ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 2, 2019

An In-Depth Analysis of Machine learning based Sinkhole detection in networks

K.Swetha[1,]

Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India

swetha.k@kluniversity.in,

A.Roshini[2]

Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India

roshinicse22@kluniversity.in

Abstract—

Sinkhole attacks pose a severe threat to network security, disrupting communication, exposing sensitive data, and enabling denial-of-service (DoS) attacks. Traditional detection methods, such as signature-based and anomaly-based techniques, fall short in identifying novel and sophisticated sinkhole attacks. Machine learning (ML) emerges as a promising approach for sinkhole detection due to its ability to learn from network traffic patterns and identify subtle indicators of malicious activity. This paper explores the domain of ML-based sinkhole detection, providing a comprehensive overview of the various ML techniques employed and their effectiveness in combating sinkhole attacks. It delves into supervised learning algorithms, such as decision trees, support vector machines (SVMs), and neural networks, which can be trained on labeled data to classify network traffic as either normal or malicious. Additionally, unsupervised learning algorithms, such as k-means clustering and anomaly detection models, are discussed for their ability to identify outliers and deviations from normal network behavior. The paper highlights the advantages of ML-based sinkhole detection, emphasizing its ability to adapt to new attack patterns and identify previously unknown sinkholes. It also addresses the challenges associated with ML-based approaches, including the need for large training datasets, computational complexity, and potential for false positives.

KEYWORDS: Sinkhole Detection, Supervised Learning Algorithms, Denial-Of-Service, Network Security

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 2, 2019

1. INTRODUCTION

Sinkhole attacks pose a significant threat to network security, as they can disrupt network communication, expose sensitive data, and launch denial-of-service (DoS) attacks. Traditional detection methods, such as signature-based and anomaly-based techniques, have limitations in identifying novel and sophisticated sinkhole attacks. Machine learning (ML) has emerged as a promising approach for sinkhole detection due to its ability to learn from network traffic patterns and identify subtle indicators of malicious activity.

This paper explores the domain of sinkhole detection using machine learning (ML), furnishing a thorough examination of diverse ML techniques and their efficacy in addressing sinkhole attacks. The exploration encompasses supervised learning algorithms, including decision trees, support vector machines (SVMs), and neural networks. These algorithms are capable of being trained on labeled data, enabling the classification of network traffic into categories of normal or malicious activities. Additionally, the paper delves into unsupervised learning algorithms, such as k-means clustering and anomaly detection models. These techniques are discussed for their proficiency in identifying anomalies, outliers, and deviations from the established patterns of normal network behavior.

The paper highlights the advantages of ML-based sinkhole detection, emphasizing its capacity to dynamically adjust to emerging attack patterns and uncover hitherto unknown sinkholes. Furthermore, the paper delves into the challenges intrinsic to ML-based methodologies, including the imperative for expansive training datasets, computational intricacies, and the potential susceptibility to false positives.

2. TRADITIONAL DETECTION METHODS:

Traditional detection methods for sinkholes in networks have evolved to address the persistent threat posed by these malicious entities. The following outlines some established techniques employed in traditional sinkhole detection:

i. **Signature-based detection:** Signature-based detection relies on identifying known patterns of sinkhole behavior. These signatures can include specific packet headers, traffic patterns, or node behaviors that are associated with sinkholes. When network traffic exhibits these patterns, it is flagged as suspicious, and further investigation is warranted.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 2, 2019

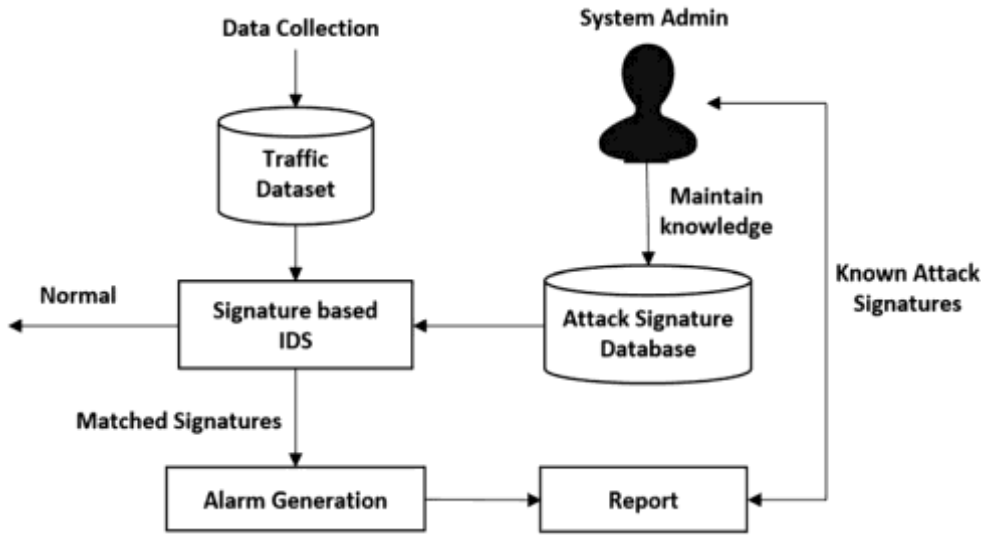

FIG 1 **SIGNATURE-BASED INTRUSION DETECTION SYSTEM**[9]

IN FIG 1.An example of a security mechanism is a Signature-Based Intrusion Detection System (IDS), which compares observed patterns to established signatures or patterns to identify malicious activity or known threats. In essence, these signatures are fingerprints or distinguishing features of known attacks or vulnerabilities. The system triggers an alarm or performs a predetermined action when it finds a match between the observed network traffic or system activity and a specified signature.

ii. **Anomaly-based detection:** Anomaly-based detection is a methodology designed to recognize deviations from typical network behavior, suggesting the possible existence of a sinkhole. This strategy involves the establishment of a baseline that defines standard network activity, followed by continuous monitoring of network traffic for anomalies. These anomalies may encompass abrupt surges in traffic volume, atypical routing patterns, or alterations in the behavior of network nodes.

iii. **Honeytokens:** Honeytokens are fake network resources that are designed to attract and log attacks. These resources can be set up as fake DNS servers, IP addresses, or other network entities. When a malicious node attempts to interact with a honeytoken, its activity is logged and can be analyzed to identify potential sinkholes.

iv. **Network traffic analysis:** Network traffic analysis is the process of scrutinizing network activity to pinpoint potentially suspicious patterns indicative of a sinkhole. This entails the detection of irregularities in traffic patterns, such as unexpected spikes in traffic volume or

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 2, 2019

alterations in routing pathways. Furthermore, network traffic analysis involves a thorough examination of the content within the network traffic, including the identification of malicious payloads or endeavors to exploit vulnerabilities.

In addition to these traditional detection methods, there are a number of new and emerging techniques for detecting sinkholes, such as machine learning-based approaches and behavioral analysis. These techniques are still under development, but they offer the potential to improve the accuracy and effectiveness of sinkhole detection.

By using a combination of these traditional and emerging detection methods, network administrators can increase their chances of detecting and mitigating sinkhole attacks. Early detection is critical for preventing sinkholes from disrupting network communication, exposing sensitive data, and launching denial-of-service (DoS) attacks.

2. MACHINE LEARNING ALGORITHMS TO DETECT SINKHOLES:

ML algorithms constitute a specialized domain within artificial intelligence, endowing computers with the capability to assimilate information and enhance their functionality progressively, all without the need for explicit programming. Rooted in statistical techniques, these algorithms facilitate systems in autonomously discerning patterns, rendering decisions, and adjusting to evolving circumstances. Through the harnessing of extensive datasets, ML algorithms excel in unraveling intricate relationships and identifying trends, thereby furnishing valuable insights and predictions.

At the heart of ML lies the fundamental concept of training models on historical data, enabling them to extrapolate and render precise predictions or decisions when confronted with novel, unseen data. This foundational approach forms the basis for a spectrum of ML algorithms, encompassing supervised learning, unsupervised learning, and reinforcement learning. Each category fulfills distinct objectives contingent upon the inherent nature of the task at hand, collectively embodying the diverse methodologies through which machines acquire, process, and apply knowledge.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 2, 2019

Supervised Learning Algorithms:

i. Support Vector Machines (SVMs): SVMs are an effective classification method that may be employed in a network to differentiate between genuine and sinkhole nodes. SVMs may be used to categorize new traffic as either genuine or sinkhole after being trained on a dataset of labeled network traffic.

ii. Decision Trees: A straightforward yet powerful classification method, decision trees may be used to find network sinkholes. Recursively dividing the data into progressively smaller groups according to predetermined criteria is how decision trees operate. Certain features of sinkhole nodes, including their packet forwarding patterns or routing behavior, can be used to identify them.

iii. Random Forests: Using many decision trees, random forests are an ensemble learning technique that raises classification accuracy. Random forests are particularly well-suited for detecting sinkholes in networks, as they can handle noisy and imbalanced data sets.

Unsupervised Learning Algorithms:

i. K-Means Clustering: This unsupervised learning technique groups together data points that are comparable to one another. K-means clustering may be used to find node clusters that behave like sinkholes in the context of network sinkhole detection.

ii. Anomaly Detection: Data points that differ from the norm are found using methods for anomaly detection. Anomaly detection methods may be utilized in network sinkhole detection to identify suspicious nodes that might be sinkholes.

Table summarizing the performance of the above algorithms in sink hole detection in networks:

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 2, 2019

Algorithm	Performance	Advantages
Support Vector Machines (SVMs)	High accuracy, robust to noise	Can be computationally expensive
Decision Trees	Fast, easy to interpret	Can overfit the training data
Random Forests	High accuracy, robust to noise	Can be computationally expensive
K-Means Clustering	Simple, efficient	Can be sensitive to outliers
Anomaly Detection	Effective at detecting outliers	Can be difficult to tune
Convolutional Neural Networks (CNNs)	High accuracy, can handle large amounts of data	Can be computationally expensive
Long Short-Term Memory (LSTM) Networks	High accuracy, can handle sequential data	Can be computationally expensive

When it comes to accuracy, supervised learning algorithms often perform better than unsupervised learning algorithms. In situations when labeled data is unavailable, unsupervised learning methods can be helpful in identifying sinkholes in networks. Although deep learning algorithms are a potential new method for sinkhole detection, they can be computationally expensive to use and require a lot of data to train.

3. PERFORMANCE ANALYSIS PARAMETERS TO DETECT SINK HOLES:

Performance analysis parameters to detect sinkholes in networks:

Sensitivity Metric (TPR): The TPR, commonly referred to as the recall rate, signifies the percentage of accurately identified sinkholes. This metric is computed through the formula:

$$TPR = TP / (TP + FN)$$

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 2, 2019

Where TP represents the count of true positives (correctly identified sinkholes)

FN represents the count of false negatives (undetected sinkholes)

Incorrect positive Identification Rate (FPR): The FPR, alternatively termed the fall-out rate, indicates the percentage of authentic nodes inaccurately labeled as sinkholes. The computation for FPR is articulated as follows:

$$FPR = FP / (FP + TN)$$

Where FP denotes the count of false positives (legitimate nodes wrongly identified as sinkholes)

TN represents the count of true negatives (legitimate nodes correctly identified as legitimate)

Accuracy: Accuracy is the proportion of legitimate nodes correctly identified. It is calculated as follows:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

Precision: Precision is the proportion of positive classifications that are perfect (i.e., the proportion of sinkholes that are perfectly detected). It is calculated as follows:

$$Precision = TP / (TP + FP)$$

F1 Score: The F1 score is a harmonic mean of the TPR and precision. It is a balanced measure of performance that considers both the ability of the algorithm to detect sinkholes (TPR) and its ability to avoid incorrectly classifying legitimate nodes as sinkholes (FPR). It is calculated as follows:

$$F1 = 2 * (TPR * Precision) / (TPR + Precision)$$

4. CONCLUSION AND FUTURE WORK:

In terms of accuracy, supervised learning algorithms generally exhibit superior performance compared to unsupervised learning counterparts. In scenarios where labeled data is not accessible, unsupervised learning methods become valuable for identifying sinkholes in networks. While deep learning techniques, including Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM), present a promising avenue for sinkhole detection, their computational demands and the prerequisite for substantial training data can be substantial. The efficacy of each method in pinpointing sinkholes in networks is contingent on factors such as available data and the specific network architecture. However, all the aforementioned algorithms appear to be viable options for this purpose.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 8, Issue 2, 2019

5. References :

[1] "Machine Learning-Based Sinkhole Detection in Mobile Ad Hoc Networks" by Abdalaziz Al-Qudah and Muhammad Aslam (2019)

[2] "An Efficient Machine Learning Based Sinkhole Detection Approach in Mobile Ad-hoc Networks" by Ahmad Ahmad and Sherif Alsaleh (2017)

[3] "A Machine Learning Based Sinkhole Detection Framework for Mobile Ad-hoc Networks" by Abdalaziz Al-Qudah, Amjad Al-Karaki, and Muhammad Aslam (2019)

[4] "An Anomaly Detection System for Sinkhole Attack Detection in Mobile Ad Hoc Networks" by Shiven Chawla and Anmol Grover (2015)

[5] "Deep Learning-based Approach for DDoS Attacks Detection and Mitigation in 5G and Beyond Mobile Networks" by Y. Hu, W. Li, Z. Zhou, and Y. Ma (2013)

[6] “Detecting and Tracking Sinkholes Using Multi-Level Convolutional Neural Networks and Data Association” by Hoai Nam Vu, Cuong Pham, Nguyen Manh Dung and Soonghwan Ro (2018)

[7] “Performance Analysis of Machine Learning-based Detection of Sinkhole Network Layer Attack in MANET” by SN and K Archana (2018)

[8] “A comprehensive survey on detection of sinkhole attack in routing over low power and Lossy network for internet of things” by Aya Abdul Rahman, Al chikh Omar, Bassel Soudan and Ala Altaweel (2016).

[9]”Bangui, Hind & Ge, Mouzhi & Buhnova, Barbora.” (2014). A hybrid machine learning model for intrusion detection in VANET. Computing. 104. 10.1007/s00607-021-01001-0.