



LSTM Autoencoder-Based Insider Abnormal Behavior Detection Using De-Identified Data

Seo-Yi Kim¹ ^a and Il-Gu Lee^{1,2} ^b

¹Department of Future Convergence Technology Engineering, Sungshin Women's University, Seoul, South Korea

²Department of Convergence Security Engineering, Sungshin Women's University, Seoul, South Korea

Keywords: De-Identification, LSTM Autoencoder, Abnormal Behavior Detection.

Abstract: Leakages of national core technologies and industrial secrets have occurred frequently in recent years. Unfortunately, because most of the subjects of confidential data leaks are IT managers, executives, and employees who have easy access to confidential information, more sophisticated theft is possible, and there is a risk of large-scale data leakage incidents. Insider behavior monitoring is being conducted to prevent confidential data leaks, but there is a problem with personal information being collected indiscriminately during this process. This paper proposes a security solution that protects personal privacy through a process of de-identifying data, while maintaining detection performance in monitoring insider aberrations. In the abnormal behavior detection process, a long short-term memory (LSTM) autoencoder was used. To prove the effectiveness of the proposed method, de-identification evaluation and abnormal behavior detection performance comparison experiments were conducted. According to the experimental results, there was no degradation in detection performance even when data was de-identified. Furthermore, the average re-identification probability was approximately 1.2%, whereas the attack success probability was approximately 0.2%, proving that the proposed de-identification method resulted in low possibility of re-identification and good data safety.

1 INTRODUCTION


Today, as science and technology increasingly become a competitive edge among nations, data leakages and theft incidents between countries or organizations occur more frequently. If national core technologies are leaked overseas, it can have a fatal impact on national security and the economy, lowering national competitiveness and further leading to cyber warfare between countries. Additionally, if the internal secrets of an organization are leaked, it can cause significant damage to the image of the organization and lead to loss of profits and competitive advantage, thereby hindering corporate sustainability (Goryunova et al., 2020). For this reason, countries and organizations are trying to protect confidential data and minimize damage caused by the leakage of industrial secrets.


However, recently, cases of internal data leaks by insiders have increased, emerging as a global security problem. Because insiders already have access

to the network and internal services, they can easily access confidential data, and thus, more sophisticated theft is possible (Abiodun et al., 2023). To solve the problem of confidential leakage by insiders, research is actively being conducted to detect abnormal behavior among insiders.

According to the 2023 Insider Threat Report (Gurukul, 2023), in 2022, approximately 35% of the total respondents reported that they experienced insider attacks 1 to 5 times, whereas approximately 8% reported that they experienced insider attacks more than 20 times. Figure 1 shows the percentage of insiders ranked by cybersecurity experts as posing the greatest security risk to their organizations. Privileged IT users/admins were approximately 60%, privileged business users/executives were approximately 53%, and other IT staff were 24%, indicating a high proportion of IT managers, employees, and executives with extensive access rights. These are all groups of users that have easy access to confidential or sensitive information within an organization.

For example, in May 2022, there was a case wherein a researcher who was working at Company A at the time transferred to Company B, a competitor,

^a  <https://orcid.org/0009-0004-4890-1972>

^b  <https://orcid.org/0000-0002-5777-4029>

and leaked the confidential information of Company A to their new workplace. This researcher was a senior manager in an advertising related product team in Company A who, after receiving a job offer from Company B, stole approximately 570,000 pages of source code and confidential product-related information using a personal external device in only a few minutes. A few weeks later, Company A, which became aware of the leak, filed a lawsuit against the researcher who took possession of the confidential information of Company A through a personal external device until they were ordered to stop.

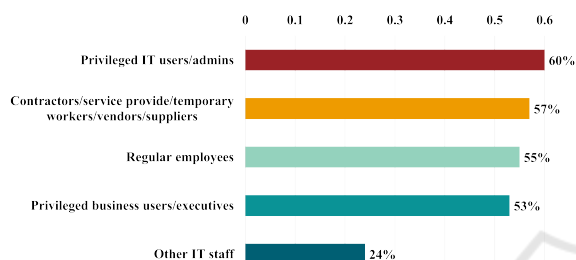


Figure 1: Types of insiders that pose the greatest security threat to an organization(Gurucul, 2023).

In July 2021, cybersecurity Company C was robbed of confidential sales support data by a former employee. The employee stole confidential data using a personal USB device before moving to competitor Company D. Company C built and used its own data loss prevention (DLP) solution but did not block internal staff from accessing, downloading, and sharing critical documents to external storage devices. A few months later, Company C discovered the data leak and sued the employee, but at that point, the leak may have already proven useful to the channel sales power of Company D, which recorded an increase in sales following the incident.

As the number of cases and scale of damage caused by confidential data leaks by insiders gradually increase, security solutions to prevent such incidents are increasingly being introduced within organizations. One of these methods is to monitor insider behavior using a user behavior analytics (UBA) tool. According to the 2023 Insider Threat Report, approximately 86% of organizations monitor insider behavior; however, their most utilized method is to monitor access logging only. The next most utilized method, employed by approximately 25% of organizations, is to monitor all actions of insiders 24/7. Figure 2 provides a graph of whether and to what extent insider behavior is monitored.

However, although monitoring all actions of insiders can be effective at detecting abnormal behaviors among insiders, it raises privacy infringement con-

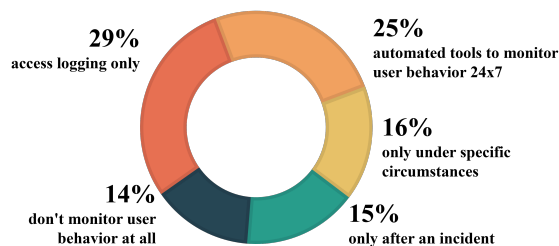


Figure 2: Percentages of organizations by whether and to what extent insider behavior is monitored (Gurucul, 2023).

cerns. In the process of monitoring all actions, sensitive information that can seriously harm the human rights of the users may be included, creating another type of security problem. Therefore, organizations must protect the privacy of insiders while quickly and accurately detecting abnormal behaviors among them to minimize the risk of leakage of confidential data and damage to the company.

In this paper, we aim to present a security solution that can protect the privacy of insiders through data de-identification in the process of detecting abnormal behavior among the insiders. After insider data are de-identified using the ARX Data Anonymization Tool (Koll et al., 2022), abnormal actors are detected using a long short-term memory (LSTM) autoencoder, an algorithm suitable for anomaly detection. Three attack models were used to evaluate the level of de-identification of the de-identified data, and the level of re-identification was evaluated by detecting insiders belonging to specific departments. The proposed method protects the privacy of insiders through de-identification while providing a similar abnormal behavior detection rate to that of the conventional method in the detection of abnormal actors.

The main contributions of this study are summarized as follows:

- After the insider data are de-identified to make it difficult to identify individuals, a high degree of de-identification, as evaluated by applying three attacker models, was obtained, as indicated by a probability of re-identification of approximately 1.2% and a probability of successful attack of approximately 0.2%. Thus, the ability of the proposed de-identification method to ensure the safety of the dataset was proven.
- Based on the results of detecting abnormal behavior among insiders using the LSTM autoencoder on the de-identified data, an accuracy of 94% and F1-score of 97% were achieved, showing similar abnormal behavior detection results to those provided by the corresponding identifiable data (original data).

This paper is structured as follows. Section

2 introduces background knowledge related to de-identification and analyzes existing research on abnormal behavior detection using machine learning. Section 3 explains the proposed technology, and then Section 4 describes the experimental environment in which to evaluate the de-identification level and abnormal behavior detection performance of the proposed technology. Section 5 provides a presentation and analysis of the experimental results. Finally, Section 6 presents the conclusions of the study.

2 RELATED WORK

2.1 De-Identification

As the transition to a digital society accelerates, the need for new regulations and innovations to protect personal privacy emerges. All areas of society are being digitized, and personal information and privacy data are being collected online and transmitted, used, and stored through networks. Various digitalized services provide convenience, but at the same time, they cause security problems such as personal information theft and privacy threats (Yun et al., 2023). Thus, concerns regarding the need to protect individual privacy continue to be raised, and laws in each country are being revised accordingly.

In Europe, the General Data Protection Regulation (GDPR) was enacted in 2019, stipulating that personal privacy data should be protected in all transactions occurring within EU countries (Li et al., 2023). GDPR specifies that non-identification measures such as pseudonymization and anonymization must be taken when personal information is used. Pseudonymization can be used for research and statistical purposes by ensuring that a specific individual can no longer be identified based on the data without additional information. In the United States, California's Consumer Privacy Act (CCPA) will be implemented starting in 2020, to which both domestic companies with business establishments in California and companies headquartered overseas are to be subject (Naim et al., 2023). CCPA defines "de-identification" as the use of technical/administrative protection measures to prevent re-identification, where the de-identified information is not included in personal information. As such, the data protection laws of many countries identify de-identification as a necessary step in collecting and utilizing personal information. Through appropriate de-identification, personal privacy can be protected while effectively utilizing personal information.

De-identification is the process of modifying or

replacing personal identifiers to hide some information from a public perspective (Chomutare, 2022). Non-identification will be subjected to a preliminary review process to review which data correspond to personal information. Subsequently, it includes a follow-up management process such as re-identifiability monitoring and safety measures after an adequacy assessment process to determine whether an individual can be easily identified when the data are combined with other information. Non-identification methods include kana processing, total processing or average value substitution, data deletion, categorization, and data masking. Table 1 shows a list of traditional methods of de-identification. Through this approach, data containing sensitive information are processed and then used for research or statistical indicators. Information loss must be minimized by selecting an appropriate de-identification method according to the data type and purpose. In this study, de-identification was performed using data reduction and data masking methods. In the "adequacy evaluation," which evaluates the possibility of re-identification after de-identification, methods such as k -anonymity, l -diversity, and t -accessibility are typically used.

- k -anonymity. k -anonymity is a non-identification model that prevents the identification of specific individuals by ensuring that there are more than k identical record values in the entire data set. If some of the information used is combined with other information that is publicly available to identify an individual, a linkage attack problem may occur. To compensate for this vulnerability, the k -anonymity model is used. In this way, attackers will not be able to find out exactly which record the attack target is from the de-identified data (Ito and Kikuchi, 2022).
- l -diversity. l -diversity is a de-identification model used to complement the vulnerabilities of k -anonymity. Records that are de-identified together must have at least l different pieces of sensitive information. For context, k -anonymity is vulnerable to identity attacks because it does not consider the diversity of information during de-identification. Additionally, it is vulnerable to attacks enabled by background knowledge because it does not consider the background knowledge of the attacker, other than the provided data. Therefore, it must be ensured that the de-identified data have more than l different pieces of data, to enable some degree of defense even in situations where the attacker possesses background knowledge (Rai, 2022).

Table 1: De-identification methods.

Method	Explanation
Pseudonymization	Replacing key identifying elements with other values to make it difficult to identify an individual. Examples include heuristic pseudonymization, encryption, and exchange methods.
Aggregation	Preventing individual data values from being exposed by replacing them with the total value of the data. Examples include total processing, partial totals, rounding, and rearrangement.
Data reduction	Deleting values that are unnecessary or serve as personal identifiers among the values included in the dataset depending on the purpose of data collection and the level of sharing and openness. Examples include deleting or partially deleting identifiers, deleting records, and deleting all identifiers.
Data suppression	Replacing data values with category values without directly exposing them. Examples include hiding, random rounding, range method, and control rounding.
Data masking	Replacing data values with category values without directly exposing them. Examples include hiding, random rounding, range method, and control rounding.

2.2 Abnormal Behavior Detection Using Machine Learning

Al-Mhiqani et al. (Al-Mhiqani et al., 2022) proposed a multi-layer framework for insider threat detection. In the first step, the levels of performance of nine machine learning models were evaluated using multi-criteria decision making (MCDM) to select a model optimized for insider threat detection. As a result of simulations, random forest and k-nearest neighbors (KNN) were selected. Based on these results, for the second step, hybrid insider threat detection (HITD), consisting of a misuse insider threat detection (MITD) component based on random forest and an anomaly insider threat detection (AITD) component based on KNN, was proposed. To evaluate its performance, the CERTr4.2 dataset was employed to test unknown and known insider attack scenarios. The evaluation indicators used were recall, accuracy, precision, area under the curve (AUC), F-score, and true negative rate (TNR). In terms of these measures, the proposed HITD method demonstrated the best performance. However, although the proposed method showed significant improvement in terms of detection performance, it did not consider the overhead and waiting delays that may occur when adopting a hybrid method. In addition, the original data were used as is after preprocessing. Because of this, there is a limitation in that there is a risk of infringing on individual privacy when the proposed method is applied to actual situations.

Cui et al. (Cui et al., 2021) observed that traditional federated learning, while effective for privacy protection and low latency, lacks stability because of non-uniform data distribution among distributed clients. Therefore, they proposed a blockchain-based

distributed asynchronous federated learning model for anomaly detection in an Internet of things (IoT) environment. The model mitigates the problem of imbalanced data distribution because it is stored on the blockchain rather than a central server, ensuring that all clients share the same model regardless of data distribution or quantity during model updates. Simultaneously, it efficiently addresses privacy concerns by storing only update information on the blockchain, without exposing sensitive data directly to a central server. For performance evaluation, a generative adversarial network (GAN) algorithm was used; the model demonstrated superior performance in terms of accuracy and convergence speed compared to those of traditional federated learning models. However, the model does have a number of limitations, such as significant accuracy variations in the learning rate settings, as observed in the IoT device learning rate comparison graph, and additional computational effort and time overhead in the process of setting the optimal learning rate.

Jamshidi et al. (Jamshidi et al., 2024) proposed a privacy enhancement model using an autoencoder structure to efficiently de-identify personal sensitive information when collecting data from providers during the anomaly detection model learning process. The data are first compressed using an encoder; the confidential and non-confidential attributes are separated and then passed through a classifier to weaken the correlation. Among them, the confidential attributes are de-identified by adding appropriate noise based on differential privacy and combined with the non-confidential attributes through a decoder, thus creating original data. To evaluate its performance, experiments were conducted using image datasets and categorical datasets, and on both datasets, the pro-

posed model exhibited better accuracy, precision, recall, and F1-score compared to those of the conventional autoencoder algorithms CelebA-G-M, CelebA-G-S, and CelebA-G-C. Their proposed model demonstrated excellent performance in the performance evaluation. Additionally, in a differential privacy-based noise optimization experiment, an appropriate value was obtained for an efficient de-identification parameter, which resulted in high accuracy. However, the process of adding noise to de-identify data may also increase the model complexity and computational overhead compared to those of a conventional autoencoder model.

According to a survey of previous studies related to detecting abnormal behavior using machine learning, there are two challenging aspects that must be considered: on one hand, privacy issues often occurred when data were not de-identified, and on the other hand, overhead was generated when de-identification was performed. In this paper, we would like to present an abnormal behavior detection solution that protects privacy by de-identifying data, while maintaining high detection performance by using an LSTM autoencoder.

3 LSTM AUTOENCODER-BASED INSIDER ABNORMAL BEHAVIOR DETECTION

Here in, an LSTM autoencoder-based abnormal insider behavior detection technology that uses de-identified data is proposed. This security solution can solve the problem of insider privacy infringement when abnormal-behavior monitoring is performed within an organization. When insider activity is monitored, the actions of all devices connected to the network are recorded. Before these raw data are used, they are subjected to a de-identification process to ensure that they do not contain sensitive information. Abnormal behaviors among insiders are then detected from the de-identified data using the LSTM autoencoder.

LSTM is an algorithm that complements the limitation of recurrent neural networks (RNN), which operate effectively only in short sequences, making it difficult to model dependencies in long sequences. By contrast, LSTM improves gradient propagation performance by adding cell-state to the state of the hidden layer. In addition, the four layers interact and operate, and in this process, the short-term state and long-term state are learned separately and undergo a merging and prediction process (Nguyen et al., 2021).

LSTM can process relatively long time-series data without performance degradation and can efficiently use memory by deleting data that are less relevant to prediction (Ashraf et al., 2020).

On the other hand, an autoencoder is a type of artificial neural network (ANN) model for image data compression. It consists of an encoder and decoder and creates a model in the form of compressed data by repeating the encoding and decoding process (Roelofs et al., 2021). The autoencoder is an unsupervised learning model that is actively considered in situations such as anomaly detection where there are only small amounts of labeled data (Thill et al., 2021).

An LSTM autoencoder is a model that applies the LSTM algorithm to the encoder and decoder of an autoencoder and is used for the dimensionality reduction and anomaly detection of a dataset (Said Elsayed et al., 2020). Because the detection of abnormal insider behavior is performed on large datasets with long time-series data, the LSTM autoencoder is suitable for the purpose (Nam et al., 2020).

Figure 3 shows a flowchart illustrating how abnormal insider behavior is detected using an LSTM autoencoder that uses de-identified data. After data generated from all PCs connected to the network of an organization are collected, preprocessing is performed to replace the collected data with numerical data suitable for machine learning use. Among the collected data, data that are personally identifiable and have a high possibility of violating privacy are deleted, whereas sensitive information are de-identified. After de-identification, the data are subjected to a risk assessment process to ensure that they are de-identified to an appropriate level. The data processed in this way are used as input data for the LSTM autoencoder to detect abnormal behavior. If abnormal behavior is determined, the abnormal actor is traced through a re-identification process.

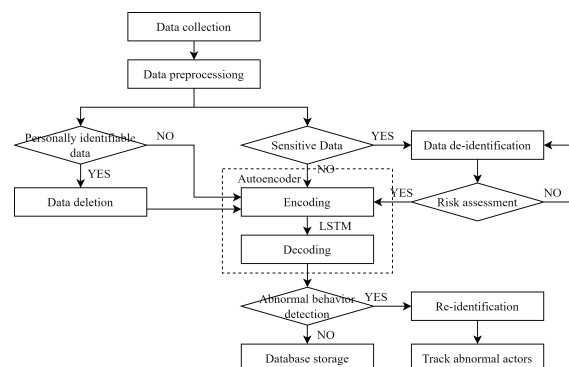


Figure 3: Flowchart of abnormal-insider-behavior detection based on LSTM autoencoder using de-identified data.

Table 2: CERT dataset configuration.

File	Contents	Characteristics
Logon .csv	PC login or logoff	<ul style="list-style-type: none"> • Fields: ID, Date, User, PC, Activity (Logon/Logoff) • 1,000 insiders each have an assigned PC. • The following items appear similar among users. <ul style="list-style-type: none"> - Start time (slight error) - End time (slight error) - Length of work day (slight error) - After-hours work: Most users do not log on outside of working hours.
Http .csv	Internet access history	<ul style="list-style-type: none"> • Fields: id, date, user, pc, url, content • URL includes the domain name and path. Words included in the URL are generally related to the content of the web page. • Each web page can contain multiple pieces of content.
File .csv	Copy files to removable media devices	<ul style="list-style-type: none"> • Fields: id, date, user, pc, filename, content • content: Consists of a hexadecimal encoded file header followed by a space-separated list of content keywords. • Each file can contain multiple topics. • File header is related to file name extension. • Each user has a normal number of file copies per day (deviations from these normal numbers can be used as an important indicator).
Email .csv	Incoming and outgoing emails	<ul style="list-style-type: none"> • Fields: id, date, user, pc, to, cc, bcc, from, size, attachment_count, content • Some noise edges are introduced. • A small number of insiders send emails to outsiders. • There may be multiple recipients. • Email size indicates the number of bytes of the message, excluding attachments (email size and number of attachments have no correlation to each other).
Device .csv	External device input or output	<ul style="list-style-type: none"> • Fields: id, date, user, pc, activity (connect/disconnect) • Some users use flash drives. • If the user shuts down the system before removing the drive, the disconnect record is missing. • Users are assigned a typical average number of flash drive uses per day (deviations from the normal number of uses can be used as an important indicator).
LDAP	Personnel information of insider	<ul style="list-style-type: none"> • Fields: employee_name, user_id, email, role, business_unit, functional_unit, department, team, supervisor • Data for approximately 1 year and 6 months exist by month from 2009-12 to 2011-05. • There is a significant difference in the numbers of emails received and sent, depending on the role. • role - ITAdmin: Systems administrators with global access privileges

4 EVALUATION ENVIRONMENT

4.1 Dataset

The CERT Insider Threat Test Dataset (Institute, 2013) was used in the experiment. The CERT dataset was created by the Carnegie Mellon University Software Engineering Institute in collaboration with ExactData and LCC, and with support from DARPA I2O. It was created for the purpose of researching insider threat behavior, and currently, six releases have been updated to 10 versions. The dataset includes

data on 1,000 insiders and malicious actors executing five malicious behavior scenarios. In this experiment, CERT r4.2 was used, because it was judged to be suitable for the experiment for its higher rate of malicious behavior data than in other versions. Table 2 provides a description of the files included in the CERT dataset. In this experiment, logon.csv, http.csv, file.csv, email.csv, device.csv, and LDAP were used. In the case of the CERT r4.2 data set, there are 100 malicious actors among 1,000 insiders, including thirty malicious actors in scenario 1, thirty in scenario 2, and ten in scenario 3. Table 3 provides

descriptions of the scenarios used in this experiment on the CERT r4.2 dataset.

4.2 ARX

The open source software ARX, which is used for data de-identification, was used on the original data. ARX supports a variety of anonymization tools and privacy models and can analyze risks by applying attacker models. Its de-identification process is illustrated in Figure 4. After the data to be de-identified are imported, the type of each attribute is set as either Identifying, Quasi-Identifying, or Sensitive / Insensitive. In the case of the Identifying and Quasi-Identifying types, the value of each attribute is replaced with “*”. The privacy model to be applied is then set. On the other hand, in the case of the Sensitive type, a separate privacy model must be set for each. After all settings for data de-identification are completed, the appropriate de-identification level is determined via the Explore process. Afterward, the input data and output data are compared to evaluate the de-identification performance. The Analyze utility function can be used to check the classification performance and quality model according to the target variable. On the other hand, the Analyze risks function can be used to evaluate the risk level according to various attacker models.

4.3 Evaluation Method

Experiments were conducted to demonstrate and evaluate the performance of the proposed technology. Figure 5 shows the simulation model. All records of operations by 1,000 insiders from PCs connected to the internal network of the organization are transmitted to the central monitoring server. The monitoring server subjects the data to preprocessing to use the collected data for machine learning. Data divided into files such as logon.csv, http.csv, file.csv, email.csv, device.csv, and LDAP are reclassified based on insider ID and then undergo processing for missing values and the removal of outliers.

Two procedures are performed to de-identify the preprocessed data. In the first step, a new item is created by combining up to three appropriate items to avoid using items that may contain content directly related to the actions of the insider and their personal privacy, such as date, content, and url. For example, through the use of the logon and logoff records of the logon data, it is possible to determine the time work was performed within designated working hours or the time work was performed outside working hours. These newly created items can be combined

again with http and device data to create data such as records of Internet access outside working hours and records of files copied to external devices.

The second step is de-identification using the ARX tool. First, the type of each field is set. In this experiment, the type of user was set to Identifying, that of sessionid was set to Quasi-identifying, and those of role, f_unit, dept, team, ITAdmin, and n_email related to position and department information were set to Sensitive. The field n_email shows the sum of incoming and outgoing e-mails, and as specified in Table 3, there is a large difference in the amounts of e-mails received and sent, depending on the role, and thus it was classified as Sensitive-type data. The types of all remaining fields were set to Insensitive.

Next, the generalization hierarchy method to be applied to the Sensitive-type data is set. In this experiment, character masking was applied to the six Sensitive-type data fields. Character masking is a general purpose mechanism and is the most widely available method for data anonymization. After all values were aligned to the left, masking was performed from right to left. After the length of the longest string in each field was set to Max.characters, padding was added for all values to have a length of Max.characters. The settings are specified in Table 4.

Domain size refers to the number of different possible values for each field. For example, in the case of role, the possible values for the field, that is, the domain, are 0 to 41, and thus, the domain size is 42. Max.characters indicates the length of the longest integer type in each field.

Therefore, Max.characters is set to 1 for f_unit and ITAdmin, which have domain sizes of 10 or less, and 2 for the remaining fields, which have domain sizes of 100 or less. Padding is then added according to alphabet size for all values to have the same length. After the privacy model was set to l-diversity, l was set to 2. Abnormal behavior is detected by using the de-identified data as input data to the LSTM autoencoder. If abnormal behavior is detected during this process, the abnormal actor is tracked through a re-identification process. If abnormal behavior is not detected, the data are stored in the database.

In this experiment, 20% of the overall dataset constituted the learning data set, whereas the remaining 80% constituted the validation dataset. We evaluated the degree of de-identification by the proposed technology and compared the detection rate with those of conventional technologies.

This study used log data collected over a short period of time. This makes it difficult to reflect bias and errors that may occur when massive amounts of log

Table 3: CERT experiment scenario descriptions.

Scenario	Description
Scenario 1	User who did not previously use removable drives or work after hours begins logging in after hours, using a removable drive, and uploading data to wikileaks.org. Leaves the organization shortly thereafter.
Scenario 2	User begins surfing job websites and soliciting employment from a competitor. Before leaving the company, they use a thumb drive (at markedly higher rates than their previous activity) to steal data.
Scenario 3	System administrator becomes disgruntled. Downloads a keylogger and uses a thumb drive to transfer it to the machine of their supervisor. The next day, he uses the collected keylogs to log in as his supervisor and send out an alarming mass email, causing panic in the organization. He leaves the organization immediately.

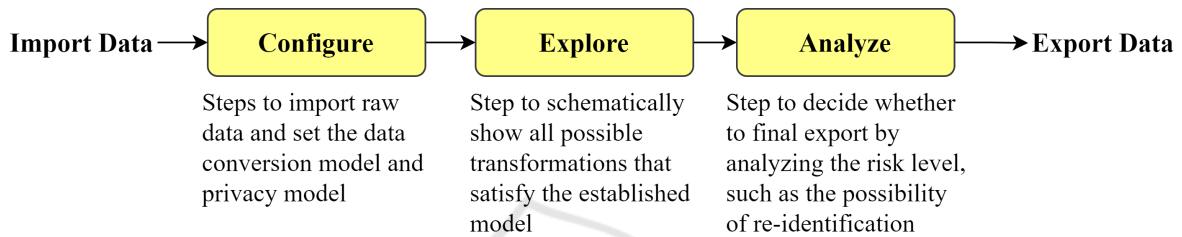


Figure 4: ARX de-identification process.

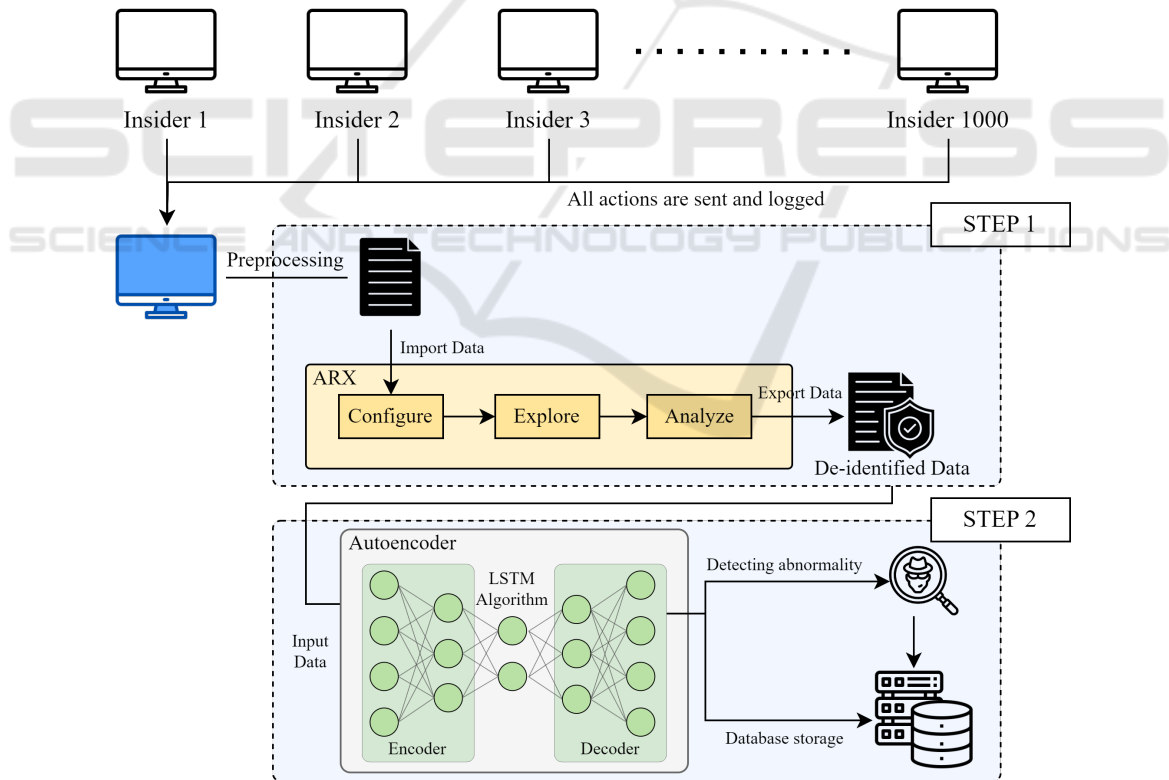


Figure 5: Simulation of abnormal-insider-behavior detection based on LSTM autoencoder.

data are collected in real situations. Therefore, in order to use it in an actual intrusion detection system, learning must be done using sufficient sample data.

5 PERFORMANCE EVALUATION AND ANALYSIS

We analyzed the results of the experiments that we conducted to evaluate the performance of the pro-

Table 4: Demonstration of data de-identification using ARX.

Fields	Type	Hierarchy	Domain size	Alphabet size	Max. characters	Privacy model
user	Identifying	Delete	-	-	-	-
sessionid	Quasi-id.	Delete	-	-	-	-
role	Sensitive	Masking	42	10	2	Distinct-2-diversity
f_unit	Sensitive	Masking	7	7	1	Distinct-2-diversity
dept	Sensitive	Masking	23	10	2	Distinct-2-diversity
team	Sensitive	Masking	39	10	2	Distinct-2-diversity
ITAdmin	Sensitive	Masking	2	2	1	Distinct-2-diversity
n_email	Sensitive	Masking	37	10	2	Distinct-2-diversity

posed technology. In this assessment, the levels of performance obtained using identifiable data (original dataset), using the conventional de-identification method, and using the de-identification method of the proposed technology were compared. The conventional de-identification method is to select sensitive data and then to delete the data.

5.1 De-Identification Performance Evaluation

The safety of the de-identified data was analyzed using the ARX Analyze risk function. For safety analysis, it is important to consider what kind of and how much knowledge the attacker has. The more knowledge an attacker has that is necessary for an attack, the easier the attack can be. In the risk analysis process of this experiment, to analyze safety, we employed three attacker models: the Prosecutor attacker model, the Journalist attacker model, and the Marketer attacker model. The evaluation indicators used included Records at risk, Highest risk, and Success rate. Records at risk indicates the percentage of risk above the standard value, whereas Highest risk indicates the highest risk for a single record. Meanwhile, the Success rate indicates the percentage of records that can be re-identified on average.

The Prosecutor attacker model assumes that an attacker targets a specific individual. The attacker knows that the data regarding their target individual are included in the dataset. Table 5 shows the results of risk analysis for when the Prosecutor attacker model was applied on the original dataset and de-identified datasets. In the case of the original dataset, all evaluation indicators, i.e., Records at risk, Highest risk, and Success rate, were at approximately 100%. On the other hand, when the conventional de-identification method was applied, the risk was at its lowest because sensitive data were completely deleted. When the proposed de-identification method was applied, the highest risk was only in the 1%

Table 5: Risk analysis results for when the Prosecutor attacker model was applied.

Data type	Records at risk	Highest risk	Success rate
Identifiable data	100%	100%	100%
De-identified data (conventional)	0%	0.02%	0.02%
De-identified data (proposed)	0%	1.30%	0.23%

range, indicating safety of the dataset.

By contrast, the Journalist attacker model assumes a situation wherein an attacker targets, but has no background knowledge about, a specific individual. Table 6 shows the results of risk analysis for when the Journalist attacker model was applied. When the original dataset was used, all indicators of risk were at 100%, whereas when the conventional de-identification method was employed, all indicators were at 0%, showing that the dataset was safe. When the proposed de-identification method was applied, the highest risk was in the 1% range.

Table 6: Risk analysis results for when the Journalist attacker model was applied.

Data type	Records at risk	Highest risk	Success rate
Identifiable data	100%	100%	100%
De-identified data (conventional)	0%	0.02%	0.02%
De-identified data (proposed)	0%	1.30%	0.23%

The Marketer attacker model aims to help attackers re-identify multiple individuals rather than targeting specific individuals. An attack is considered successful when a large number of records can be re-identified. Table 7 shows the results of risk analysis for when the Marketer attacker model was applied. Even in this case, the proposed de-identification method resulted in excellent safety, whereas the suc-

cess rate, which was 100% in the case of the original dataset, was in the 0% range for both the conventional and proposed de-identification methods.

Table 7: Risk analysis results for when the Marketer attacker model was applied.

Data type	Success rate
Identifiable data	100%
De-identified data (conventional)	0.02%
De-identified data (proposed)	0.23%

As a result of risk analysis by applying three attack models, the dataset de-identified using the conventional method was found to produce the lowest re-identification probability. The dataset de-identified using the proposed method was also found to produce a sufficiently low re-identification probability that is considered sufficiently safe.

5.2 Detection Rate Performance Evaluation

Subsequently, an experiment was conducted to compare abnormal behavior detection performance with respect to whether and how data de-identification was applied. The loss value obtained using verification data is compared with a randomly set threshold value, and if the loss value is greater than the threshold value, it is judged to be an abnormal behavior. In this experiment, thresholds were set to 1.1. Figure 6 shows a graph comparing the accuracy and F1-score results of the LSTM autoencoder-based abnormal behavior detection experiment with respect to whether and how de-identification was applied.

In the case of identifiable data, the accuracy was 0.950. By contrast, in the case of data de-identified using the conventional method, the accuracy was 0.788, showing significant deterioration in detection performance. Conventional de-identification causes loss of information because it deletes data without replacing or generalizing sensitive data. This resulted in a decrease in detection performance. On the other hand, in the case of data de-identified using the proposed method, the accuracy was found to be 0.954, which was a slight improvement from that obtained using the identifiable data.

With regard to the F1-score, that for the identifiable data was 0.975, that for the data de-identified using the conventional method was 0.881, and that for the data de-identified using the proposed method was 0.976. It can also be observed from the F1-scores that the data de-identified using the conventional method resulted in significantly deteriorated detection performance, whereas the data de-identified using the pro-

posed method resulted in a slight performance improvement compared to that obtained using the identifiable data.

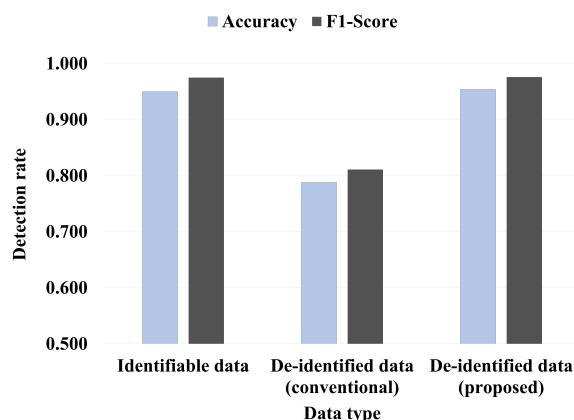


Figure 6: Comparison of LSTM autoencoder-based abnormal behavior detection accuracy and F1-score results with respect to de-identification application.

Through experiments to evaluate de-identification, it was confirmed that the proposed de-identification leads to a low possibility of re-identification and therefore good safety for the de-identified dataset. In addition, as a result of comparing the LSTM autoencoder-based anomaly detection performance obtained with the identifiable data and that obtained with the data de-identified using the proposed method, it was confirmed that the proposed de-identification resulted in a slight performance improvement. On the other hand, conventional de-identification provided the lowest possibility of re-identification and, therefore, superior data safety, but because significant information loss occurred during the de-identification process, the abnormal behavior detection performance deteriorated. We demonstrated that the proposed LSTM autoencoder-based insider abnormality detection technology that uses de-identified data provides safety in terms of personal privacy and leads to high detection performance.

6 CONCLUSION

In this paper, we present a security solution that protects individual privacy by applying data de-identification when monitoring abnormal behavior among organization insiders, while maintaining abnormal behavior detection performance similar to that obtained using existing identifiable data. In the abnormal behavior detection process, we attempted to effectively process long time-series data using an

LSTM autoencoder, an algorithm suitable for abnormal behavior detection.

To prove the effectiveness of the proposed method, de-identification evaluation and abnormal behavior detection performance comparison were conducted. In the de-identification evaluation, risk analysis was conducted by applying three attacker models, and it was proven that the de-identified dataset had only a low possibility of re-identification and was therefore safe. On the other hand, in the abnormal behavior detection performance comparison experiment, the de-identified data resulted in slightly improved performance and a higher detection rate than those obtained using the identifiable data.

In follow-up research, we plan to conduct further studies to expand the scope of application of anomaly detection solutions using de-identified datasets by setting various anomaly detection situations and providing anomaly detection solutions tailored to each situation.

ACKNOWLEDGEMENTS

This work was partly supported by the Korea Institute for Advancement of Technology (KIAT) grant funded by the Korean Government (MOTIE) (P0008703, The Competency Development Program for Industry Specialists) and MSIT under the ICAN (ICT Challenge and Advanced Network of HRD) program (No. IITP-2022-RS-2022-00156310), supervised by the Institute of Information Communication Technology Planning and Evaluation (IITP).

REFERENCES

- Abiodun, M. K., Adeniyi, A. E., Victor, A. O., Awotunde, J. B., Atanda, O. G., and Adeniyi, J. K. (2023). Detection and prevention of data leakage in transit using lstm recurrent neural network with encryption algorithm. In *2023 International Conference on Science, Engineering and Business for Sustainable Development Goals (SEB-SDG)*, volume 1, pages 01–09. IEEE.
- Al-Mhiqani, M. N., Ahmad, R., Abidin, Z. Z., Abdulka-reem, K. H., Mohammed, M. A., Gupta, D., and Shankar, K. (2022). A new intelligent multilayer framework for insider threat detection. *Computers & Electrical Engineering*, 97:107597.
- Ashraf, J., Bakhshi, A. D., Moustafa, N., Khurshid, H., Javed, A., and Beheshti, A. (2020). Novel deep learning-enabled lstm autoencoder architecture for discovering anomalous events from intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems*, 22(7):4507–4518.
- Chomutare, T. (2022). Clinical notes de-identification: Scoping recent benchmarks for n2c2 datasets. *Stud Health Technol Inform*, pages 293–6.
- Cui, L., Qu, Y., Xie, G., Zeng, D., Li, R., Shen, S., and Yu, S. (2021). Security and privacy-enhanced federated learning for anomaly detection in iot infrastructures. *IEEE Transactions on Industrial Informatics*, 18(5):3492–3500.
- Goryunova, V., Goryunova, T., and Molodtsova, Y. (2020). Integration and security of corporate information systems in the context of industrial digitalization. In *2020 2nd International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA)*, pages 710–715. IEEE.
- Gurucul (2023). Insider threat report: 2023 cybersecurity survey. Technical report, Gurucul.
- Institute, C. M. U. S. E. (2013). Cert insider threat test dataset. <https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=508099>. Accessed: 2023-09-14.
- Ito, S. and Kikuchi, H. (2022). Estimation of cost of k-anonymity in the number of dummy records. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–10.
- Jamshidi, M. A., Veisi, H., Mojahedian, M. M., and Aref, M. R. (2024). Adjustable privacy using autoencoder-based learning structure. *Neurocomputing*, 566:127043.
- Koll, C. E., Hopff, S. M., Meurers, T., Lee, C. H., Kohls, M., Stellbrink, C., Thibeault, C., Reinke, L., Steinbrecher, S., Schreiber, S., et al. (2022). Statistical biases due to anonymization evaluated in an open clinical dataset from covid-19 patients. *Scientific Data*, 9(1):776.
- Li, Z., Lee, G., Raghu, T., and Shi, Z. (2023). Does data privacy regulation only benefit contracting parties? evidence from international digital product market.
- Naim, A., Alqahtani, H., Muniasamy, A., Bilfaqih, S. M., Mahveen, R., and Mahjabeen, R. (2023). Applications of information systems and data security in marketing management. In *Fraud Prevention, Confidentiality, and Data Security for Modern Businesses*, pages 57–83. IGI Global.
- Nam, H.-S., Jeong, Y.-K., and Park, J. W. (2020). An anomaly detection scheme based on lstm autoencoder for energy management. In *2020 international conference on information and communication technology convergence (ICTC)*, pages 1445–1447. IEEE.
- Nguyen, H. D., Tran, K. P., Thomassey, S., and Hamad, M. (2021). Forecasting and anomaly detection approaches using lstm and lstm autoencoder techniques with the applications in supply chain management. *International Journal of Information Management*, 57:102282.
- Rai, B. K. (2022). Ephemeral pseudonym based de-identification system to reduce impact of inference attacks in healthcare information system. *Health Services and Outcomes Research Methodology*, 22(3):397–415.

- Roelofs, C. M., Lutz, M.-A., Faulstich, S., and Vogt, S. (2021). Autoencoder-based anomaly root cause analysis for wind turbines. *Energy and AI*, 4:100065.
- Said Elsayed, M., Le-Khac, N.-A., Dev, S., and Jurcut, A. D. (2020). Network anomaly detection using lstm based autoencoder. In *Proceedings of the 16th ACM Symposium on QoS and Security for Wireless and Mobile Networks*, pages 37–45.
- Thill, M., Konen, W., Wang, H., and Bäck, T. (2021). Temporal convolutional autoencoder for unsupervised anomaly detection in time series. *Applied Soft Computing*, 112:107751.
- Yun, S.-W., Lee, E.-Y., and Lee, I.-G. (2023). Selective layered blockchain framework for privacy-preserving data management in low-latency mobile networks. *Journal of Internet Technology*, 24(4):881–891.

