processes and capabilities will be supported or required to optimize their use.

To effectively engage this process, we suggest focusing on 4 key vetting phases as part of a Crown Jewels-based approach:

1. Identify	• What are your critical business units and processes? • Where is AI being used to enable these business processes? • What type of information does your AI model store and process (e.g., employee PII, customer data, intellectual property)? • Are you selling AI related products (B2B or B2C)?
2. Threats	• What external and internal threats would impact your AI Crown Jewels the most? • What type of threat actor would be the most likely to target your AI asset? • What motivates an attacker to target your AI asset?
3. Vectors	• What attack vectors are you inherently vulnerable to? • What would the impact of compromise be to your business? • Are these impacts related to financial, operational, compliance, or reputational matters?
4. Countermeasures	• What prevention, detection, response, and countermeasures do you have in place to minimize your risks? • Are these countermeasures based on attacker tactics, techniques, and procedures or others?

First, organizations should outline their key business capabilities as they relate to revenue generation, operational stability, and resource management to identify and prioritize the key technologies and partners that support those capabilities, which if compromised or no longer available, would prevent the organization from staying a going concern. As that understanding develops, organizations should further enrich the information to help elevate which of their systems are considered to be Crown Jewels. To achieve this, we often determine a Crown Jewel across several dimensions and by answering the following questions:

- **Is it unique?:** Often unique and irreplaceable, a Crown Jewel contains sensitive information or proprietary processes that cannot be easily replicated.
- **How critical is it?:** Loss or compromise of a Crown Jewel would have a severe impact on the organization's operations, finances, or brand/reputation.
- **Is it an attractive target?:** Due to its high value, Crown Jewels are prime targets for cyberattacks and other malicious activities.

Some examples of Crown Jewels often include:

- **Customer data:** Customer records, purchasing history, and personally identifiable information (PII).
- **Financial data:** Financial records, intellectual property, and trade secrets.
- **Operational data:** Critical infrastructure data, production processes, and internal communications.

By taking a measured approach to determine the Crown Jewels, an organization can then validate whether AI is seen as a Crown Jewel for the organization (e.g., market facing product) or limited to a business-enabler technology.

The above process also provides organizations with a unique perspective that stitches business imperatives together with their supporting technologies, but also helps identify third party partners and dependencies that may also play a role in the technology support model.

Effectively applying this best practice

Let’s focus on an instance in which a large language model system (LLM) is identified as a Crown Jewel. After the identification phase is completed, the remaining three Crown Jewel phases are used to surface the associated risks and to identify the countermeasures needed to account for specific threats vectors that could potentially be used to misappropriate an LLM system.

To accomplish this, it is critical to develop an understanding of your LLM use cases, general architecture of its development pipeline, the data used to train the model and processed by the model, and determination of whether it’s a predictive or generative learning model. This is realized by examining the components of an LLM system (Figure 1).

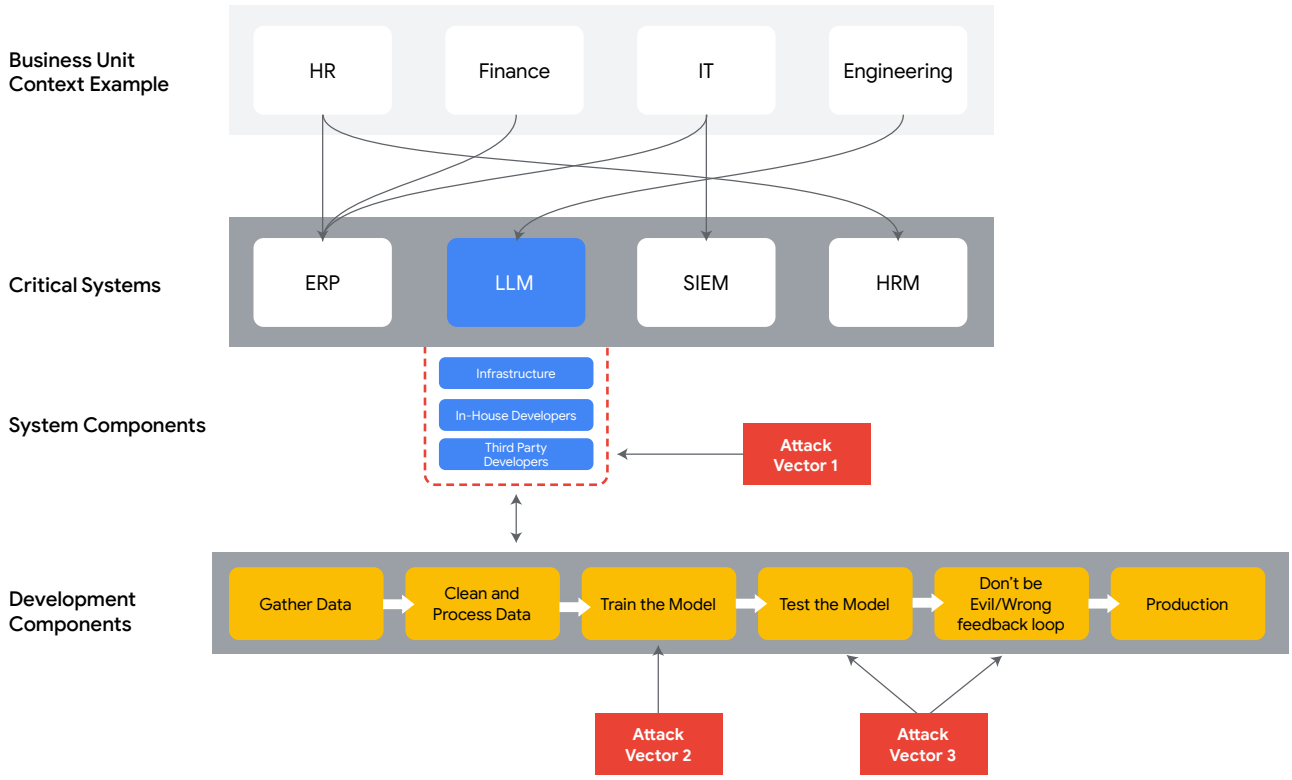


Figure 1. General Workflow for Crown Jewels with Generic Threat Overlay

Understanding how the LLM is being used, plus how it's developed and managed, helps to provide focus areas for identifying unique threat vectors. For example, if there is a strong dependence on third party developers, it's critical to not only take into account internal practices, but also research how third parties are supporting LLM development, their access to the data used to train the model, the type of data they handle, their third party library use, etc.

Some examples of techniques that attackers can use to target **LLM systems** include:

- **Prompt injection:** Attacker bypasses controls resulting in unintended behavior
- **Poisoning:** Attacker masks invalid data as valid data
- **Reverse engineering:** Attacker assesses model or training data sets
- **Extraction (model inversion):** Attacker clones existing model by studying model outputs
- **Energy latency:** Attacker deliberately slows down compute capability of model
- **Supply side:** Attacker compromises libraries used by the model

As threats and associated vectors are understood and defined, organizations should then design and implement countermeasures and controls. Some examples of controls that organizations should consider for **LLM systems** include:

- **Access controls:** Prevent direct model access (if possible)
- **Auditing and logging:** Comprehensive auditing and logging to track activities and changes within the pipeline
- **Monitoring:** Detect anomalies in LLM behavior, which could indicate attacks, security breaches, or unexpected model behavior
- **Rate limiting:** Control the frequency of requests that can be made
- **Input and output controls:** Validation of inputs and filtering of outputs

Some examples of **AI Crown Jewels asset** infrastructure controls that organizations should consider include:

- **Access controls and authentication:** Implement role-based access, authorization controls, and multi-factor authentication
- **Data security and privacy:** Encrypt data at rest and in transit
- **Logging:** Provide comprehensive logging for all components of the AI pipeline
- **Network segmentation:** Isolation of the AI pipeline and its components within secure network segments to limit the potential attack surface
- **Intrusion detection and prevention systems:** Monitor traffic and connections to the AI pipeline to detect anomalies
- **Patching and updates:** Ensure regular patching of operating systems, libraries, and software used in the pipeline
- **Vulnerability scanning:** Regular scanning of the AI pipeline infrastructure to identify and mitigate vulnerabilities before they are exploited
- **Incident response plan:** Create a well-defined incident response plan to guide actions in the event of a security breach

Foundational controls recommended to reduce AI-related business risks

Mandiant encourages organizations to maintain foundational controls across their overall AI security program as well. The following three principles from SAIF provide added value as key tenets for organizations to consider as part of their countermeasure program.

1. Extend detection and response to bring AI into an organization's threat universe
2. Automate defenses to keep pace with existing and new threats
3. Harmonize platform level controls to ensure consistent security across the organization

The malleability of SAIF allows organizations to extend governance and technology countermeasures to the AI pipeline, some examples include:

ACCESS 	**Access controls should be configured and implemented properly to prevent unauthorized access. Users and systems requesting privileged access should receive additional scrutiny.**	• Password policies should be stringent, such as meeting the recommended character length, complexity, etc. with additional controls like multi-factor authentication (MFA). • Tiered administration should be enforced, ensuring dedicated secondary administrator accounts are used for privileged access. • Privileged access workstations should be used for remote access to Crown Jewel systems, while additional capabilities, such as privileged session manager, should be leveraged. • Privileged accounts should be managed by a third-party privileged account manager tool. • Implement a "Least Privilege" policy, ensuring that only those who need access to Crown Jewel systems have it.
DETECT 	**A comprehensive set of defenses should be utilized to ensure preventive and detective controls are in place.**	• Ensure endpoint security control agents are deployed on all endpoints. • Verify network security tools, such as firewalls and intrusion detection and prevention systems, are utilized to analyze network traffic for unusual activity. • Confirm that your SIEM is receiving all necessary event types and logging details. • Implement and operationalize relevant use cases within your SIEM to effectively detect and alert on unusual behavior.
RESPOND 	**Companies should have incident response efforts in place. These efforts should prepare and guide security analysts on response actions that escalate events in a timely manner to minimize the incident impact.**	• Ensure an incident response plan is in place and well understood by responsible teams. • Verify playbooks and runbooks are up-to-date, available to staff, and address evolving response procedures. • Confirm network and application diagrams are current and accessible, at a minimum, for Crown Jewel systems. • Remediation instructions are clearly documented, available, and address critical systems. • Execute tabletop exercises on a regular basis to ensure staff awareness of where procedural documents are stored, how they should be utilized, and the associated responsibilities of staff during an incident.

RECOVER 	**Companies should implement a business continuity plan to safeguard critical data and minimize operational disruptions. These efforts should be conducted and tested routinely to minimize major data loss or extended downtime.**	• Create and implement a disaster recovery plan. • Identify all systems and data that are considered to be Crown Jewels. • Ensure backup data is stored in an offline or immutable storage format. • Implement local emergency access accounts for critical recovery applications.

Conclusion

As organizations continue to adopt new technologies, most notably AI technologies where consumption is propagating across organizations and industries with alarming velocity, it becomes more critical for organizations to prioritize their security controls investments. By taking a Crown Jewels-based approach, organizations can keep pace with technology consumption while in parallel elevating their safeguards to protect intellectual property, minimize disruption and abuse to services like LLM systems, and reduce the likelihood of financial losses or reputational damage. Frameworks such as Google's SAIF and NIST's AI Risk Management Framework are examples of how organizations can take a risk-based approach to their AI investments to amplify their safeguards when integrated with a prioritized approach to identifying critical assets and relevant threat vectors to the business.

Why Connected Devices Are Expanding Manufacturing Attack Surfaces

A blessing and a curse

Mandiant continues to see a transformative shift in the manufacturing threat landscape. Industry 4.0 is a major component, which represents the fourth industrial revolution, characterized by the integration of technologies into manufacturing processes like the Industrial Internet of Things (IIoT), artificial intelligence (AI), and cloud computing. This technological fusion creates smart factories where machines, sensors, and systems communicate and cooperate, enabling data-driven optimization and greater production efficiency. Industry 4.0 is reshaping traditional manufacturing, leading to increased automation, predictive maintenance, and the ability to customize products on a mass scale.

The integration of IIoT devices undoubtedly offers significant manufacturing advantages. Smart sensors collect real-time data on equipment performance, enabling predictive maintenance and minimizing downtime. Connected machines optimize production processes, leading to increased output and reduced waste. Remote monitoring capabilities allow for enhanced oversight and control of operations, even from afar. Although, the connectivity that fuels these benefits also creates a new set of risks and vulnerabilities that broaden attack surfaces for malicious actors targeting Industrial Controls Systems (ICS) and Operational Technology (OT)—ultimately posing a significant cybersecurity threat to these critical operations.

This broadened attack surface stems from several factors. First, the sheer number and diversity of connected devices introduce various entry points for cyber attacks. Each device, with its unique software and security posture, represents a potential weak link in the overall security chain. Unpatched vulnerabilities, weak authentication protocols, and insecure communication channels create attack vectors that hackers can leverage to gain access.

Also, the complex and often aging nature of ICS also presents challenges. Unlike traditional IT systems, ICS often includes legacy infrastructure with limited security capabilities. Integrating these disparate systems with newer IIoT devices can create compatibility issues and further complicate security management. The implementation of robust security controls is further hindered by the operational constraints of manufacturing environments where downtime can be costly.

Real-world impact of expanded attack surfaces

In 2021, the Colonial Pipeline attack affected fuel supplies across the East Coast of the United States, demonstrating the crippling effect cyber attacks can have on critical infrastructure. Mandiant has seen a steady rise in the number of vulnerabilities being discovered in network connected tools and sensors that are found on the manufacturing floor. These vulnerabilities increase risk of IIoT devices being rendered inoperable, which could lead to production shutdowns and revenue losses. Additionally, undetected compromises of this nature, could provide attackers with a clear path to alter data in a way that causes damage to the product being assembled or potentially renders it unsafe because it does not meet quality standards.

Another contributing factor to the growing attack surface in these environments is the required infrastructure to support the widening array of IIoT devices in use, many of which require wireless communications to be effective. Mandiant has seen an increase of wireless network technologies leveraged in industrial environments and quite often deployed with minimal configuration, weak encryption, and default credentials - which opens opportunities for attackers to access these systems at a distance and bypass network access controls implemented between IT and OT networks.

The expanded use of IIoT often leverages Manufacturing Engineering Systems (MES) and Enterprise Resource Planning (ERP) systems, where there can be significant and indirect impacts from a compromise on enterprise IT environments, stunting the manufacturing floor's abilities that rely on these systems to operate.

Reduce connected device attack surfaces

To effectively navigate this expanding problem, manufacturers should prioritize cybersecurity and implement robust, relevant security measures. Mandiant suggests a layered approach that encompasses several key elements:

- Segment and secure the connections to IT infrastructure, including servers, workstations, network devices, MES/ERP systems, and system data that supports OT to control access; contain the spread of malware; and limit the impact of a compromise
- Identify, track, and manage assets in the OT environment, including physical and logical assets, to understand the attack surface and support effective vulnerability management and patching

- Build robust detection and response capabilities that enable threat hunting, early detection, and rapid containment of compromise
- Conduct regular vulnerability assessments and penetration tests to help identify and address weaknesses before they are exploited by attackers
- Educate and train staff on cybersecurity best practices and foster a culture of security awareness to prevent human error and phishing attacks

Conclusion

The future of industrial cybersecurity will demand continuous adaptation and adoption of emerging technologies. Although regardless of technological advancements, one thing remains clear: prioritizing cybersecurity and implementing a comprehensive security strategy is paramount for manufacturers to harness the full potential of IIoT and mitigate the growing risks associated with an expanded attack surface. By taking proactive steps, manufacturers can ensure secure and resilient operations of their critical infrastructure, safeguarding not only their operations but also the well-being of society at large.

Tuning Your Cybersecurity Communications to Support SEC Compliance

With the implementation of the new U.S. Securities and Exchange Commission (SEC) cybersecurity rule, publicly traded companies are now subject to new regulations that impact how they report a material cybersecurity incident, and how they communicate with investor audiences about the details of their security program. The rule is intended to benefit investors by providing more timely and consistent disclosures that give transparency around material cybersecurity incidents and easily accessible information regarding cybersecurity risk management, strategy, and governance practices.

As with any change, there's an opportunity for organizations to rethink their approach and improve their overall response process, while incorporating the new reporting requirement to provide consistent information to all impacted stakeholders.

Understanding the new SEC cybersecurity rule

The rule requires registrants to disclose, within four business days, cybersecurity incidents that are deemed "material." Although the SEC did not define "material," it notes that aspects of materiality include "nature, scope, and timing, as well as its material impact or reasonably likely material impact on the registrant." The SEC recognized it can take time for a company to determine whether an incident is "material", which can affect the timing of a mandatory disclosure.

The new rule also requires that registrants annually disclose their processes for assessing, identifying, and managing material risks or likely material risks from cybersecurity threats and previous incidents. Additionally, registrants must disclose relevant expertise of any members of management or committees that are responsible for assessing and managing registrants' material cybersecurity risks and describe board governance and oversight of such risks and threats.

As expected, the four day disclosure period has changed the way organizations communicate the details of an event. Previously, public companies would use a variety of factors to determine when and what to communicate around a cybersecurity incident. These factors included: geo-specific data breach laws, brand and customer impact, threat actor behavior, and effects to business operations.

Now, the four day rule provides a more straightforward guideline for public companies, although based on initial cases, this is happening sooner than before with less information shared during the initial disclosure statement.

Adjusting the response process to alert on materiality

The determination of materiality is the domain of the legal department and is specific to each organization. However, once materiality guidelines have been established, it is important for the organization to review its existing cybersecurity program to ensure the response process has been tuned to surface materially significant investigative details during an incident. Once materiality has been determined, the communications process should be ready to move quickly—to accurately report the incident to the SEC and other potentially impacted stakeholders.

Two common challenges

Mandiant often sees two common challenges on the communications response side, which are related to impacted data types and customer contracts.

Impacted data: When malicious actors access a corporate environment, they move laterally to seek out and exfiltrate valuable data. By proactively identifying which systems hold materially significant information, it's possible to surface early warning signs that may cause an event to become material if initial forensic evidence indicates those systems were accessed by a malicious actor. Depending on the situation, this may not be considered material, but it allows additional scrutiny and initial preparation to take place in the event it does become material.

There's no fundamental change to the investigative process or materiality, however by ensuring both the technical investigative and legal partners understand what indicators of materiality to look for, organizations can improve their overall response.

Customer contracts: It's not uncommon for large organizations to include cybersecurity reporting requirements in their contracts with vendors and business partners. These companies often possess sophisticated cybersecurity teams that evaluate their risk when one of their business partners experiences an incident. The business partner can be in breach of contract if they do not provide sufficient information about their event.

Public companies should understand if there are specific reporting requirements within their business contracts that would mandate additional reporting to their business partners. If so, there are two considerations to build into the process:

1. Does the reporting require additional details beyond what was shared publicly, and if so, would that represent non-public material information?
2. Will the business be potentially impacted by the loss of revenue as part of the contractual disclosure? For example, to a customer that represents a significant portion of the company's revenue, would the impact be material?

It's important to consider what the baseline of reporting requirements are across all contracts and legal and regulatory commitments—to in turn ensure your communications meet that baseline.

This is also a good time to review any contracts that may have cybersecurity reporting requirements and build those responses into your overall business response plan.

Prepare the whole organization to respond effectively

Cyber incidents have a bigger impact across the organization, therefore it's important to consider all stakeholders who may need information and establish an official process to respond to their requests.

During a recent Mandiant response engagement, a public company filed an 8-K following a ransomware attack, with material impact to their operations. The company did not provide guidance to employees on how to properly respond to customer inquiries—one employee shared inaccurate information with a customer regarding when operations would resume. This inaccurate timeline quickly made its way to social media and then the media, requiring a clarification on a statement from their own employee.

As stated above, in an attempt to be helpful or maintain a key business relationship, employees can take it upon themselves to communicate inaccurate information that unintentionally puts the business at risk.

With less time to develop a comprehensive communications plan, it's important to enter an incident with previously identified stakeholders, accompanied by an outline of their specific communication needs, as part of the overall communications response plan.

Conclusion

The good news is that cybersecurity incidents are survivable events for public companies. However, the way a company responds can have a significant impact on their business operations and brand reputation. A well-managed event with effective communications can minimize business disruptions and reduce the harmful impact to brand reputation—ultimately helping to speed up the return to normal operations.

Disruptive Cyber Operations Used as a Political and Military Tool

Russia's use of disruptive cyber operations since its full scale invasion of Ukraine in February 2022 altered our perspective on these operations as primarily a tool of peacetime operations and demonstrated their application as a component of battlefield operations in a wartime context. Disruptive cyber operations during peacetime have been a means to support a specific political objective, whereas during wartime they can support political objectives or be executed in parallel to ongoing military operations. Mandiant assesses that disruptive cyber operations are a political shaping tool for peacetime that continue to be a key part of shaping the battlefield during wartime.

Cyber operations as a peacetime shaping tool

Disruptive cyber operations conducted during peacetime by nation state actors are uncommon; however, these operations have often been seen as high-profile, plausibly deniable, and geopolitically-driven. Though discourse around disruptive cyber operations often considers them to be a tool for signaling, they are viewed by their sponsoring nations as having wider political and military utility. In support of this notion, scholarship like Ben Buchanan's "The Hacker and the State", argues that cyber attacks are most effective as a tool for shaping, rather than signaling.[1]

[1] Buchanan, Ben. *The Hacker and the State: Cyber Attacks and the New Normal of Geopolitics.* Cambridge, MA: Harvard University Press, 2022.

One of the best examples of this was seen in the 2018 disruptive attack widely dubbed as OlympicDestroyer. This attack, later attributed to Russian military intelligence actor SANDWORM, was driven by suspension of the Russian Olympic Committee and the sanctioning of multiple Russian athletes. OlympicDestoyer had multiple layers of deniability that attempted to distance the Russian government from the attack against the PyeongChang Winter Olympic Games. Additional disruptive cyber operations during peacetime, such as the attack on Sony by North Korean hackers, and the Stuxnet operation, both similarly have layers of deniability built into the technical aspects of these attacks and were directly connected to furthering political interests.

Sony

In November 2014, a group calling itself the "Guardians of Peace" compromised Sony Pictures Entertainment, stealing personal information on thousands of employees, internal communications, and several unreleased feature films. The group went on to deploy and execute the DESTOVER wiper malware, resulting in the permanent deletion of data from thousands of computers, hard drives, and servers. As part of this operation, the group demanded that Sony withdraw its then-upcoming film "The Interview", and made lethal threats to U.S theaters which resulted in the cancellation of the planned theatrical release. A subsequent public FBI statement indicated actors with a North Korean nexus were responsible for the incident.

North Korea likely intended the Sony attack to demonstrate its willingness to respond disproportionately to even small provocations. Rebuilding computer systems, reduced film revenue, and loss of brand reputation associated with controversial company leadership communications showcases the considerable impact of disruptive cyber operations. As Buchanan discusses, however, the success of this operation through a lens of signaling is mixed: the North Koreans did, in fact, find more success in coercing theaters to no longer show the film once they threatened violence.

SHAMOON

Another example of peacetime signaling was seen in a series of disruptive attacks that took place between 2012 and 2017 that targeted the greater Middle Eastern oil and gas industry, government agencies, and other critical industry sectors leveraging variants of the SHAMOON wiper malware; malware that has been attributed to Iranian actors. The SHAMOON malware capabilities include the ability to destroy data on logical and physical elements of the hard disk and render an operating system inoperable by wiping the master and volume boot record.

In the August 2012 attack alone, an estimated 35,000 computer systems were rendered inoperable, resulting in the need to source replacement hard drives, increasing hard drive prices worldwide, incurring substantial financial costs, and the temporary disruption of Saudi Aramco operations.

Mandiant assesses that Iran may have intended these operations to signal its displeasure with Saudi Arabia's continuing cooperation with the West by targeting Saudi Aramco, Iran's preeminent competitor in crude oil production.

Cyber operations as a wartime domain

Disruptive cyber operations during wartime are unlikely to be viewed as deniable operations. Rather than shaping wider adversary behavior, the objectives and targeting of these operations during wartime are more often related to discrete military objectives. Additionally, these operations may be executed in parallel with others, both psychological and kinetic. Compounding effects of operations across multiple domains may seek to degrade societal support for continued war, degrade military or industrial capabilities, undermine trust in current government and services, or seek to create general disorder. From Mandiant's vantage point after responding to many Russian cyber attacks on Ukraine since the invasion of Ukraine began, disruptive cyber operations during wartime primarily focused on shaping the battlefield.

For instance, Russian military intelligence (GRU) cyber-enabled influence operations include wiper operations, often targeting government services or critical infrastructure and hacktivist personas, which amplify and project the damaging effects of cyber attacks. Ukrainian communications, whether civilian or military, have been a key target for Russian disruptive cyber operations since its full scale invasion. Key examples of this are the disruptive operation targeting Viasat modems at the outset of the war, as well as the recent targeting of the Kyivstar telecommunications company. The disruption of communications networks can provide battlefield advantage while simultaneously denying key infrastructure to civilians.

Russian disruptive operations in Ukraine

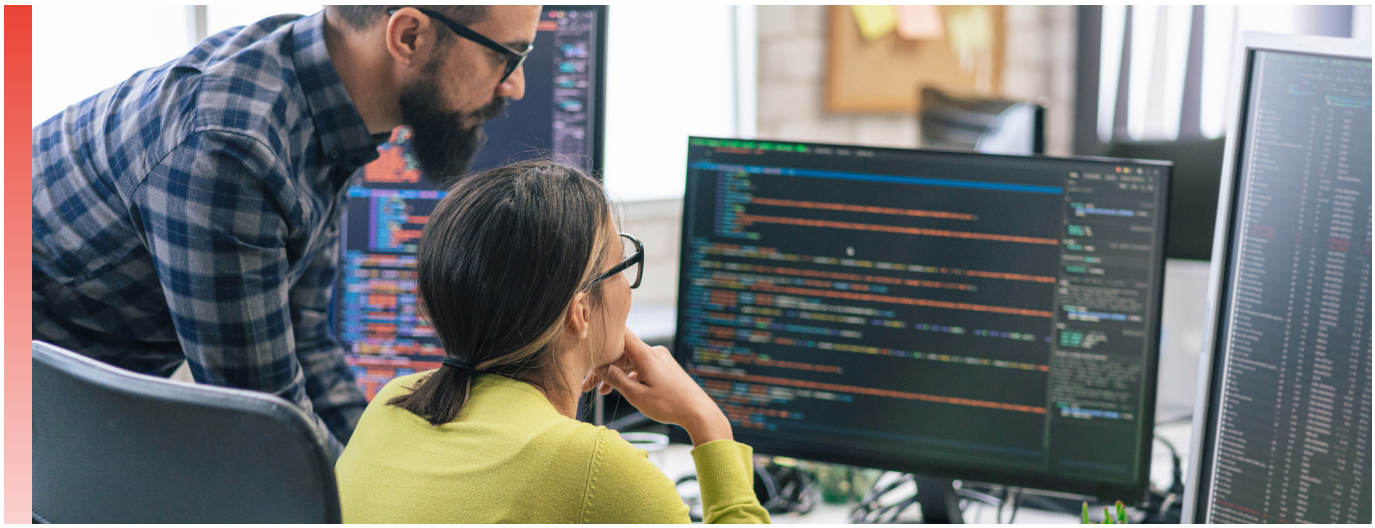
Russian activity in Ukraine at wartime has been a balance of espionage and disruptive activity by each of Russia's three main intelligence agencies. Russian disruptive cyber operations, conducted exclusively by the GRU, are deployed in a multifaceted environment where deniability is not a priority. Mandiant has observed the GRU operate a standard, repeatable playbook to pursue its objectives. These cyber-enabled operations, which have used both wiper malware and fake-ransomware to achieve a disruptive effect and hacktivist personas to amplify those effects, target mainly government, civilian, and critical infrastructure targets. Additional disruptive operations have targeted services used for Ukrainian military communications, such as Viasat modems, suggesting coordination between disruptive cyber operations and kinetic operations. SANDWORM's use of ransomware to add a thin veil of deniability to their disruptive operations is a playbook they have used in the past and have reused since the 2022 invasion of Ukraine.

At the outset of the war, SANDWORM utilized a multitude of multifaceted wipers; however, as the war progressed and their capabilities continued to be outed and detected, they began to shift their tooling to fit the continued war and widespread loss of capability. Over time, SANDWORM's wiper operations have become relatively consistent; the group prefers using increasingly lightweight tools, including modified publicly available tools and ransomware, to achieve a disruptive effect. They have fine-tuned their disruptive playbook to efficiently conduct repeatable, fast-paced operations with psychological effects.

Forward looking

As Russia's invasion of Ukraine enters its third year, it is important we continue to anticipate and track adaptations in SANDWORM's operations, both in Ukraine and abroad. As the GRU's primary cyber attack unit, SANDWORM represents the spearhead of Russia's concept of information confrontation, and the lessons we have learned so far are imperative to defending against future operations. It is also important we assess additional threats outside of Ukraine. China's Volt Typhoon is a group that has garnered a great deal of public attention for their targeting of US critical infrastructure. As opposed to other nation state groups such as SANDWORM, Volt Typhoon has not yet publicly been associated with any disruptive or destructive operations, although their targeting patterns likely demonstrate this intent. While signaling is notoriously difficult in cyber operations due to implicit deniability and challenges in attribution, as well as technical challenges in understanding motive until the late stages of the attack lifecycle, patterns of Volt Typhoon activity suggest the group's motive may deviate from pure espionage in the future.

To learn about effectively mitigating the impact of destructive and disruptive attacks like these with proactive hardening and detection methods, read our blog.



Revolutionizing Malware Analysis in the Age of AI

The digital landscape is a constant battleground between cyber defenders and their attackers. As malware evolves at breakneck speed, traditional analysis methods often struggle to keep pace. This is where Artificial Intelligence (AI) emerges as a potent weapon in the fight against malicious software.

VirusTotal, the largest crowd sourced threat intelligence suite, leverages the power of multiple AI engines—including its own Code Insight that was launched in April 2023 for advanced analysis of suspicious scripts. Both Hispasec and NICS Lab engines, integrated through the crowdsourced AI initiative, further enhancing analysis capabilities for Microsoft Office and Powershell files, respectively. VirusTotal tested these AI engines against hundreds of thousands of files, helping to explore the strengths and limitations of AI for malware analysis. Below are some of our topline findings.

AI as a turbo boost for malware hunters

For decades, cyber defenders have relied on a tried-and-true arsenal of tools to battle malware: signature matching, sandboxing, and manual code analysis. These methods, while effective, can be slow and resource-intensive, leaving defenders on their heels against the evolving tide of malicious scripts. AI is a game changer that not only complements these traditional tools, but also supercharges them with new capabilities.

One of AI's most significant cybersecurity-related contributions is the ability to identify malicious scripts, even heavily obfuscated ones. Traditional methods often struggle with complex code cloaked in layers of deception, but AI can achieve up to 70% better detection rates compared to traditional methods alone, effectively unmasking hidden threats and saving cyber defenders precious time.

AI's impact goes beyond identifying malicious scripts. The ability to analyze code behavior and exploit patterns makes it a champion for uncovering vulnerabilities within scripts—improving exploit identification by 300%.

Time saved by AI's superior detection and identification capabilities is invaluable. Cyber defenders can spend less time sifting through mountains of data and more time focusing on critical tasks like incident response and threat hunting.

The integration of AI into the malware analysis arsenal is a paradigm shift. VirusTotal believes the “universal” code analysis capabilities that AI engines demonstrate can help to avoid blind spots that other security solutions may have at the moment, especially for all non-endpoint detection and protection. File type detection for text files is another unexpected advantage.

A potential double-edged sword

While AI's potential to bolster cyber defenses is undeniable, there is still an open question regarding its potential misuse. The development of AI-powered malware generation and execution remain a specter, as concrete evidence of its existence is difficult to find. Although experts have found different malware families using AI themes for distribution. This is not surprising, given the opportunistic nature of attackers for trending topics.

The potential for AI to enhance the sophistication of social engineering attacks seems to be one of the most likely short-term scenarios. Malicious actors can leverage AI to craft personalized phishing emails, generate believable fake news articles, or even manipulate human interactions in real-time.

The lack of definitive evidence presents both a challenge and an opportunity. While it's crucial to remain vigilant and actively research the potential misuse of AI in malware, it's equally important to avoid sensationalizing unconfirmed threats.

A future powered by AI

The VirusTotal experience with AI engines has shed light on its incredible potential to transform the landscape of malware analysis. AI’s ability to tackle some of the most time-consuming and challenging tasks, like deobfuscation and malware behavior explanation, offers a glimpse into a future where analysis is faster, more accurate, and accessible.

One of the most significant shifts in AI, is its ability to provide a comprehensive explanation of its verdicts and findings. This transparency empowers analysts, enabling them to understand the reasoning behind detections, challenge them if necessary, and ultimately gain deeper insights into the malicious intent of scripts. This also serves

as an interim solution for the global cybersecurity workforce deficit. AI's ability to make analysis more accessible could potentially help bridge the cyber skills / talent gap, equipping more individuals to contribute to the fight against cyber threats.

The results around CVE and obfuscation detection are particularly promising, showcasing AI's superior performance over traditional techniques. Its code analysis capabilities offer a broader perspective, helping to avoid blind spots that may exist in other security solutions, especially for non-endpoint detection and protection.

Looking ahead

Enriching the context provided to AI engines and exploring customized prompting techniques are key areas for improvement. This will help AI engines better align with specific analysis criteria and bridge potential gaps between the results and those of traditional solutions.

Even though the evidence for AI-powered malware generation or execution remains inconclusive, the potential for misuse is undeniable. The possibility of an underground market for "uncensored" AI engines and its potential impact on social engineering tactics must be closely monitored and addressed through responsible development practices and ethical guidelines.

AI is a catalyst for revolution in malware analysis. By responsibly harnessing its potential, cybersecurity professionals can unlock a future of faster, more accurate analysis that is accessible to a wider range of individuals—ultimately creating more resilient security programs to combat evolving cyber threats. The journey ahead is filled with challenges and uncertainties, but the promise of a safer digital world fueled by AI is undeniable.

Contributors

Trisha Alexander
Vicente Diaz
Jennifer Guzzetta
Neil Karan
Kerry Matre
Muhammad Muneer
Nick Richard
Gabby Roncone
Paul Shaver
Dan Wire
John Wolfram

Mandiant Cybersecurity Consulting
Mandiant provides frontline expertise and deep understanding of global attacker behavior to help your organization effectively defend against compromise—before, during, and after an incident.

Google Cloud Security
Google Cloud acts as your security transformation partner by providing a secure-by-design foundation—a shared fate model for risk management—supported by products, services, frameworks, best practices, controls, and capabilities.

Google Cloud

For more information, visit cloud.google.com/security
A-EXT-RT-EN-US-000518-01