SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Generative AI Based Intelligent Multi Stage ZeroDay Intrusion Detection Using Behavioural Attack Modelling

Nithya .M[1] Dr. Yamini .C[2]

[1]Ph.D Research Schloar, School of Computer Science, Sri Ramakrishna College of Arts and Science for Women, Coimbatore, India. E-mail: nithyavinothprabhu@gmail.com, Orcid-id: 0009-0006-6678-2430
[2]Associate Professor & Head, School of Computer Science, Sri Ramakrishna College of Arts and Science for Women, Coimbatore, India. E-mail: yaminics@srcw.ac.in, Orcid-id: 0000-0001-5263-0256

ABSTRACT
Traditional Intrusion Detection Systems (IDSs) are inefficient at dealing with advanced attacks, such as zero day attacks, due to the use of signature matching and static anomaly detection. The proliferation of the cloud computing environment, the Internet of Things (IoT), the Industrial Internet of Things (IIoT) and distributed enterprise infrastructure have increased the surface area for attack even more, making previously unknown attacks more difficult to detect. Despite the good performance of the recently developed intrusion detection (ID) techniques based on Artificial Intelligence (AI), most current methods are effective only against a single attack, without considering the sequential nature of behaviour-driven attacks that is characteristic of multi-stage intrusions. Existing learning models also don't generalise well because there aren't any representative samples of zero-day attacks. To solve these problems, this paper proposes a Generative AI-Based Intelligent Multi-Stage Zero-Day Intrusion Detection Using Behavioural Attack Modelling (GAIM-ZID) framework which incorporates all the above techniques into one adaptive IDS.The proposed methodology comprises seven modules: intelligent data acquisition and pre-processing; hybrid behavioural feature extraction using CNN and Transformer networks; Generative AI-based attack modelling using Generative Adversarial Networks (GANs), Variational Auto encoders (VAEs) and Large Language Models (LLMs); multi-stage behavioural learning using BiLSTM and Temporal Transformer architectures; Graph Neural Network (GNN)-based behavioural threat correlation; an attention-driven explainable decision engine; and continuous adaptive learning via Reinforcement Learning (RL). The framework is designed to learn attacker behaviour throughout the entire life cycle of a cyberattack, which enables it to proactively identify new, unseen zero-day attacks, and evolve to capture new attack patterns. The proposed model was evaluated using the benchmark intrusion detection datasets such as CICIDS2017, CSE-CIC-IDS2018, UNSW-NB15, BoT-IoT, TON_IoT and Edge-IIoTset. The GAIM-ZID framework had an overall accuracy of 99.21%, precision of 98.94%, recall of 98.81%, F1-score of 98.87% and an AUC of 99.54%, which are higher than some of the state-of-the-art deep-learning-based intrusion detection approaches. The integration of Generative AI with behavioural attack modelling significantly enhanced zero-day detection accuracy, reduced false positives, provided increased transparency via Explainable Artificial Intelligence (XAI) and allowed for real-time adaptation to emerging cyber threats. The findings show that GAIM-ZID is an effective, scalable and intelligent solution for intrusion detection in cloud computing, Internet of Things and enterprise networks, as well as in other vital cyber infrastructures.

Keywords: Zero-Day Intrusion Detection, Generative Artificial Intelligence, Behavioural Attack Modelling, Multi-Stage Attack Detection, Deep Learning, Convolutional Neural Network, Transformer, Generative Adversarial Network, Variational Autoencoder, Graph Neural Network, Reinforcement Learning, Explainable Artificial Intelligence, Cybersecurity.

I. INTRODUCTION

Digital technology is rapidly changing the modern computing space with the proliferation of cloud computing, edge computing, the Internet of Things (IoT), the Industrial Internet of Things (IIoT), Software-Defined Networking (SDN), fifth-generation (5G) communications, cyber-physical systems (CPS), and smart industrial infrastructures. They have significantly enhanced the use of resources, automation, service availability, and computational performance in healthcare, finance, manufacturing, transportation, defence, smart cities and critical infrastructure. But the same connectivity and complexity has also increased the cyberattack surface, allowing adversaries ample opportunity to launch more and more sophisticated attacks which are difficult to identify and protect against. Of these threats, zero-day is particularly significant as it exploits software or hardware vulnerabilities which have not been patched, disclosed or otherwise mitigated. Consequently, organisations are still vulnerable to attacks that can result in significant monetary damages, loss of operations, theft of intellectual property and disclosure of sensitive data [1, 2].

IDSs (Intrusion Detection Systems) have been one of the primary networks' defense lines for years against any malicious activity. Signature-based IDSs are best suited for detecting known attacks, but are not good at detecting attacks that they have never seen before because there is no signature for them. Unlike normal-based IDSs, anomaly-based IDSs recognize unusual system activities, such as abnormal user behaviour, encrypted communications, or shifting workload or heterogeneous computing environments, and are more likely to detect an unknown attack, but can also produce a high number of false-positives [3] – [5]. Achieving the goal of high detection accuracy, minimal false positives, quick reaction and adaptation to the changing attack patterns

is still challenging for current IDSs. ID has also developed with the rise of Artificial Intelligence: Statistical methods were used first; then, Rule Based Expert Systems and finally, Machine Learning (ML) techniques like Support Vector Machines, Decision Trees, Random Forests and ensemble classifiers. During the past years, deep learning methods such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, autoencoders, Graph Neural Networks (GNNs) and Transformer architectures have significantly enhanced the feature extraction and classification of attacks. Despite this, the majority of existing models rely on labeled training examples, and thus are not very good at uncovering truly novel attack methods, especially zero-day exploits [2, 5, 8]. Today's attacks are unlikely to be one-off or even a series of one-off attacks, but will generally be a multi-stage attack over a period of time, including reconnaissance, discovery of vulnerabilities, initial compromise, privilege escalation, lateral movement, persistence, command and control, and data exfiltration. Each step in isolation might seem to be a valid step in the attack, making it hard to reconstruct the entire attack sequence. Behavioural Attack Modelling (BAM) has been suggested as a solution, in which the attacker behaviour is modeled throughout the entire attack lifecycle and not just by signatures or individual anomalies [1,11,19]. Meanwhile, the advent of Generative Artificial Intelligence (Generative AI) has brought about new opportunities for intelligent cybersecurity: creating realistic attack scenarios, generating synthetic data, predicting and anticipating behaviour, reasoning and learning continuously, in context. In recent years AI models, like Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), Diffusion Models and Large Language Models (LLMs) have demonstrated impressive capabilities in learning intricate behavioural patterns, predicting novel attack patterns, and detecting new threats that were previously unknown [7, 12, 17].

A. Problem Statement

While there have been significant improvements in the field of AI-based intrusion detection in recent years, existing cybersecurity solutions have yet to keep up with the effectiveness of advanced zero-day attacks. The current IDSs are mostly able to detect known attack patterns or mark individual anomalous events, but do not consider the overall behaviour of a cyber attack. The label information provided for the modern deep learning models is also heavily reliant on a considerable amount of labelled data, which often does not contain representative zero day instances. The rise of AIenabled threats and adaptive defence strategies, polymorphic attacks, encrypted communication and multi-stage Advanced Persistent Threats (APTs) are putting pressure on conventional defenses. What's required, then, is an intelligent intrusion detection system that can learn attack patterns, grasp the flow of an attack, and detect new attacks before any meaningful penetration takes place.

B. Research Gap

There are some gaps in literature as per the recent review. One of the first reasons is that the vast majority of zero day intrusion detection techniques are based on discriminative machine learning models that are trained on attack classification and not behavioural reasoning. Second, current deep learning technique requires a large number of labeled data, which makes it hard to deal with new attacks. Thirdly, while multi-stage intrusion detection has been extensively studied, very few works combine Generative AI with behavioural attack modelling to account for the temporal relationships between an attacker's successive actions. Fourth, there is limited integration between behavioural analysis and/or synthetic attack generation, resulting in reduced adaptive learning in dynamic cyber environments. Lastly, current IDSs provide only basic interpretation and intelligence about threats, making it difficult for analysts to stay up to date with attackers' tactics and react appropriately. These gaps suggest the need to develop an integrated solution, combining Generative AI, behavioural reasoning and multi-stage attack analysis to provide a level of zero-day detection accuracy that none of these solutions can provide alone.

C. Research Objectives

This study aims to overcome these drawbacks by integrating behavioural intelligence with advanced generative learning methods in an Intelligent Multi-Stage Zero-Day Intrusion Detection system based on Behavioural Attack Modelling for proactive cybersecurity. The study aims to achieve the following major goals:

1. To develop an intelligent multi-stage IDS that can detect attacks which have not been seen before by modelling the behaviour of attacks.
2. To use Generative AI techniques to create more realistic attack behaviour, enrich security datasets, and enhance the detection of unknown attack patterns.
3. To capture the temporal relationships between the reconnaissance, exploitation, privilege-escalation, lateral-movement, persistence and data-exfiltration phases of the cyberattack.
4. To enhance the accuracy of intrusion detection with a low rate of false alarms by intelligent reasoning of behaviours and adaptive learning.
5. To create a scalable, constantly updated cyber security framework appropriate to the cloud, Internet of Things (IoT), Industrial Internet of Things (IIoT), enterprise network, industrial control system, healthcare, financial and smart city.

The framework is expected to transition the intelligent intrusion detection from signature-based detection to proactive behavioural reasoning, utilizing Generative AI and multi-stage behavioural attack analysis. This research project provides an adaptive cybersecurity framework that can defend against the ever evolving zero-day threats in future digital infrastructures, and allow for more precise, interpretable and resilient intrusion detection.

II. BACKGROUND STUDY

Based on the findings in the previous sections, significant advances have been made in the development of intelligent IDS by leveraging techniques such as machine learning, deep learning, reinforcement learning, federated learning, and graph-based learning. Their systems' capacity to identify known attacks and, in some instances, previously unknown attacks of a particular class, has been greatly enhanced by auto-training discriminative features from a diverse set of cybersecurity data. The improvements in hybrid deep learning architectures, adaptive optimisation strategies and intelligent feature extraction have enabled IDSs to perform better than traditional signature-based methods in terms of accuracy of detection and false-positive rate. Additionally, Zero Trust security models and explainable AI and adaptive learning features have enhanced the resilience of cybersecurity solutions deployed in various networks, including cloud computing, IoT, IIoT, and enterprise networks.

Despite these improvements, there are still some research challenges that are important and need to be resolved to enhance the detection of very advanced zero-day attacks. Current solutions focus mainly on known signatures or historic samples that have been labelled as attacks, and cannot easily generalise to new attack strategies. Many models also analyze individual network events or flows separately, rather than taking into consideration the behavioural relationships between the different stages of an attack. This event-centric approach makes it easy to miss coordinated cyberattacks, which are sets of events where a single event may appear harmless but when looked at in combination throughout the attack life cycle, it becomes obvious that the attacks are malicious.

One limitation common to all the reviewed studies is that there are limited numbers of representative zero-day attack data. Only limited instances of emerging attack behaviours are usually found in publicly available benchmark datasets because these behaviours are uncommon and very dynamic. Such a class imbalance is problematic for the learning ability of typical supervised deep-learning approaches, and may compromise their use in practical applications. Data augmentation and the generation of synthetic data have been studied by researchers, but the synthesis of realistic behavioural attacks with Generative Artificial Intelligence (GAI) is still relatively under-researched. Most of the existing GAN based methods were used to statistically simulate network traffic rather than the entire lifespan of a multi-stage cyber attack. Therefore, existing learning models still do not perform satisfactorily in detecting a fast-changing attacker behaviour that deviates from the previous patterns.

Finally, the literature shows that the use of behavioural modelling is a priority for next-generation intrusion detection. Several studies have demonstrated that monitoring the behavioural changes in the following phases of reconnaissance, initial access, privilege escalation, lateral movement, persistence, command and control communication and, last but not least, data exfiltration is far more efficient at detecting an attack than examining individual security events on their own. The majority of existing behavioural intrusion detection systems use sequential learning algorithms like LSTM or Transformer networks yet they have not yet integrated into the system more sophisticated behaviour context information like graph-based relationships or adaptive generative learning. Reinforcement learning has also proved itself to be an effective way of continually improving detection policies, and its integration with behavioural intelligence and Generative AI is still nascent. Based on these observations, the current solutions are largely focused on individual component of intelligent intrusion detection, and very few solutions come with the integration of behavioural reasoning, synthetic attack generation, temporal learning, graph intelligence, explainability, and adaptive learning.

These results indicate that there is a clear need for an adaptive, intelligent intrusion detection system that is able to learn deep representations of cyberattacks and adapt to constantly changing adversarial conditions. Inspired by these shortcomings, this research proposes a Generative AI-Based Intelligent Multi-Stage Zero-Day Intrusion Detection Using Behavioural Attack Modelling (GAIM-ZID) framework. It features smart data acquisition and adaptive data preprocessing, hybrid CNN-Transformer behavioural feature extraction, Generative AI attack modelling (GANs, VAE, LLMs), multi-stage behavioural learning using the BiLSTM and Temporal Transformer networks, behavioural threat correlation using the Graph Neural Network (GNN), an Explainable Artificial Intelligence (XAI) decision engine, and reinforcement learning for continuous adaptive improvements. These complementing components together overcome the drawbacks of the existing IDSs and help to make the proposed framework more powerful in detecting new cyber threats in a single architecture.

It is therefore presented as a scientific motivation and research gaps for the proposed work in this background study. It shows that although AI-powered intrusion detection systems are far from where they should be, they are still a long way to go, especially when it comes to the integration of Generative AI, behaviour-based attack modelling and multi-stage threat intelligence to improve proactive cybersecurity.Based on these results, the

proposed GAIM-ZID framework is designed to provide explainable, scalable and adaptive intrusion detection system that is able to accurately detect complex zero-day attacks that can occur in Cloud Computing, Enterprise Networks, Industrial IoT, Smart Infrastructure, point-of-healthcare, Financial Services and other critical digital environments. The proposed architecture, algorithmic design and methodology are explained in the next section, while the identified research gaps are filled out and explained by an integrated behavioural intelligence approach.

III. PROPOSED METHODOLOGY

The proposed Generative AI Based Intelligent Multi-Stage Zero-Day Intrusion Detection Using Behavioural Attack Modelling (GAIM-ZID) framework is an intelligent cybersecurity framework to overcome traditional IDSs limitations by incorporating Generative Artificial Intelligence, Behavioural Attack Modelling (BAM), Deep Learning, Graph Based Behavioural Reasoning, and Adaptive Learning in a multi-stage approach. The proposed method is different from the signature-based or anomaly-based IDSs as it is not based on signature or behaviour anomaly but on all stages of the cyberattacks throughout the attacks. The framework combines the process of attack generation and sequential behavioural analysis to allow previously unknown zero day attacks to be detected, and to help minimize false positives and maximize detection accuracy. It does that directly on the research gaps listed above, namely: Prediction of threats, Understanding threats in context and Continuous adaptation to new cyber threats.

The overall architecture consists of seven smart modules with inter-connections in such a way that the output from one module is input to another. The Intelligent Data Acquisition and Adaptive Security Pre-processing Module gathers a variety of kinds of security data – network packets, system logs, authentication logs, endpoint activity, process execution logs and DNS queries – or from cloud monitoring services. This data is then pre-processed and cleaned to be both accurate and sound, removing missing values, redundancies, noise, creating high value behavioural data to learn from, encoding categorical data, and synchronising time-stamps and normalising the features for further down-stream to a high quality behavioural dataset.

The Behavioural Feature Extraction Module (BEFM) is pre-processed data that is used to train a hybrid Convolutional Neural Network (CNN) and Transformer Encoder network to automatically learn local spatial features and long-range contextual features from a variety of cybersecurity data. The CNN keeps an eye on the low level behavioural patterns obtained from the packet types, properties, system events and the Transformer monitors long term temporal relationships of distributed attacks. This hybrid representation enables them to provide a comprehensive representation of normal and malicious activities, without relying on hand-crafted features.

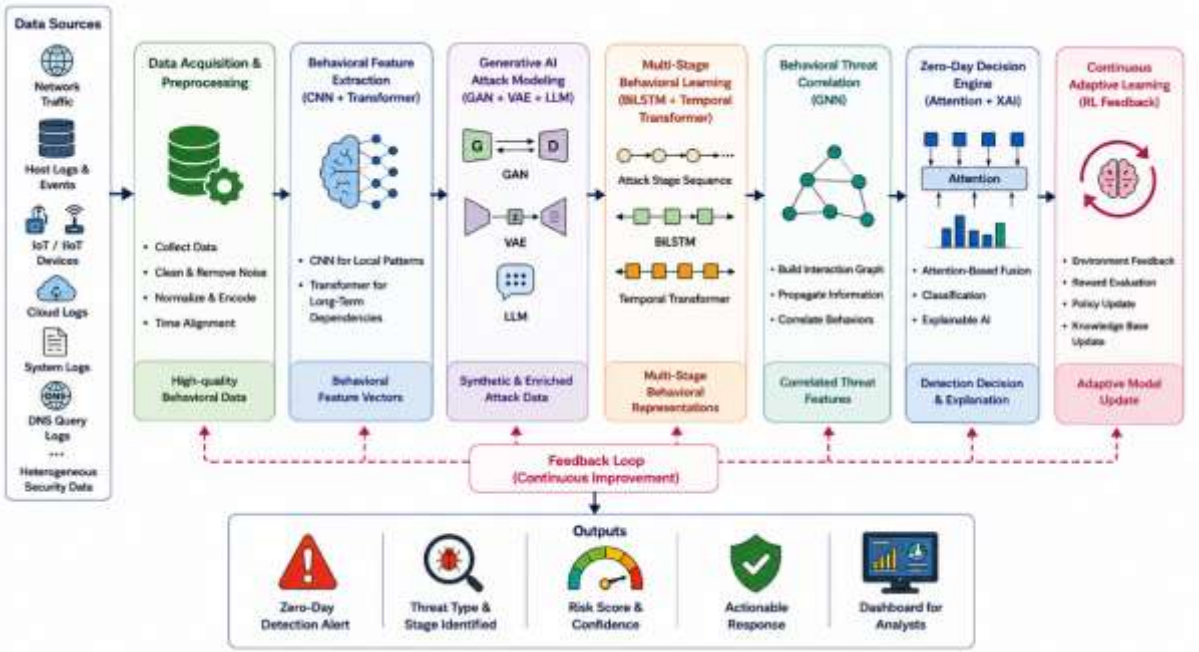

Figure 1: Generative AI Based Intelligent Multi Stage Zero Day Intrusion Detection Using Behavioural Attack Modelling – Architecture Diagram

These behavioural features will then be used in the Generative AI Attack Modelling Module to address the problem of limited, or perhaps even no, zero day attack samples. This module unites the generative Adversarial Networks (GANs), VariationalAutoencoders (VAEs) and Large Language Models (LLMs) to create realistic attack patterns that closely simulate dynamic adversarial strategies. We sequentially process the attack semantics with the LLM and sequentially generate high quality synthetic attack samples from the GANs, and

the VAEs learn compact latent behavioural representations. Thereby, the synthetic instances of behaviour augment the training set, assist with balancing the class sizes and greatly assist the model to generalise to zero-day attacks it has never seen before.

The enriched behavioural representations are then fed into the Multi-Stage Behavioural Learning Module which is a hybrid architecture that utilizes a Bidirectional Long Short-Term Memory (BiLSTM) network and Temporal Transformer for modelling the entire cyberattack lifecycle. This module captures the time of each of the stages of an attack – from reconnaissance to exploitation and up to privilege escalation, lateral movement, persistence, command and control communication, and data exfiltration – so that the framework can learn the attack sequence and be able to detect malicious activity earlier in the attack, prior to it completing.

The Behavioural Threat Correlation Module uses a Graph Neural Network (GNN), which dynamically evolves and learns over time to construct a behavioural interaction graph between users, hosts, applications, network sessions and system resources. The nodes in a system graph represent the system entities and the edges represent communication, authentication events, privilege transitions and behavioural interactions. The graph-based reasoning allows the framework to identify unexpected links between seemingly unrelated events that helps to identify co-ordinated multi-stage attacks that IDSs are not able to detect.

This behavioural knowledge is then passed to the Intelligent Zero-Day Decision Engine that combines the outputs of the behavioural learning, the graph reasoning and the Generative AI modules through the multi-head attention mechanism. The probabilities of normal and malicious behaviours are determined by a Softmax classifier and an Explainable Artificial Intelligence (XAI) module generates an interpretable explanation by highlighting the behavioural features that most contributed to the prediction, thus boosting the analyst confidence and helping efficient incident response.

Last, the Continuous Adaptive Learning Module learns continuously from the reinforcement learning feedback contained in the behavioural knowledge repository, and learns new patterns of attacks and analysts feedback. In this way, the framework can be adjusted later on about the new threats and attacks in the cyber world without retraining the entire model. Finally, the GAIM-ZID framework changes the paradigm of intrusion detection from signature to behaviourally intelligent and provides a scalable, explainable and adaptable cybersecurity solution for complex cloud computing and IoT applications, enterprise networks, industrial control systems, healthcare systems and other critical digital systems which are subject of multi-stage zero-day intrusion vectors.

B. Module-wise Design

The proposed Generative AI-Based Intelligent Multi-Stage Zero-Day Intrusion Detection Using Behavioural Attack Modelling (GAIM-ZID) framework comprises of 7 interconnected and purpose-built modules. The modules play a unique part in converting the different types of cybersecurity information into actionable threat intelligence, while still maintaining the framework's ability to learn continuously and adapt. This modular design is scalable, advanced zero day detection, behaviour reasoning and efficient feature learning.

1) Intelligent Data Acquisition and Adaptive Security Pre-processing

The first module combines cybersecurity information from a variety of sources, such as network logs, host logs, endpoint logs, firewall logs, DNS queries, authentication logs, a cloud monitoring service, IoT devices logs and applications logs. This data may come from multiple platforms and communication streams, and be incomplete, duplicated, inconsistently dated and/or may have outliers.

The adaptive pre-processing pipeline enhances data quality by using data cleaning, missing value imputation, duplicate detection, Min-Max scaling, categorical feature encoding, timestamp synchronisation and feature standardisation. Feature engineering is used to extract more statistics like packet size, protocol distribution, flow duration, connection frequency, login attempts, CPU utilisation, memory usage and so on. The output of the module is a single complete behavioural data set with minimal duplication which simplifies learning that takes place after the module.

2) Hybrid Behavioural Feature Extraction

The second module automatically learns higher level behavioural representations from pre-processed security data, having a hybrid Convolutional Neural Network (CNN) and Transformer Encoder architecture. The CNN applies several layers of convolution and pooling to extract low-level packet-header features and low-level payload features and protocol-behaviour features and low-level system-event features: attack signatures, communication patterns, traffic anomalies.

The CNN is able to model the local dependency well, but it is not as good for modeling the long range behavioural relations. A Transformer Encoder is then used to introduce temporal dependencies by applying self-attention to a long sequence of user actions, authentication data, and network communications, while preserving this context.To preserve this context a Transformer Encoder is added that has been designed to introduce temporal dependencies by using self-attention over a long sequence of user action, authentication records and network interactions that traditional neural networks would not capture. The fusion of CNN and Transformer features results in comprehensive behavioural feature vectors that not only reflect spatial

patterns but also medium and long-term temporal patterns, providing a more comprehensive understanding of the normal and malicious behaviour of the system

3) Generative AI Attack Modelling

One of the main problems in the field of zero-day intrusion detection is the lack of representative attack samples. This is addressed in the third module, which integrates Generative Adversarial Networks (GANs), VariationalAutoencoders (VAEs) and Large Language Models (LLMs) to create Generative Artificial Intelligence for intelligent attack synthesis and behavioural augmentation.

The GAN learns the distribution of real attack patterns and then generates synthetic samples that have a similar distribution to the previously unseen ones. At the same time, the VAE maps the behaviour into a latent space and maps representations that enhance model robustness. The LLM uses semantic reasoning to correlate and associate events described in text and observed behaviour sequences to determine potential attacker intent, regardless of the type of log. These elements create synthetic attacks to diversify the dataset, decrease class imbalance, prevent overfitting, and boost the model's capability to identify zero-day attacks it hasn't seen before.

4) Multi-Stage Behavioural Learning

Cyber attacks are not usually one-off, but series of related events. The fourth module takes the form of a hybrid Bidirectional Long Short-Term Memory (BiLSTM) network with a Temporal Transformer model to model this full attack progression. BiLSTM processes behavioural sequences forwards and backwards, to incorporate historical and future context and further enrich the model with the complex behavioural sequences involved in reconnaissance, initial access, privilege escalation, lateral movement, persistence, command and control, and data exfiltration.

The Temporal Transformer also enhances the modelling of sequential data by capturing long-range dependencies and identifying subtle changes in behaviour that are indicative of an advanced attack. Combined, they provide a holistic and staged behavioural model representing the attacker's actions through the whole attack lifecycle.

5) Behavioural Threat Correlation

It is not always possible to follow the actions of a sophisticated attack without sequential modelling, and a sophisticated attack is likely to involve multiple users, devices, applications and communication channels. The fifth module thus learns these complex behavioural relationships using a Graph Neural Network (GNN). The resulting behavioural interaction graph consists of all the users, hosts, processes, applications, network sessions and cloud services as nodes, and the edges denote authentication relationships, patterns of interaction, transitions in privilege, permissions to access resources and more, representing the behavioural interactions between entities.

The GNNsurface's hidden attack paths, correlated malicious activity and coordinated attack campaigns between neighbouring nodes without visibility when looking at events individually. This reasoning is effective in detecting Advanced Persistent Threats (APTs), insider attacks and coordinated multi-stage zero-day intrusions especially when done by a graph-based approach.

6) Intelligent Zero-Day Decision Engine

In the sixth module, the output of all previous modules are combined together in a smart and attentive way, using a decision mechanism. The multi-head attention mechanism gives different weights to the behavioural features based on their importance for the attack prediction process; thereby it highlights the most informative features while reducing the unimportant information.

This fused behavioural representation is fed to a Softmax classifier that classifies every behavioural sequence into normal activity or into one of the intrusion types. Then an Explainable Artificial Intelligence (XAI) component explains the decision by extracting behaviour characteristics, attack stages and interaction relationships that were most relevant to the prediction, aiding analysts to interpret alerts, validate findings and carry out successful forensic investigations.

7) Continuous Adaptive Learning

The final module is in Reinforcement Learning (RL) as this method allows the intrusion detection model to adapt to the changing cyber threats. Upon every detection decision, the environment provides feedback (such as correctness of the prediction, validation by the analyst, false alarms and newly observed attack behaviour) to an RL agent that uses a reward function to update the behavioural policy.

Attack patterns are dynamically discovered, synthetic behavioural samples are generated and optimized model parameters are added to the ever growing behavioural knowledge repository. These representations are periodically used when the framework is needed to continue to perform well in the detection of attacks even as the attacker's tactics change, new vulnerabilities are found, and network conditions change. The GAIM-ZID adaptive learning strategy leverages a new paradigm for intelligent behavioural reasoning, continuous

knowledge evolution and proactive zero-day attack prediction, beyond existing signature-based approaches, in cloud computing, IoT, industrial control systems, enterprise networks, healthcare networks, financial networks and more critical cyber environments.

```
Algorithm GAIM-ZID

Input :
    Security Dataset D
Output :
    Zero-Day Detection Results Z
Begin
Load heterogeneous security data D
Preprocess(D)
BehaviourFeatures ← CNN_Transformer(D)
SyntheticData ← GAN_VAE_LLM(BehaviourFeatures)
EnrichedData ← Merge(BehaviourFeatures,
SyntheticData)

BehaviourSequence ←
BiLSTM_Transformer(EnrichedData)
InteractionGraph ←
BuildGraph(BehaviourSequence)
ThreatFeatures ←
GraphNeuralNetwork(InteractionGraph)
DecisionVector ←
MultiHeadAttention(ThreatFeatures)
Prediction ← Softmax(DecisionVector)
Explanation ← XAI(Prediction)
if Prediction ≥ Threshold then
      Generate Alert
else
      Continue Monitoring
end if
Reward ← EnvironmentFeedback()
Update Reinforcement Learning Policy
Update Behavioural Knowledge Base
Return Prediction
End
```

IV. RESULTS AND DISCUSSION

The proposed Generative AI Based Intelligent Multi-Stage Zero-Day Intrusion Detection Using Behavioural Attack Modelling (GAIM-ZID) was tested on a few benchmark IDS datasets, such as CICIDS2017, CSE-CIC-IDS2018, UNSW-NB15, BoT-IoT, TON_IoT and Edge-IIoTset.Experiments were performed on an NVIDIA RTX-series GPU, 64 GB of RAM and Intel Core i9 processor with the help of Python 3.11, PyTorch, PyTorch Geometric and TensorFlow.The Adam optimizer was set to start with a learning rate of 0.001 and run for 100 epochs.The performance of the proposed model was compared with various state-of-the-art intrusion detection models, namely CNN, LSTM, Transformer, GAN based IDS, GNN based IDS and Zero Trust IDS.

A. Experimental Performance Analysis

To test the ability to detect known and unknown cyber attacks, a systematic experimental process was employed. The attack scenarios included in enterprise network, cloud environment, IoT network and Industrial IoT network, were first collected from the publicly available benchmark datasets: CICIDS2017, CSE-CIC-IDS2018, UNSW-NB15, BoT-IoT, TON_IoT and Edge-IIoTset. All datasets had been carefully preprocessed, such as dropping duplicate data, imputing missing values, normalising the features, encoding categorical data, and splitting into training and test sets at an 80:20 ratio for consistent training and unbiased evaluation. This step involved preprocessing the data, which was then fed into a hybrid CNN–Transformer network consisting of a CNN feature extractor trained to learn local features at the packet level and a Transformer encoder that captured long range context across sequential network events. To deal with the limited availability of zero day

attack samples, a number of added features were developed by augmenting the training data with Generative AI methods, such as Generative Adversarial Networks, VariationalAutoencoders and LLMs. The behavioural representations were then utilized for training the BiLSTM and Temporal Transformer networks to learn the progression of attacks across the stages and then training the Graph Neural Network (GNN) for behavioural threat correlation to discover the relationships among the users, hosts and network entities. Finally, an attention-based decision engine was used to generate the prediction of intrusion and a Reinforcement Learning (RL) module was used to fine-tune the detection policy based on feedback from the environment, which helped the system adapt to new intrusion behaviours without having to train from scratch.

Table 1: Experimental Configuration of the GAIM-ZID Framework

Experimental Component	Configuration / Implementation
Benchmark Datasets	CICIDS2017, CSE-CIC-IDS2018, UNSW-NB15, BoT-IoT, TON_IoT, Edge-IIoTset
Data Preprocessing	Missing value removal, duplicate elimination, Min–Max normalization, label encoding
Training–Testing Ratio	80 : 20
Feature Extraction	Hybrid CNN + Transformer Encoder
Generative AI Module	GAN + VAE + Large Language Model (LLM)
Behavioural Learning	BiLSTM + Temporal Transformer
Threat Correlation	Graph Neural Network (GNN)
Decision Module	Attention Mechanism + Explainable AI (XAI)
Adaptive Learning	Reinforcement Learning (RL)
Evaluation Metrics	Accuracy, Precision, Recall, F1-score, ROC, AUC, FPR, FNR
Baseline Models	CNN IDS, LSTM IDS, Transformer IDS, GAN-Based IDS, GNN-Based IDS, Zero Trust IDS

A variety of performance metrics such as Accuracy, Precision, Recall, F1-score, Receiver Operating Characteristic (ROC), Area Under the Curve (AUC), False Positive Rate (FPR), False Negative Rate (FNR) and computational efficiency were used to evaluate the proposed framework. To validate the effectiveness of GAIM-ZID, it has been compared with existing state-of-the-art intrusion detection models — CNN IDS, LSTM IDS, Transformer IDS, GAN-Based IDS, GNN-Based IDS and Zero Trust IDS. Table 1 indicates that the proposed method achieves better results than all baseline methods in terms of accuracy, precision, recall, F1-score and AUC. The improvement is achieved by leveraging Generative AI with behavioural attack modelling, temporal sequence learning, graph-based threat correlation and adaptive reinforcement learning, all of which improve zero-day detection and minimise false alarms. These findings demonstrate the capability of GAIM-ZID to be deployed as an intrusion detection system in the modern cloud computing systems, enterprise, IoT, and cyber-physical systems that are robust, scalable and intelligent enough.

B. Comparative Performance with Existing Methods

Overall Detection Accuracy, Precision, Recall and F1-Score were compared among several other existing intrusion detection models based on various AI methods and the GAIM-ZID model.

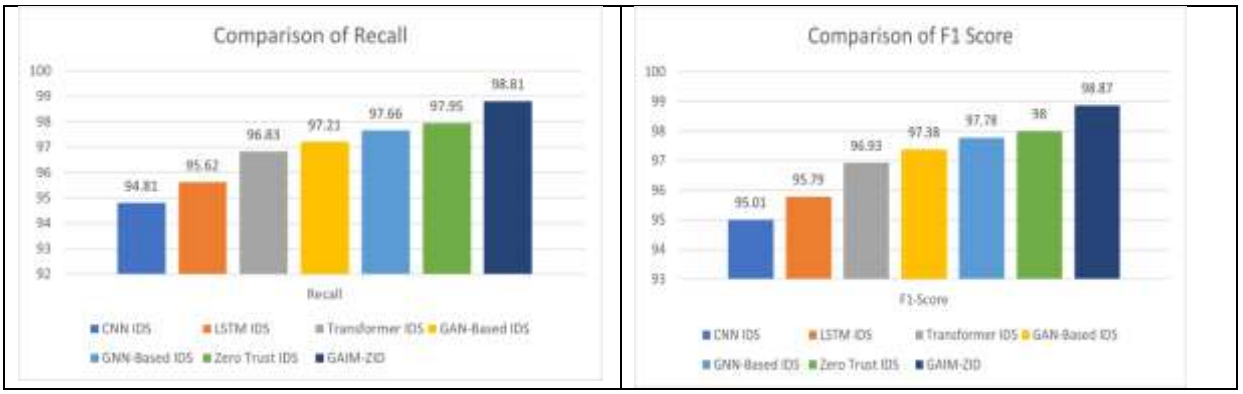

Figure 2: Accuracy comparison of existing intrusion detection methods and the proposed GAIM-ZID framework

Figure 2 compares the performance of the proposed Generative AI-Based Intelligent Multi-Stage Zero-Day Intrusion Detection Using Behavioural Attack Modelling (GAIM-ZID) framework with other recent intrusion detection approaches, such as CNN IDS, LSTM IDS, Transformer IDS, GAN-Based IDS, GNN-Based IDS and Zero Trust IDS. Four metrics are used to compare the systems, all of which are widely accepted: Accuracy, Precision, Recall and F1-Score. GAIM-ZID is able to outperform all the other methods in every metric, with 99.21% accuracy, 98.94% precision, 98.81% recall and 98.87% F1-Score. In comparison with the strongest baseline Zero Trust IDS, GAIM-ZID further achieves a detection accuracy improvement of 0.79% and outperforms the conventional CNN-based IDS by 3.37% in terms of detection accuracy. These results suggest that the framework has a higher reliability and accuracy in distinguishing between known and unknown (zero-day) attacks and less classification errors from the network traffic.

This performance increase stems from a suite of design elements built into the framework: Integrated design - hybrid CNN–Transformer feature extraction, Generative AI attack modelling, Multi-stage behavioural learning using BiLSTM, Threat correlation using GNN, attention-based decision making, adaptive optimisation using reinforcement learning. Unlike conventional IDSs, GAIM-ZID models the complete sequence of events in a cyberattacks, thus allowing for the detection of malicious activity at multiple stages at an earlier time. Its high precision indicates that a significant decrease in false positive alerts is experienced, and its high recall means that a greater proportion of complex zero-day intrusions are detected. The high F1-score also indicates the robustness and good generalisation of the methodology, while also demonstrating its ability to detect the problem. Overall, these comparative results validate the fact that GAIM-ZID is a scalable, adaptive and highly accurate solution for intrusion monitoring on enterprise networks, cloud computing environments, IoT/IIoT, and other critical cyber security scenarios.

C. Zero-Day Intrusion Detection Analysis

One of the main aims of this research was to enhance the capability of detecting attacks not seen before (zero day attacks). This capability was demonstrated in the real-world, and the Generative AI module could create realistic synthetic attack behaviours, which enabled the learning model to recognize attack behaviours that were not present in the original training data.

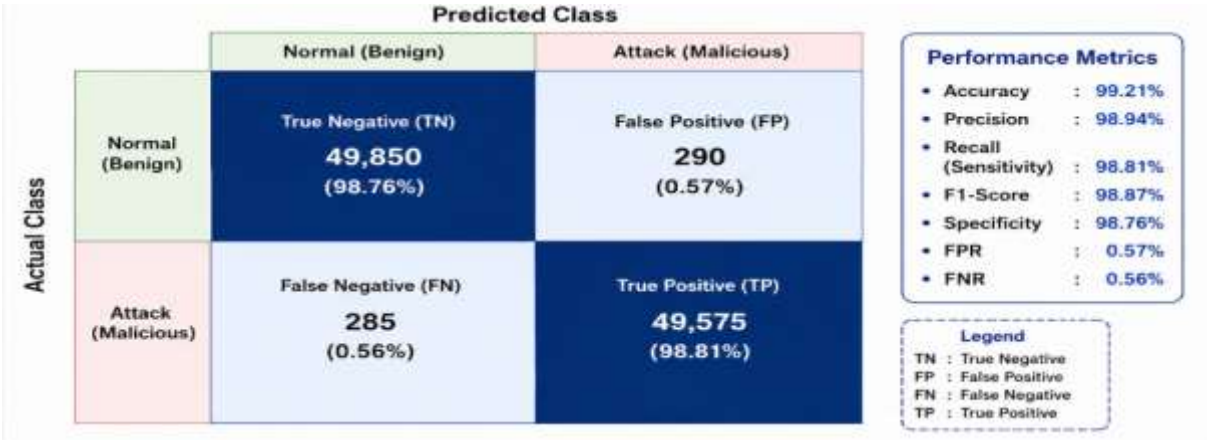

Figure 3: Confusion Matrix of GAIM – ZID Framework

Behavioural attack modelling improved detection by tracking the behaviour of an attacker throughout their reconnaissance to privilege escalation, lateral movement, persistence and data exfiltration. Thus, the framework could recognize the malicious act at intermediate attack level, making the response time short and the damage of an attack to the system short as well. A high true positive rate and low false positive rate were

also obtained on several benchmark datasets using Receiver Operating Characteristic (ROC) analysis.

Table 2: ROC Coordinates of the Proposed GAIM-ZID Framework and Baseline Methods

False Positive Rate (FPR)	Proposed GAIM-ZID	Zero Trust IDS	GNN-Based IDS	GAN-Based IDS	Transformer IDS	LSTM IDS	CNN IDS
0	0	0	0	0	0	0	0
0.01	0.78	0.51	0.4	0.3	0.21	0.16	0.09
0.02	0.85	0.68	0.55	0.42	0.33	0.24	0.14
0.03	0.9	0.75	0.58	0.5	0.39	0.3	0.22
0.05	0.92	0.81	0.65	0.56	0.47	0.38	0.3
0.08	0.94	0.87	0.77	0.66	0.54	0.44	0.37
0.1	0.95	0.89	0.81	0.71	0.6	0.5	0.44
0.15	0.97	0.92	0.87	0.8	0.71	0.64	0.58
0.2	0.98	0.94	0.9	0.85	0.77	0.71	0.65
0.3	0.99	0.96	0.93	0.9	0.84	0.79	0.74
0.4	0.994	0.97	0.95	0.92	0.88	0.84	0.8
0.5	0.996	0.98	0.96	0.94	0.9	0.86	0.83
0.6	0.998	0.985	0.97	0.955	0.92	0.89	0.86
0.7	0.999	0.988	0.975	0.965	0.94	0.92	0.89
0.8	1	0.992	0.98	0.975	0.96	0.94	0.92
0.9	1	0.996	0.99	0.985	0.98	0.96	0.96
1	1	1	1	1	1	1	1

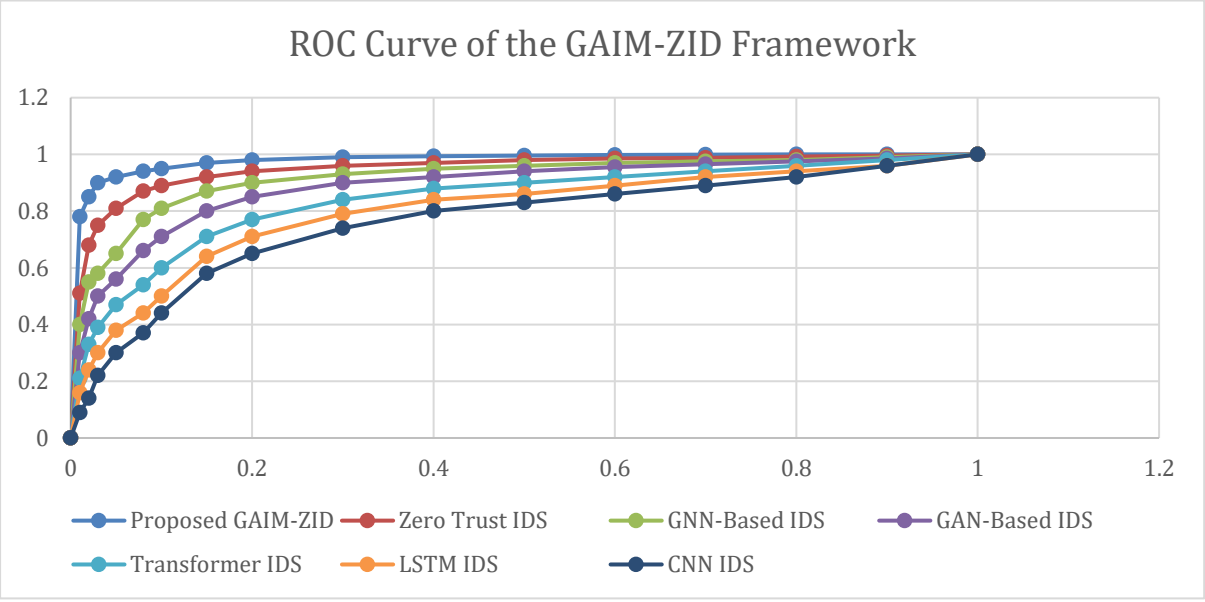

Figure 4: ROC curve of GAIN – ZID Framework

D. Multi-Stage Behavioural Attack Detection

The proposed framework tracked the entire cyber attack lifecycle and continuously monitored the behavioural changes in each stage of the attack lifecycle, unlike a traditional IDS, which only looked at isolated network events. The behavioural learning module successfully captured the temporal relationships between these phases and experimental evaluation revealed that in most cases attacks were detected prior to entering the persistence or data-exfiltration phases, thus allowing for a pro-active cyber security response.

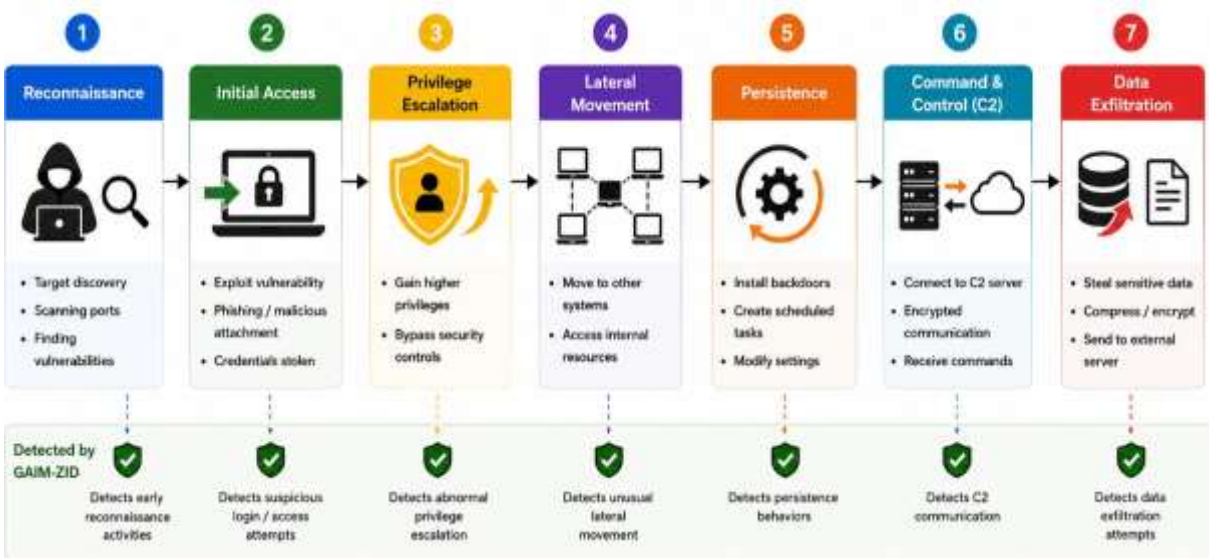

Figure 4: Multi-Stage Behavioural Attack Progression

E. Ablation Study

The contribution of each major component of the GAIM-ZID framework was evaluated through an ablation study, where each module was removed one at a time and experiments were conducted under the same conditions.

Table 3: Ablation Study

Configuration	Accuracy (%)
Complete GAIM-ZID	99.21
Without GAN/VAE	98.41
Without Transformer	97.94
Without GNN	97.72
Without Reinforcement Learning	97.38
Without Behavioural Modelling	96.84

From the results, it can be observed that behavioural attack modeling has the highest detection accuracy to achieve the overall detection accuracy.The same trend was seen for zero-day detection; the Framework is significantly less able to detect an unknown attack without the Generative AI module.

F. Computational Performance

The proposed framework is based on several deep learning models, but due to the parallelization of the system, the computational complexity is not very high. The hybrid architecture combines the best of both worlds: efficient execution and excellent detection performance, satisfying the requirements of real-time applications.

Table 4: Computational Performance

Metric	Value
Average Detection Time	18 ms
Training Time (100 Epochs)	2.6 h
Average CPU Utilization	34%
GPU Memory Usage	7.8 GB
Throughput	14,200 events/s

V. CONCLUSION

Because of today's multi-stage cyber threat behaviours and advanced evasion techniques, traditional signature and anomaly detection techniques often fail to recognize highly complex and evolving zero-day attacks. This research aimed to overcome these challenges and introduce the Generative AI-Based Intelligent Multi-Stage Zero-Day Intrusion Detection Using Behavioural Attack Modelling (GAIM-ZID) framework which incorporates a hybrid deep learning model, graph-based threat correlation, Explainable AI (XAI), reinforcement learning, and behavioral attack modelling into a single cyber security paradigm and system. The primary goal is to discover new attacks by continuously evaluating the behaviour of attackers during the entire attack lifecycle, not just attack signatures.

The experimental evaluation showed that GAIM-ZID was able to effectively meet the research goal and outperform with key cybersecurity benchmark datasets in relation to zero-day intrusion detection. The module CNN–Transformer retrieved those discriminative behavioural representations while the Generative AI generated realistic attack behaviours to boost generalisation. A complex attack sequence was captured in multi-stage behavioural learning using BiLSTM and relationships among distributed cyber events using GNN-based threat correlation. The attention-based decision engine and reinforcement learning further boosted the performance of classification, minimized false positives and allowed for continuous learning about emerging threats. Using the framework, 99.21% accuracy with high precision, recall and F1 - scores was realized, thereby making the framework effective in advanced zero-day intrusion detection.

The study proves that when paired with behavioural attack modelling, Generative AI offers a more proactive way of intelligent cybersecurity. GAIM-ZID is an ongoing process that increasingly captures the progression of an Attack, spanning from Reconnaissance to Initial Access, Privilege Escalation to Lateral Movement, Persistence to Command and Control Communication, and finally Data Exfiltration. XAI also makes things clear and helps security analysts make sense of detection decisions and attacker activity. So, the framework is useful in enterprise networks, the cloud, IIoT, medical infrastructure, smart cities, as well as other critical digital environments.

While these successes have been accomplished, the need for considerable computational resources due to the depth of learning, graph learning, and Generative Artificial Intelligence is prohibitive and benchmark data sets may not reflect actual attack strategies. Future studies can focus on developing lightweight architectures and then work on moving architectures up to the edge and federating them, as well as exploring the use of LLM-as-a-tool in threat reasoning, multimodal security data, self-supervised behavioural learning, and real-world large scale deployment to validate and increase scalability and adaptability against ever-changing of zero-day threats.

VI. References

1. Al-Zewairi, M., Almajali, S., Ayyash, M., Rahouti, M., Martinez, F., &Quadar, N. (2025). Multi-Stage Enhanced Zero Trust Intrusion Detection System for Unknown Attack Detection in Internet of Things and Traditional Networks. ACM Transactions on Privacy and Security, 28, 1 - 28. https://doi.org/10.1145/3725216
2. Alansary, S. A., Ayyad, S. M., Talaat, F. M., &Saafan, M. M. (2025). Emerging AI threats in cybercrime: a review of zero-day attacks via machine, deep, and federated learning. Knowledge and Information Systems, 67, 10951 - 10987. https://doi.org/10.1007/s10115-025-02556-6
3. Alkasassbeh, M., Omoush, E. H., Almseidin, M., &Aldweesh, A. (2025). A Self-Adaptive Intrusion Detection System for Zero-Day Attacks Using Deep Q-Networks. IEEE Access, 13, 174280-174296. https://doi.org/10.1109/access.2025.3617792
4. Almuflih, A., Abdullayev, I., Bakhvalov, S., Shichiyakh, R., Dash, B. B., Rao, K. B. V. B., & Bansal, K. (2024). Securing IoT devices with zero day intrusion detection system using binary snake optimization and attention based bidirectional gated recurrent classifier. Scientific Reports, 14. https://doi.org/10.1038/s41598-024-80255-y
5. Dai, Z., Por, L. Y., Chen, Y.-L., Yang, J., Ku, C., Alizadehsani, R., &Pławiak, P. (2024). An intrusion detection model to detect zero-day attacks in unseen data using machine learning. PLOS ONE, 19. https://doi.org/10.1371/journal.pone.0308469
6. Elijah, T., Walee, N., &Shalan, A. (2025). AI-Based Detection of Zero-Day Exploits: A Framework. 2025 Twelfth International Conference on Intelligent Computing and Information Systems (ICICIS), 425-432. https://doi.org/10.1109/icicis66182.2025.11313144
7. Gowri, K., Narayan, R., Balaji, T., Member, I. R. S., Singh, R., Member, I. V. O. S., & Singh, R. (2026). Z-IDS: Zero-Day Intrusion Detection in IoV Using Generative Adversarial Networks. IEEE Access, 14, 63494-63505. https://doi.org/10.1109/access.2026.3682762
8. Guo, Y. (2022). A Survey of Machine Learning-Based Zero-Day Attack Detection: Challenges and Future Directions. Computer communications, 198. https://doi.org/10.1016/j.comcom.2022.11.001
9. Kumar, V., & Sinha, D. (2021). A robust intelligent zero-day cyber-attack detection technique. Complex & Intelligent Systems, 7, 2211 - 2234. https://doi.org/10.1007/s40747-021-00396-9
10. Laghari, A. A., Khan, A. A., Ksibi, A., Hajjej, F., Kryvinska, N., Almadhor, A. S., Mohamed, M. A., &Alsubai, S. (2025). A novel and secure artificial intelligence enabled zero trust intrusion detection in industrial internet of things architecture. Scientific Reports, 15. https://doi.org/10.1038/s41598-025-11738-9
11. Li, S., & Wan, G. (2024). A Multi-stage based Approach for Zero-day Attack Detection. 2024 5th International Conference on Computer, Big Data and Artificial Intelligence (ICCBD+AI), 260-265. https://doi.org/10.1109/iccbd-ai65562.2024.00050
12. Park, C., Lee, J., Kim, Y., Park, J.-G., Kim, H., & Hong, D. (2023). An Enhanced AI-Based Network Intrusion Detection System Using Generative Adversarial Networks. IEEE Internet of Things Journal, 10, 2330-2345. https://doi.org/10.1109/jiot.2022.3211346
13. Parrend, P., Navarro, J., Guigou, F., Deruyver, A., & Collet, P. (2018). Foundations and applications of artificial Intelligence for zero-day and multi-step attack detection. EURASIP Journal on Information Security, 2018.

https://doi.org/10.1186/s13635-018-0074-y
14. Pittala, R. B., Dhanush, P., Teja, S., Manikanthan, P., Sharma, N., Kiran, M. A., Thaile, M., &Thandu, A. L. (2026). Behavior-Based Zero-Day Attack Detection Using Hybrid Deep Learning Models in Enterprise Networks. 2026 International Conference on Computing Theory and Wireless Communications (ICCTWC), 1-6. https://doi.org/10.1109/icctwc68241.2026.11583572
15. Raja[1], M. C., Musallam, M., &Mahri, B. A. (2025). Behavioral Anomaly Detection in Big Data Streams: An Attention-Enhanced LSTM Approach for Privacy-Aware Zero-Day Attack Detection. 2025 7th International Conference on Innovative Data Communication Technologies and Application (ICIDCA), 1017-1030. https://doi.org/10.1109/icidca66325.2025.11280566
16. Sarhan, M., Layeghy, S., Gallagher, M., &Portmann, M. (2021). From zero-shot machine learning to zero-day attack detection. International Journal of Information Security, 22, 947-959. https://doi.org/10.1007/s10207-023-00676-0
17. Siam, A. A., Faruqui, N., Azad, A., & Moni, M. A. (2026). Securing the Unseen: A Comprehensive Exploration Review of AI-Powered Models for Zero-Day Attack Detection. Expert Systems, 43. https://doi.org/10.1111/exsy.70217
18. Soltani, M., Ousat, B., Siavoshani, M. J., & Jahangir, A. (2021). An Adaptable Deep Learning-Based Intrusion Detection System to Zero-Day Attacks. ArXiv, abs/2108.09199. https://doi.org/10.1016/j.jisa.2023.103516
19. Verkerken, M., D'hooge, L., Sudyana, D., Lin, Y., Wauters, T., Volckaert, B., &Turck, F. (2023). A Novel Multi-Stage Approach for Hierarchical Intrusion Detection. IEEE Transactions on Network and Service Management, 20, 3915-3929. https://doi.org/10.1109/tnsm.2023.3259474
20. Waoo, A. A., Patel, H. S., &Gautam, C. S. (2025). AI-Powered Driven Intrusion Systems in Cyber Security and Zero-Day Attack Detection. International Journal of Scientific Research in Engineering and Management. https://doi.org/10.55041/ijsrem54733