Autonomous Cyber Systems Using AI - Approach on How to Improve Detection and Response

“aham brahmāsmi (अहं ब्रह्मास्मि)- Brhadāranyaka 1.4.10”

Badri S.
Vice President (MIS) - Bajaj Auto Ltd.
B.E (Elec & Comm) – PESIT; PGDGM –NMIMS

Abstract: Industrialization of Cyber Attacks is leading to the increase in their severity and complexity. Organizations around the globe are moving to adopting AI to improve their InfoSec posture. However, despite the measures taken the protective mechanisms are not adequate. The industrialization has lead to the emergence of easily available hacking tools and services online as well. State sponsored, motivated activists and young enthusiasts are constantly using these tools & services. Data breaches and Ransomwares have grown exponentially. Ransoms paid do not necessarily lead to complete decryption. Payouts are demanded through stealth crypto currency like Monero. Innocent clicks by ignorant internal users in organizations cause major business breakdown no matter how complicated the cyber defenses that are implemented. There is initial work which has been done on Autonomous Cyber Defense (ACyD / ACD) by researchers to explore possible mechanisms of defense. However, there are certain limitations in the approaches. This research proposal is focused on strengthening the existing research that has been done, through some proven mechanism and techniques. These autonomous actions described in these researches uses AI, inputs for which are based on ML, DL, CNN, NLP and Visual AI. The outcomes however can be improved to protect data at rest, in motion and during consumption through techniques discussed in this research proposal. Data in the form of IP addresses, SIEM data, User & Entity Behavior Data, Existing Malwares etc. are necessary to evolve some models. Better Data classification, Clustering etc would help improve existing some models such that the accuracy to identify anomalies and taking necessary self-defending actions are better. Both InfoSec service providers / organizations, would then be enabled to defend & respond better, and this research proposal is focused on that.

Keywords: AI; EDR; Cyber Attack; InfoSec Posture; Cyber Defense; ML; UEBA; Vulnerability, ACyD, ACD, ACO, AI, DL, CNN, GAN, NLP

1. Introduction

Some data points[1]

Enhancing Cyber Security Posture using AI has become imperative for organizations.
- 69% of organizations are intending to use AI to respond to cyberattacks.
- Nearly 67% think that AI will help identify critical threats.

Speed of adoption of AI Cyber-Security is increasing.
- Nearly 20% organizations used AI pre-2019.
- However, adoption is poised to skyrocket, with almost 67% organizations planning to employ AI by 2023.

Hackers are using AI for cyber-attacks. It is important organizations have an AI-enabled response cyber response posture:
- 75% of the incidents since 2020 have warranted faster action. This is because of the increase of Ransomware attacks.
- There has been an increased proliferation of hacking services like ReVIL (Ransomware Evil; also known as Sodinokibi). Other Ransomware-as-a-Service (RaaS) kit examples are Locky, Goliath (Locky for beginners), Shark (Customizable Kits), Stampado, Encryptor and
variety, there is a need to upgrade to an AI system that is fully automated. With the advancements in technologies, it is projected that the future certainly holds the reality of Artificial Intelligence-driven cyber-attacks, where malware can self-propagate through a series of autonomous decisions and intelligently tailor itself to the parameters of the compromised system in order to become stealthier to dodge detection. Discussing the autonomous cyber defense system where algorithms will fight against algorithms on the battleground of enterprise networks, only the best AI will win. To defend against future AI-driven attacks, autonomous cyber AI can be a revolutionary asset. Performing as a cyber immune system, this AI is capable of learning what is normal and abnormal for the digital business, without relying on prior knowledge of threats. The technology can not only detect never-before-seen threats but also autonomously reply to detach the attack before damaging any system.

Readiness required
Therefore, there is a need to counter AI cyber threat actors and cyber threat vectors using Machine Learning, Artificial Intelligence, making threat intelligence robust in Cyber Security allowing systems to identify data patterns, User and Entity Behavior Analytics (UEBA) such that the organization’s security system understand and learn from past experiences, including system breaches and other forms of cyber threats. Moreover, these technologies help in reducing the time taken for incident response. The use of AI in Cyber Security increases Effectiveness of Control, Reduces Threat Surface & Exposure, Predicts Breach Risks, significantly increases Incident Response.

Cybersecurity threats surged since 2021, with Colonial Pipeline and several water treatment facilities in the United States among the targets whose systems were attacked. Recent study shows that ransomware increased 105% from 2020 to 2021, with BFSI/Manufacturing becoming the most targeted industries. 2021 has also seen some of the most impactful supply chain attacks to date. From the SolarWinds and Microsoft Exchange Server exploits to Apache Log4j vulnerabilities, high-profile attacks filled news feeds, raising awareness—and alarm—among business leaders and their customers.

In brief, the pandemic accelerated digital transformation, amplifying both opportunities and risks. Now there are substantially more remote workers. More cloud users. More cloud services. Essential systems integrations with third-party partners. An astounding number of edge devices passing IoT data to the cloud. All interconnected and interdependent, delivering sophisticated connectivity and creating value at speeds and scales impossible just a few years ago.

2. Research Problem and Questions

[2] Probably the first place to look for AI-generated hacks is in financial systems, since those rules are designed to be algorithmically tractable. High-frequency trading algorithms are a primitive example of this and will become much more sophisticated in the future. We can imagine equipping an AI with all the world’s financial information in real time, plus all of the world’s laws and regulations, plus newsfeeds and anything else we think might be relevant, then assigning it the goal of “maximum legal profit” or maybe “maximum profit we can get away with.” The guess is that this isn’t very far off, and that the result will be all sorts of novel and completely unexpected hacks. And there will probably be some hacks that are simply beyond human comprehension, which means we’ll never realize they’re happening. (Source: wired.com)

Going forward, we’re more likely to see collaborative AI–human hacks. An AI could identify an exploitable vulnerability that would potentially be a hack, and then an experienced accountant or tax attorney would use their experience and judgment to figure out if that vulnerability could be profitably exploited.

For almost all of history, hacking has exclusively been a human activity. Searching for new hacks requires expertise, time, creativity, and luck. When AIs start hacking, that will change. AIs won’t be constrained in the same ways or have the same limits as people. They won’t need to sleep. They’ll think like aliens. And they’ll hack systems in ways we can’t anticipate.

Computers have accelerated hacking across four dimensions: speed, scale, scope, and sophistication. AI and automation will exacerbate these trends even more. (Source: A Hacker's Mind: How the Powerful Bend Society’s Rules, and How to Bend them Back by Bruce Schneier).

Therefore, the question and problem is clearly that, would we be able to make timely responses with well informed decisions with human-AI, or AI alone action such that defenses are stronger, informed, and accurate? Thereby the human energies & efforts are conserved on progressing business, human race and society on a whole.

What’s the pressing need now [8]
[14] As cyber-attacks are rising in number, complexity, and variety, there is a need to upgrade to an cyber defense system that is fully automated. With the advancements in technologies, it is projected that the future certainly holds the reality of Artificial Intelligence-driven cyber-attacks, where malware can self-propagate through a series of autonomous decisions and intelligently tailor itself to the parameters of the compromised system in order to become stealthier to dodge detection. Discussing the autonomous cyber defense system where algorithms will fight against algorithms on the battleground of enterprise networks, only the best AI will win. To defend against future AI-driven attacks, autonomous cyber AI can be a revolutionary asset. Performing as a cyber immune system, this AI is capable of learning what is normal and abnormal for the digital business, without relying on prior knowledge of threats. The technology can not only detect never-before-seen threats but also autonomously reply to detach the attack before damaging any system.

3. A Preliminary Review of the Literature

In a study done by Michael Oreyomi & Hamid Jahankhani [6] on Autonomous Cyber Defence (ACyD) Systems against cyber attacks, they demonstrated that with the help of AI (Artificial Intelligence), ML (Machine Learning) and DL (Deep Learning) they could detect a MITRE Attack successfully. They used consolidated WFCM-ANN (Weighted Fuzzy C-Clustering and Artificial Neural
Network) for this purpose. However, they also pointed out limitations through a quote by Kevin Mitnick (one of the most notorious hackers of all time), “A company can spend hundreds of thousands of dollars on firewalls, intrusion detection systems and encryption and other security technologies, but if an attacker can call one trusted person within the company, and that person complies, and if the attacker gets in, then all that money spent in technology is essentially wasted”.

Therefore, they emphasized the need on ‘User Education, Awareness and Learning’. However accidental or careless clicks could also result in major calamities.

**AI Cyberattack Case**[3]

However AI can beat human learning too with a classic example quoted earlier when a group of fraudsters made off with $35 million after using forged email messages and deep fake audio to convince an employee of a United Arab Emirates company that a director requested the money as part of an acquisition of another organization, according to a US federal court request filed in 2021. (Source: Dark Reading)

Therefore, the need of UEBA (User And Entity Behavior Analysis) is necessary to ensure the best action from the event data collected from SIEMs (Security information and event management) systems. A detailed study on UEBA has been done by S. R. &. N. R. V. Rahma Olaniyan[7] showed that one of the many use cases of its application is Identification of Attacker Behaviors.

**There are two main areas of which have gained an impetus post Covid-19**

The two main areas which have been impacted post Covid-19 have been Data Privacy and Company data locking (Ransomware).

**Data Privacy**

Data breaches have been on the rise for a number of years, and sadly, this trend isn't slowing down. 2021/22 have been littered with thefts of sensitive information. Data breaches have affected companies and organizations of all shapes, sizes, and sectors, and they're costing US businesses millions in damages.

The widely-covered T-mobile data breach that occurred last year, for instance, cost the company $350 million in 2022 – and that's just in customer pay outs. This puts more onus than ever on businesses to secure their networks, ensure staff have strong passwords, and train employees to spot the telltale signs of phishing campaigns.

Figure 3 is a compiled list of significant, recent data breaches (and a couple of important data leaks) that have taken place since January 1, 2022, dated to the day they were first reported in the media.

**Company data locking (Ransomware)[6]**

Ransomware attacks have risen exponentially.

**Rise of cryptocurrencies**

The rise of cryptocurrencies has made it easier for hackers and bad actors to profit from ransomware. Before crypto,
Attackers had to resort to telling a victim to go to a corner store, buy a $100 gift card, and send them the code. In this scenario, tracing fiat currencies was easy for law enforcement. Today, because of the inherently anonymous and decentralized nature of crypto wallets, authorities are having a hard time tracing the money taken as ransomware payments.

Bitcoin transactions can be traced, but scammers and hackers are aware of this. So they move their illicit profit through hundreds, even thousands of transactions across a dozen or so wallets. They employ ‘mixers’, which take an amount of crypto, break that up into smaller transactions, and ‘mix’ those with transactions from other people in the blockchain – further masking the paper trail.

Hackers have started using pseudonymous (only shields a user’s true identity with a generated alphanumeric address that’s traceable through financial forensics).

Example: Monero, which promises a “private, decentralized cryptocurrency that keeps your finances confidential and secure.” It is virtually untraceable and, according to their explainer video, helps citizens “escape government repressions and nosy neighbors or crooks.” Monero employs various methods to hide the identities of their senders and receivers, such as:

1. **Stealth addresses**: one-time generated addresses to obscure public blockchain transactions
2. **Ring signatures**: a digital signature that “can be performed by any member of a group of users that each have keys.”

Other cryptocurrencies that have robust built-in privacy features include:

- **Zcash**: implements Zero-Knowledge Proof, a cryptographic tool that obscures transaction amounts and allows participants to transact without seeing each other’s addresses.
- **DASH**: has a PrivateSend feature that obfuscates the origin of the transaction’s funds with a blockchain mixing protocol.
- **Horizen**: offers both public T-Addresses and privacy shielded Z-Addresses.
- **Verge**: protects user identities through The Onion Router’s (TOR) technology and the Invisible Internet Project’s (I2P) encryption.
- **Beam**: a security-focused token where all transactions are private by default and no private information from the sender or receiver is stored on the blockchain.

**Global Quarterly Ransomware Attacks**

Online services - Ransomware-as-a-Service (RaaS)

RaaS is a DIY (Do it Yourself) model has also spurred this increase in ransomware attacks.

Curious Newbies, Inquisitive Students, Small Crooks looking to make fast bucks et al are free to avail this service. ‘Hall Of Ransom’ is one such site in the Dark Web which hosts many such hacking tools too, making it easy for anyone to acquire them. These tools have clear instruction of usage enabling anyone with little or some knowledge to use them.

Some of the most notable RaaS providers are Cerber, which earned $2.5M annually and was known as the largest RaaS ring in 2016, and REvil (aka Sodinokibi), which zeroed in on large businesses and allegedly earned $100M within a year.

These are sophisticated outfits that offer 1:1 bespoke services and have established reputations within the cyber underworld. By filling the RaaS niche, they have opened up the potential for ransomware attacks to a much broader field of actors – essentially, to anyone with money and means.
Top 5 Ransomware families

![Ransomware Pie Chart](chart.png)

Figure 4: (Source: Dropsuite)

National Safe Havens for Attackers

Ransomware growth has also been influenced by the fact that certain countries and territories have become “safe havens” for attackers.

Russia, in particular, ignores threat actors in their country, as long as they do their ‘work’ outside of Russia’s borders. According to the New York Times:

“Cybersecurity experts say the ‘don’t work in .ru’ stricture, a reference to Russia’s national domain suffix, has become de rigueur [demanded by custom] in the Russian-speaking hacking community, to avoid entanglements with Russian law enforcement. The Russian authorities have made it clear they will rarely prosecute cybercriminals for ransomware attacks.”

North Korea is also considered a safe haven for ransomware hackers and cybercriminals alike. According to The New Yorker:

“North Korea’s cybercrime program is hydra-headed, with tactics ranging from bank heists to the deployment of ransomware and the theft of cryptocurrency from online exchanges.”

The White House has accused China of abetting ransomware hackers in July of 2021, along with its allies in the United Nations, in an effort to condemn China’s “malicious cyber activity.”

In a CNBC interview, Michael Orlando, acting director of the National Counterintelligence and Security Center, listed other countries that protect hackers from within their borders, and how this problem is part of their mission to counteract ransomware incidents.

“We do know that countries like Russia and China, Iran and others certainly create safe havens for criminal hackers as long as they don’t conduct attacks against them. But that’s a challenge for us that we’re going to have to work through as we figure out how to counter ransomware attacks.”

National safe havens for ransomware attackers – and the explicit encouragement, even direct business of, governments – allow attackers to thrive. They now have places where they can build large offices, recruit teams, access powerful infrastructure – all without fear of reprisal.

It takes a thief to catch a thief

To beat some in their own game, you must think like them. If they are fast, you must be fast; if they are cutting-edge, you must be cutting edge. To counter the threats posed by cybercriminals, organizations ought to be faster. It requires to do away with traditional security measures and embrace new age, automation-driven practices that could put us ahead of any hacker. The regular practice includes securing only mission-critical parts within an infrastructure. This leaves room open for hackers to target non-critical components. Therefore, organizations must implement comprehensive and robust cybersecurity procedures that cover every component within an infrastructure. Further, organizations should align themselves with the practice of leveraging automated scripts to facilitate continuous monitoring and reporting in real-time.

However attacks are not limited to companies alone.

With cyber attacks on essential services like in the case of the ransomware attack on
1) Colonial Pipelines which runs oil pipe along 5,500 miles along the East Coast in North America, it halted both consumers and airlines.
2) Hacker remotely killed car engines after breaking into GPS tracking apps, and that too across countries in different continents, e.g. South Africa, Morocco, India, and the Philippines.

The above two examples enumerate the dangers that we can foresee especially possible techniques of Cyber Terrorism.

Hence the review done by Sanyam Vyas, John Hannay, Andrew Bolton, Professor Pete Burnap on [9] have been able bring out that cybercriminals have curated a variety of organised and resolve cyber attacks within a range of cyber
systems including AI, leading to consequential ramifications to private and governmental institutions.

Sanyam, John, Andrew and Pete Pete establish the need of automated defence and attack agents and Autonomous Cyber Operation (ACO).

Though Michael Oreyomi & Hamid Jahankhani [6] bring out the need and mechanisms of ACyD, there are certain aspects which could strengthen the development of ACyD.

1) One clearly is the use of NLP (Natural Language Processing). In a study Andrew Golczynski and John A. Emanuello establish that Neural IDS have limitations and demonstrate that the effectiveness could be increased by End-To-End Anomaly Detection for Identifying Malicious Cyber Behavior through NLP-Based Log Embeddings [10].

2) The second important aspect is accuracy of detection. Anson Pinhero, Anupama M. L, Vinod P, C.A. Visagio, Aneesha N., Abhijith S., Anantha Krishnan through a Malware detection employed by visualization and deep neural network paper [13] investigate the effectiveness of a new approach that uses malware visualization, for overcoming the problems related to the features selection and extraction, along with deep learning classification, whose performances are less sensitive to a small dataset than machine learning. The experiments carried out on twelve different neural network architectures and with a dataset of 20,199 malware, demonstrate that the proposed approach is successful as produced an F-measure of 99.97%.

4. Possible research approach[7]

The study would involve analysis of email gateway-based malwares, firewall based breach attempts (botware etc), CTI information, many other threat vectors and SIEM data. With the resultant data, use Michael Oreyomi & Hamid Jahankhani [6] technique to identify a plausible attack. Then use Andrew Golczynski and John A. Emanuello [10] methodology to narrow the accuracy. Post which Identify two models to run it on a GAN to arrive at the final detection risk score. Both qualitative and quantitative outcomes are expected in the approach as ML models would lean towards improving risk scores and AI models a mix of qualitative inputs e.g. cognitive actions to self-respond objectively. Therefore, the intent is to identify mechanisms to defend the perimeter and UEBA for the user endpoints.

The ideal role of AI in cybersecurity is the interpretation of the patterns established by machine learning (ML) algorithms. AI at its core is concentrated on “success” with “accuracy” carrying less weight. Natural responses in elaborate problem-solving are the ultimate goal. In a true execution of AI, actual independent decisions are being made. Its programming is designed for finding the ideal solution in a situation, rather than just the hard-logical conclusion of the dataset.

ML proceeds intending to learn from a task-focused dataset. It concludes by finding the most optimal performance of the given task. It will pursue the only possible solution based on the given data, even if it’s not the ideal one. With ML, there is no true interpretation of the data, which means this responsibility still falls on human task forces. The following therefore a very important:

Data classifying
Data classifying works by using preset rules to assign categories to data points. Labeling these points is an important part of building a profile on attacks, vulnerabilities, and other aspects of proactive security. This is fundamental to the intersection of AI and cyber security.

Data clustering
Data clustering takes the outliers of classifying preset rules, placing them into “clustered” collections of data with shared traits or odd features. For example, this can be used when analyzing attack data that a system is not already trained for. These clusters can help determine how an attack happened, as well as, what was exploited and exposed.

Possibility synthesis
Possibility synthesis allows for the synthesizing of brand-new possibilities based on lessons from previous data and new unfamiliar datasets. This is a bit different from recommendations, as it is concentrating more on the chances that an action or the state of a system falls in line with similar past situations. For example, this synthesis can be used for a preemptive probing of weak points in an organization’s systems.

Predictive forecasting
Predictive forecasting is the most forward-thinking of the ML component processes. This benefit is achieved by predicting potential outcomes by evaluating existing datasets. This can be used primarily for building threat models, outlining fraud prevention, data breach protection, and is a staple of many predictive endpoint solutions.

Following are some use case areas for the above approach in Cyber Security with actions in an automated manner[10]

- Possible threat identification
- Cyber incident response
- Home security systems
- CCTV cameras and crime prevention
- Credit card fraud detection and risk reduction
- Border control security
- AI-powered biometrics technology
- Fake customer reviews identification and handling
- Anti-money laundering
- Email spam filters
- Securing authentication
- Detecting zero-day malware
- Cloud security automation
- Enhancing sensitive information security

There are however some the issues of AI in Autonomous Cyber Defense.

5 top challenges that prevent the successful implementation of AI/ML for cybersecurity.

- Non-aligned internal processes
  Most companies have optimized their infrastructure, especially its security components, by investing in tools
and platforms. Yet, we see that they face security hurdles and fail to safeguard themselves against an external attack. This is a result of a lack of internal process improvements and cultural change that prevents capitalizing the investments in security operation centers. Further, the lack of automation and fragmented processes creates a less robust playground to defend against cybercriminals.

- **Decoupling of storage systems**
  Most organizations do not leverage data broker tools like Rabbit MQ and Kafka to initiate analytics of the data outside the system. They do not decouple storage systems and compute layers, which doesn’t allow AI scripts to execute effectively. Further, a lack of decoupling of storage systems increases the possibilities of vendor lock-ins in case of a change in the product or platform.

- **The issue of malware signature**
  Signatures are like fingerprints of malicious code that assist security teams in finding the malware and raising an alert. The signatures do not match the growing number of malware every year. The concern is that any change in the script of the virus makes the signature invalid. In short, signatures will only help debug malware if the code is pre-established by security teams.

- **The increasing complexity of data encryption**
  The rise in the use of sophisticated and advanced data encryption strategies are making it difficult to isolate an underlying threat. The most common way to monitor external traffic is via deep packet inspection (DPI) that helps filter external packets. However, these packets consist of a predefined code characteristic that can be weaponized to infiltrate in the system by the hackers. Further, the complex nature of DPI puts pressure on the firewall, slowing down the infrastructure speed.

- **Choosing the right AI use cases**
  More than 50% of the AI implementation project fails in the first go. This is because organizations try to adopt AI on a company-wide level. They often neglect the importance of baby steps – narrowing down on AI-based use cases. Thus, they miss out on initial learning curves and fail to absorb critical hiccups that often jeopardize the AI projects.

**Control and governance is necessary on AI in any field especially Cybersecurity [13]**
AI enabled cyber defense could be the future implications of Quantum computing or high processing computers. This enhancement to support data-processing may increase the efficiency of algorithms. Algorithms are key components of running AI and may be tailored to counter complex cyber threats. An algorithm is a set of step-by-step instructions given to a computer to accomplish a specific task. AI may push this technology to another level, to achieve intelligent autonomous algorithms. To illustrate these research challenges, Facebook recently abandoned an AI experiment after ‘chatbots’ invented their own language which was not understandable by humans. Computer machines had demonstrated better skills than humans in playing chess or poker. This breakthrough technology is likely to be disruptive in many ways nobody can predict today.

Following is a view of AI in Cyber Security - Market size Global [4]

![Global Artificial Intelligence in Cybersecurity Market $B](source: Grandview Research)


![Asia Pacific Artificial Intelligence in Cybersecurity Market SB, CAGR 32.2%](source: Precedence Research)
5. Conclusion

Even the most successful security organization is a work-in-progress. The dynamic nature of operations and the continuous emergence of new threat vectors require organizations to prioritize readiness and resilience. It is not a matter of if an organization will be breached, but when and to what extent. Similarly, recognition that AI models must continue to learn, and security teams must keep feeding them with new performance insights.

This commitment to constant learning influences the outcomes organizations can achieve. For organizations, security performance is impacting both operational efficiency and business value, while creating a more empowered, more adaptable work environment for security analysts. Whether an organization is piloting these capabilities for the first time or expanding the functionality of existing applications, three recommendations can guide these efforts.

Benchmark performance across key security metrics

Prioritization of security improvements that deliver the most value and align with top security goals

Develop key enablers for security improvement initiatives

References