International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Geometric Erudition Intended for Variance Recognition within Shade Server Organization: A Multi-Order Markov Series Structure

[1] Dr. S. Sree Hari Raju, [2] Dr. K. Srinivas Babu, [3] Dr. K. Rameshwaraiah, [4] Priyanka Pandarinath, [5] Mittapalli Sonia

Submitted: 05/05/2024 **Revised**: 17/06/2024 **Accepted**: 23/06/2024

Abstract*:* To enhance the security of IT infrastructure, especially in cloud computing platforms that host entire applications and data, anomaly detection plays a critical role. This paper introduces a multi-order Markov chain-based framework for anomaly detection, expanding upon the traditional Markov chain model. Our approach incorporates both high-order Markov chains and multivariate time series, embedded within algorithms designed under a statistical learning framework. To address time and space complexity, our algorithms employ non-zero value tables and logarithmic values in initial and transition matrices. For validation purposes, we utilized system calls and their return values from the DARPA intrusion detection evaluation dataset to construct a two-dimensional test input set. The test results indicate that our multi-order approach produces more effective indicators. Both the absolute values derived from single-order models and the changes in rankings across different-order models are strongly correlated with abnormal behaviors, thereby enhancing the detection of anomalies.

Keywords: *Markov processes, training methodologies, cloud computing, algorithm design and analysis, mathematical modeling, equations, and server management*

1. Introduction

A growing number of academic and industrial users are increasingly relying on cloud computing servers to host entire applications and storage solutions. These distributed and open-structured cloud services have become prime targets for potential threats. Consequently, ensuring the security and availability of both business and personal data on these servers is crucial. This safety is vital for both individuals and society. However, during catastrophic events such as intrusions, crashes, or breakdowns, anomalies must be detected promptly to apply effective remedies. Early-stage issues often lack clear indicators, leading to delayed responses and potentially irreversible damage.

Fortunately, server behavior generally follows consistent statistical patterns, which underpins anomaly detection algorithms using machine learning or data mining. These algorithms extract patterns from extensive training datasets and identify deviations to trigger alerts. Anomaly or intrusion detection methods are typically classified into three categories: statistical approaches, machine learning techniques, and data mining methods.

Statistical Approaches: These methods observe the behavior of objects to create statistical distributions during the training phase, forming a set of trained profiles. During detection, new profiles are compared against these trained profiles, and anomalies are identified if the profiles do not match. Incidents that deviate significantly from statistical norms trigger alarms.

Machine Learning Approaches: These methods minimize the need for extensive supervision during the training phase. Machine learning systems autonomously learn and improve their performance, adapting based on feedback from execution results, such as system call sequence analysis, Bayesian networks, and Markov models. Techniques like neural networks and Hidden Markov Models have proven effective.

Data Mining Approaches: These methods uncover hidden rules and patterns by analyzing large datasets collected either online or offline. Anomaly

[1]*Associate Professor*
[2]*Professor*
[3]*Head & Professor,*
[4]*Assistant Professor*
[5]*Assistant Professor*
Department of Computer Science & Engineering
Nalla Narasimha Reddy Education Society's Group of Institutions

detection systems benefit from additional insights such as hidden patterns, associations, changes, and events found in the data. Common data mining techniques include classification, clustering, and outlier detection. Methods like K-nearest neighbor, clustering, and association rule discovery are commonly applied in anomaly detection.

In this work, we focus on statistical and machine learning approaches. We utilize classic Markov model theories to detect anomalous patterns in systems, leveraging the ordering property of events. The high-order Markov chain, as introduced by Ju and Vardi, is a key component, and we address the high computational costs with methods such as hybrid models and support vector machines.

Time series of system calls are recognized as powerful tools for identifying system behavior. System calls, due to their privileged nature, have been extensively researched for intrusion detection by analyzing and modeling audit data. In 1998, MIT Lincoln Laboratory’s Cyber Systems and Technology Group conducted a seven-week simulation of intrusions, releasing the data as the DARPA Intrusion Detection Evaluation Data Set.

Our major contribution is a multi-order Markov chain approach, which significantly improves upon single-order models at a reasonable cost. This approach, initially used in rainfall modeling, is effective for anomaly detection, with the relative ranking of probabilities from multi-order models serving as an indicator for anomalies. Ascending order indicates normal behavior, while descending order signals anomalies. Our approach differs from recent models that use n-gram transitions to analyze HTTP requests rather than system calls. We also consider a new input category: the return values of system calls, forming a two-dimensional model to enhance the effectiveness of the multi-order Markov chain-based approach. Traditional methods that use system calls to identify behavior are limited as they do not consider execution outcomes. Recent studies have begun incorporating return values to better detect and interpret anomalies.

Cloud computing has become an indispensable part of networked computer systems, despite the well-known risks associated with its security and reliability. Many academic and industrial users are now relying exclusively on cloud computing servers to host entire applications and storage solutions. This reliance on cloud computing services, which are inherently distributed and open in structure, makes them clear targets for potential threats. As a result, ensuring the security and availability of both business and personal data on these servers is crucial. The invulnerability of these servers is of paramount importance to individuals and society alike. System calls and their corresponding return values are extracted from the classic DARPA intrusion detection evaluation dataset to create a two-dimensional test input set. Testing results demonstrate that the multi-order approach generates more effective indicators: in addition to the absolute values provided by individual single-order models, changes in the ranking positions of outputs from different-order models also closely correlate with abnormal behaviors. We focus on statistical and machine learning-based approaches. Classic Markov model theories are applied to detect anomalous patterns in the system, using the ordering property of events as proposed by Ju and Vardi. They introduce the high-order Markov chain as an extension. Several approaches are then introduced to mitigate the high computational cost associated with this model, including hybrid models and support vector machines.

Overview:

In statistical approaches, anomaly or intrusion detection systems typically observe the behaviors of objects to create statistical distributions during the training phase, forming a set of trained profiles. These systems then compare this set of trained profiles against a new set of profiles of observed objects during the detection phase. An anomaly or intrusion is detected if these two sets of profiles do not match. Generally, any incident whose occurrence frequency exceeds standard deviations from statistical normal ranges triggers an alarm.

2. Literature Survey

A Program-Based Anomaly Intrusion Detection Scheme Using Multiple Detection Engines and Fuzzy Inference

This work introduces a hybrid anomaly intrusion detection scheme that utilizes program system calls. It combines a hidden Markov model (HMM) detection engine with a normal database detection engine to exploit their respective strengths. A fuzzy inference mechanism is implemented to create a soft boundary between anomalous and normal behavior, which can be difficult to distinguish when they overlap. To address the high computational cost associated with HMM training, an incremental HMM training method with optimal parameter initialization is proposed.

Alert Correlation in Collaborative Intelligent Intrusion Detection Systems—A Survey

Since it is impossible to prevent all computer attacks, intrusion detection systems (IDSs) are essential for minimizing damage from various attacks. IDS methods include misuse-based and anomaly-based detection. A collaborative intelligent intrusion detection system (CIIDS) that integrates both methods is proposed, as individual detection engines often fall short. This survey reviews the architecture and alert correlation algorithms of current collaborative intrusion detection systems (CIDSs). It explains and compares different CIDS architectures, highlighting the issues and challenges in alert correlation when multiple security systems are used. Various techniques for alert correlation are discussed, focusing on CIIDS alerts. Computational intelligence approaches and their applications in IDSs are reviewed, with soft computing methods providing understandable and autonomous solutions.

Interpreting Chance for Computer Security by Viterbi Algorithm with Edit Distance

This work highlights the importance of chance discovery in computer security. Traditional methods often detect anomalies without interpreting them. This study uses the Viterbi algorithm to analyze state sequences and assess the distance between the standard model of anomaly types and the state sequence of discovered anomalies. Since environmental factors can cause inconsistencies in state sequences, the edit distance is used to measure the distance effectively.

A Multi-Order Markov Chain-Based Scheme for Anomaly Detection

This research introduces a scheme for anomaly detection in server systems based on multi-order Markov chains. It utilizes high-order Markov chains and multivariate time series, along with detailed training and testing algorithms. System calls and their corresponding return values from the DARPA Intrusion Detection Evaluation Data Set form a two-dimensional input set. The results demonstrate that this method effectively identifies anomalies. In addition to the absolute values provided by single-order models, the changes in ranking positions from different-order models also closely correlate with abnormal behaviors.

Spectrogram: A Mixture-of-Markov-Chains Model for Anomaly Detection in Web Traffic

This work presents Spectrogram, a machine learning-based statistical anomaly detection sensor designed to protect against web-layer code-injection attacks such as PHP file inclusion, SQL injection, and cross-site scripting, as well as memory-layer exploits like buffer overflows. Unlike malcode samples, statistical anomaly detection sensors are driven by the data being protected. This work introduces a new model and sensor framework that balances accuracy and false-positive rates. Spectrogram dynamically reconstructs content flows from assembled packets and learns to recognize legitimate web-layer script input using a mixture of Markov chains and a corresponding training algorithm.

Modeling Rain Risk: A Multi-Order Markov Chain Model Approach

This study explores the optimal frequency description of a chain-dependent Markov process for daily precipitation simulation. It compares a mixed-order model to a first-order model to assess its effectiveness in pricing a rainfall index put option. Using Bayesian information criteria to analyze rainfall data, the study determines the monthly varying order of the mixed-order model. The findings show that while the mixed-order model offers a slight improvement in representing rain statistics compared to the first-order model, significant differences emerge when comparing daily simulations to the burn method for pricing a rainfall index put option.

Evaluating Intrusion Detection Systems: The 1998 DARPA Off-Line Intrusion Detection Evaluation

An intrusion detection evaluation test bed was created to simulate normal traffic akin to that on a government site, involving hundreds of users and thousands of hosts. Over seven weeks of training data and two weeks of test data, more than 300 instances of 38 different automated attacks were launched against UNIX hosts. Six research groups participated in a blind evaluation, with results analyzed for probe, denial-of-service (DoS), remote-to-local (R2L), and user-to-root (U2R) attacks. The findings indicated that the best systems detected old attacks at moderate rates but struggled with new and novel R2L and DoS attacks.

Profiling Program Behavior for Anomaly Intrusion Detection Based on the Transition and Frequency Property of Computer Audit Data

Intrusion detection is a critical component of the defense-in-depth strategy for network security. This paper introduces two methods for profiling normal program behavior using computer audit data: Hidden Markov Models (HMM) and Self Organizing Maps (SOM). The HMM method leverages the transition properties of events, while the SOM method focuses on frequency properties. Both methods were evaluated using CERT synthetic Sendmail system call data and Live FTP system call data. The HMM method demonstrated strong detection performance but required significant computational resources. Conversely, the SOM method, designed for real-time intrusion detection, efficiently processed large data volumes with minimal computational overhead.

3. Existing Methods for Anomaly or Intrusion Detection

Anomaly or intrusion detection is typically implemented using various approaches, classified into three main categories: statistical, machine learning, and data mining approaches.

Statistical Approaches

Statistical methods for anomaly detection involve observing the behavior of objects to generate statistical distributions during the training phase. These distributions form a set of trained profiles, which are compared against new profiles of observed objects during the detection phase. An anomaly or intrusion is detected when there is a discrepancy between the trained and new profiles. Generally, any incident whose occurrence frequency deviates significantly from the normal statistical range triggers an alarm.

Machine Learning Approaches

Machine learning methods aim to minimize the supervisory requirements associated with training phases in statistical methods. These systems are designed to autonomously learn and improve their performance. They generally function within a framework that continuously adapts its strategies based on feedback, which may include analyses of system call sequences, Bayesian networks, and results from Markov models. Techniques such as neural networks and Hidden Markov Models (HMMs) have demonstrated effectiveness in this area.

Data Mining Approaches

Data mining techniques focus on discovering hidden patterns and rules by analyzing extensive datasets collected both online and offline. These systems benefit from additional insights provided by identifying hidden patterns, associations, changes, and events within the data.. Data Mining Approaches Data mining approaches focus on uncovering hidden patterns and rules from large datasets gathered both online and offline. Key technologies used in these approaches include: - Classification: Categorizing data into predefined classes. - Clustering: Grouping similar data points together. - Outlier Detection: Identifying data points that significantly differ from the norm.

- Association Rule Discovery: Finding relationships between variables in large datasets.

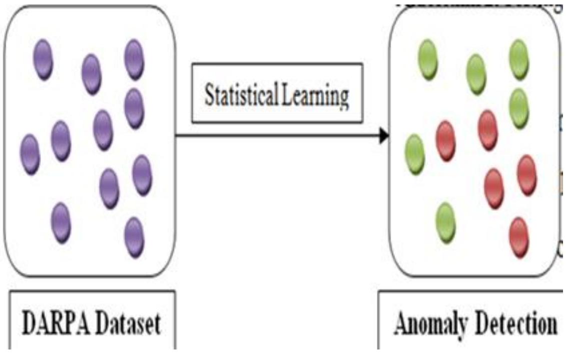
Common algorithms used in data mining include K-nearest neighbor and various clustering techniques.

Disadvantages of Existing Methods

A notable drawback of current anomaly detection methods is their often limited accuracy in identifying anomalies.

4. Proposed System

The proposed system leverages a multi-order Markov chain framework for detecting anomalies, integrating high-order Markov chains and multivariate time series within a statistical learning model. The system's algorithms are designed to address time and space complexity efficiently by employing non-zero value tables and logarithmic values in the initial and transition matrices.



To validate the approach, system calls and their return values from the DARPA Intrusion Detection Evaluation Data Set are used to create a two-dimensional test input set.

Advantages of the Proposed System

The main advantage of the proposed system is its enhanced accuracy in identifying anomalies.

5. Implementation

Load Dataset:

1. Module Description

This module handles the DARPA Intrusion Detection Dataset, which includes a variety of training sequences. The system utilizes these sequences for both training and testing purposes to identify anomalies.

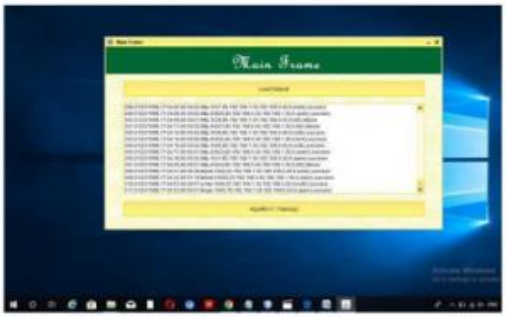
2. Training

3. Algorithm

1—designated as "Training"—processes the training dataset to generate two key matrices: the initial probability distribution matrix and the transition probability distribution matrix. The transition matrix is the main output of this training phase.

4. Testing

Algorithm 2—referred to as "Testing"—uses the transition probability matrices generated during training and applies them to the test dataset to calculate the likelihood of occurrence according to the trained model. Both algorithms rely on two functions: "Increase Transition Matrix" and "Get Transition Matrix." These functions are responsible for updating transition probabilities and retrieving values from the matrices, respectively.



6. Conclusion

This work introduces a multi-order Markov chain-based framework for anomaly detection, which enhances the detection process by examining the relative relationships among results from various-order models. This method provides a new and effective means of identifying anomalies. Given the regular and periodic nature of cloud server behaviors, if the probability of a test set under a lower-order model surpasses that of a higher-order model, it indicates the potential presence of unusual events, prompting further investigation.

Incorporating multi-dimensional, interrelated sequences into a unified multivariate model could improve detection sensitivity. The integration of return value series with traditional system call data, as previously demonstrated, adds another layer of analysis. The Training and Testing algorithms are designed to be efficient in terms of time and space, reducing the likelihood of the system itself generating anomalies and supporting online (or real-time) detection. The training phase completes in under 15 seconds for a dataset of 1.6 million records, even with up to third-order models combined

7. Future Enhancements

To enhance efficiency, strategies like constructing equivalent spaces by introducing an artificial state or employing binary representation for sparse matrices could substantially decrease space complexity.

References

[1] Improving product marketing by predicting early reviewers on E-Commerce websites

[2] S. Kodati, M. Dhasaratham, V. V. S. S. Srikanth, and K. M. Reddy, “Improving product marketing by predicting early reviewers on E-Commerce websites,” Deleted Journal, no. 43, pp. 17–25, Apr.

2024, doi: 10.55529/ijrise.43.17.25.

[3] Kodati, DrSarangam, et al. “Classification of SARS Cov-2 and Non- SARS Cov-2 Pneumonia Using CNN.” Journal of Prevention, Diagnosis and Management of Human Diseases(JPDMHD)2799-1202,vol.3,no.06,23 Nov. 2023, pp. 32–40,

[4] journal.hmjournals.com/index.php/JPDMHD/article/view/3406/2798, https://doi.org/10.55529/jpdmhd.36.32.40. Accessed 2 May 2024.

[5] V. Srikanth, “CHRONIC KIDNEY DISEASE PREDICTION USING MACHINE LEARNING ALGORITHMS,” IJTE,pp.106–109,Jan.2023,[Online].

[6] Available: http://ijte.uk/archive/2023/CHRONIC-KIDNEY-DISEASE-PREDICTION-USING-MACHINE-LEARNING-ALGORITHMS.pdf

[7] V. SRIKANTH, “DETECTION OF PLAGIARISMUSINGARTIFICIAL NEURAL NETWORKS,” International JournalofTechnologyandEngineering,vol. XV, no.I, pp.201–204,Feb.2023, [Online].

[8] Available: http://ijte.uk/archive/2023/DETECTION-OF-PLAGIARISM-USING-ARTIFICIAL-NEURAL-NETWORKS.pdf

[9] V. SRIKANTH, “A REVIEW ON MODELING AND PREDICTING OF CYBER HACKING BREACHES,” IJTE, vol.XV,no.I,pp.300–302,Mar.2023,[Online]. Available:

[10] http://ijte.uk/archive/2023/A-REVIEW-ON- MODELING-AND-PREDICTING-OF- CYBER-HACKING-BREACHES.pdf

[11] S. Kodati, M. Dhasaratham, V. V. S. S. Srikanth, and K. M. Reddy, “Detection of fake currency using machine learning models,”DeletedJournal,no.41,pp.31–38, Dec. 2023, doi: 10.55529/ijrise.41.31.38.

[12] “Cyberspace and the Law: Cyber Security.” IOK STORE, iokstore.inkofknowledge.com/product-page/cyberspace-and-the-law. Accessed 2 May2024.

[13] “DataStructuresLaboratoryManual.” IOK STORE, www.iokstore.inkofknowledge.com/product-page/data-structures-laboratory-manual. Accessed 2 May 2024.

[14] Data Analytics Using R Programming Lab.” IOK STORE, www.iokstore.inkofknowledge.com/product

[15] -page/data-analytics-using-r-programming-lab. Accessed 2 May 2024.

[16] V. Srikanth, Dr. I. Reddy, and Department of Information Technology, Sreenidhi Institute of Science and Technology, Hyderabad, 501301, India, “WIRELESS SECURITY PROTOCOLS (WEP,WPA,WPA2&WPA3),”journal-article, 2019. [Online]. Available: https://www.jetir.org/papers/JETIRDA0600 1.pdf

[17] V. SRIKANTH, “Secured ranked keyword search over encrypted data on cloud,” IJIEMR Transactions, vol. 07, no. 02,pp.111–119,Feb.2018,[Online]. Available: https://www.ijiemr.org/public/uploads/paper /1121_approvedpaper.pdf

[18] V. SRIKANTH, “A NOVEL METHOD FOR BUG DETECTION TECHNIQUES USING INSTANCE SELECTION AND FEATURESELECTION,”IJIEMR

[19] Transactions, vol. 06, no. 12, pp. 337–344, Dec. 2017, [Online]. Available: https://www.ijiemr.org/public/uploads/paper /976_approvedpaper.pdf

[20] 12 . SRIKANTH MCA, MTECH, MBA, “ANALYZING THE TWEETS AND DETECT TRAFFIC FROM TWITTER ANALYSIS,”Feb.2017.[Online].

[21] Available: http://ijmtarc.in/Papers/Current%20Papers/I JMTARC-170309.pdf