DOI: 10.24996/ijs.2025.66.3.28

ISSN: 0067-2904

Initial Disclosure of Insiders Based on Expert Rules

Buraq Almusawy, Ali Abdulkarem Habib Alrammahi[*]

Faculty of Computer Science & Mathematics, Kufa University, Najaf, Iraq

Received: 14/10/2023 Accepted: 31/3/2024 Published: 30/3/2025

Abstract

Recently, the number of insiders of computer networks in companies has increased. Some insiders can be detected when they perform outside activities, such as sending a file or opening blocked websites, to enter the company network. Another class of insiders is the company's employees, and it is difficult to identify them. And here lies the problem—the taxonomies of insiders and insider threats. The survey method is one of the most widely used approaches to constructing a taxonomy of insiders. This method is based on the analysis of materials from investigations of computer security incidents conducted by computer security specialists. Based on the incidents investigated, it is possible to categorize studies using technical and psychosocial data. User data on networks requires detailed preliminary analysis to study user behavior and identify insiders more accurately. In this research, we proposed a model to detect insiders when making any events on networks like the open blocked site, sending emails, and logging into the network at a suitable time. The proposed algorithm needs to analyze a user data set based on the No SQL language and then define expert rules to determine the degree of risk for insiders. Analysis of the proposed algorithm regarding time, accuracy, and correctness of the insider classification led to satisfactory results.

Keywords: CERT Insiders Dataset, NoSQL, Prediction model, Expert rules.

الكشف التمهيدي عن المطلعين بناءً على قواعد الخبراء

براق الموسوي, علي عبد الكريم حبيب الرماحي[*]

قسم علوم الحاسوب كلية علوم الحاسوب والرياضيات، جامعة الكوفة، النجف، العراق

الخلاصة

في الآونة الأخيرة، زاد عدد المطلعين على شبكات الحاسوب، في الشركات أشخاص من الداخل يمكن اكتشافهم عند قيامهم بأنشطة خارجية، مثل إرسال ملف أو فتح المواقع المحجوبة للدخول إلى شبكة الشركة . فئة أخرى من المطلعين، وهم موظفي الشركة ومن الصعب التعرف عليهم، وهنا تكمن المشكلة .تصنيفات المطلعين والتهديدات الداخلية .حاليًا، إحدى الطرق الأكثر استعمالاً على نطاق واسع لإنشاء تصنيف للمطلعين هي طريقة الدراسة .تعتمد هذه الطريقة على تحليل المواد الناتجة عن التحقيقات في حوادث أمن الحاسوب التي أجراها متخصصون في أمن الحاسوب .بناءً على الحوادث التي تم التحقيق فيها، من الممكن تصنيف الدراسات واستعمال البيانات التقنية والنفسية الاجتماعية .تتطلب بيانات المستخدم الموجودة على الشبكات تحليلًا أوليًا تفصيليًا لدراسة سلوك المستخدم حتى نتمكن من تحديد المطلعين بشكل أكثر دقة .في هذا البحث قمنا باقتراح نموذج لاكتشاف المطلعين بعد إجراء أي أحداث على الشبكات مثل فتح موقع

[*] Email: alia.alramahi@uokufa.edu.iq

محظور ، إرسال بريد وتسجيل الدخول إلى الشبكة في وقت غير مناسب .تحتاج الخوارزمية المقترحة إلى تحليل
مجموعة بيانات المستخدم بناءً على لغة No SQL ثم تحديد قواعد الخبراء لتحديد درجة خطورة المطلعين .
أدى تحليل الخوارزمية المقترحة من حيث الوقت والدقة ودرجة صحة التصنيف الداخلي إلى نتائج مرضية.

1. Introduction

Modern insider attacks are complex and use a variety of implementation methods and attack vectors to gain unauthorized access and compromise information objects on the internal network. An insider can be any network user. Therefore, performing procedures for analyzing and monitoring user actions in attack protection systems, called user behavior profiling, is necessary.

Existing research and development refers to these procedures as user behavior analytics (UBA) and user and entity behavior analytics (UEBA) [1, 2, 3]. Formally, UBA and UEBA systems belong to the same class of systems, but there is one fundamental difference between them. UBA systems use information containing only data about user activity. Therefore, they focus on users and their roles. UEBA systems and the data used in UBA systems consider information about the system environment (network traffic, storage systems, workstations, and software). This allows UEBA systems to profile not only users but also the state of software and hardware. This allows UEBA systems to recognize a broader class of threats [4, 5]. To implement UBA and UEBA, you need a database management system (DBMS) that can quickly scale and has high query processing speeds. Currently, people use NoSQL DBMSs, not just SQL ones, for this purpose [6, 7, 8, 9]. NoSQL-based solutions provide a scalable and flexible way to solve problems previously managed by relational databases. An example of a NoSQL DBMS is Orient DB [10, 11], which combines the capabilities of document-oriented and graph-oriented databases (DBs). It has full graphics capabilities and features typically found only in document databases.

2. Related Work

Taxonomies of Insiders and Insider Threats The survey method is one of the most widely used approaches to constructing a taxonomy of insiders. This method is based on the analysis of materials from investigations of computer security incidents conducted by computer security experts. Based on the investigated incidents, it is possible to categorize the studies using technical and psychosocial data.

Let's look at a list of surveys related to insider threats. Salem [12, 13] introduced a taxonomy of attackers, dividing them into two categories based on their knowledge of the target system: traitors and masquerades. When reviewing the literature on insider detection, one can divide the work into three types based on the approaches: host-based user profiling approaches, network-level approaches, and integrated approaches. Network layer and host-based profiling can have a high probability of detecting traitors, while host-based user profiling can successfully identify hidden threats. The authors argue that malicious insider activities occur at the application and business process levels. Hunker and Probst [14] proposed a research categorization based on combining psychosocial source data with technical data. The resulting categories consist of three types of approaches to detecting threats from insiders:

1) sociological, psychological, and organizational.
2) social-technical
3) technical.

The authors emphasized that successful insider threat detection techniques require combining different approaches.

Pfleeger [15] defines an insider as an individual with legal access to an organization's computers and networks. RAND Corp.

The term unintentional insider threat, according to [16], is defined as a current or former employee, contractor, or another business associate who: 1) has or had authorized access to an organization's network, system, or data; and 2) has no malicious intent associated with its actions (or inactions) that caused harm or significantly increased the likelihood of serious harm to the confidentiality, integrity or availability of information or information security of the organization in the future. According to [17], an unintentional insider threat is defined as inattentive, complacent, or untrained people who require permission and access to IS to do their job.

Network data monitoring and analysis must be integrated with decision-making algorithms that can adequately detect unusual occurrences to overcome insiders. Combining these methods may improve cybersecurity and network protection. EHO-FDMM, a new model, was suggested in this work. This framework includes capture, logging, pre-processing, and a novel EHO-FDMM-based IDS technique. The NSL-KDD and UNSW-NB15 datasets evaluate this methodology. Statistical analysis of network data helps identify the best model that matches the data. The EHO-FDMM-based intrusion detection approach has a lower FPR and higher DR than the other three robust methods. The EHO-FDMM and accurate interval of confidence constraints enabled the recommended approach to identify minute differences between lawful and attack routes. Correlations and proximity metrics are inadequate against modern attacks that mimic daily acts [18].

Philip A. Legg [19] aimed to overcome the main insider danger concern by verifying the system in a real business utilizing actual data. Only 44,000 of the 750,000 daily submissions were valid. Deployments lasted 31 days each. The author noted that high notifications from people each day caused many false positives. The algorithm identified a company-watch-listed employee as a danger. This indicated that the system may be accurate. The training data set determines the machine's insider threat detection. It has demonstrated the ability to reduce the search area of an event, yet it cannot supplant a human analyst. A visual analytic dashboard with four views—User Selection, Projection, Detail, and Feature—was included.

Kumar, V. S. [20] combined network data monitoring and analysis with decision-making algorithms that can correctly detect unusual events. Combining these approaches may improve cybersecurity and network protection. A new model, EHO-FDMM, is proposed in this work. This framework includes capture and recording, pre-processing, and a new IDS technology based on EHO-FDMM. The NSL-KDD and UNSW-NB15 datasets evaluate this methodology. Statistical analysis of network data helps determine the best model that matches the data. The EHO-FDMM-based intrusion detection method has a lower FPR and higher DR than the other three robust methods. EHO-FDMM and the precise confidence constraint interval enabled the recommended approach to identify the differences between legal and attack methods. Correlations and proximity metrics are insufficient against modern attacks that mimic everyday business.

J. Kim *et al.* [21] proposed insider-threat detection techniques built on anomaly detection algorithms and user behavior modeling. The daily activity summary of the user, the topic distribution of email contents, and the user's weekly email communication history are the three types of datasets that the researchers created using user log data. Next, in order to identify malicious activity, we used four anomaly detection methods in concert with each

other. According to experimental findings, the suggested architecture can function effectively with unbalanced datasets that lack domain experts' expertise and have limited insider risks.
N. Garba et al. [22] outlined a technique for detecting insider threats using anomaly detection algorithms and email user behavior in order to get around this restriction. The CERT r6.2 dataset is used, together with natural language pre-processing modules, to create email content based on the IT administrator job. To identify fraudulent email contents, anomaly detection algorithms use a vector space created by topic modeling the dataset. The suggested model has an 89% detection rate advantage over the baseline model, as shown by the experimental data. For 1%, 5%, 10%, 15%, 20%, 25%, and 30% cut-off values of anomaly scores, a combination of K-means and PCA anomaly detection.

W. Jiang et al. [23] proposed a user behavior analysis model by aggregating user behavior over a period of time, comprehensively characterizing user attributes, and then detecting internal attacks. Firstly, the user behavior characteristics are extracted from the multi-domain features extracted from the audit log, and then the XGBoost algorithm is used to train. The experimental results on a user behavior dataset show that the XGBoost algorithm can be used to identify insider threats. The value of F-measure is up to 99.96%, which is better than SVM and the random forest algorithm.

Q. Ma and N. Rastogi [24] introduced a novel approach that uses system logs to detect insider behavior using a special recurrent neural network (RNN) model. Ground truth is established using DANTE and used as a baseline for identifying anomalous behavior. For this, system logs are modeled as a natural language sequence, and patterns are extracted from these sequences. We create workflows of sequences of actions that follow natural language logic and control flow. These flows are assigned various categories of behaviors—malignant or benign. Any deviation from these sequences indicates the presence of a threat. We further classify threats into one of the five categories provided in the CERT insider threat dataset.
Through experimental evaluation, they show that the proposed model can achieve 93% prediction accuracy.

Insider threats are a significant issue for all corporations. Designing an effective mitigation strategy to combat the problem has been a recurring research issue. Machine learning techniques have been proffered to solve similar issues like anomaly detection (AD) and network intrusion detection (NID). This paper has reviewed non-machine learning and machine learning techniques for insider threat detection and concluded that the machine learning technique is a promising solution for insider threat detection.

3. Methodology

3.1. Dataset Analysis Using NoSQL

The research analyzes a data set on insider attacks in NoSQL format to detect intruders in the computer system [25, 26]. Based on the total information analyzed, they are then used to create user behavior profiles and identify behaviors that differ from regular ones. Then, based on this information, you can identify possible insiders and how they can implement unauthorized actions. The main goal is to show the possibility of creating and using an aggregated model for representing data about insider attacks in NoSQL format that takes into account user behaviors for subsequent use of this model to detect information security violators.

The Computer Emergency Response Team published the data set at Carnegie Mellon University in the file R4.2.tar.bz. [27]. This dataset contains information about usernames, computer names, URL requests with HTTP session timestamps, login data, devices used, and a list of modified files from 1000 employees over more than 17 years. Figure 1 presents a block diagram for collecting user attributes from collected information in the dataset, including the timestamp of logon, emails, websites (HTTP), files, and devices.

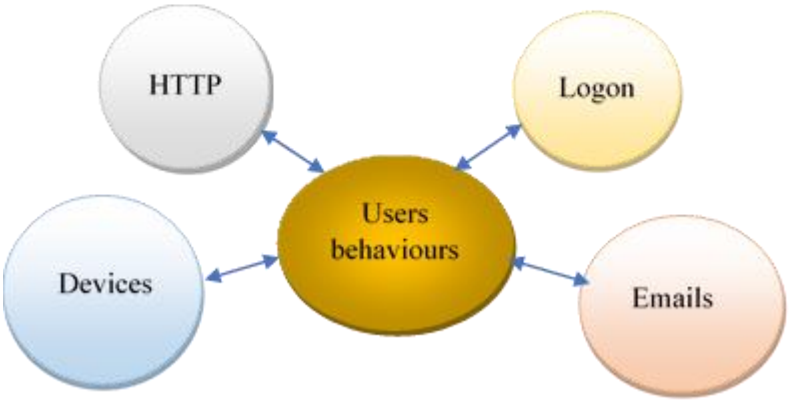

Figure 1: Set of attributes for user behavior

The formal form of the model for representing the dataset on insider attacks is as follows:

$$M = \langle A, I \rangle \quad (1)$$

where A represents attributes of user behaviors, and *I* is the insider model and criteria, which allow the current user to be classified as an insider.

Let us present the selected attributes of user behaviors and their relationships formally:

$$A = \langle Logon, Email, HTTP, Devices\rangle \quad (2)$$

Let's list the elements included in this tuple:

- ***Logon*** = ⟨*id, data, user, pc, Activity*⟩.

Logon to the network, each element representing user ID, login date, user name, device name, and activity type.

- ***Email*** = ⟨*id, data, user, pc, to, cc, from, size, attachments, content* ⟩.

Email sent, each element of which represents user ID, login date, user name, device name, email sent to, carbon copy, sender's email, size of the message, attachment files, and message content.

- ***HTTP***= ⟨*id, data, user, pc, URL, content*⟩.

Website, each element of which represents, respectively, user ID, login date, user name, device name, website link, and website content.

- ***Devices***= ⟨*id, data, user, pc, URL, content*.

External elements each represent, respectively, devices, user ID, login date, user name, device name, and activity type.

$$I = \langle R, L \rangle \quad (3)$$

Where R: attribute criteria consist of a set of characteristics by which a decision is made to classify a user as a set of insiders (for example, a regulated work schedule, permissible load on the network, and assessment of work with information resources). L: access levels that define user rights in the CS, a violation of which would mean potential insider activity.

Expert rules, listed in Table 1, apply the aggregation rules to the dataset using NoSQL, resulting in a CSV file for each user. This file includes the sites visited, the emails sent, the official login and logout times, and the external devices connected. This will be a preliminary step for use in the proposed algorithm that uses expert rules.

Table 1: Expert rules for users in a computer network

Collection	**Rules**
Logon	1- logon time (From 18:00 To 00:00) or 2- logon time (From 01:00 To 08:00) Activity should check when logon and when logoff
HTTP	Visiting dangerous/prohibited sites - in accordance with the list of banned URL addresses of the organization. http://wikileaks.org/Julian_Assange/assange/The_Real_Story_About_DTAA/Gur_Erny_Fgbel_Nobhg_QGNN1528513805.php, http://monster.com/WboUhagvat1180707852.html, http://lockheedmartin.com/WboUhagvat1636367808.htm , http://careerbuilder.com/WboUhagvat660170997.htm, http://hp.com/WboUhagvat1944152218.jsp, http://boeing.com/WboUhagvat1904327536.htm, http://raytheon.com/WboUhagvat343187784.jsp, http://hp.com/WboUhagvat1944152218.jsp, http://jobhunt.org/WboUhagvat919122234.html, http://northropgrumman.com/WboUhagvat572113271.aspx, http://harris.com/WboUhagvat1919385663.htm, http://simplyhired.com/WboUhagvat57469130.jsp, http://aol.com/jobs/WboUhagvat1963819229.jsp, http://yahoo.com/hotjobs/WboUhagvat752138490.php, http://jobhuntersbible.com/WboUhagvat1258877042.aspx
Device	User Connected or Disconnect To external devices From (10:00 To 17:00)
Email	Sent email to blocked mail Stephanie_C_Wells@raytheon.com , Jaime_Carey@raytheon.com, Adrienne-Osborne@boeing.com, ANG91@harris.com, Hayfa_Newman@raytheon.com, Kylie.H.Scott@northropgrumman.com, Beck-Grant@raytheon.com,GKP3@northropgrumman, Wang_Cohen@raytheon.com, Lee_R_Tyler@lockheedmartin.com, ALM5@raytheon.com, Aphrodite_B_Case@raytheon.com, BLG798@raytheon.com, Slater.Nigel@raytheon.com, Huber-Hyatt@harris.com, Harper-Sexton@harris.com, Middleton-Sean@northropgrumman.com, Justine_Pate@northropgrumman.com, EBB6218@lockheedmartin.com, Francis.P.Edwards@harris.com, Fiona.C.Parrish@lockheedmartin.com, Odette_D_Davis@northropgrumman.com, Paul-Acton@harris.com, Ivy_Shaw@hp.com, Key-Janna@hp.com, BKM6137@hp.com, BSM953@northropgrumman.com, Beasley-Flynn@hp.com, CIC4@northropgrumman.com, Selma-Burch@harris.com, Francis.Brian.Armstrong@dtaa.com, Frances.Alisa.Wiggins@dtaa.com,

To explain how the models work, we will give examples of displaying an insider (UI) and a legitimate user (UL) in them. Suppose the first user's insider activity is subsequent visits to suspicious sites (for example, to send processed data outside the organization’s perimeter). Accordingly, the second user can do the same, but in significantly smaller volumes - to conduct legal activities (for example, within the framework of official duties).

This type of attack can be detected in the insider model (described in terms of the NoSQL database), and the following criteria will be used:

Table 2: Example of data analysis presentation model fields in NoSQL format for an insider and a legitimate user

Key	Document
1	{"ID": "1", "date": "01-01-2020 18:40:00", "Sites": ["monster.com," "yahoo. iq", "google.iq"]}
2	{"ID": "2", "date": "01-01-2020 16:00:00", "Sites": ["yahoo. iq", "google.iq"]}

As seen from Table 2, the user associated with ID:2 (row with key 2) matches user UL because it does not satisfy insider's criteria: visits allowed sites. On the other hand, the user associated with ID: 1 (row with key 1) corresponds to user UI, since it meets the insider criteria: visits prohibited sites (for example, monster.com).

3.2 Insider detecting algorithm using expert rules

An algorithm based on expert rules was developed to detect insiders in computer science. The prerequisite for creating expert rules was a version of the standard security policy in the organization (in relation to user behaviors in computer networks), created based on experts' opinions in the field. The above security policy is just one possible example. However, the proposed approach to forming an algorithm can be applied to most actual policies.

The policy consists of a set of basic rules, each associated with criteria for determining compliance with the rules and the degree of criticality of non-compliance in Table 3.

Table 3: The Effectiveness of the Insider Based on the Degree of Risk

Degree of risk	Effectiveness of the insider
4	User shared info and damaged the network.
6	Anomalous session
8	Gathering information
10	Information leak
>10	Critical

Note: The first column in Table 3 displays a value that varies based on the user's activity. This suggests that the user may not complete all activities in a single session, leading to the provision of a description for each value that exceeds the risk threshold. An essential consequence of rules and criteria is that in each rule, compliance with one part of the criteria can be determined accurately, and compliance with another part can be determined with some degree of probability. Multiple users breaking a security policy need to keep these kinds of detections from happening, so insider detection algorithms look at a lot of computer network parameters that are hard to figure out and aren't related to the subject at hand. Based on the proposed information security policy and the proposed approach, the following algorithm was developed based on expert rules: The algorithm As shown in Figure 2, one can divide the algorithm into (1) the main execution branch, which selects the inspection area for applying the rules, and (2) subroutines, which detail individual rules.

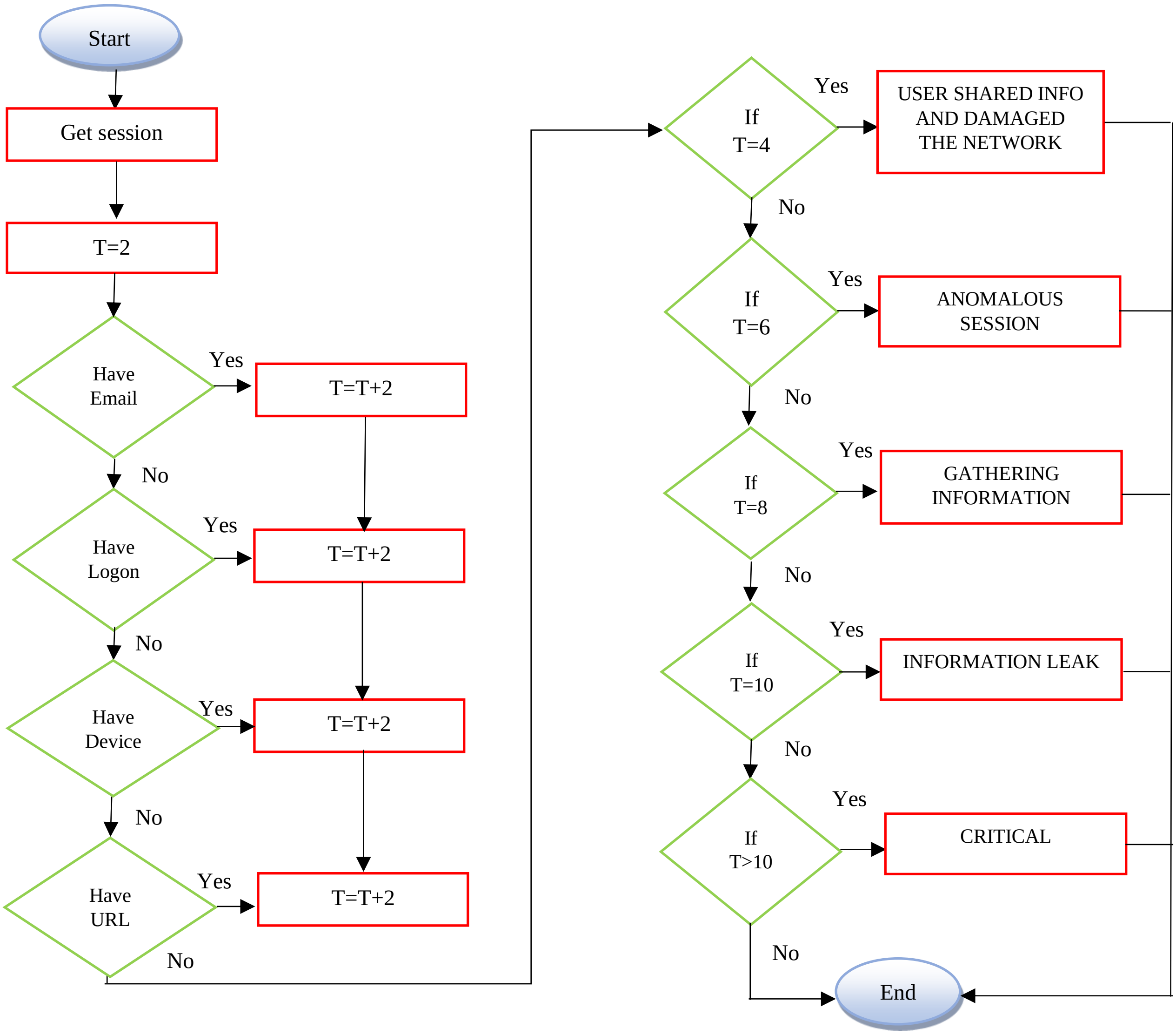

Figure 2: Flowchart of the main branch of the algorithm

Algorithm: Insider Detection using Expert Rules

```
STEP:1 Start
STEP:2 T = 2
STEP:3 Divide the session into a set of collections
STEP:4 If we have new sent emails then
Goto:10
STEP:6 If we have new logon then
Goto:11
STEP:7 If we have new device activity then
Goto:12
STEP:8 If we have new enter URL then
Goto :13
STEP:9 If user sent a message with an attached file to a blocked email then
T = T + 2
STEP:10 If user started logon out of work time, then
T = T + 2
STEP:11 If user has activity on external device, then
```

```
T = T + 2
STEP:12 If user entered a blocked website, then
T = T + 2
STEP:13 If T = 4 then
Output (USER SHARED INFO AND DAMAGED THE NETWORK)
STEP:14 If T = 6 then
Output (ANOMALOUS SESSION)
STEP:15 If T = 8 then
Output (GATHERING INFORMATION )
STEP:16 If T = 10 then
Output (INFORMATION LEAK)
STEP:17 If T >10 then
Output (CRITICAL)
Alarm
```

4. Results and Discussion

This paper used the CERT Insiders Dataset, which has 1000 users on a computer network [28]. NoSQL analyzes the dataset using four collections, aggregating each collection to identify suspension users, and then tests the results on both insiders and regular users, as illustrated in Table 4. We use the proposed algorithm to classify all users as insiders or not and determine the degree of risk for each.

Table 4: Sample of Suspicious Users

User Suspicious	Suspicious events
AAM0658	logon, {K3V4-Y4OK65SI-1583GEOQ},10/23/2010 01:34:19,AAM0658,PC-9923,Logon device, {H1L0-X7RH83FI-5967VUQY},10/23/2010 06:18:48,AAM0658,PC-9923,Connect HTTP,{Y4Q9-U5VQ11UG-1279ZPTL},10/23/2010 06:26:01,AAM0658,PC-9923,http://wikileaks.org/Julian_Assange/assange/The_Real_Story_About_DTAA/Gur_Erny_Fgbel_Nobhg_QGNN1528513805.php,spy bait bait distort evade surveillance deceit distort surveillance report subterfuge evade covert confidential covert top-secret clandestine Israel Russia lies deceit china forgery covert europe covert aisa confidential lie forgery clandestine currency surveillance europe bait surveillance confidential restricted spy currency handler restricted evade conspiracy subterfuge top-secret evade report isreal china 2010 handler spy forgery aisa chronicle Israel middle-east surveillance forgery device, {W7C0-G1SW41KB-1991BUTL},10/23/2010 06:26:48,AAM0658,PC-9923,Disconnect.
BBS0039	email,{Z1N2-F1FA87YB-4221ITSF},08/12/2010 10:24:05,BBS0039,PC-9436,Frances.Alisa.Wiggins@dtaa.com,Bevis.Brady.Sheppard@dtaa.com,21796,1,i may leave fed up complaints i work weekends too much company will suffer i work weekends i may leave my work not appreciated i work weekends i may leave i work holidays no gratitude complaints i work after-hours too much no gratitude my work not appreciated i work holidays complaints my work not appreciated fed up my work not appreciated i work weekends i work holidays fed up i work holidays my work appreciated no gratitude i work holidays my work not appreciated i work after-hours too much i work weekends i work after-hours company will suffer i may leave i work after-hours i work holidays company will suffer i work weekends no gratitude complaints
	logon, {T6E2-L5ZU34KY-2977OOSY},08/12/2010 19:08:21,BBS0039,PC-5866,Logon
	device, {H3R7-R1DD50PC-7819BOXS},08/12/2010 19:14:20,BBS0039,PC-5866,Disconnect
	logon, {L7X9-E0UA22KX-2664XTWN},08/13/2010 18:50:00,BBS0039,PC-5866,Logoff

Let's calculate the accuracy of insider detection algorithm results using the accuracy evolution model [29], as shown in Table 5.

Table 5: Confusion Matrix

True Positive (TP): • **Reality: Insider** • **Class: Insider** • **Number of TP results:68**	**False Positive (FP):** • **Reality: Legitimate** • **Class: Insider** • **Number of TF results:29**
False Negative (FN): • **Reality: Insider** • **Class: Legitimate** • **Number of FN results:2**	**True Negative (TN):** • **Reality: Legitimate** • **Class: Legitimate** • **Number of TN results:901**

Accuracy $=\frac{TP+TN}{TP+TN+FP+FN}=\frac{68+901}{68+901+29+2}=\mathbf{0.96}$

A detailed analysis of positives and negatives is needed to gain insight into our proposed algorithm's performance. Of the 1000 users on computer networks, 901 are legitimate (901 TN, 29 FP), and 70 are insiders (68 TP, 21 FN 2). Of the 70 insider users, the algorithm correctly classified 68 as insiders. That's a perfect performance. However, of 930 legitimate users, the proposed algorithm correctly determined 901 as legitimate and 29 as insiders. That's also good. The proposed algorithm has not been previously used for the purpose of insider detection; however, its results can be compared with machine learning and deep learning algorithms.

The accuracy of the internal detection algorithm proposed in our paper can reach 1% in many cases, compared to the artificial intelligence and anomaly detection algorithms mentioned in the literary overview, which rarely reach this accuracy, as shown in Table 6. Moreover, this algorithm is based on determining the user's risk.

Table 6: Compare results

Algorithm	Accuracy
anomaly detection algorithm	0.0710
anomaly detection algorithms and email user behavior	89%
recurrent neural network	93%
Insider detection based on Expert Rules	96%

Conclusion

In this paper, the dataset was analyzed using NoSQL to aggregate users' information on computer networks. The dataset used includes 1,000 users who have made multiple effective changes over a 12-month period. The use of expert rules by any company adds security to the network by restricting the user and increasing the accuracy of insider detection when any user breaks one of the rules or a group of them. We proposed a detection algorithm based on expert rules. This algorithm checks the user's behavior; whenever the user violates one of the rules set by specially established experts, he will be considered an insider. After that, the algorithm performs a set of tests to determine the insider's degree of risk by measuring the degree of risk, which was set by default to a maximum value of 10. The proposed algorithm represents the beginning of discovering a good group of insiders and classifying them to begin a final stage of decisive classification using artificial intelligence algorithms. The paper's goal was reached after applying the algorithm and analyzing the results.

Future work

In future work, it is expected that the algorithm proposed in our current research will be an expert system that contributes to supporting supervised machine learning algorithms. Machine learning algorithms are not devoid of a possible error rate, even if it is 1%. Moreover, classification algorithms can reveal to users whether they are insider threats or not, and they cannot distinguish the risk level for each insider. We can combine the results of the proposed algorithm with artificial intelligence algorithms through several methods, namely union, intersection, or taking the results of each algorithm separately.

References

[1] M. A. Salitin and A. H. Zolait, "The role of User Entity Behavior Analytics to detect network attacks in real-time," *2018 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, Sakhier, Bahrain, 2018, pp. 1-5, Doi: 10.1109/3ICT.2018.8855782.

[2] O. Polyakov, "UEBA (User and Entity Behavior Analytics) for when traditional Cyber Security can't protect your network," *Northforge Innovations*, pp. 1-7, 7 December 2017.

[3] Laiba Siddiqui, "User and Entity Behavior Analytics (UEBA) For Enterprise Security*," SPLUNK PRODUCTS & SOLUTIONS*, 24 October 2023. [Online]. Available: https://www.splunk.com/en_us/data-insider/user-behavior-analytics-ueba.html.

[4] Augusto Barros and Anton Chuvakin, "A Comparison of UEBA Technologies and Solutions," 29 March 2017. [Online]. Available: https://www.gartner.com/en/documents/3645381.

[5] J. Graves, "How machine learning is catching up with the insider threat," *Cyber Security: A Peer-Reviewed Journal*, vol. 1, no. 2, p. 127–133, 2017.

[6] **t**Kunda, D., & Phiri, H., "A Comparative Study of NoSQL and Relational Database," *Zambia ICT Journal*, vol. 1, No. 1, pp. 1–4, Dec. 2017. https://doi.org/10.33260/zictjournal.v1i1.8

[7] Ali Davoudian, Liu Chen, and Mengchi Liu, "A Survey on NoSQL Stores," *ACM Computing Surveys (CSUR),* vol. 51, no. 2, pp. 1–43, 2018, Doi: 10.1145/3158661.

[8] Garba, Musa, and Hassan Abubakar, "A Comparison of NoSQL and Relational Database Management Systems (RDBMS)," *KASU Journal of Mathematical Science*, vol. 1, no. 2, pp. 61-69, 2020.

[9] S. Venkatraman, K. Fahd, S. Kaspi and R. Venkatraman, "SQL Versus NoSQL Movement with Big Data Analytics," *International Journal of Information Technology and Computer Science (IJITCS)*, vol. 8, no. 12, pp. 59-66, 2016.

[10] C. Tesoriero, "Getting started with OrientDB," Birmingham, England: Packt Publishing, 2013.

[11] SAP developer , "OrientDB Manual - version 3.1.20," SAP, 09 Aug 2022. [Online]. Available: https://orientdb.com/docs/3.0.x/.

[12] M. Salem, S. Hershkop, S. Stolfo, "A survey of insider attack detection research," in *Insider Attack and Cyber Security. Advances in Information Security*, Boston, Springer, 2008, pp. 69-90, https://doi.org/10.1007/978-0-387-77322-3_5.

[13] M. Salem, S. Stolfo, "Masquerade attack detection using a search-behavior modeling approach," Columbia University, Computer Science Department, Columbia, Technical Report CUCS-027-09, 2009.

[14] J. Hunker, C. Probst, "Insiders and insider threats an overview of definitions and mitigation techniques," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications,* vol. 2, no. 1, pp. 4-27, 2011.

[15] S. L. Pfleeger, J. B. Predd, J. Hunker and C. Bulford, "Insiders Behaving Badly: Addressing Bad Actors and Their Actions," in *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 1, pp. 169-179, March 2010, Doi: 10.1109/TIFS.2009.2039591.

[16] M. Collins, M. Theis, R. Trzeciak, J. Strozer, J. Clark, D. Costa, T. Cassidy, M. Albrethsen, and A. Moore, "Common Sense Guide to Mitigating Insider Threats, Fifth Edition," *Carnegie Mellon University, Software Engineering Institute's Digital Library*. Software Engineering Institute, Technical Report CMU/SEI-2016-TR-015, 21-Dec-2016 [Online]. Available: https://doi.org/10.1184/R1/12890918.v1. [Accessed: 15-Nov-2023].

[17] D. Liu, X. Wang, L. Camp, "Mitigating inadvertent insider threats with incentives," in *International Conference on Financial Cryptography and Data Security*, Berlin, Heidelberg, 2009. https://doi.org/10.1007/978-3-642-03549-4_1.

[18] Kumar, V. S, "A Big Data Analytical Framework Intrusion Detection Based On Novel Elephant Herding Optimized Finite Dirichlet Mixture Models," *International Journal of Data Informatics and Intelligent Computing*, vol. 2, no. 2, pp. 11-20, 2023.

[19] P. Legg, "Human-Machine Decision Support Systems for Insider Threat Detection," in *Data Analytics and Decision Support for Cybersecurity. Data Analytics*, pp. 33–53, Springer, Cham, 2017.

[20] V. S. Kumar, "A Big Data Analytical Framework for Intrusion Detection Based On Novel Elephant Herding Optimized Finite Dirichlet Mixture Models.," *International Journal of Data Informatics and Intelligent Computing*, vol. 2, no. 2, p. 11–20, 2023.

[21] Kim, Junhong, Minsik Park, Haedong Kim, Suhyoun Cho, and Pilsung Kang, "Insider Threat Detection Based on User Behavior Modeling and Anomaly Detection Algorithms," Applied Sciences, vol. 9, no. 19, p. 4018., 2019. https://doi.org/10.3390/app9194018

[22] N. Garba, S. Rakshit, C. D. Mang, and N. R. Vajjhala, "An email content-based insider threat detection model using anomaly detection algorithms," in *Proceedings of the International Conference on Innovative Computing & Communication (ICICC)*, 2021.

[23] W. Jiang, Y. Tian, W. Liu, and W. Liu, "An insider threat detection method based on user behavior analysis," in *Intelligent Information Processing IX: 10th IFIP TC 12 International Conference, IIP 2018, Nanning, China, October 19-22, 2018, Proceedings 10*, 2018, pp. 421-429: Springer.

[24] Q. Ma and N. Rastogi, "DANTE: Predicting Insider Threat using LSTM on system logs," *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, Guangzhou, China, 2020, pp. 1151-1156, Doi: 10.1109/TrustCom50675.2020.00153.

[25] T. O. Oladimeji, C. K Ayo1 and S.E Adewumi, "Review on Insider Threat Detection Techniques," in *Journal of Physics: Conference Series*, vol. 1299, p. 012046, 2019. DOI 10.1088/1742-6596/1299/1/012046

[26] P. Martins, F. Sá, F. Caldeira, and M. Abbasi, "NoSQL: A Real Use Case," in *Advances in Intelligent Systems and Computing*, Cham: Springer International Publishing, 2022, pp. 231–243.

[27] M. Razu Ahmed, M. Arifa Khatun, M. Asraf Ali, and K. Sundaraj, "Literature review on NoSQL database for big data processing," *J. Eng. Technol*, vol. 7, no. 2, pp. 902-906, 2018.

[28] B. Lindauer, "Insider Threat Test Dataset," Carnegie Mellon University, 30-Sep-2020, 10.1184/R1/12841247.V1.

[29] H. M and S. M.n, “A review on evaluation metrics for data classification evaluations,” *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, no. 2, pp. 01–11, 2015.