Impact of Emerging Cybersecurity Trends – AI and IoT

Zero Day Security, Ltd.

Colin Cooper
R00240295
Emerging Cyber Trends (COMP 8063) 2022/23 - Assignment 2

1. Executive Summary

Zero Day Security (ZDS) is a cybersecurity company that provides threat detection and remediation services to its customers. This makes it both a potential valuable target for malicious actors as well as making the financial and reputational cost of a breach extremely high. Recent developments such as the decision to provide €2.5m of funding for a Data Science team to develop AI-based threat detection capabilities, and the deployment of "smart" facilities infrastructure in the company offices create challenges and opportunities from a cybersecurity perspective.

There are significant opportunities as well as threats associated with the adoption of artificial intelligence (AI) in the field of cybersecurity. AI tools can scale the capabilities of human-based threat detection, put the power of experience into the hands of more junior threat analysts, and help make more consistent decisions about whether patterns of behaviour are malicious or not. AI has the potential to identify real threat patterns more quickly in very large data sets and to improve the efficacy of incident response through automation.

However, AI tools can also enable malicious actors to accelerate the development of new threats and open new targets for attacks such as the AI models and model development processes themselves. For ZDS, investment in a Data Science team provides opportunities to improve existing threat detection capabilities and automate incident response actions. However, ZDS needs to take care while creating this capability to ensure that a secure AI development infrastructure is in place that minimises the risk of training data contamination or compromise of any developed models used for threat detection. We recommend the creation of secure-by-design AI development infrastructure that includes secure development and coding practices, supply-chain management, and regular security auditing.

The deployment of IoT infrastructure to help improve physical security and reduce the energy costs of running the company's office creates security challenges that were not fully considered by the Facilities team who funded and installed these systems. A security catch-up is required to ensure that such devices are up to date, securely configured, and do not create a backdoor to compromise other critical systems – including the new AI development environments. It's recommended to carry out a comprehensive review of the IoT environment to assess the current configuration and vulnerability status; to put in place NIPS and firewall devices into the environment to prevent and detect attacks; to physically separate the system from mission-critical infrastructure; to improve procurement process for future IoT purchases; and to revise the contingency plans to take account of these systems.

Table of Contents

2. Introduction

2.1. Background

Zero Day Security ("ZDS") is a Managed Detection and Response (MDR) provider specialising in outsourced cybersecurity monitoring, threat detection and remediation services to its customers. ZDS's typical customers are small enterprises with between 50 and 250 employees for whom a dedicated in-house Security Operations Centre (SOC) would not be financially feasible. Through a shared service model, ZDS enables dilution of SOC operational costs (such as physically secure SOC environments, hardware, networking, threat analyst staff costs, and threat detection and forensics software licensing costs) among its customer base to offer a service that scales to the needs and budgets of small organisations.

ZDS had gross revenues of €15.2m in 2021 from 350 customers based in the UK, Ireland, and Sweden from which it made an operating profit of €3.6m in the same period. The company has three centres of operation: in Cork (Ireland), Reading (UK) and Stockholm (Sweden). All three sites have fully operational interconnected Security Operations Centres.

In late 2022, the ZDS board approved a €2.5m investment into its Artificial Intelligence and Machine Learning capabilities. The purpose of this investment is to help address the challenges of growing the business in an environment of increasing threat volumes while cybersecurity talent is scarce and expensive. The board realised that to enable the business to grow, internal human cybersecurity talent would need to be augmented by technology. This investment will be used to create an R&D Data Science team to initially work on developing threat detection technologies that would be used both by the internal security team, as well as by customer-facing SOC employees.

The CISO has also become concerned that while ZDS has a strong security posture for traditional IT infrastructure and devices such as laptops and servers, there has been a recent growth in non-traditional "smart" infrastructure such as office security cameras, building temperature control systems and smart energy and lighting systems that were installed under a Facilities department investment to reduce the company's energy costs and carbon footprint.

ZDS recognises the potential reputational damage of a cybersecurity breach, and that the nature of its business makes it a significant target for attackers.

2.2. Purpose

The purpose of this report is to:

- Critically evaluate the opportunities and threats associated with the emergence of Artificial Intelligence (AI) and Machine Learning (ML) technologies, particularly as these pertain to ZDS.
- Analyse the impact of increased digitisation and connectivity on ZDS, particularly regarding recent investments by the Facilities department in smart networked technologies for security and building energy control.

3. Evaluation of Emerging Enterprise Cybersecurity Trends

3.1. Artificial Intelligence (AI)

AI refers to the field of computer science and statistical mathematics associated with the study of how non-organic machines can simulate human consciousness to solve problems and perform tasks that have traditionally required biological intelligence. Machine Learning (ML) is a subset of AI that seeks to create decision-making models that are trained using large amounts of data. These generalised models can then be applied to new previously unseen data to make accurate predictions or decisions.

Developments in AI theory, improved computing power and the availability of large publicly accessible datasets to train models have meant on the one hand that developing powerful models has become progressively easier and cheaper, and on the other that extremely powerful AI models capable of convincing mimicry of human intelligence (such as that behind ChatGPT) have become feasible. The advancement of AI tools creates both opportunities and threats in the field of cybersecurity.

Opportunities

- **Scaling Threat Detection:** With the growth in threat surfaces, ever-more sophisticated threats, and limitations in the availability of cybersecurity professional skills, threat analysts are finding that their ability to analyse ever-growing volumes of data in SIEM and EDR solutions to detect anomalies and threats is constrained. It will become important for existing cybersecurity analysts to use AI technology to identify threats more quickly [1].

- **Broadening Expertise:** Integrating models that have been trained with industry-leading threat detection knowledge into cybersecurity solutions makes that expertise available to all users of those solutions [2]. This helps to shorten the learning curve for more junior threat analysts.

- **Removing Human Error:** With AI, we are usually attempting to replicate a small number of decision-making processes that are just a fraction of thousands of such processes performed by humans daily. In the context of cybersecurity threat detection, the only decision a model may need to make could be as simple as "*does this process behaviour pattern look malicious*." The model that will make that decision will make it solely based on the data features it has been trained to include – such as genealogy of the process, the set of actions being performed, and the target of those actions. The model is not influenced in the complex ways that humans are when we make decisions, such as whether it is late on Friday before a long weekend or whether it had just had a difficult meeting with a head of department. Human decision-making is affected by short-term factors such as fatigue, mood, or distraction. A well-trained model will not make mistakes due to becoming tired, frustrated, or distracted.

For ZDS, the investment into a Data Science team provides opportunities for improving existing threat detection capabilities as the company has access to significant quantities of data from customer telemetry. A well-trained AI model has the potential to more quickly identify real threat patterns that are difficult for a human to recognise within large quantities of data. AI-directed threat detection and investigation should mean that threat analysts are less distracted by false positives generated by basic pattern-matching rules, and able to react to real threats more quickly. Current automated threat detection in the company's EDR solution is based on event pattern matching which does not consider customer-specific baselines for normal behaviour. For example, one ZDS customer had an IT-managed PowerShell script running on all Windows devices that periodically wrote to sensitive Windows Registry locations. This behaviour matched patterns associated with known malware techniques and so while the pattern-matching rule was valid and working correctly, it raised a substantial number of false positives for one customer. A well-trained AI model could overcome this by learning customer-specific behaviour patterns.

The future roadmap for the Data Science team should also seek to use AI to automate Incident Response actions such as containment and remediation of threats.

Threats

- **Adversary Enablement:** The same tools available to cybersecurity defenders are also available to malicious actors who can use AI tools to accelerate the development of new threats. This can enable attackers to more quickly create and adapt threats and discover zero-day vulnerabilities. According to McKinsey, "*over the next several years, [attackers] will be able to expedite—from weeks to days or hours—the end-to-end attack life cycle, from reconnaissance through exploitation. For example, Emotet, an advanced form of malware targeting banks, can change the nature of its attacks. In 2020, leveraging advanced AI and machine-learning techniques to increase its effectiveness, it used an automated process to send out contextualized phishing emails that hijacked other email threats*" [3]

- **Model Vulnerabilities:** As part of an information systems and cybersecurity infrastructure, the models themselves are attractive targets for attack. The decision-making process of a machine-learning model such as one underpinned by an artificial neural network can be opaque. If the model itself can be compromised by an attacker, it can be difficult for that compromise to be detected and remediated.

- **Model Quality:** A decision-making model can only be as good as the training that underpins it. Bad training data or a poor training process (e.g., model overfitting and underfitting) will lead to models that make poor decisions. Training data itself can contain record and feature selection bias that limit the effectiveness of the model in real-world scenarios. Detecting quality problems with an AI model can also be challenging due to complex and opaque processes that underpin it and the dynamic nature of constantly trained systems. This view is endorsed by NIST who warned that:

> *"AI systems, for example, may be trained on data that can change over time, sometimes significantly and unexpectedly, affecting system functionality and trustworthiness in ways that are hard to understand. AI systems and the contexts in which they are deployed are frequently complex, making it difficult to detect and respond to failures when they occur."* [4]

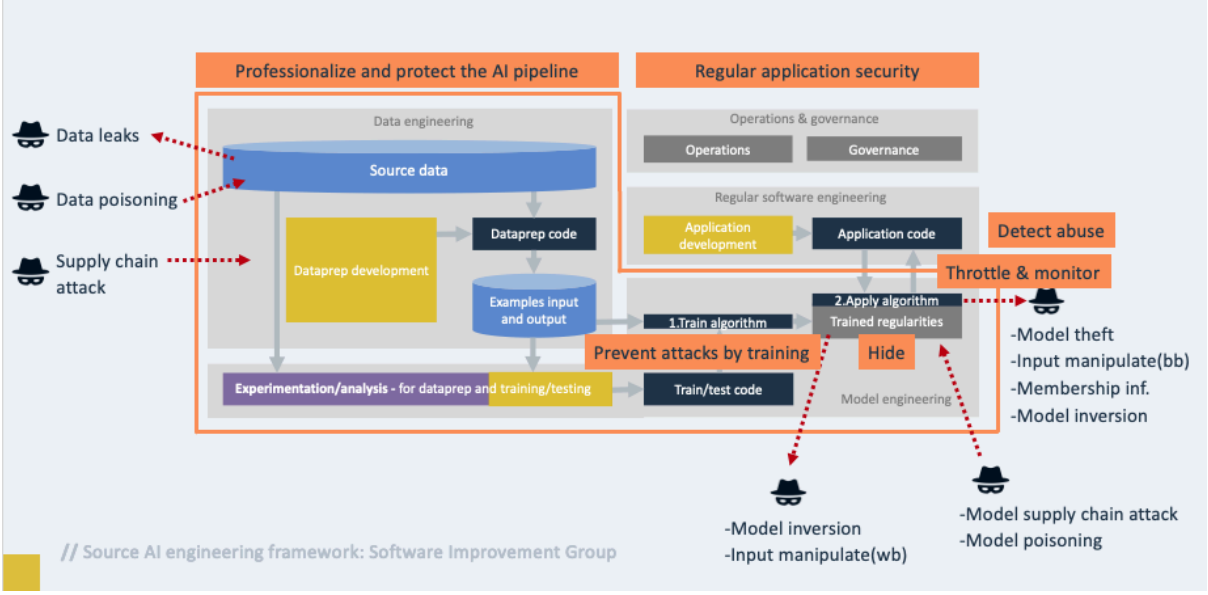
For ZDS, the adoption of AI both internally and by malicious actors outside of the organisation also poses significant challenges. Models need to be well-trained and regularly validated to minimise false-negative (missed detection) and false-positive (bad detection against normal behaviour) outcomes. Care must be taken to not over-depend on AI systems that are themselves open to malicious manipulation and bypass. In fact, it should be recognised that an Information Security AI system is itself a valuable target for attack through contamination of training data, base models, and other parts of the model supply chain.

3.2. Implications for Building a Cybersecurity Program

ZDS has already committed funding to develop a Data Science team with the aim of using the huge quantities of data collected from internal and customer sources to develop AI models that will supplement and support human threat analysts. This investment is designed to support projected growth in new business and increased volumes of threat telemetry coming from these customers.

It will be critical for the models developed by the new team to be secured from malicious tampering throughout the models' development and deployment.

The OWASP group have developed a generic model [5] of the attack surface for AI models that are useful in planning the security response to AI threats.



- **Training Data Attacks:** To develop AI models, data will be needed to train these models. In some cases, human intervention may be required to label records (supervised training). This leaves open the potential for data poisoning or manipulation of the data used to train the model. For example, if an attacker could manipulate the labels of records associated with an attack to teach the model that these are normal, then the model could be trained to treat related behaviours as safe.

- **Supply Chain Attacks:** Data Scientists often use publicly available base models as starting points for custom fine-tuned model development to accelerate development. An attacker could manipulate the base model and thereby corrupt any derivative models developed from it.

- **Model Theft:** An attacker who can directly get a copy of a model (or simulate it by training a new model to replicate the outputs of an existing one) can use that model to probe for weaknesses that could be exploited in a real-world attack.

The ramp-up of the Data Science team provides an opportunity for a security-by-design approach to AI model development that mirrors traditional secure development best practices. The following specific recommendations need to be implemented during the build-out of the team and model development environment:

1. **Secure infrastructure for model development:** The AI development environment should be on a separated network from other operational and IT environments, with its own firewall, IDS/IPS, threat detection and anti-malware solutions in place that minimise the attack surface. Data repositories used for model training should be within the secure environment with access based on the principle of least privilege – i.e., only those with responsibility for loading data into these repositories should have write-access, and only those needing the data for analysis and training having read-access. Data should be encrypted both at rest and in transfer using approved secure encryption methods.

2. **Controlled supply chain**: Base models, development tools and libraries to be used need to be specified and approved through the information security team. This will help ensure only appropriately licenced and up-to-date component versions are used in the model development.

3. **Secure coding practices and training:** ZDS should invest during the hiring of the Data Science team – particularly those who will be involved in programming data preparation and model training – in secure development practices to avoid issues such as storing credentials, certificates, or sensitive data in source control systems; and usage of unapproved third-party libraries.

4. **Security Audits:** Regular security audits of the AI development environment and models should be conducted to ensure that security controls are working as intended and any issues identified and corrected quickly.

4. Analysis of the IoT Impact on ZDS's Security Posture

4.1. Introduction

The last decade has seen an acceleration in the convergence of computing power and networking technologies with traditional devices to enable those devices to collect, process, transmit, and receive data. Whereas a traditional building heating system could only be configured manually at physical interface points with little-or-no data recording or decision-making, "smart" building systems can now run a scaled-down webserver that is accessible remotely and be used to collect information from and send information to dozens of sensors across a collection of physically separate buildings. These sensors can have networking capabilities and run their operating systems and applications.

Such innovations create tremendous opportunities for businesses to optimise their operations in manufacturing, service provision and cost reduction. In the case of ZDS, the main application of IoT has been in managing the security and environmental control within the company's offices and data centres with:

- Installation of new IP-based security cameras that enable remote monitoring and cloud-based storage of captured video.
- Integration of the door entry systems with the HR database to provide personalised access to buildings and areas within those buildings based on role.
- Installation of environmental sensors in all buildings that monitor and record data such as temperature, humidity, brightness, and human presence within monitored zones.
- Deployment of a centralised remotely-accessible web-based console to allow monitored zones to be configured and managed – for example, a data centre zone can be set to maintain constant temperature throughout the day, while hot-desk zones can be set to allow a wider range of temperatures during periods of low-occupancy.

4.2. Practical Implications

IoT devices are in effect small computing devices that run operating systems and applications. For example, the energy control console is a Node.js web application that runs on an Apache Tomcat web server that runs in a Java VM which in turn runs on a customised Ubuntu operating system. All these layers as well as the device firmware and the third-party open source libraries used by these layers (e.g., OpenSSL) have had multiple security updates within the past 12 months.

The installation and deployment of smart systems and their integration with existing information systems such as the HR database has increased the attack surface available for malicious actors and created additional dependencies on IoT vendors. By adopting IoT technology, ZDS is putting a large amount of trust in its IoT vendors and their software

supply chain. This level of trust is not one that such vendors have earned by default. According to the 2021 Nokia Threat Intelligence Report IoT devices "*remain vulnerable to botnets, often serving as gateways for more sophisticated attackers to gain a foothold into a network"*. The same report showed that *"IoT devices still account for 32% of all infected devices*" [6].

I strongly recommend the following actions to be taken to improve the security posture of the company with regards to its new IoT infrastructure:

1. **Security Review:** An immediate security review should be conducted on all IoT devices to assess and document device configurations, operating system and application software inventory, versions, and vulnerabilities. A common vulnerability with IoT devices is the use of insecure default configurations including administrator usernames and passwords that are easily guessed by attackers.

2. **Network Segregation:** Because of the limitations in being able to control and update the OS and application patch levels on IoT deployments, these environments must be considered as a "dirty" network that is physically separated from other ZDS networks such as the AI model development network and the customer EDR cloud infrastructure. Implementation of network firewall and NIPS devices onto the IoT network is also recommended to provide control and visibility of anomalous traffic. Furthermore, it is recommended that the door entry systems be disconnected from the live HR database and that instead, a one-way replica of the HR database be created to reduce the risk of data leakage and compromise of the live system.

3. **Patch Management:** A comprehensive patch management process should be implemented for all IoT devices in use by ZDS. This should include regular vulnerability scans and prioritisation of high-risk vulnerabilities for immediate patching. Any IoT device that cannot be patched or is no longer capable of receiving security updates should be replaced as soon as possible.

4. **Update Business Continuity and Incident Response Plans:** The IoT infrastructure can be the forgotten component of business continuity and incident response. The Facilities teams that operate this infrastructure need to be included in contingency planning for security incidents with assigned roles and responsibilities.

5. **Work with Procurement:** For future IoT product and service purchases, a security questionnaire should be included in the Request for Proposal (RFP) system that requires vendors to answer questions about their secure development practices, software update standards and SLAs for patch management. It is critical that products are not selected based just on price or service, but also on their ability to securely develop their product and provide timely updates.

5. Overall Conclusions

Emerging technologies constantly create opportunities for ZDS to improve its strategic capabilities, improve customer service and reduce operational costs. The fields of Artificial Intelligence, Machine Learning and IoT were recently identified by separate business units as areas of investment to achieve these goals.

However, insufficiently planned adoption of these technologies creates risks for the business that could result in sub-optimal investment outcomes and in the worst-case scenario, create a path to compromising critical business IT infrastructure.

The early-stage investment into AI provides an excellent opportunity to shape that programme from a security perspective and build-in best security practices that both help to avoid security incidents, but also to maximise the return on investment.

In contrast, the investment by Facilities in IoT technology was done without adequate consideration of security implications that now requires a catch-up programme to document, assess, and if necessary – correct - the security posture of this infrastructure. The process failures in this regard provide lessons that must be used to improve future procurement programmes through collaboration between Facilities, IT, Information Security and Procurement teams.

6. Bibliography

[1] R. Das and R. Sandhane, "Artificial Intelligence in Cyber Security," *Journal of Physics: Conference Series,* pp. 1,2, 2021.

[2] M. Dsouza, "How will AI impact job roles in Cybersecurity," Packthub, 25 September 2018. [Online]. Available: https://hub.packtpub.com/how-will-ai-impact-job-roles-in-cybersecurity/. [Accessed 21 April 2023].

[3] J. Boehm, C. Lew, K. Li and D. Wallance, "Cybersecurity trends: Looking over the horizon," McKinsey, 10 March 2022. [Online]. Available: https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/cybersecurity/cybersecurity-trends-looking-over-the-horizon. [Accessed 18 04 2023].

[4] National Institute of Standards and Technology, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," U.S. Department of Commerce, 2023.

[5] OWASP, "OWASP AI Security and Privacy Guide," [Online]. Available: https://owasp.org/www-project-ai-security-and-privacy-guide/. [Accessed 16 04 2023].

[6] "Nokia Threat Intelligence Report," Nokia OYJ, Espoo, 2021.