AKSU Journal of Engineering and Technology, **2026**
Vol. 1, No. 1, Pages 17-33
Published Online April 2026 in AKSUJET (http://www.aksujet.org) ISBN: 978-978-53006-97

A comprehensive analysis of multi-stage cyber-attack detection and prevention Using Machine Learning approaches: A review.

Akpasam J. Ekanem[1]
Department of Electrical and Electronic Engineering,
Akwa Ibom State University, Ikot Akpaden.
akpasamekanem@aksu.edu.ng

Okon Nsa Ufot[2]
Department of Computer Engineering Technology,
Akwa Ibom State Polytechnic, Ikot osurua,
ufot.okon@akwaibompoly.edu.ng

Abstract
The increasing sophistication of cyber threats has made multi-stage cyber-attacks a critical concern for modern networked systems. Unlike single-stage attacks, multi-stage attacks involve a sequence of coordinated actions that evolve over time, making their detection significantly more complex. This paper presents a comprehensive review of Machine Learning-based approaches for the detection and prevention of multi-stage cyber-attacks. The study categorizes existing techniques into supervised, unsupervised, semi-supervised, and reinforcement learning paradigms, and evaluates their effectiveness based on detection accuracy, adaptability, and computational efficiency. Furthermore, this review highlights key challenges such as data imbalance, lack of standardized datasets, high false positive rates, and limitations in detecting zero-day attacks. A comparative analysis of existing models is presented to identify research gaps and performance trade-offs. Finally, future research directions are proposed, emphasizing the need for hybrid intelligent systems, improved datasets, and advanced learning frameworks capable of handling evolving attack patterns. This study provides valuable insights for researchers and practitioners aiming to design robust and scalable cyber defense systems.

1.**Introduction**

Cybersecurity is a major concern for countless organizations, institutions, corporations, and individuals globally. Buczak and Guven [1] provide a precise definition of cyber security as the encompassing array of technologies and methodologies employed to oversee and thwart unauthorized entry, modification, abuse, and disruption of computer networks and resources. This also includes the ability to control and authorize access to sensitive information and critical infrastructure that can be reached through a network. Most networks are highly interconnected through the Internet, enabling the interchange of data, information, knowledge, software, and hardware. The computer networking paradigm has enabled the exchange of crucial resources to enhance operational efficiency. However, it has also facilitated the widespread dissemination of malware, resulting in an increase in cyber-attacks in the digital domain.

The expansion of potential risks is a consequence of the growing impact of cyber capabilities, which are gradually penetrating and impacting all parts of home, commercial, and industrial processes. Akyazi in [2] asserts that cyber-attacks pose a risk by virtue of their ability to modify system or database parameters, which can have a kinetic effect that may increase the attacks and perhaps result in the destruction of confidential information.

To protect against cyber-attacks, it is essential to employ both proactive and reactive tactics. The strategies, known as active and passive, are relevant in the given context of application. They encompass proactive defense measures or mitigation strategies against cyber threats. Denning [3] argues that the significance of cyber defense measures resides in their capacity to efficiently counteract both active and passive threats, which have become widespread in the cyber realm.

The differentiation between single-stage attack detection and multi-stage attack detection is substantial. A single-stage attack effectively exploits the target system by carrying out simple and indiscriminate attack attempts within a short period of time. Nevertheless, because of the repetitive and indiscriminate manner in which the single-stage assault is carried out, it rapidly creates a track of evidence that can be readily identified by most inline or endpoint protection systems. Multi-stage attacks differ from single-stage assault detection in that they employ a more intricate and protracted offensive approach when compared to single-stage attacks. For example, to circumvent the standard security measures, the time intervals of the multi-stage attack range from a few minutes to several months. To properly detect and respond to the multi-stage attack, the administrator must carefully monitor and correlate individual attack alerts from different machines and attack scenarios, even if identifying each specific attack stages are not very difficult. Hence, identifying multi-stage attacks is extremely difficult without previous awareness of such instances.

Understanding the research gaps in current cyber security approaches is crucial. The present study will analyse the Machine Learning approaches available in the public realm, clarifying the advantages and disadvantages of each approach. The next parts will analyse cyber security strategies in terms of detection, prevention, and challenges.

Attack detection and prevention can be achieved using a range of methods, such as Machine Learning and evolutionary algorithms,

AKSU Journal of Engineering and Technology, **2026**
Vol. 1, No. 1, Pages 17-33
Published Online April 2026 in AKSUJET (http://www.aksujet.org) ISBN: 978-978-53006-97

statistical techniques, association rules, similarity-based approaches, causal correlation, structural-based approaches, case-based approaches and mixed approaches [4,5,6]. Similarly, most strategies employed to thwart attacks include scrutinizing network data to identify and eradicate (or limit) malicious activities. This study will focus on Machine Learning approaches of multi-stage. Figure 1 depicts this.

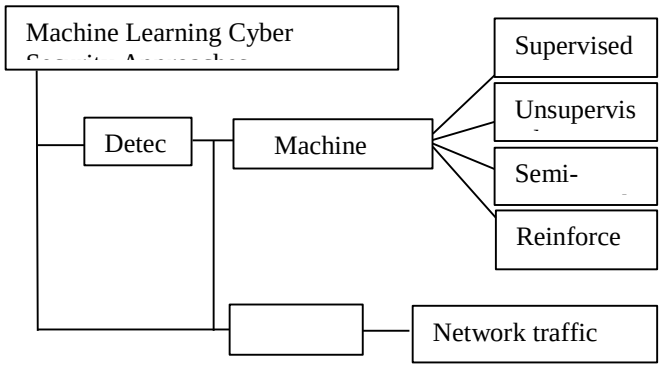

Figure 1: Machine Learning cyber security approaches

2. Methodology of Review

This study adopts a structured literature review methodology to analyze existing research on multi-stage cyber-attack detection and prevention using machine learning approaches.

2.1 Data Sources and Search Strategy

Relevant literature was collected from reputable academic databases, including IEEE Xplore, SpringerLink, ScienceDirect, and Google Scholar. These sources were selected due to their extensive coverage of peer-reviewed journals and conference proceedings in cybersecurity and machine learning.

A combination of keywords was used to retrieve relevant studies, including *"multi-stage cyber-attacks," "intrusion detection systems," "machine learning in cybersecurity," "anomaly detection,"* and *"zero-day attacks."* Boolean operators such as AND and OR were used to refine the search results.

2.2 Inclusion and Exclusion Criteria

To ensure the quality and relevance of the reviewed studies, the following inclusion criteria were applied:

- Studies focusing on multi-stage cyber-attack detection
- Research involving machine learning or hybrid approaches
- Peer-reviewed journal articles and conference papers
- Studies with clearly defined methodologies and evaluation metrics

The exclusion criteria included:

- Non-peer-reviewed articles and grey literature
- Studies not related to cybersecurity or intrusion detection
- Papers lacking sufficient experimental or methodological detail

2.3 Time Range of Selected Studies

The review focuses on studies published between 2013 and 2023, capturing recent advancements in machine learning and cybersecurity. Earlier foundational works were also included where necessary to provide theoretical background.

2.4 Data Extraction and Analysis

Selected studies were carefully analyzed based on key parameters, including methodology, dataset used, performance metrics, contributions, and identified limitations. The extracted information was synthesized to compare different approaches and highlight research gaps in multi-stage attack detection.

2.5 Limitations of the Review Methodology

Although this review follows a structured approach, it is limited by the availability of publicly accessible datasets and published studies. Additionally, variations in evaluation metrics and experimental setups across different studies make direct comparison challenging.

3. Review of Existing Literature

This section focuses on the detection of multi-stage cyber-attacks using machine learning approaches.

3.0.1 Detection using Machine Learning Algorithms

Machine Learning methods have been increasingly popular in recent years for identifying cyber-attacks. Machine Learning is a powerful tool for examining data and making accurate predictions about the outcomes of certain events. It accomplishes this by employing sample inputs to create an appropriate model that enables precise decision-making [1]. Classifying and predicting the presence or absence of a given sample using

AKSU Journal of Engineering and Technology, **2026**
Vol. 1, No. 1, Pages 17-33
Published Online April 2026 in AKSUJET (http://www.aksujet.org) ISBN: 978-978-53006-97

training data is the primary objective of Machine Learning algorithms. The application of Machine Learning in the current context of cyber-attack detection has greatly improved the precision of the detection process.

This study examines four discrete Machine Learning methodologies. The strategies include supervised, unsupervised, semi-supervised, and reinforcement learning.

i. **Supervised Learning**

Supervised or supervised learning refers to an educational method that incorporates the active involvement and guidance of a supervisor or mentor. Supervised learning is a subset of pattern recognition that employs a set of labelled instances, referred to as training data, together with their corresponding desired output. During the training phase, a predictive model is constructed using the labelled cases to classify fresh datasets. This is achieved by feeding the cases that have been assigned labels into a designated Machine Learning algorithm. Machine Learning techniques covered in [1] include Artificial Neural Networks, K-Nearest Neighbour (KNN), Support Vector Machines (SVM), Hidden Markov Models (HMM), Decision Trees and Naïve Bayes.

a. Decision Trees: This algorithm categorizes events based on the values of the feature. Each feature in a classified event is represented by a node in the model, and the branches represent the possible values for that feature. Events are categorized beginning from the top node and organized according to their feature values. At each level of the decision tree, the algorithm selects the feature that best separates the events into subclasses, using methods such as entropy or information gain [7].

b. **Support Vector Machine (SVM):** The SVM translates data into higher dimensions and identifies the optimum hyper-plane for data separation. It employs the concept of margin to determine the maximum margin of the dataset. The concept of a "margin" with carrying sides that separate two data classes is central to SVMs. It increases the margin, resulting in the greatest feasible separation between separating hyper-planes [7].

c. Artificial Neural Network: Neurons are used to create a neural network that represents the human brain [8]. It has hidden layers that can perform data processing and transfer the output to the output layer. A pattern that is generated from the data is used to train the Neural Network (NN). The output is retrieved and checked, and if it is correct, the next pattern is supplied as input. When a mistake occurs, there is an error that is propagated backward to the input layer (back propagation algorithm), the weights are adjusted to obtain the output for all training patterns [9]. Some of the advantages of Artificial Neural Network includes the ability to comprehend and stimulate complex and nonlinear relationships, ability to generalize the model and anticipate data that have yet to be observed and partially resistant to harm [10]. However, due to the enormous number of parameters to be set, optimizing the network might be difficult and large neural networks require significant computational time.

d. Random Forest: this classifier is an ensemble classifier that makes predictions. A portion of the training dataset is selected via substitution to construct trees (a bagging approach). In other words, certain samples can be used several times, and others may never be chosen. The model requires defining the number of decision trees (Ntrees) and the number of variables to be selected for optimal splitting. [11, 12]. The strength of Random Forest is that it enhances the accuracy of classification and works well with datasets with a large number of input variables [10]. However, Random Forest is quick to train, but it is slow to make predictions once it has been trained. Also, the evaluation is time consuming and interpretation is difficult [10].

e. XGBoost: this model is based on a decision tree ensemble Machine Learning system that applies a gradient boost algorithm to a known dataset before classifying it. The execution speed and the model's execution are two of the benefits of using XGBoost. When XGBoost is compared with other gradient-boosting implementations, the results reveal that it is quicker [13].

f. **K-Nearest Neighbour:** The K-nearest neighbor technique works by classifying fresh data utilizing previously classified data. A test sample refers to data that is uncertain about its class, while a training sample is data that has already been categorized. KNN approach determines the test sample and the training sample distances and it selects the k-nearest training samples that are most similar to the test sample. If the majority of these selected k samples belong to a specific class, the test sample is assigned the same. [14]. k-Nearest-Neighbour is easier to deploy and training is completed more quickly. The disadvantage of the K-Nearest Neighbour is that finding the nearest neighbor in the training data, which is massive, takes a long time, making it slow. Additionally, it requires large storage and lacks transparency in the representation of knowledge.

g. **Naïve Bayes**: The model is a statistically driven classification system that assigns labels to data. It is commonly employed in categorization tasks due to its simplicity. In Bayesian classification, the goal is to compute the conditional probability of the class to which the data belongs, aiming to estimate the probability of a class based on the provided data [15]. Naïve Bayes is simple to use and the outcome is more accurate due to the higher probability value but there is a strong assumption about the form of data distribution and there is a loss of precision.

h. **Logistic Regression (LR):** This model determines the link between a large number of variables that are either independent or dependent. LR is now used in social science because of the inefficiency of the least squats method (LSM) in a multivariate model with discrimination of dependent and independent variables. In LR, prediction is based on the chances associated with the two possible values of the dependent variable [15]. In logistic Regression, continuous outcomes are impossible to forecast, it is susceptible to overconfidence (i.e., the models may

AKSU Journal of Engineering and Technology, **2026**
Vol. 1, No. 1, Pages 17-33
Published Online April 2026 in AKSUJET (http://www.aksujet.org) ISBN: 978-978-53006-97

appear to have greater predictive potential than they do, leading to overfitting). Also, to get consistent findings, a large sample size is needed [10].

Osarumwense et al. [16] developed a probabilistic inference method to anticipate multi-stage attacks originating from malicious IP addresses, employing a supervised Machine Learning technique referred to as a Causal Network, or Bayesian Belief Network (BBN) Model. This approach forecasts multi-stage attacks by utilizing a joint probability density function, facilitating the creation of a Bayesian attack graph. This graph assigns particular probabilistic values to each attack and attack mode, enabling the calculation of the likelihood of both single-stage and multi-stage attacks. This model is engineered for use in computer network infrastructures, offering essential information to bolster network protection through enhanced prediction and detection of multi-stage attacks, blacklisting of malicious IPs, and overall improvement of network security.
The system exhibits considerable progress in multi-stage attacks prediction and detection; yet, it possesses limitations. It fails to anticipate multi-stage attacks employing MAC addresses or devices utilizing VPNs, it cannot also detect zero-day attacks, potentially creating deficiencies in its comprehensive security coverage. The Bayesian Belief Network model serves as an essential instrument in enhancing defenses against progressively intricate cyber dangers.

Verkerken et al. [17] have suggested a revolutionary multi-stage approach for hierarchical intrusion detection, demonstrating substantial progress in network security. The experimental solution was meticulously assessed utilizing the CIC-IDS-2017 and CSE-CIC-IDS-2018 network intrusion datasets, exhibiting significant adaptability without necessitating the retention of any classifiers. This adaptability facilitates an n-tier deployment approach, significantly minimizing bandwidth and processing demands while preserving the ability to identify zero-day assaults.

This method prioritizes the preservation of privacy throughout the training and operational phases of hierarchical deployment. This aspect is becoming increasingly essential as firms strive to protect sensitive data while executing effective security protocols. The results demonstrate that several current models trained on network Intrusion Detection System (IDS) datasets frequently exhibit insufficient generalization capabilities, hence constraining their efficacy in practical applications.
The experimental execution of this multi-stage methodology depends exclusively on network characteristics as input; nevertheless, the proposed architecture is sufficiently adaptable for application in host-based and hybrid IDS environments. This versatility is essential for firms pursuing complete security solutions in diverse situations.
The innovative multi-stage method adeptly reconciles the trade-off between attaining high recall for zero-day attacks, minimizing bandwidth consumption, and preserving classification efficacy. The outcomes are remarkable, attaining a weighted F1 score of 0.9875 and a balanced accuracy score of 0.9342. The advanced method for multi-stage intrusion detection achieved scores of 0.9383 and 0.8550, underscoring the superior efficacy of Verkerken et al.'s approach. This research highlights the potential for novel strategies to improve intrusion detection systems, especially in responding to rising threats while maximizing resource efficiency.

The study by [18] presents a new Kill Chain State Machine (KCSM) aimed at improving the detection of multi-stage attacks through the analysis of clustered alert data. This novel approach markedly decreases the quantity of warnings by means of efficient alert correlation and attack contextualization, which is essential in cybersecurity, where analysts frequently encounter an excessive volume of notifications.
Analysts may now triage events using multi-stage scenario graphs produced by the KCSM algorithm, rather than sifting through hundreds of thousands of singleton alerts. The system uses correlated meta-alerts in conjunction with unclustered single alerts to create Advanced Persistent Threat (APT) scenario graphs, offering critical context about potential multi-stage attacks within the network. This contextualization allows analysts to concentrate on the most pertinent warnings, optimizing the incident response process.
The algorithm exhibits significant efficacy in reducing alerts. In a simple configuration, it produced 642 alerts from an original total of 12,735 alerts over a ten-day span, resulting in a drop of merely 5.04%. In a high-security configuration, the system diminished an astounding 446,458 alerts to merely 700, achieving a notable reduction of 0.16%. A reduction of two to three orders of magnitude in alerts results in a manageable amount for human analysts, hence enhancing the efficiency of threat identification and response.
The KCSM approach's primary feature is its flexibility, as it formulates stage deductions based on network direction without necessitating specific information inherent in the underlying alerts. This functionality enables the algorithm to be widely adaptable across diverse network-based alert systems.
Furthermore, it is capable of processing additional stage-specific warnings, integrating the results into the APT scenario graphs to enhance the contextual comprehension of the attack.
Nonetheless, it is important to highlight that the algorithm presently handles just network-level information and is incapable of including host-level and user identity circumstances. This constraint may limit the complexity of the created scenario graphs. The viability and efficacy of the KCSM technique were assessed through various experiments utilizing the CSE-CIC-IDS2018 dataset, validating its promise as a robust instrument for enhancing the analysis and response to multi-stage cyber threats.

ii. Methodology for Unsupervised Learning

Unsupervised learning entails the detection of patterns in a

AKSU Journal of Engineering and Technology, **2026**
Vol. 1, No. 1, Pages 17-33
Published Online April 2026 in AKSUJET (http://www.aksujet.org) ISBN: 978-978-53006-97

dataset that lacks any form of labelling. Subsequently, these patterns are employed to provide precise classification determinations for novel occurrences. Usually, this technique involves using clusters to identify the categories that examples belong to. In their study, Song et al. [19] investigated an anomaly detection system that employed unsupervised learning to automatically adapt and optimize parameter values. This was done to enhance the system's ability to classify events as either attack strings or normal connections.

The suggested methodology conducts instance classification following the training phase, which encompasses activities such as filtering, clustering, and modelling. Filtering is employed to extract the essential subset of regular data, which is then partitioned into k clusters. The k clusters represent common patterns observed in the traffic data, including HTTP, SMTP, and FTP. Every typical cluster undergoes the one-class SVM algorithm, which produces k-SVM models, also referred to as k-hyper-spheres, for classification. Subsequently, each k-model is compared with new instances to assess whether the instance lies within the predetermined hyper-sphere. If a connection satisfies certain criteria, it is categorized as a normal connection; otherwise, it is designated as an attack state.

By employing unsupervised learning in this approach, a highly effective technique is used to classify new instances by setting a threshold to distinguish between malicious and regular data throughout the model's development. Presently, a significant drawback of the approach is readily apparent since typical connections vary across different networks, which can considerably hinder the creation of precise profiles of normal behaviour. The significant variation in the behavioural patterns and characteristics of one network compared to other networks might lead to an inefficient model, requiring extensive parameter adjustment and optimization to match the specific network environment.

According to Abduvaliyev et al., [20] and Butun et al., [21] investigated several forms of attacks on wireless sensor networks (WSNs) and the potential influence of attack detection systems on the expanding threat landscape. However, as stated by [20], additional actions must be taken to protect a Wireless Sensor Network (WSN) from various forms of attacks, including denial of service (DoS), sinkhole/blackhole attacks, selective forwarding, node replication attacks, and wormhole attacks, in order to minimize their harmful consequences. The second line of defense utilizes a clustering technique that is categorized as an unsupervised learning process. This technique facilitates the detection of anomalous traffic in Wireless Sensor Networks (WSNs). The model is constructed by utilizing twelve network traffic patterns, which are subsequently employed throughout both the training and testing phases. During the phase of training the model, a technique called fixed-width clustering is used to create clusters in the feature space. Anomalies in clusters are identified when the samples being analysed have a reduced training size.

Anomalies are detected when traffic samples exceed a specific threshold. In addition, the testing phase involves the identification of anomalous patterns by linking certain traffic samples with a cluster set. A significant drawback of this technology is the considerable processing requirements imposed on sensor nodes, resulting in major additional expenses for the main network.

Aparicio-Navarro et al. [22] presented a novel intrusion detection system (IDS) that utilizes contextual information, particularly pattern-of-life (PoL) data, to evaluate anticipated network behavior. This IDS is engineered to identify multi-stage assaults (MSA) in real time without any prior training. Findings demonstrate that the integration of contextual information markedly enhances the system's efficiency, increasing the detection rate of MSAs by 58% in real-time situations. This detection method enhances an unsupervised, anomaly-based IDS framework by adding a fuzzy cognitive map (FCM) to incorporate contextual information into the detection process.

The results demonstrate that the use of contextual data significantly improves the efficacy of Intrusion Detection Systems in detecting Malicious Software Activities. The incorporation of an FCM resulted in a 58% enhancement in the detection rate (DR) relative to the IDS lacking this component. Nonetheless, the architecture of the FCM is exceedingly context-dependent, constraining its generalizability. A 5-step MSA was specifically designed for testing in the study, with the FCM adapted for this context. Thus, implementing the model in different situations or MSAs necessitates the creation of a new FCM tailored to those circumstances. The mechanism employed to capture temporal correlations among MSA phases is not transferable to other forms of multi-stage assaults. The temporal parameters established in this work are context-specific, and the IDS is incapable of autonomously adapting its detection methodology for differing MSAs.

Shin et al. [23] introduced an innovative architecture for multi-stage attack detection, comprising two primary phases: the creation of detection rules and the actual detection phase. This framework functions by creating precise detection rules designed for multi-stage assaults and subsequently evaluating incoming network data against these criteria. In contrast to conventional approaches, it does not necessitate prior knowledge of single-stage assault behaviors, rendering it adaptive and proficient in detecting diverse multi-stage attack patterns even in the absence of specific information on individual attack stages. The framework exhibited robust performance when assessed,

AKSU Journal of Engineering and Technology, 2026
Vol. 1, No. 1, Pages 17-33
Published Online April 2026 in AKSUJET (http://www.aksujet.org) ISBN: 978-978-53006-97

even when processing substantial amounts of intricate multi-stage attack data. It precisely recognized all multi-stage attack patterns inside the DARPA LLS DDoS dataset, validating its efficacy in a demanding test environment. Furthermore, in evaluations utilizing the CTU-13 datasets, characterized by extensive sophisticated assault patterns, the framework attained a peak F1 score of 0.9380, signifying a substantial degree of precision and recall.
The study emphasizes that a considerable percentage of network attacks are multi-stage, characterized by coordinated and intricate series of operations that occur over prolonged durations. These attacks are methodically divided into several discrete single-stage operations, rendering them challenging to identify when examined independently. Thus, recognizing these multi-stage patterns is crucial for comprehending the distinct behavioral traits of sophisticated network threats and for strengthening network security protocols.

iii. Methodology for Semi-supervised Learning
Ashfaq et al., [24] suggest that semi-supervised learning integrates both labeled and unlabeled samples to enhance the classifier's performance. Moreover, [25] argues that a semi-supervised Machine Learning approach use a pre-annotated dataset to imitate common patterns of behaviour. Semi-supervised learning combines the strengths of supervised and unsupervised learning techniques to create a model capable of classifying new data points in a dataset.

However, [25] suggested a two-stage semi-supervised statistical method for identifying network anomalies. The technique constructs a probabilistic model by utilizing pre-classified normal instances. Afterwards, this model is used to evaluate departures from the normal behaviour by using a predetermined threshold. The second phase involves implementing an iterative approach to reduce the frequency of false alarms. This is accomplished by employing a similarity distance and dispersion rate that are derived from the initial classifications of the probabilistic model [25].

Although the strategy has achieved high detection rates and low false positive rates, beating the Naïve Bayes algorithm in both true positive and false positive rates, it is still limited by the constraints of the anomaly detection method outlined in [26].

The author in [27] proposed a semi-supervised learning approach to address co-resident attacks in cloud-based systems. The approach incorporated a safeguarding mechanism that substantially enhances the computational expense required for a co-resident attack to achieve success in a virtual machine within a cloud computing system. The problem was framed as a 2-player security game, in which users were classified through the utilization of clustering analysis and semi-supervised Support Vector Machines (SVMs). Users are classified into three categories: high risk (malicious), medium risk (uncertain), and low risk (legal), depending on the modifications made to the virtual machine allocation method. This contributes to raising the overall expense for an attacker to execute a computationally demanding attack action, hence strengthening the defensive mechanism.

The technique successfully mitigated co-residence attacks by increasing the attacker's overall cost by a factor of 100. However, employing a single data center to execute the method is not feasible due to the requirement of addressing diverse scenarios across several data centers, which may involve co-location and co-resident attacks.

The paper [28] examines cyber-attacks in computer networks using semi-supervised approach, offering an approach that integrates honeypots and network traffic manipulation to identify and prevent attacks proactively by anticipating the attacker's subsequent activities. Honeypots are recognized as essential instruments for assessing attacks, owing to the diverse range of implementations, resources, and insights derived from previous encounters. They are very effective at detecting and analyzing zero-day assaults. Honeypots are intentionally designed to monitor network traffic, regarding any incoming packets as potentially hostile, as any entity interacting with the honeypot is likely an attacker.
This method, however, has difficulties, particularly with faked traffic, as evidenced by major DDoS assaults on the Czech Republic's internet infrastructure. In this instance, both legitimate sites and honeypots inside the network were utilized as reflectors, resulting in false positives that erroneously identified authentic targets as sources of malicious traffic. Notwithstanding the deceptive character of this traffic, it nonetheless contributed to the overall attack pattern.
A flow-based monitoring tool named Honeyscan was created, implemented, and evaluated on a live network to assess honeypot traffic. Acknowledging that not all phases of an assault are identifiable or transpire within the network, an anti-phishing framework, PhiGARO, was instituted for early detection in application-specific contexts. Moreover, prolonged network flow monitoring improved the identification of network spying activities. By examining application-level data instead of depending exclusively on the conventional 5-tuple flow record, the system attained enhanced detection accuracy.
The integration of an HTTP parser enhanced large-scale monitoring, allowing the system to trace HTTP requests aimed at the monitored network. This indicated that application-level scanning, especially via repeated HTTP queries, was a notable evidence of reconnaissance activities. By concentrating on the pattern of HTTP requests instead of the parameters utilized in each scan, the system effectively discovered scanning activity

AKSU Journal of Engineering and Technology, **2026**
Vol. 1, No. 1, Pages 17-33
Published Online April 2026 in AKSUJET (http://www.aksujet.org) ISBN: 978-978-53006-97

that may have otherwise remained unnoticed, especially those aimed at certain web apps.

Multi-stage assaults, encompassing both malevolent and innocuous stages, underscore the difficulty of detecting novel, intricate threats. The authors in [29] proposed a complete method employing a substantially Boosted Neural Network model to effectively detect multi-stage attack situations. The model exhibited significant predictive accuracy: 94.09% for the Quest model, 97.29% for the Bayesian Network, and 99.09% for the Neural Network. Upon assessment with the Multi-Step Cyber-Attack Dataset (MSCAD), the suggested Extremely Boosted Neural Network attained an impressive accuracy of 99.72% in forecasting multi-stage cyber-attacks.
The proposed model was constructed utilizing distinct Machine Learning methods in Python, implementing the QUEST, Bayesian Network, and Neural Network models in the preliminary phase to forecast multi-stage cyber-attacks in a cloud context. The model's performance was evaluated against multiple attack types, including Brute Force, HTTP DDoS, ICMP Flood, Normal traffic, Port Scanning, and Web Crawling. This thorough methodology highlights the model's capability to anticipate and mitigate a wide range of cyber-attacks with significant precision.
Industrial Control Systems (ICS) require extensive and efficient protection, particularly when they support vital infrastructure. Vasilomnolakis et al. [30] presented a novel honeypot aimed at identifying multi-stage assaults specifically directed at ICS networks. Upon the identification of a multi-stage attack, this honeypot is capable of generating attack signatures, enabling misuse-based Intrusion Detection Systems (IDSs) to thwart analogous attempts utilizing the HOSTaGe honeypot. Given that honeypots have no other function, any engagement with them is intrinsically deemed suspicious, leading to an absence of false positives. A fundamental characteristic of honeypots is their capacity to remain undetected, masquerading as authentic gadgets instead than just decoys.

Shodan, an internet-connected device search engine, has devised techniques to identify honeypots by doing a series of probes and assessments, finally providing a score to each examined device. Shodan can ascertain the likelihood of a system being a honeypot based on this score, which may diminish its efficacy, as malware operators can leverage this information to evade recognized honeypots.
The research findings indicate that the honeypot and its produced signatures attain a high level of detection accuracy. Furthermore, these signatures can be incorporated into the Bro IDS, allowing it to effectively thwart future assaults. The honeypot methodology is feasible for practical implementation and is compatible with current IDS frameworks. Shodan's identifying capabilities may diminish the usefulness of honeypots, as malware could circumvent well-known honeypots to avoid detection.

The authors in [31] proposed a comprehensive framework integrating anomaly and signature-based methodologies to proficiently detect both established and novel cyber threats. The framework analyzes incoming data packets using a Stacked Autoencoder, categorizing them as benign or malicious. The Grey Wolf Optimization technique extracts the most significant characteristics from packets designated as harmful. The system was evaluated using two prominent datasets, UNSW-NB15 and CIC-IDS-2017, attaining notable accuracy rates of 90.94% and 99.67%, respectively. This dual methodology, integrating statistical methodologies and deep learning techniques, exhibits robust proficiency in threat identification and classification.

Nonetheless, the signature-based element of the architecture has limits. Although it is proficient at addressing known threats, it encounters difficulties with novel or unrecognized attacks. Moreover, updating the signature database to incorporate the most recent attack definitions is a complicated and time-consuming endeavor, which may impede real-time threat detection for new cyber threats.
Hachimi et al. [32] developed a multi-stage Machine Learning-based intrusion detection system (ML-IDS) designed for 5G Cloud Radio Access Networks (C-RAN), focusing on the detection and classification of four types of jamming attacks: constant jamming, random jamming, deceptive jamming, and reactive jamming. The Wireless Sensor Network Dataset (WSN-DS), designed for wireless intrusion detection, was utilized to assess the system's efficacy. The multi-stage detection methodology, incorporating both supervised and deep learning classifiers, aims to decrease undiscovered attacks while reducing false negatives and false positives, attaining a detection and classification accuracy of up to 94%.
Notwithstanding its superior performance, this solution possesses limitations. It is incapable of detecting certain advanced jamming tactics, including shot noise-based intelligent jamming. The system also fails to address various other attack vectors aimed at C-RAN architecture, such as eavesdropping, primary user emulation, and impersonation attacks, which are outside its detection capabilities.
Multi-stage attacks can evolve considerably, resulting in considerable losses and damages for businesses. This research [33] introduces a framework that forecasts multi-stage assaults through a hybrid methodology, combining IP information evaluation and a process query system (PQS). This method is deemed useful for detecting multi-stage attacks; nevertheless, it necessitates prior knowledge of attack patterns (sequences), which can be difficult, as recognizing novel, intricate attacks frequently require time.

The identity checker, which is based on IP information, was assessed independently through a metrics-based methodology, resulting in excellent performance with no false positives and a

AKSU Journal of Engineering and Technology, **2026**
Vol. 1, No. 1, Pages 17-33
Published Online April 2026 in AKSUJET (http://www.aksujet.org) ISBN: 978-978-53006-97

high detection rate. Nonetheless, it is incapable of identifying multi-stage assaults if the associated IP addresses are not classified as malicious. The PQS technique is augmented by including additional models into its components, which are subsequently assessed independently using a metrics-based evaluation. To enhance overall system performance, analogous models will be amalgamated, minimizing the total number of models where feasible.

iii. **Reinforcement Learning-Based Approach**
Reinforcement learning is a form of Machine Learning that allows a software agent, such a sensor node, to gain information by actively interacting with its environment. The author in [34] stated that reinforcement learning is essential for pattern recognition since it allows software agents to learn from their interactions with the environment and make optimum decisions that maximize long-term rewards. Moreover, according to [35], reinforcement learning agents communicate within an initially unfamiliar environment and use the gained information to modify their action methods in order to maximize their rewards.

The discourse in [35] centers on reinforcement algorithms. The authors suggest that reinforcement learning is very suitable for addressing sequential problems, which may be described as Markov decision processes (MDPs), and therefore suitable for comprehending learning control problems. Supervised learning systems frequently encounter difficulties in comprehending these situations due to their intrinsic intricacy
The author in [36] employed fuzzy Q-learning to detect and prevent intrusions in Wireless Sensor Networks (WSNs). The concept utilizes a combination of cooperative game theory and fuzzy Q-learning algorithms to detect DDoS attacks. The methodology emulates sinkholes, a central station, and a malicious entity in a 3-player strategic game, wherein the game is initiated upon the transmission of a substantial volume of packets towards a certain node. Presently, the packets received in the Wireless Sensor Network (WSN) are evaluated against a pre-established threshold to detect alarm occurrences. Once the threshold is surpassed, the technique initiates collaborative defense tactics to safeguard the sinkhole and the base station.

The NS-2 simulator was used to undertake a performance evaluation of the low energy adaptive clustering hierarchy (LEACH) technique through simulation. The purpose of this simulation was to demonstrate the precision of the technique in identifying and safeguarding against potential dangers. The method's architecture allows for the sink hole and base station to adaptively modify their approach in order to effectively detect and respond to an abrupt assault. The Intrusion Detection and Prevention System (IDPS) uses fuzzy Q-learning to continuously update its learning parameters in order to detect and prevent future attacks. This technique enables continuous learning from past attack patterns. Focusing solely on DDoS flooding attacks may make it challenging to determine its efficacy against other types of attacks. Hence, the model requires a thorough upgrade to strengthen its decision-making capabilities, specifically in identifying and mitigating novel forms of attacks.

AKSU Journal of Engineering and Technology, 2026
Vol. 1, No. 1, Pages 17-33
Published Online April 2026 in AKSUJET (http://www.aksujet.org) ISBN: 978-978-53006-97

Table 1
Comparison of multi-stage attack detection methods.

TITLE [REF]	DATA SET	METHODOLOGY/ TOOLS	CONTRIBUTION	RESEARCH GAP
MARS: Multi-stage attack recognition system [37,38]	LLS DDOS Data set	Supervised Approach/misuse	Multi-stage attack detection based on correlation of knowledge-based and statistical model	Vulnerable to zero-day attack due to dependency on pre-defined details
Applications of hidden Markov models to detecting multi-stage network attacks [39]	Self generated dataset	Supervised approach	Applying HMM to multi-stage attack detection showing the best performance among C4.5, *k*-NN and HMM	Vulnerable to zero-day attack due to
A multi-stage attack mitigation mechanism for software-defined home networks [40]	Private data set	Supervised approach	Applying SDN/NFV environment data to multi-stage attack detection	Vulnerable to zero-day attack due to dependency on pre-defined details Experiment with private data only
Multi-stage Jamming Attacks Detection using Deep Learning combined with kernelized support vector machine in 5G Cloud Radio Access Network (CRAN) [32]	Wireless sensor network dataset (WSN-DS)	Developing a new Machine Learning intrusion detection system (ML-IDS) based on supervised and deep learning classifiers model.	The system reduces the number of attacks missed, decrease the system's false negative and false positive rate. High detection and classification accuracy up to 94% as compared to only Multilayer Perceptron (MLP)	It only detects four types of network jamming attacks (constant, random, deceptive and reactive) It fails to detect other types of jamming attacks like shot noise-based Intelligent jamming. It fails to detect zero-day attacks and other types of Cloud Radio Access Network (CRAN) such as eavesdropping, primary user emulation and social engineering attacks
Early detection and mitigation of multi-stage network attacks [28]	Not stated	Uses honeypots and network traffic manipulation approach.	Flow-based monitoring of honeypot Early attack detection in application specific domains. Application-level network scanning detections Can handle zero-day attacks	Not effective in a spoofed network traffic. Some application level scans attacks such as repeating HTTPS request, would avoid detection Varying accuracy and inconsistent against different attacks.

Published Online April 2026 in AKSUJET (http://www.aksujet.org) ISBN: 978-978-53006-97

				High computational requirements and false positive rates
Extremely boosted neural network for more accurate multi-stage cyber-attack prediction in cloud computing environment [29]	MSCAD Dataset	Extremely boosted neural network model	The model achieves 99.72% accuracy as compared to Quest model with 94.09%, Bayesian network with 97.29% and neural with 99.09%	Varying accuracy with limited test data The model is limited to Brute force, HTTP, DDOS, ICMP-flood normal, port scan, web-crawling attacks. Vulnerable to some zero-day attacks
Multi-stage intrusion detection system aided by grey wolf optimization algorithm [31]	UNSSW-NB15 and CIC-IDS-1017	It utilizes both signature based and anomaly-based techniques with grey wolf optimization algorithm	It can detect both unknown and known attacks with high accuracy	Inconsistency against different attacks with varying accuracy. High computational complexity according to time Vulnerable to zero-day attacks
Unsupervised multi-stage attack detection framework without details on single-stage attack [23]	DARPA LLS DDOS and CTU-13	Unsupervised approach	It can detect known and zero-day attacks without knowing pre-defined details on single-stage attack activities. Low false positive rate	Increasing computational complexity according to time Low understandability of multi-stage attack behavior.
Multi-stage attack detection and signature generation with ICS honeypots [30]	BRO-IDS	HOSTaGe honeypot development for attack detection and signatures generation for Bro IDS	The system supports existing IDS infrastructure. It can detect zero-day attacks. Not detected by Shodan search engine	The system becomes ineffective if the attacker avoids honeypot High computational complexity
Multi-stage attack detection via kill chain state machines [18]	CSE-CIC-IDS 2018	Uses a kill chain state machine that operate on clustered alert data to identify states and transitions of multi-stage attacks. Supervised approach	Substantially reduces the false positive alert correlation and attack contextualization It can be applied to any network-based alerts. It can detect complex attacks such as advance persistent threats (APT)	The algorithm only processes network level information and cannot handle host-level and user identity context attacks. Vulnerable to zero-day attack
A casual network-based system for predicting multi-stage attack with malicious IP [16]	DARPA'S Grand challenge problem GCP	Uses a probabilistic inference system for predicting multi-stage attack with malicious	Good in multi-stage attack prediction, detection and blacklisting of malicious IP	It cannot predict attacks on devices that utilizes VPN.

AKSU Journal of Engineering and Technology, 2026
Vol. 1, No. 1, Pages 17-33
Published Online April 2026 in AKSUJET (http://www.aksujet.org) ISBN: 978-978-53006-97

		IP based on Bayesian belief network (BBN) model	addressing and computer network security in general.	It cannot predict attacks using MAC address. Vulnerable to zero-day attack.
Multi-stage attack detection using contextual information [22]	Not stated	It is based on IDS exploits of contextual information in the form of point of life model and information related to expert judgment on the network behaviour by the use on FUZZY cognitive map (FCM)	It is able to efficiently detect the presence of a multi-stage attack in real time without prior training process. Can detect zero-day attack	The design of an FCM is very context-specific and may not easily generalized. Increased computational complexity
A novel multi-stage approach for hierarchical intrusion detection [17]	CIC-IDS 2017 and CSE-CIC-IDS 2018	Hierarchical intrusion detection approach is proposed	High adaptability without the necessity to retrain any of the classifiers. It can detect zero-day attack Hierarchical deployment ensures that privacy is preserved during training and operational service phases.	Cannot handle network level attacks High false positive alerts Varying accuracy and inconsistency against different attacks

4. Challenges in identifying multi-stage assaults

Identifying multi-stage attacks has several challenges. One of the challenges we face is dealing with the diverse problems related to wide-ranging intrusion detection. An obstacle that arises is the immense intricacy of contemporary network data, rendering the identification of pertinent security information arduous. In addition, the most dangerous attacks occur seldom, leading to a small number of attack occurrences in each dataset. Ourston et al. [39] offer additional elucidation on the Rare Data Problem, a difficulty encountered in intrusion detection research, along with other statistical concerns pertaining to security data, such as skewed distributions or imbalanced data sets.

The gravity of these challenges is intensified when considering multi-stage attacks. For a single-step attack, the trace often includes all the relevant information and is linked to a vulnerability in a system. The attack can be examined independently and contrasted with prior occurrences. As the number of processes engaged rises, it becomes more difficult to study and characterize the similarity between attacks. When conducting a multi-stage attack, it is essential to determine the specific attributes of each individual step as well as the correlation between them. While we may possess the ability to identify the steps involved, we may nevertheless fail to notice the attacking strategy.

Here, we will present a brief summary of some challenges faced while identifying multi-stage attacks.

i. The individual stages of a multi-stage attack may seem innocuous.

ii. An attacker can develop multiple strategies to carry out an attack [41]. Additionally, the execution of a plan may come to an end if the attacker loses interest or is incapable of taking use of the vulnerabilities in the network [42]. Also the interval between consecutive episodes of an assault can vary significantly, spanning from hours to days or even months [43].

iii. Technical limitations of network devices or their deployment or design may lead to the inability to recognize some processes, as mentioned in reference [43].

iv. An attacker is not required to follow a precise order when executing a multi-step assault [44], so the possible sequences of actions might be extremely complex.

AKSU Journal of Engineering and Technology, **2026**
Vol. 1, No. 1, Pages 17-33
Published Online April 2026 in AKSUJET (http://www.aksujet.org) ISBN: 978-978-53006-97

v. Intrusion detection systems (IDS) often lack comprehensive information about the root causes of a problem [45, 46], making it difficult to identify contextual details.

vi. The majority of the features observed in traces are categorical, indicating that there is no inherent order or correlation among the possible values. This hinders the application of mathematical methods for creating attack scenarios [46].

vii. There is a lack of standardized datasets that may be used to evaluate the effectiveness of multi-stage attack detection systems. Furthermore, public research is limited in its access to a substantial fraction of the methodology and datasets employed by other researchers or commercial enterprises.

viii. A considerable proportion of the analysed methods for detecting multi-stage attacks depend on Intrusion Detection System (IDS) alerts as their main source of data. The production of IDS alerts is connected with several difficulties, including [47, 48, 49].

ix. Authentic notifications are frequently mixed with false positives and irrelevant ones.

5. Strategies for Cyber-Attack Prevention

The prevention of assaults is a proactive measure that swiftly discovers and addresses possible risks within a network. Prevention is highly pertinent in the process of mitigating cyber assaults. Most detection methodologies are reactive and are implemented only after significant damage has occurred in the affected area. Numerous intrusion prevention systems (IPSs) have been suggested to enhance cybersecurity.
Patil and Meshram [49] examined a strategy for thwarting network intrusions, focusing on varieties of Denial of Service (DoS) attacks, including flooding, IP spoofing floods, and ping of death attacks. The approach is designed to be platform-independent by utilizing the Java Virtual Machine (JVM). The packet sniffer is constructed with the Jpcap library, while the mitigation of malicious traffic invading the internal network is accomplished through the Linux iptables command. The utilization of the Jpcap library necessitates the prior installation of the WinPcap library for Windows and the libpcap library for Unix or Linux systems.
In this instance, attack prevention is executed by analysing inbound packets intercepted with Jpcap in promiscuous mode. When packets are analysed and the SYN flag is activated, consistently targeting the same destination address amongst ongoing network traffic, the system infers a SYN flood assault. The identified attack data is recorded in the log file, and subsequent measures are implemented to discard the packet using the iptables command (Linux) or net-filter (Windows). The suggested approach can also avoid smurf attacks (ICMP packets), SYN-FIN attacks, XMAS attacks, fraggle attacks (UDP packets), and all flag assaults. The method provided an expedited procedure for attack mitigation, albeit primarily reliant on iptables regulations. Furthermore, if an attack is not identified promptly by an expedited rule-matching procedure, it may penetrate the network and inflict significant harm to internal resources.
One of the most elusive forms of cyber-attack is the Distributed Denial of Service (DDoS) flooding attack, which employs botnets to obstruct services intended for legitimate users of a system or network. Botnets are frequently utilized to inundate several computers, perhaps on a global scale, with malicious packets via the Internet by exploiting weaknesses in these systems.
These botnets consist of elements referred to as masters, handlers, and bots. The attackers are the orchestrators, who interact with the operatives or bots through intermediaries. The attackers utilize compromised systems for command-and-control operations. Zombies are compromised computers that constitute an attack force, systematically infiltrating other vulnerable systems along the attack vector to amplify the assault until all computing resources are rendered inoperative, potentially resulting in significant and irreversible damage. Countering DDoS attacks has proven to be an arduous endeavour. Ideally, numerous computers connected to Internet-enabled networks, together with devices such as smartphones and personal digital assistants (PDAs), should be safeguarded against all types of vulnerable services and ports. Nonetheless, these susceptible services and ports result from unpatched systems, for which the majority of security updates transmitted to devices via proxy servers are neither implemented nor timely executed. This results in numerous unpatched and insecure devices that can be exploited in a DDoS flooding assault utilizing the command and control (C & C) functionality of botnets.
Therefore, [50] asserts that techniques such as source address authentication, capabilities, and filtering are crucial for mitigating DDoS flooding attacks. Internet service providers should collaborate by utilizing technologies such as cloud computing and IoT to prevent and mitigate threats, leveraging the closeness of the Internet as a significant advantage. Moreover, while the classification provided has taken into account various sources and results of DDoS assaults, little focus has been placed on the correlation of warnings, which could potentially identify the attack's source when a significant number of cases are present. When evaluating flooding assaults based just on the volume of traffic produced, without examining the causal linkages among traffic occurrences, identifying the

AKSU Journal of Engineering and Technology, **2026**
Vol. 1, No. 1, Pages 17-33
Published Online April 2026 in AKSUJET (http://www.aksujet.org) ISBN: 978-978-53006-97

source of the attack in real time might be challenging. A software-defined intrusion prevention system, referred to as SDNIPS, is proposed for mitigating cloud-based assaults in [51]. The methodology incorporates a detection element that integrates a Snort-based intrusion detection system with Open vSwitch (OVS). The architecture of SDNIPS is further streamlined, enabling cloud resources to create traffic that traverses the SDNIPS agent. The traffic is subsequently compared to the Snort rules, and any matches are flagged as alerts that enhance the log file. The SDNIPS daemon captures this warning information and transmits it to the JSON server at the controller's end. During the processing of alert information, the alert interpreter analyses the data and extracts essential details, such the attack type, source and destination IP addresses, and TCP port, among others. During the further processing of alert information, the OVS modifies the flow table utilizing the OpenFlow rule entries from the rule's generator. Any dubious traffic corresponding to the revised flow table entries is subsequently addressed by implementing the requisite countermeasures in the data plane. This method truncates an attack, so safeguarding cloud resources from compromise.
The methodology exhibits an effective preventative system; nevertheless, the reliance on Snort necessitates complete dependency on specialist knowledge for rule definition, which might be time-consuming. This strategy will likely result in increased use of computing resources, as traffic must continuously traverse the IPS, and delays in matching each traffic pattern to the established rules will considerably affect the host system's resources.
As organizational security requirements escalate, increasingly sophisticated solutions are necessary to safeguard extensive resource allocations. A crucial element in attaining a viable security solution is the availability of a cost-effective, flexible, and scalable product or methodology. This is the emphasis of [52] in their suggested real-time methodology for identifying and mitigating assaults.
The methodology, grounded on the software engineering framework such as requirement analysis, design, implementation, and testing, configures Snort in inline mode to facilitate intrusion prevention. Configuring Snort in inline mode enables the Intrusion Prevention System (IPS) to position its sensors for the interception and elimination of suspicious packets that are likely to contain attack payloads. The discarded packets are ultimately recorded in Splunk. Despite this, the incapacity of signature-based intrusion detection and prevention systems, such as Snort, to identify unknown attacks, along with its inadequate performance under excessive network traffic, constitutes a significant limitation of this method [53].

6. Evaluations, obstacles, and future prospects

Table 1 presents a comparison of various multi-stage attacks detection strategies. The comparison is based on the methodology/tool used, the training/testing data set employed, as well as the contributions and research gaps identified. Out of the thirteen detection strategies analysed in the table, nine of them utilize a combined total of nine publicly accessible data sets for both training and testing the model. One scheme utilizes a private dataset, with one plan employing a self-generated data set, while for the remaining schemes, the specific data sets are not specified. Furthermore, these nine public data sets originate from various domains such as malware and network attacks, and they vary in terms of their feature space.
The evaluated strategies demonstrate potential in effectively identifying multi-stage attacks through various methodologies. Nevertheless, the majority of detection techniques frequently demonstrate significant fluctuations in their accuracy when faced with various types of attacks. Hence, it is challenging to quantitatively compare various schemes due to the differences in evaluation data sets, assessment measures, and comparison schemes. The available training/testing data sets are constrained and isolated, and there is a notable absence of a comprehensive, representative data collection at a significant scale. Various methods are commonly assessed using distinct sets of performance indicators. Because implementing comparison schemes requires a substantial amount of effort, the examined schemes are typically compared to only a few rival systems before drawing conclusions. The outcomes are often inconsistent and difficult to interpret due to the use of non-representative data sets and a restricted number of model comparisons.
The multi-stage detection systems listed in Table 1 can be classified into four categories: Unsupervised, semi-supervised, supervised and reinforcement Learning methodologies. Unsupervised learning refers to a category of Machine Learning algorithms that extract patterns from data that lacks explicit labelling. The objective of the assault's detection model is to acquire knowledge about a condensed representation of regular data in order to identify attacks. Supervised and semi-supervised attack detection approaches employ either supervised learning or a combination of supervised and unsupervised learning techniques. Supervised and hybrid attack detectors can provide precise detection by utilizing representative training data sets. Regrettably, the data collection does not include any instances of certain types of assaults, such as zero-day attacks. The detection techniques must make the assumption that zero-day attacks exhibit similar behaviour to known attacks, a hypothesis that has not yet been verified. The speed of training and detection is a crucial component in multi-stage attack detection. Although training often requires more time than detection, the examined approaches are all capable of completing both training and detection within a suitable timeframe, despite variations in pace across different methods.

AKSU Journal of Engineering and Technology, **2026**
Vol. 1, No. 1, Pages 17-33
Published Online April 2026 in AKSUJET (http://www.aksujet.org) ISBN: 978-978-53006-97

6.1 Identified Research Gaps

Despite the significant advancements in machine learning-based approaches for multi-stage cyber-attack detection, several critical research gaps remain.

1. Lack of Zero-Day Attack Detection

One of the major limitations of existing models is their inability to effectively detect zero-day attacks. Most supervised and hybrid learning approaches rely heavily on labeled datasets containing known attack patterns, making them less effective against novel threats [16], [31]. As a result, these models struggle to identify previously unseen or evolving attack strategies, leaving systems vulnerable to emerging cyber threats.

2. Dataset Limitations

The performance of machine learning models is highly dependent on the quality and diversity of training datasets. However, many existing studies rely on limited or domain-specific datasets such as CIC-IDS and DARPA datasets [17], [23]. These datasets often fail to represent real-world network environments, which reduces the reliability and generalizability of detection models in practical deployments.

3. High Computational Complexity

Advanced machine learning and deep learning models often require significant computational resources for training and real-time detection. Techniques such as ensemble learning, deep neural networks, and optimization-based models introduce high processing overhead, limiting their applicability in real-time and resource-constrained environments [31], [32].

4. Lack of Generalization

Many proposed models demonstrate high accuracy when evaluated on specific datasets but fail to generalize across different network environments. This limitation arises due to overfitting and the dependency on specific training data distributions, which reduces their effectiveness in dynamic and heterogeneous network conditions [17], [22].

6.2. Prospects for the future

Additional endeavours are necessary to tackle the difficulties in formulating efficient multi-stage attack detection techniques. To resolve the problem of insufficient zero-day attack information in the training data set, one can utilize a honeypot [296] to gather the zero-day attack data prior to their discovery. Utilizing domain expertise to perform feature engineering is an additional method to enhance the accuracy of detection. Attackers can evade detection if the characteristics of their attack closely resemble those of lawful activities. Incorporating the expertise of domain specialists is crucial for effectively incorporating their knowledge into the process of feature engineering. This ensures that the newly developed attacks will be detectable within the specified feature space.

Detection techniques are advancing rapidly. Utilizing the most recent breakthroughs in Machine Learning and implementing the latest Machine Learning models is an additional method to protect against certain types of attacks, such as zero-day attacks. Reinforcement Learning (RL) is a form of Machine Learning where an agent, or decision maker, learns in an interactive environment through trial and error, using feedback from its own actions and experiences. The agent must only have access to induced feedback, without the necessity to possess knowledge of all the components that determine these feedbacks. Reinforcement Learning (RL) is especially suitable for multi-stage cyber assault challenges that involve unknown vulnerabilities and attack targets.
The creation of a multi-stage attacks detection benchmark will help overcome many obstacles that impede research and development progress. An extensive benchmark suite with standardized data sets, a wide range of representative models, and automated testing and assessment capabilities will significantly accelerate the development of multi-stage attack detection systems.

7. Summary

Multi-stage attacks frequently occur and result in significant financial losses, as well as potential damage to the reputation of both organizations and individuals. Machine Learning-based detection is the most promising and successful approach for detecting multi-stage attacks.
This paper presents a thorough examination of multi-stage attack detection methods. The review focuses on the different types of Machine Learning methods used, such as supervised, un-supervised, semi-supervised and reinforcement approaches. Figure 1 provides a visual representation of these methods. Nevertheless, the Machine Learning-based detection method encounters a fundamental difficulty in that it does not have the ability to represent zero-day attacks in the datasets. The restricted and isolated datasets, along with the incomplete range of features, significantly diminish the accuracy, resilience, and dependability of the models, falling short of the necessary level. In order to make progress, we suggest utilizing the most recent developments in Machine Learning

research and integrating the expertise of domain specialists more effectively into the building of the Machine Learning model. Furthermore, the creation of a comprehensive and standardized benchmark that contains abundant data will greatly aid in the ongoing enhancement of multi-stage attack detection models. Additionally, addressing the research gaps mentioned is essential for developing more robust, scalable, and adaptive multi-stage attack detection systems. Future research should focus on designing models that can generalize across environments, efficiently detect zero-day attacks, and operate with reduced computational overhead while leveraging more realistic and comprehensive datasets.

REFERENCES

[1] A. L. Buczak and E. Guven, "A survey of data mining and Machine Learning methods for cyber security intrusion detection", IEEE Communications Surveys & Tutorials, vol. 18, no. 2, **(2016)**, pp. 1153-1176.

[2] U. Akyazi, "Possible scenarios and maneuvers for cyber operational area", In European
Conference on Cyber Warfare and Security, Academic Conferences International Limited, Greece, **(2014)** July 3-4.

[3] D. E. Denning, "Framework and principles for active cyber defense", Computers & Security, vol. 40, **(2014)**, pp. 108-113.

[4] G. Kim, S. Lee and S. Kim, "A novel hybrid intrusion detection method integrating anomaly
detection with misuse detection", Expert Systems with Applications, vol. 41, no. 4, **(2014)**, pp. 1690-1700. International Journal of Security and Its Applications Vol. 12, No. 4 (2018) 28 Copyright ⓒ 2018 SERSC Australia

[5] S. Yoo, S. Kim, A. Choudhary, O. P. Roy and T. Tuithung, "Two-phase malicious web page
detection scheme using misuse and anomaly detection", International Journal of Reliable Information and Assurance, vol. 2, no. 1, **(2014)**, pp. 1-9.

[6] M. S. Rani and S. B. Xavier, "A Hybrid Intrusion Detection System Based on C5. 0 Decision
Tree Algorithm and One-Class SVM with CFA", International Journal of Innovative Research in Computer, vol. 3, no. 6, **(2015)**, pp. 5526-5537.

[7] H. Sugumaran, M., and Balasaraswathi, V. R. (2016). Ids using Machine Learning- current state of art and future directions. *British Journal of Applied Science and Technology*, *15*(3).

[8] M. Jacob, and Wanjala, M. Y. (2018). A Review of Intrusion Detection Systems. *Global Journal of Computer Science and Technology.*

[9] S. Biswas (2018). Intrusion detection using Machine Learning: A comparison study. *International Journal of pure and applied mathematics*, *118*(19), 101-114.

[10] A. Choudhury and D. Gupta (2019). A survey on medical diagnosis of diabetes using Machine Learning techniques. In *Recent developments in Machine Learning and data analytics* (pp. 67-78). Springer, Singapore.

[11] A. Ghosh, E. Fassnacht, Joshi, P. K., Koch, B., 2014. A framework for mapping tree species combining hyperspectral and LiDAR data: role of selected classifiers and sensor across three spatial scales. Int. J. Appl. Earth Obs. Geoinf. 26, 49-63.

[12] S. Kiran, Devisetty, R. K., Kalyan, N. P., Mukundini, K., and Karthi, R. (2020). Building an intrusion detection system for iot environment using Machine Learning techniques. *Procedia computer science*, *171*, *2372-2379*

[13] X. Liu, Yang, Y., Choo, K. K. R., and Wang, H. (2018). Security and privacy challenges for internet-of-things and fog computing.

[14] M. Chen, Wan, J., and Li, F. (2012). Machine-to-machine Communications: Architectures, Standards and Applications. *KSII Transactions on Internet and Information Systems*, *6*. https://doi.org/10.3837/tiis.2012.02.002

[15] A. Tabassum, and Lebda, W. (2019). Security Framework for IoT Devices against Cyber- Attacks. arXiv preprint arXiv: 1912.01712.

[16] A. Osarumwense, E. Oghenerukevbe (2020) "A casual network-based system for predicting multi-stage attack with malicious IP". International journal of Academic multidisciplinary research (IJAMR), vol 4, issue 5, may, 2020. Page 1-8.

[17] V. Miel, L. D'hooge, T. Wauters, B. Volckaert and F. De turck "A novel Multi-stage approach for Hierarchical intrusion detection"

[18] W. Florian, F. Ortmann, S. Hass, M. Vallentin, M. Fischer (2021),' Multi-stage Attack Detection via kill chain state machine'.

[19] J. Song, H. Takakura, Y. Okabe and K. Nakao, "Toward a more practical unsupervised anomaly detection system", Information Sciences, vol. 231, **(2013)**, pp. 4-14.

[20] A. Abduvaliyev, A. S. K. Pathan, J. Zhou, R. Roman and W. C. Wong, "On the vital areas of intrusion detection systems in wireless sensor networks", IEEE Communications Surveys & Tutorials, vol. 15, no. 3, **(2013)**, pp. 1223-1237.

[21] I. Butun, S. D. Morgera and R. Sankar, "A survey of intrusion detection systems in wireless sensor networks", IEEE communications surveys & tutorials, vol. 16, no. 1, **(2014)**, pp. 266-282.

[22] K. Kyriakopoulos, F. Aparicio-Navarro, I. Ghafir, S. Lambotharan,and J. Chambers. 2019. "Multi-stage Attack Detection Using Contextual Information". figshare. https://hdl.handle.net/2134/34219

[23] S. Jinmyeong, C. Seok-Hwan, P. Liu and C. Yoon-Ho (2019) 'Unsupervised multi-stage attack detection framework without details on single-stage attacks; Future generation computer systems 100(2019) 811-825

[24] R. A. R. Ashfaq, X. Z. Wang, J. Z. Huang, H. Abbas and Y. L. He, "Fuzziness based semi-supervised learning approach for intrusion detection system", Information Sciences, vol. 378, **(2017)**, pp. 484-497.

[25] N. B. Aissa and M. Guerroumi, "Semi-supervised statistical approach for network anomaly detection", Procedia Computer Science, vol. 83, **(2016)**, pp. 1090-1095.

[26] G. Kim, S. Lee and S. Kim, "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection", Expert Systems with Applications, vol. 41, no. 4, **(2014)**, pp. 1690-1700.

[27] Y. Han, T. Alpcan, J. Chan, C. Leckie and B. I. Rubinstein, "A game theoretical approach to defend against co-resident attacks in cloud computing: Preventing co-residence using semi-supervised learning", IEEE Transactions on information Forensics and Security, vol. 11, no. 3, **(2016)**, pp. 556-570.

[28] H. Martin (2015) "Early detection and mitigation of multi-stage network attacks PhD thesis in Masarykova unwerzita fukuita informality January, 2015

[29] D. Surjeet, P. Manoharan, L. Kumar, B. Seth, D. Alsekait, S. Simaiya, M. Hamdi, K. Raahemifar (2023) "Extremely boosted Neural network for more accurate multi-stage cyber-attack prediction in cloud computing environment." Journal of cloud computing: Advanced systems and applications, 2023.

[30] V. Emmanouil, S. Srinivasa, C. Garcia Cordero, M. Muhlhauser (2016) Multi-stage Attack Detection and Signature Generation with ICS Honeypots'

[31] C. Somnath, V. Shaw and R. Das (2021) , 'Multi-stage Intrusion Detection System aided by grey wolf optimization algorithm.' Spriger nature 2021. http://doi.org/10.21203/rs.3.rs-2680915/v1

[32] H. Marouane, G. Kaddoum, G. Gagnon and P. Illy (2020) "multi-stage jamming attacks detection using deep learning combined with kernelized support vector machine in 5g cloud radio access network 2020 IEEE international symposium on networks computers and communications (ISNCC'20), October, 2020.

[33] Z. Abdulrazaq, J. flint, D. parish (2015)," predicting multi-stage Attacks based on hybrid approach," international journal for information security Research (IJISR), VOL 5, issue 3, September 2015, pp582-590.

[34] M. A. Alsheikh, S. Lin, D. Niyato and H. P. Tan, "Machine Learning in wireless sensor networks: Algorithms, strategies, and applications", IEEE Communications Surveys and Tutorials, vol. 16, no. 4, **(2014)**, pp. 1996-2018.

[35] X. Xu, L. Zuo and Z. Huang, "Reinforcement learning algorithms with function approximation: Recent advances and applications", Information Sciences, vol. 261, **(2014)**, pp. 1-31.

[36] S. Shamshirband, A. Patel, N. B. Anuar, M. L. M. Kiah and A. Abraham, "Cooperative game theoretic approach using fuzzy Q-learning for detecting and preventing intrusions in wireless sensor networks", Engineering Applications of Artificial Intelligence, vol. 32, **(2014)**, pp. 228-241.

[37] F. Alserhani, M. Akhlaq, I.U. Awan, A.J. Cullen, P. Mirchandani, Mars: Multi-stage attack recognition system, in: 2010 24th IEEE International Conference on Advanced Information Networking and Applications, 2010, pp. 753–759, http://dx.doi.org/10.1109/AINA.2010.57,

[38] F. Alserhani, A framework for multi-stage attack detection, in: 2013 Saudi International Electronics, Communications and Photonics Conference, 2013, pp. 1–6, http://dx.doi.org/10.1109/SIECPC.2013.6550973.

[39] D. Ourston, S. Matzner, W. Stump, B. Hopkins, Applications of hidden markov models to detecting multi-stage network attacks, in: 36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the, 2003, p. 10, http://dx.doi.org/10.1109/HICSS.2003.1174909.

[40] S. Luo, J. Wu, J. Li, L. Guo, A multi-stage attack mitigation mechanism for software-defined home networks, IEEE Trans. Consum. Electron. 62 (2) (2016) 200–207, http://dx.doi.org/10.1109/TCE.2016.7514720

[41] X. Qin, W. Lee, Attack plan recognition and prediction using causal networks, in: 20th Annual Computer Security Applications Conference, IEEE, 2004, pp. 370–379. doi:10.1109/CSAC.2004.7

[42] S. J. Yang, J. Holsopple, M. Sudit, Evaluating threat assessment for multi-stage cyber-attacks, in: MILCOM 2006-2006 IEEE Military Communications conference, IEEE, Washington, D.C., 2006, pp. 1-7.doi:10.1109/MILCOM.2006.302216.

[43] B. Chen, J. Lee, A. S. Wu, Active event correlation in Bro IDS to de- tect multi-stage attacks, in: Fourth IEEE International Workshop on Information Assurance (IWIA'06), IEEE, 2006, pp. 16 pp.–50. doi: 10.1109/IWIA.2006.2.

[44] Z. Zhang, P.-H. Ho, X. Lin, H. Shen, Janus: A two-sided analytical model for multi-stage coordinated attacks, in: 9th International Conference on Information Security and Cryptology, Springer, Busan, Korea, 2006, pp.136{154. doi:10.1007/11927587_13.

[45] S. Salah, G. Maci´a-Fern´andez, J. E. D´ıaz-Verdejo, A model-based survey of alert correlation techniques, Computer Networks 57 (5) (2013) 1289–1317.

[46] K. Julisch, Mining alarm clusters to improve alarm handling efficiency, in: Computer Security Applications Conference, 2001. ACSAC 2001. Pro- ceedings 17th Annual, IEEE, 2001, pp. 12–21.

AKSU Journal of Engineering and Technology, **2026**
Vol. 1, No. 1, Pages 17-33
Published Online April 2026 in AKSUJET (http://www.aksujet.org) ISBN: 978-978-53006-97

[47] P. Ning, D. Xu, Toward automated intrusion alert analysis, Springer, 2010,pp. 175–205.
[48] D. Xu, P. Ning, Alert correlation through triggering events and common resources, in: Computer Security Applications Conference, 2004. 20th Annual, IEEE, 2004, pp. 360–369.
[49] S. Patil and B. B. Meshram, "Intrusion Prevention System", International Journal of Emerging trends in Engineering and Development, vol. 4, no. 2, **(2012)**, pp. 577-584.
[50] S. T. Zargar, J. Joshi and D. Tipper, "A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks", IEEE Communications Surveys & Tutorials, vol. 15, no. 4, **(2013)**, pp.2046-2069.
[51] T. Xing, Z. Xiong, D. Huang and D. Medhi, "SDNIPS: Enabling Software-Defined Networking based intrusion prevention system in clouds", In Network and Service Management (CNSM), 10th International Conference, IEEE, **(2014)**, pp. 308-311.
[52] P. S. Kenkre, A. Pai and L. Colaco, "Real time intrusion detection and prevention system", In Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA), Springer, Cham, **(2014)**, pp. 405-411
[53] A. Abduvaliyev, A. S. K. Pathan, J. Zhou, R. Roman and W. C. Wong, "On the vital areas of intrusion detection systems in wireless sensor networks", IEEE Communications Surveys & Tutorials, vol. 15, no. 3, **(2013)**, pp. 1223-1237.