ManageEngine
Log360

The dark side of AI:

The ultimate guide to combat its imminent threats

ManageEngine
Log360

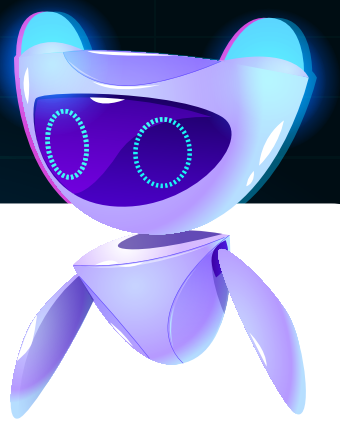


Table of contents

1. **Introduction** 1
2. **Early timeline of cyberattacks and the beginning of AI-powered threats** 1
    i. Twentieth century efforts in cyberspace 1
    ii. The internet era 2
    iii. How ML and AI are changing the cyber landscape 2
3. **Security implications of AI-powered attack scenarios** 3
    i. Automation and high scaling 3
    ii. Efficiency and adaptability 3
    iii. Highly sophisticated malware 3
    iv. Malware-infected botnets 4
    v. Generative AIs in the picture 4
4. **AI-powered attack scenarios and what they look like at each stage** 5
    i. Reconnaissance 5
    ii. Initial access 5
    iii. Privilege escalation 6
    iv. Persistence and lateral movement 6
    v. Command and control 6
    vi. Exfiltration 7
5. **What potential do AI-driven attacks hold in the future?** 7
    i. Search engine optimization and malvertising 8
    ii. Scaling up attacks with adversarial AI 8
    iii. Social engineering attacks using AI 8
6. **Economic implications of the growing cyberattack space** 9
7. **What can we do about this?** 9
    i. In the detection stage 9
    ii. In the mitigation stage 10
    iii. Compliance and health check 10
8. **What's the takeaway?** 11

We no longer live in the era where cyberattacks are solely reliant on manual efforts and the limited scope of the cyberspace. Cyberattacks are no longer solely the domain of skilled and specialized individuals. With the advent of AI, the threat landscape has taken a huge leap with an unprecedented arsenal of tools and techniques that can intelligently automate attacks. The integration of AI with the traditional cyberthreat space has allowed anyone with AI resources and basic technical skills to execute a successful cyberattack.

According to an IBM report, the average cost of a data breach hit an all-time high at USD 4.45 million in 2023. And with the growing involvement of artificial intelligence in the cyberattack space, these numbers won't be coming down anytime soon.

With the remote work continuing throughout several organizations even after the pandemic, the attack surface has widely expanded. Adversaries don't have to be part of a well-recognized threat group; even lesser known threat groups or individuals can effectively breach an organization's network by leveraging a remote application's vulnerability found through a botnet. Likewise, AI can aid every stage of a cyberattack, from reconnaissance to exfiltration.

Let's explore the methods and risks of AI-powered attacks in the evolving cyberspace and deep dive into the attack stages.

Early timeline of cyberattacks and the beginning of AI-powered threats

(i) Twentieth century efforts in the cyberspace

To get hold of the exponentially growing cyberspace and the associated threats, let's go back in time a little to the birth point of the Trojan Horse and other worms. Cyberattacks came into the main frame in the 1980's as an alarmingly frequent threat. With the debut of new worms and viruses, a war emerged, marking the emergence of antivirus software.

The advent of the internet in the 1990s opened the doors to a lot of cyberthreats. A significant threat development was the polymorphic viruses—a code that mutates as it spreads through computing systems. These viruses simultaneously maintain the original algorithm while mutating. To combat these, new ways to secure communications were devised and encryption standards were set. Secure Sockets Layer (SSL) was developed to secure internet connections by encrypting data between two parties.

(ii) The internet era

Entering the 21st century with a broader availability of reliable broadband, people all over the world were using the internet. Eventually, the amount of vulnerabilities and the number of new infections in the cyberspace increased. Modern malware began to take shape. This malware did not require any downloads and was able to spread through email with the help of social engineering. On top of this, there was a huge surge in hacking of credit cards. This led to companies establishing a defensive cybersecurity foothold, with several of them launching open-source antivirus software. Innovations like ODSecurity (an operating system with built-in security features) and Android antivirus also came to prominence through the decade.

With continuous digital developments, the 2010s saw adversaries pull ahead of cybersecurity efforts, costing businesses and governments huge amounts of money. Some notable high profile breaches were the global payment systems data breach in 2012, the Yahoo data breach in 2013-14, and WannaCry ransomware in 2017. Another major event was the stock market shutdown in New Zealand due to multiple DDoS attacks in 2019.

During this time, on the security front, vendors started developing different approaches like multi-factor authentication and network behavioral analysis to scan for behavioral anomalies in files.

(iii) How ML and AI are changing the cyber landscape?

Meanwhile on the other side, the cybersecurity industry has witnessed a gradual but substantial impact of AI and machine learning (ML) developments. While AI and ML have been present in the field from the 1950s, their involvement in cyberattacks was not initially perceived as a prominent threat. It was only with the development of human intellects like Apple's Siri that neural networks started making a presence in the industry.

However, as AI and ML technologies continue to advance, their role in the cybersecurity space has becoming increasingly significant. With AI, the landscape is seeing a new breed of attacks, whose capabilities are evolving, enabling threat actors to automate malicious activities, tailor their strategies, and exploit vulnerabilities with greater efficiency. Consequently, the once-underestimated role of AI in cyber attacks has emerged as a significant concern for the security industry.

Most commonly, AI is being used in the form of text-based generative AI, through which adversaries can explore the endless possibility of attack methods and automate models to evade defenses. Some of the notable AI-powered cyberattacks in 2018 include the TaskRabbit cybersecurity breach, the Nokia breach and the Wordpress data breach. With several developments on the defensive front of cybersecurity, the industry is proactively working towards innovative defense strategies like SIEM solutions to safeguard organizational networks.

Before delving into defensive strategies, it's crucial to comprehend the various methods by which AI attacks can infiltrate networks and the severity of the risks they pose. Therefore, let's examine the current cyberspace landscape, recent developments in AI due to the introduction of generative AI, and the implications of these developments.t

Security implications of AI-powered attack scenarios

However sophisticated and advanced AI developments have become, machines still cannot launch attacks on their own. But still, AI-assisted attacks have far more potential to devastate victims compared to traditional methods. This is because of the unique advantages that AI and machine learning offer—advantages that manual efforts alone cannot replicate.

(i) Automation and high scaling

Using AI, cyberattackers can automate various stages of the attack process, including reconnaissance, vulnerability scanning, and exploitation. A recent incident, the MOVEit ransomware attack in the UK, provides evidence of likely automation, as forensic analysis of IIS records indicated there were breaches observed on two separate clients within a mere 24 seconds. Automation enables attackers to target many systems simultaneously, making AI-powered attacks highly scalable. Traditional attacks, however, require more manual effort and are limited in terms of scale.

(ii) Efficiency and adaptability

Speed makes most of the difference when AI attacks are compared to traditional attacks. AI systems can analyze vast amounts of data in real time, helping attackers identify vulnerabilities rapidly and adapt their attack strategies accordingly. AI tools like PassGAN, a generative adversarial network-based password-cracking tool that uses ML to generate password guesses, are already causing mayhem due to the speed at which cyberattacks can be carried out.

The AI algorithms in these tools use vast training data to evade defense mechanisms that organizations have in place, which makes malware so adaptable. We can even pin the increase in zero-day attacks to the rise in artificial intelligent systems, since they significantly reduce the time available for defenders to deploy patches and countermeasures on those zero-day vulnerabilities.

(iii) Highly sophisticated malware

AI-powered attacks can employ sophisticated evasion techniques to bypass traditional security measures. Modern malware is highly evolved and is capable of bypassing server filters and continuously mutating to evade analysis by defenders. A notable example is IBM's DeepLocker, a proof-of-concept malware variant that navigates the attacker mindset of leveraging machine learning algorithms to launch cyberattacks.

(iv) Malware-infected botnets

Malware is not just confined to infecting individual systems anymore. The rise of botnets, networks of malware-infected devices, has become increasingly prevalent. These botnets have the capability to scan the entire internet in search of vulnerabilities rather than targeting specific organizations or industries.

In botnet attacks, like the WordPress data breach in 2018, around 20,000 sites were attacked, ensuring that the malware variant was able to attack as many sites as possible. AI algorithms in these botnets can help optimize the command-and-control infrastructure, making the malware more resilient and harder to trace.

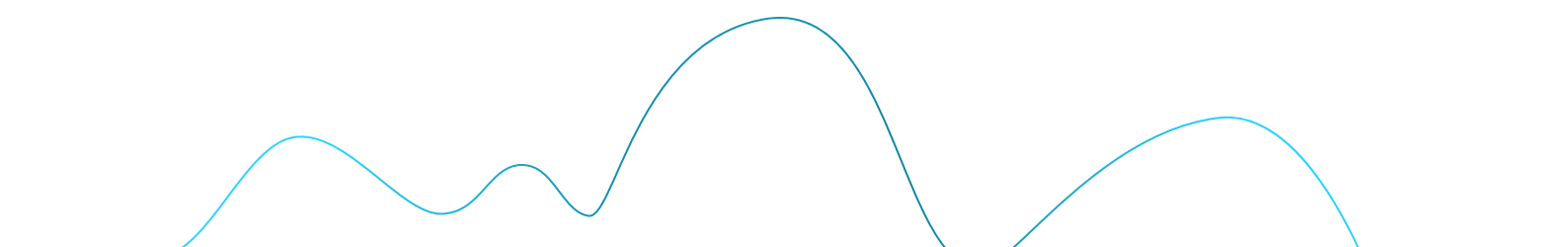
(v) Generative AIs in the picture

By analyzing large amounts of personal information and social media profiles, AI systems can generate tailored phishing emails and voice messages that appear legitimate to deceive individuals or gain unauthorized access to systems.

When generative AI entered the picture, it showcased the revolution of AI and its impact on cyberattacks. While its primary purpose is to assist users with information, its immense potential is also being exploited by threat actors to streamline their attack strategies and craft targeted social engineering schemes. This AI model can also scour all through the internet and its different contents—e-books, articles, websites, and posts, including personal information obtained without consent, which can be used to target and profile victims.

Among the many things we can do with generative AI, unfortunately exploitation is one of them. If AI can write code, it can write malware pseudo-code, too. Even though it refuses to respond to unethical and illegal requests, through intelligent prompt engineering, generative AI can be tricked into breaking down any attack scenario under the guise of developing a proactive defensive strategy against it. Similarly, breaking down the unethical requests into different parts can also lead the AI model to believe that there's nothing suspicious in the request, resulting in fulfillment of the same unethical request it once denied.

Consider this clip from the RSA Conference 2023, where Stephen Sims, an experienced vulnerability researcher and exploit developer, shares a remarkable demonstration involving ChatGPT showcasing how he utilized the model to generate code for ransomware, which was a rather alarming revelation. From writing an encryption pseudo code to verifying bitcoin addresses for ransom payments and decrypting data, ChatGPT seemed to fulfill all the requested tasks when they were broken down into separate parts.



AI-powered attack scenarios and what they look like at each stage

The weaponization of machine learning and artificial intelligence is pervasive throughout the stages of an attack, starting from the reconnaissance stages and persisting through the exfiltration stages as outlined in the MITRE ATT&CK framework. How will defenders combat this exponential growth of adversarial AI? Will we rely solely on firewalls and perimeter solutions? Unfortunately no. Defenders must adopt a comprehensive approach, including robust incident management and a sturdy risk security posture, to anticipate and mitigate emerging adversarial attacks.

(i) Reconnaissance

In the initial phase of the MITRE ATT&CK framework, the planning and reconnaissance stage, adversaries now rely on AI to automate and enhance the entire process. AI can now carry out the time-consuming tasks of profiling targets, scanning for vulnerabilities, and framing the entire attack.

AI's capability to understand, uncover, and recognize patterns within vast datasets allows for comprehensive analysis and extensive target profiling. This pattern recognition, facilitated by neural networks, enables the identification of links and correlations that may elude human analysts. AI uncovers hidden connections and vulnerabilities, helping attackers accurately identify potential attack vectors.

AI-powered bots and crawlers can quickly scour the internet, gathering publicly available information from diverse sources such as social media, business websites, forums, and leaked databases.

(ii) Initial access

Initial access is an imperative attack techniques that helps the attacker gain an initial foothold in an organization's network. It can include various social engineering exploitation methods of public-facing web servers, both of which are AI's major play area.

Long sort-term memory (LSTM) models, like DeepPhish, can produce effective synthetic phishing URLs compared to the randomly generated phishing URLs in the past. Sources claim these models improve success, with the success rate of one attack raising from 4.91% to 36.28%.

AI algorithms can also now examine huge datasets of leaked passwords and user behaviors in place of brute-force techniques, which are historically time- and resource-intensive. With intelligent password cracking models like PassGAN, the success rate of AI-powered brute-force attacks versus traditional attacks has also drastically improved.

(iii) Privilege escalation

The critical stage of an attack, wherein attackers strive to obtain higher-level access and privileges within a target network is privilege escalation. This increased access gives adversaries more confidence over the infiltrated environment, granting them the ability to carry out more damaging actions.

AI can identify user patterns that indicate privileged accounts or high-level access. After which, such specific accounts can be targeted.

AI-powered tools can automatically scan a target system or network for access control vulnerabilities. In contrast to manual techniques, these tools are more effective at finding configuration errors, giving attackers instant access to possible weak spots.

With the help of deep-reinforcement learning, there are AI models that can automate privilege escalation.

(iv) Persistence and lateral movement

This stage of a cyberattack is where the adversaries try to expand their presence within a compromised network while also ensuring that they have a strong foothold of the network. Ideally, they can do it through different methods of port scanning or vulnerability scanning to take control of active sessions or through credential access methods.

There are AI models like the one Hu and Tan proposed that use a generative adversarial network (GAN) technique to generate undetectable adversarial malware to bypass black-box detection systems.

With ensured persistence, AI-powered tools can automatically scan and map the network, identifying connected devices, services, and vulnerabilities. By analyzing network traffic and system configurations, tools can quickly discover potential entry points and vulnerable assets.

Also, just like initial access methods, ML algorithms can analyze leaked or stolen password databases, identify patterns, and accelerate the process of cracking passwords to help attackers gain unauthorized access to additional accounts.

(v) Command and control

The command-and-control (C2) stage is commonly where the adversaries establish a channel of communication and control over compromised systems without leaving any traces of detection.

AI algorithms can be used in several gateways to ensure the C2 stage of an attack is smooth. ML can be employed to generate malicious traffic or behavior that mimics legitimate patterns to obfuscate communication channels. AI enables attackers to automate responses and adapt their strategies in real time. AI algorithms can facilitate more robust and extensive encryption techniques in the C2 channel, making it harder for defenders to trace.

For example a study called DeepDGA shows how adversaries can use domain generation algorithms (DGAs) to generate stealthy DNS queries and carry out attacks in the C2 stage.

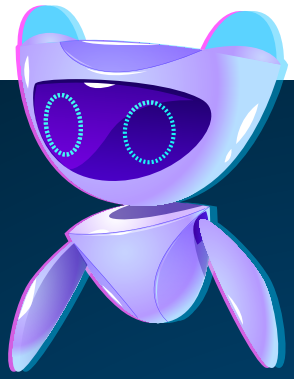
This study shows that an AI-driven C2 virus can anticipate when it will be unlocked across various types of nodes based on the target's current properties. As a result, a multi-layered AI-driven attack is capable of remotely and automatically providing access to other computer infrastructure components.

(vi) Exfiltration

Exfiltration is a method employed of stealing confidential data from an organization.

With the aid of adversarial AI, attackers no longer need extensive knowledge of data exfiltration techniques to carry out such attacks. The process of writing exfiltration codes has become more accessible, as AI can assist in generating them. In this article, the author explores data exfiltration from a network with help from ChatGPT, showing how AI can play a role in these activities.

In terms of scaling up attacks, AI can help attackers focus on extracting the most valuable information efficiently based on previous reconnaissance research. It can also help in splitting the exfiltration traffic across multiple channels or utilizing covert communication channels to mask data transfers.



What potential do AI-driven attacks hold in the future?

The road to AI's development for cyber adversaries seems to have no end, as they continually innovate and adapt their tactics to breach digital defenses. Cyber adversaries constantly come up with new and diverse tactics and techniques. This coupled with AI is setting an exponential rise in cyber attacks.

During the RSA Conference 2023, a session titled "The Five Most Dangerous New Attack Techniques" shed light on the imminent threats that AI poses, exploring emerging attack methods that exploit the transformative potential of AI in the cyber landscape. Let's look at some of these emerging attack methods below.

(i) Search engine optimization and malvertising

Katie Nickels, a certified instructor from SANS Institute, shares her insights on SEO and mavertising, which we'll discuss in this section.

Considering how web browsers can bypass perimeter detection systems, search engine optimization (SEO) has great potential to act as a medium for adversaries to gain initial foothold within the network. Threat actors leverage SEO techniques and paid advertising methods to improve page ranking and increase the chances of exposing their fake pages to the public. When these links are clicked and files are downloaded, malicious java scripts are loaded into the victim systems, gaining control of the device.

This attack type is classified under a new MITRE ATT&CK technique called malvertising. Generative AI can play a major role in this by creating fake yet believable content for spoofing websites. AI-designed websites also help adversaries mimic real websites when creating fake ones.

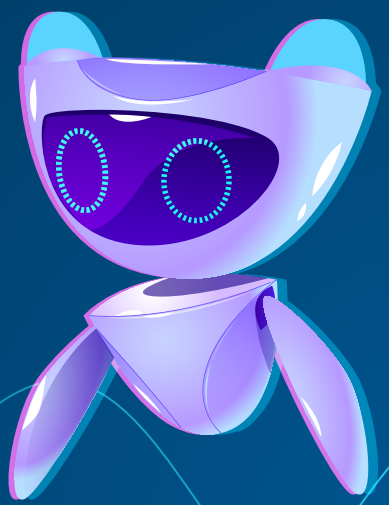
(ii) Scaling up attacks with adversarial AI

AI not only enables adversaries to conduct attacks on a large scale with relative ease but also amplifies the impact of their malicious activities. The streamlined process of creating new malware attack scripts becomes effortless for threat actors while the discovery of zero-day vulnerabilities in complex environments requires minimal effort, as emphasized by Stephen Sims, a SANS Institute instructor specializing in offensive cyber operations.

(iii) Social engineering attacks using AI

Heather Mahalik, a SANS Institute instructor, talked about the heightened risk of individuals, including friends and family, being exposed to social engineering attacks. This increased risk extends to phishing attacks in corporate organizations, a major concern in cybersecurity. With the aid of AI, adversaries can analyze vast datasets and create highly customized and persuasive phishing emails like we saw before. The success rate of such emails can be alarmingly high, as users struggle to differentiate them from legitimate communications.

Additionally, the multilingual capability of generative AI, which spans over 20 languages, presents another concern. Adversaries from various regions can now engage with the AI model seamlessly, overcoming language barriers. This expands the pool of adversaries and further amplifies the potential threats organizations face in today's interconnected world.



Economic implications of the growing cyber attack space

With the ongoing economic crisis, many organizations are facing cuts in their IT budgets. This will only result in higher success rates for adversaries who are further integrating and developing AI and ML in cyberattacks, intensifying the impact of cybercrime.

The economic impact of cyberattacks will be higher costs for remediation, recovery, and regulatory compliance. With more regional mandates going into effect across regions, such as the CCPA, the GLB, and the NYDF in the US along with the DPDP in India, organizations face hefty non-compliance penalties, further escalating the cost of breaches. The overall economic impact of cyberattacks demands attention and strategic measures to mitigate financial losses and protect organizational stability.

What can we do about this?

Looking at the optimistic side of this scenario, defenders are also developing and upgrading themselves to cope with the rapidly evolving cyberthreat landscape. Looking ahead, the future holds both challenges and opportunities. As defenders, we must adapt to the changing battlefield and equip ourselves with the latest tools and knowledge to counter AI-generated threats.

This is where Log360 comes in, a comprehensive SIEM solution that identifies the indicators of compromise (IoCs) at each stage of an AI-powered attack and helps you mitigate the scale of the attack.

In the detection stage

A SIEM solution like Log360 is exactly what you need at this critical juncture. Following the principle of "detect first and then respond," Log360's advanced analytics, anomaly detection, and behavioral analysis helps you maintain constant vigilance and stay one step ahead of AI-driven adversaries.

With Log360's correlation engine, multiple log events from different sources can be monitored simultaneously with predefined correlation rules to detect suspicious patterns or sequences of events. Real-time alert engines keep you vigilant with timely notifications, helping you take prompt action.

After scanning the IT infrastructure for potential vulnerabilities, the UEBA feature generates a risk score, enabling security teams to prioritize their response efforts by providing personalized risk scores to both individuals and assets. The risk score is based on various factors such as the criticality of the affected asset, the exploitability of the vulnerability, and the potential impact on business operations.

By mapping the logged activities to the MITRE ATT&CK Matrix, Log360 provides a clear view of potential attack vectors and the techniques adversaries might utilize to compromise systems. Users can create custom alert profiles for different attack techniques to ensure that security teams are promptly notified of any deviations from regular behavior.

In the mitigation stage

Considering the speed at which adversaries are evolving their techniques, regularly monitoring attack patterns and advanced threat intelligence strategies with Log360's incident response framework can help you completely avoid or reduce the impact of emerging cyberattacks.

Once a security incident is detected, Log360's incident response feature doesn't just notify you about it. Instead, it can automatically kick off a series of predefined workflows designed to counteract the threat. For instance, if a potentially compromised computer is detected, it can disable it. Such automated responses can drastically reduce the window of opportunity for an attacker to cause harm.

Time is of the essence when mitigating cyberthreats. Log360 offers instant notifications, ensuring that security personnel are immediately made aware of any concerning activities, allowing them to act promptly.

Each alert doesn't just sound an alarm but also initiates a ticket, automatically directing it to the right security personnel based on its nature. This ensures that the right expertise is applied to each threat, streamlining the mitigation process.

Compliance and health checks

With Log360's audit-ready report templates, you can meet your compliance needs for a wide range of policies, including the PCI DSS, SOX, HIPAA, FISMA, the GLBA, ISO 27001, the GDPR, and more.

You can simplify compliance reporting using intuitive dashboards that display metrics showing how your network is meeting compliance standards. You can fetch these reports with a single click and export them as needed. With more regulations on the way, you can also create custom compliance reports to address both external mandates and internal compliance needs effectively.

What's the takeaway?

The rise of AI-powered cyberattacks has ushered in a new era of sophisticated and relentless threats. Adversarial AI has become a formidable ally for cyberattackers, enabling them to automate, scale, and innovate their malicious activities with unprecedented efficiency. Even though some of the adversarial AI models we discussed have not been seen in wide use yet, attackers are slowly getting there.

The landscape of cybersecurity is complex and ever-changing and the advances in this space are undeniably being used on both sides of the road. As we move forward, a proactive and adaptive approach of AI leveraged by advanced security solutions will be crucial in staying ahead of the ever-evolving AI-powered cyberthreats. We can navigate the challenges posed by AI and shape a resilient and robust cyber ecosystem, where innovation and defense unite to create a safer digital realm for generations to come.

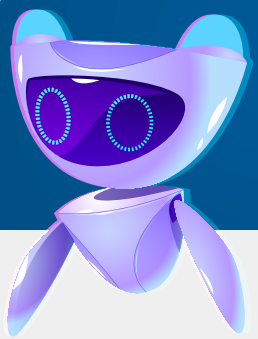
About the Authors:

Divya Narasimhan is a cybersecurity expert on ManageEngine's product marketing team. With a passion for researching and writing on emerging trends in cybersecurity, Divya is always up to date with the latest developments in the industry. She continually studies the cybersecurity market space and crafts creative, in-depth research content on emerging technologies, challenges, and their prospective solutions for businesses and cyber-enthusiasts.

Mahati Dwibhashi is a cyber security professional at ManageEngine with a talent for producing captivating content on security threats and solutions. As a cyberspace enthusiast, she loves to research constantly and write about the latest market trends. A keen writer, she helps organizations stay on top of their cybersecurity game through her research-led insights.

Calculate your ROI now!

Get a complimentary license for Log360!



Our Products

AD360 | ADAudit Plus | EventLog Analyzer | DataSecurity Plus

Exchange Reporter Plus | M365 Manager Plus

ManageEngine Log360

ManageEngine Log360 is a unified SIEM solution with integrated DLP and CASB capabilities that detects, prioritizes, investigates, and responds to security threats. It combines threat intelligence, machine learning-based anomaly detection, and rule-based attack detection techniques to detect sophisticated attacks, and offers an incident management console for effectively remediating detected threats. Log360 provides holistic security visibility across on-premises, cloud, and hybrid networks with its intuitive and advanced security analytics and monitoring capabilities. For more information about Log360, visit manageengine.com/log-management/.

$ Get Quote Download