International Journal of Safety and Security Engineering
Vol. 13, No. 2, April, 2023, pp. 365-374
Journal homepage: http://iieta.org/journals/ijsse

# Two-Steps Insider Threat Detection Protocol (2S-ITDP) Based on Computer Usage Pattern Analysis

Amnat Sawatnatee[1*], Somchai Prakancharoen[2]

[1] Department of Multimedia, Faculty of Science, Rajaphat Chandrakasem University, Bangkok 10900, Thailand
[2] Department of Computer Science, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand

Corresponding Author Email: amnat.s@chandra.ac.th

https://doi.org/10.18280/ijsse.130219

**Received:** 27 February 2023
**Accepted:** 14 April 2023

***Keywords:***
*insider threat, sequential pattern analysis, computer usage pattern, computer security protocol*

**ABSTRACT**

The server, a computer system logging on of the organization's clients, has to verify whether the log-on requester is the formals registered before accessing. Unfortunately, the common log-on protocol rarely detected the imposter requester and untrusted organizational clients. This paper presents the proactive protocol to prevent the outside intruder and detect untrusted corporate clients. This research innovates the step by step practical control flow of whole activities of proactive and preventive framework about intrusion. The suggested protocol comprises the proposed authentication protocol and the client's computer system using sequential pattern analysis. These two activities can inhibit outside intruders, and detect and eliminate the fake client (known other organizational clients' secret passwords). Otherwise, it can see whether the real clients are working correctly as prior computer usage. The evaluation presents that the "two-step insider threat detection protocol (2S-ITDP)" can proactively prevent both outside and inside intruders better than the typical authentication protocol. Furthermore, the accuracy of classification is about eighty-eight percent.

## 1. INTRODUCTION

Many computer departments mostly use the common attributes "login name" and "password" to detect whether the servers' accessing requester is the one that has already passed formal registration or not. If the "login name", and "password" are found in the secure client's "login name", and "password" entity, then the requester is identified and permitted to work in the computer system. The request will be accepted if the requester's login name and password exist in the secure client's "login name", and "password" entity. Unfortunately, this simple log-on protocol is easily attacked by simple brute force techniques. The outside intruders can repeatedly try out the new password many times, sending the log-on request until it is met in the login name & password file. Some kinds of outside intruder attacks are eliminated by more sophisticated authentication protocols [1, 2]. However, there are still other problems that are challenging to solve. These problems concern fake and real dishonest organization's clients [3]. Firstly, the fake clients are the organization's clients, the fake clients who used other clients <login name, password> to impersonate some clients. After the access permission, he could do anything under the victim access control list. The second problem is the actual organizational client, untrusted, tries to work on an unassigned list, some unauthorized or prohibited tasks [4]. Therefore, both kinds of the untrusted client must be monitored, detected, and exterminated since the information are the essence of organizational operation [5].

This research aims to suggest techniques to detect fake requesters and untrusted clients. The suggested research provides the computer usage pattern analytic technique [6] to enhance the proposed authentication protocol. Each client's pattern of computer usage [7] is presented in the data sequence pattern, which is similar to the other client's computer usage sequence pattern. If the first: client's working pattern deviates from his usual prior pattern of computer usage, the second: the sensitive data [8] are breached, the third: a client tries to neglect the organizational computer usage regulation (prohibit activities), and the fourth: client is not following to his assigned authorization rights, then this client will be suddenly purged his transaction processing. After that, his activity and evidence will be sent to his boss. In addition, this client will be set flag as a suspect insider threat. The data sequence pattern technique is implemented to detect whether the active client is a fake client. The "two-step insider threat detection protocol (2S-ITDP)" can provide proactive detection and prevention of both outside and inside intruders from computer system access.

## 2. RELATED THEORY AND RESEARCH

### 2.1 One time password

The onetime password (OTP) technique [9] is applied to enhance log-on protocol security. The server and client do not need to be responsible clients for login name and password preservation. Instead, the password will be randomly generated, then sent to both the server and client for communicate message encryption.

### 2.2 One-time-password-authenticated key exchange

The authors [10] present the OTP generating by using a

pseudo random generator based on time-dependent. The OTP is generated and performed in the task sequence based on password-authenticated key exchange protocol (PAKE). The practical OTP based on PAKE can support mutual authentication for both the client and server.

### 2.3 Authentication and password storing improvement using SXR algorithm with a hash function

Polpong and Wuttidittachotti [11] suggest four steps for user password modification. First, the user password is hashed. Second, the ratio and number of iterations are calculated from the user's secret key. Third, the hashed password is split into two pieces based on a ratio. Fourth, each piece hashed are performed the exclusive function with number rounds. Finally, these two pieces are concatenated and kept in the database as the user's new secure password.

### 2.4 Data sequential pattern mining

Data sequential pattern analysis [12] is the technique that is used to determine the pattern of sequential transactions of the individual or many entities. All transactions will be sorted by trade occurring time. Each transaction may consist of one or many items that simultaneously occurred. This set of transaction elements may be presented in n-item form, such as one item, two items, two items together, three items, and so forth.

The support of (n) item occurring must meet the minimum careful subjective defined criteria. All the combinations of support passed (n) items shall illustrate the data sequence pattern structure. These pattern structures can implement as the architecture of pattern sequences [13]. In general, the pattern sequence of each one is quite different from the other one. The derived sequence pattern presents the presence of each client's working pattern. Thus if his working pattern resembles his prior supported patterns, then it can identify as being normal working. The similarity of comparing can measure from the frequency of the pattern occurring. Two entities will be similar if they all have the same most frequent pattern [14]. The measurement of similarity or level of clients' usual (typical) working (LUW) can calculate from the total hit elements devised by total access items.

$$LUW = \frac{total\_match\_access\_item}{(total\_match\_access\_item + total\_unmatch\_access\_item)} * 100 \quad (1)$$

While; total_match_access_item is the amount of on-accessing n-item pattern that matched to prior n-item pattern, and total_unmatch_access_item is the amount of on-accessing n-item pattern that unmatched to previous n-item pattern. For example, if CL's sequence data item is <ACEBD(AC)B>. If the prior derived data pattern is "A, B, C, BC", thus there are four access items, under underline (<A C E B D (AC) B>), matched this prior data sequence pattern. Three access items, with underline <A C E B D (AC) B>, that are unmatched by the prior data sequence pattern. Therefore, LUW is 57.14% (4/(4+3)).

The threshold of LUW is the essential criterion that must be repetitive experimental resetting to reduce the problem of type I and II error classification.

### 2.5 Sensitive data

The sensitive data [15] is the organizational data that must be accessed and updated by only the authorized party. These data should cause the organization's severe effects if there are disclosures of malicious intentions. If the unauthorized party accesses an unauthorized party accesses the defined organizational sensitive data, then privacy or confidence will be reached. In a difficult situation where the unauthorized can change the sensitive data, thus the organization's data are integrity losing. Some data, databases, application programs, etc., may be interrupted an attacked, which leads to a loss of availability. Therefore, organization must carefully define the significant sensitive data in many classifications of sensitive data.

There are many types of sensitive data. The first is public or low-sensitive data, for example, staff organization, published papers, general organization news, or public activity. This kind of data is not sensitive data since there are not cause the organization's endurance. The second class is moderate data sensitive, for example, building architecture, customer records, customer service information, etc. With deep study and other coverage information, this information may present data that can accidentally reach some organization privacy data. The third class is highly sensitive or confidential data, for example, social security numbers, customer transactions, health information, business deep or secret strategic plans, etc. These data-sensitive classes are directly practical to someone, organization data privacy. It should directly harm the data owner.

The organization has to classify its information asset into three classes. After that, each interest-specific sensitive data will be deep study in its data processing scenario. All activities about this sensitive data must be carefully considered on all its activity such as creating, appending, updating, deleting, backup, and recovery. The organization must set these activities the specific rights, authority, and responsibility to whom, when, where, and what. These assignments will be used to produce the access control list: ACL assigned to the responsible CL (client). The CL's ACL can be used to monitor whether the client is performing as his given rights in ACL.

### 2.6 "The blockchain technologies in healthcare: Prospects, obstacles, and future recommendations; Lessons learned from digitalization"

Blockchain technology [16] supports healthcare data management systems in patient medical records manipulation. The advantages of Blockchain are data privacy, flexibility in data access, integrity, authentication, etc. Moreover, the recorded data can support data analytics.

### 2.7 "Security aware information classification in health care big data"

The authors [17] present word extraction, mapping into normal and sensitive data, and defines weighting. Then, the gathered information are used to classify the data if it is exposed or not.

### 2.8 Behavior analysis ("User behavior analytics for insider threat detection using deep learning")

The user's usage [18] on the fixed window was trained by one class training to construct the user behavior. Gated recurrent unit base auto encoders are used to model the user behavior per day. Thus this model is used to detect the anomaly working of specific users. The experiment dataset is

computer emergency response team r 4.2. The classification value is positive, and the true negative rate is about 79.8%.

### 2.9 "Classification of mobile customers behavior and usage patterns using self-organizing neural networks"

Today mobile network technology efficiently supports the rapid growth of many mobile applications. However, customers have different usage patterns for portable application use. Thus, the classification of customers based on customers' data from the mobile operator will present the customers' behavior segments [19]. Self-organizing-map is the technique used to show behavior. The information helps to increase efficient mobile application marketing.

### 2.10 "An insider threat detection method based on user behavior analysis"

Jiang et al. [20] have tried to extract the significant features that often cause organization information loss (attack) from much-related research. The "XGBoost" algorithm is used to identify the insider threat by aggregating the proposed significant feature with user behavior simultaneously. The F-measure of classification is about 99.96%.

Research direction: the literature review on behavior analysis research (presented on Table 1) shows that using techniques could present a client's static activities. However, it can't detect the sequence of clients' activity. Therefore, this research chose to observe the client's activities (behavior) using a generalized sequential pattern (GSP) algorithm. The computer using a pattern of each client will be used to detect whether they are working as usual. The unusual working client will be carefully observed whether he is an insider threat.

**Table 1.** The summary of related research and proposed solution

| | 2.8 | 2.9 | 2.10 | Proposed solution |
|---|---|---|---|---|
| Data | Client's Window page access log | Customer mobile usage pattern | Significant attributes about information lost | Five routine works& five non-routine works |
| Technique | Deep learning | Self-organizing-map | Xgboost | Generalized sequential pattern (GSP) |
| Classification accuracy | 79.80% | n/a | F-measure 99.96 | Level of clients' usual working (LUW) |

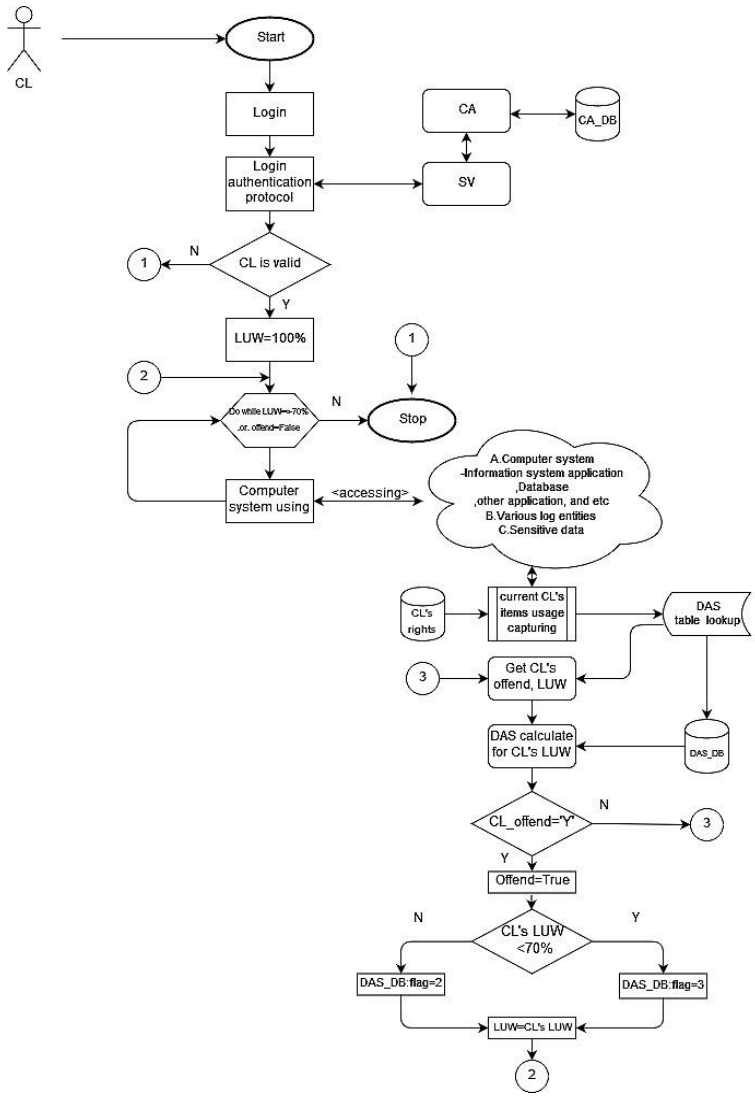


**Figure 1.** Research framework scenario

## 3. TWO-STEP INSIDER THREAT DETECTION PROTOCOL BASED ON COMPUTER USAGE PATTERN ANALYSIS DESIGN

### 3.1 Research framework

The experimental scenario is used to present the detail of the research activity framework. Figure 1 presents the CL's method of logging on to the server to ask for permission for computer system access. Two steps of "2S-ITDP" collaborate to fulfill the mission of intruder detection; 1st-prior accessing activities and 2nd-during the CL's to access the computer system.

**A.** Prior accessing activities

(1) The client applies to the organization's CA for authentication certification (Figure 2). The client must use his proper authentication public key, related parameters, and legal evidence to the organization's CA. CA verifies the requester's evidence whether there are valid or not. If all the evidence is legitimate, then Client's public key; and related parameters will be appended to the CA's DB (CA_DB) referenced (primary key) with CL's identification. The CL's identification will be used, by CL to reference CL authentication certification as CL requests to "log on" to the computer server.

(2) DAS formally submitted the CL's allocated authorized rights (Table 3) of computer system usage based on the CL's role, position, and responsibility to the CL. The CL's assigned request will be considered defined and responsible by the information system section under each entity's authorized rights (Table 2).

Each client will be assigned specific rights according to the entity's whole rights management, CL's job position, and CL's responsibility. CL must strictly work under the given authorized rights.

(3) CL consents to his assigned authorized rights and his forbiddance organizational notification about computer usage. The CL's prohibit level of the sensitive data will be presented on different prohibition levels (Table 6), from the other client, depending on the CL's assigned access list. The prohibit level is according to CL's rights, job position, etc. The information security section specified the important sensitive data.

The research organizational sensitive data list table lookup attributes, type of sensitive data, and prohibit level; {'1.Student data': '1_level', '2.Confidential Student information': '2_level','3.Financial information': '3_level', '4.Intellectual property': '4_level', '5.Strategic detail plan': '5_level', '6.Network configuration': '6_level', '7.Database architecture & rights': '7_level'}.

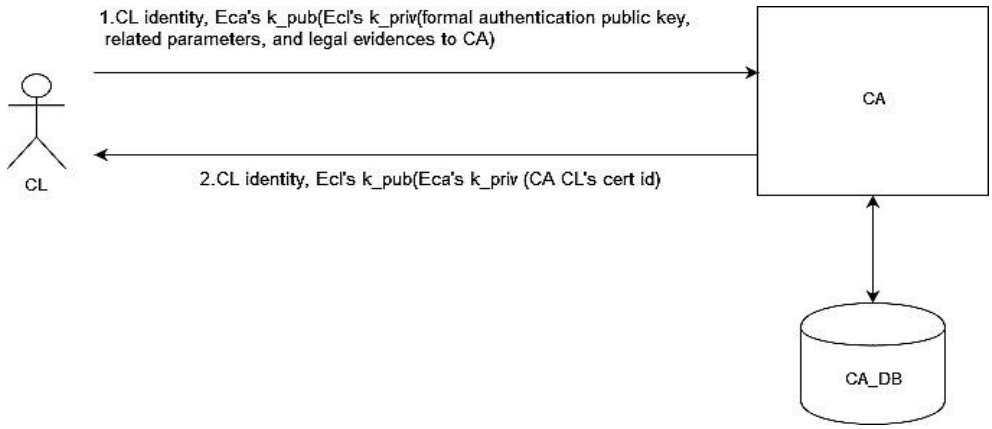


**Figure 2.** CL register to CA for authentication certify

**Table 2.** Whole authorized rights

| Attribute | Rights | | | | | | |
|---|---|---|---|---|---|---|---|
| | Append | Update | Delete | Backup-recovery | Access only | Time | Where |
| ID-student | cid-1 | cid-2 | cid-3 | cid-DBA | all | all | all |
| Name | cid-1 | cid-2 | cid-3 | cid-DBA | all | all | all |
| Address | cid-1 | cid-2 | cid-3 | cid-DBA | all | all | all |
| ID card | cid-1 | cid-2 | cid-3 | cid-DBA | cid-registrar | Working day | on office |
| G.P.A | cid-1 | cid-2 | cid-3 | cid-DBA | cid-registrar, cid-LoanDept | Working day | on office |
| Study loan | cid-1 | cid-2 | cid-3 | cid-DBA | cid-LoanDept | Working day | on office |
| … | | | | | | | |

**Table 3.** Specific CL (e.g. cid-1) authorized rights

| Attribute | Rights | | | | | | |
|---|---|---|---|---|---|---|---|
| | Append | Update | Delete | Backup-recovery | Access only | Time | Where |
| ID-student | cid-1 | - | - | - | - | all | all |
| Name | cid-1 | - | - | - | - | all | all |
| Address | cid-1 | - | - | - | - | all | all |
| ID card | cid-1 | - | - | - | - | Working day | on office |
| G.P.A | cid-1 | - | - | - | - | Working day | on office |
| Study loan | cid-1 | - | - | -- | - | Working day | on office |

**B**. During the CL's access to the computer system.

The following tasks explain the control flow of activities.

B.1 CL log-on to the server via research-suggested authentication log-on protocol.

B.2 Do while CL still access the computer system with conditions "LUW" =>70% .or. "Offend flag"=False.

B.3 (Back office–DAS's working) All CL's behavior is captured by DAS from the group of log device entities, CL's rights, and sensitive data entities. DAS constantly analyzes the current CL's behavior on accessing (today- Table 5) LUW, then compare it to his prior cumulated data usage pattern (as shown in Table 4) to check if CL works in a different pattern away from his usual work based on LUW criteria or not. The current (while CL's is working) LUW will be kept in DAS_Table lookup (on accessing LUW attribute).

B.4 If CL tries to access or do something that is not in compliance with 1.CL's assigned rights or 2. sensitive organizational data.

B.4.1 DAS set offend flag to "True" (Table 4).

B.4.2 If CL's LUW < 70% then DAS set the DAS_DB: flag= '3', else DAS set the DAS_DB: flag= '2' (Table 3)

B.4.3 DAS will immediately conceal the sensitive data presentation and suddenly purge the CL's task, then go to "B.2".

B.4.4 DAS informs CL's Boss along with CL's prohibited evidence (Offline rectification*).

B.5 End do-while.

B.6 Offline rectification*; if CL's boss afterward formal consideration on CL's issue is permission accepted, then DAS will change CL's flag status to regular status ('1') in DAS_DB and update the CL's rights.

B.7 Offline DAS_DB updating; DAS will, every day after the office's work closing, update CL's prior data sequence pattern to the current data sequence pattern.

**Table 4.** CL's cumulative data usage pattern database

| DAS_DB cumulative CL's behavior | Detail |
|---|---|
| CL-id | |
| Threat type flag* | |
| CL's status(1-normal,2-anomalous) | |
| Cumulative Monday-sequence data | |
| Cumulative Tuesday-sequence data | |
| Cumulative Wednesday-sequence data | |
| Cumulative Thursday-sequence data | |
| Cumulative Friday-sequence data | |
| 1-item | |
| 2-item | |
| 2-item (together) | |
| 3-item | |
| 3-item (together) | |
| 4-item | |
| 4-item (together) | |
| 5-item | |
| 5-item (together) | |
| *Threat type flag(1-normal user, 2-fake client, 3-suspect insider threat) | |

**Table 5.** DAS table lookup (today accessing) CL's behavior

| DAS_table lookup CL's (on accessing) behavior | |
|---|---|
| CL's ID | 1234 |
| Working day | e.g. Monday |
| List of accessing items | e.g. A,B,(AC),B |
| LUW | e.g. 83.50 |
| Offend (1=N, 2=Y) | N |

### 3.2 Actors and their responsibility

3.2.1 Client (CL)

CL requests SV's permission to access SV's facilities. Each CL will be assigned the rights to access some application, database, libraries, etc. In addition, all CLs will be informed of the prohibit activity about official computer system usages, and sensitive data.

3.2.2 Host server (SV)

SV is the computer system that provides information systems, library, devices, etc., for its member(clients).

3.2.3 Certificate authority (CA)

CA is the third-party trust department responsible for guaranteeing its members authentication [21, 22]. Members have to formally register on the CA with their legal evidence and detail about their cryptography (public key encryption algorithm and public key). CA will keep CL's information in CA's secure database. CA_DB attributes are {'customer ID', 'public key algorithm', 'customer public key', and 'related parameters'}. Every CL will use this CL's CA certificate as authentication identification. If the SV wants to check whether CL is authentication, then SA will send the CL's CA certification to CA. When CA receives the request, CA will seek the CA certification if it is present in the CA_DB. If the search is found, CA will send conventional keys to both SV and CL for further message encryption (between SV and CL). CA randomly generates this conventional key. Thus, this key acts as OTP. This research; presents the equation to generate the OTP as Eq. (2).

$$Kcon = Hash_n(Nonce \oplus CL\_ID) \quad (2)$$

While: "n" (in binary bits) is the size (chosen by CA) of Kcon, and "nonce" is the "number used once" that is randomly generated by CA. The CA sending messages will first be encrypted with CA's private key and second on the cipher text with the receiver's (SV or CL) public key. CA will not send the public key of both SV and CL. Thus SV and CL will surely be safe from the man-in-the-middle attack. All the messages transferred between both entities will be first encrypted with the sender's private key (e.g., CA's k-priv) and second encrypted with the recipient's public key (e.g., CL's k-pub). These activities will mutually support the message confidence and entity authentication.

3.2.4 Data analytic section (DAS)

DAS is the organizational section responsible for client behavior monitoring. DAS continually monitors CL's computer system behavior using all the time that CL is accessing the computer system. The data are gathered directly from various log - keeping entities such as event logs, database-transaction logs, network logs–NetLog, application logs, sensitive data, CL' right, prohibited list, etc. The accessing items of CL will be kept in the table lookup CL's behavior (Table 4). The CL's behavior is collected and kept in the cumulative data sequence. Every day after work closing, CL's usage patterns will be appended to the prior cumulative DAS_DB under C:'s ID. This accumulative sequence data will be analyzed to derive the current sequence pattern of CL's computer usage. This sequence data will be used to compare with the on-accessing CL's behavior to detect whether CL usually working as usual.

**Table 6.** Research sensitive data (for example)

| Sensitive data list | Prohibit level* | Rights | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Append | Update | Delete | Backup-recovery | Access only | time | where |
| 1.Student data | 1 | y | y | y | y | - | all | all |
| 2.Confidential student data | 1 | y | y | n | n | y | 9-15 | Local |
| 3.Financial information | 2 | y | y | n | y | y | 9-15 | Local |
| 4.Itellectual property | 2 | y | y | n | n | n | all | all |
| 5.Strategic detail plan | 2 | y | y | n | y | n | all | Local |
| 6.Network configuration | 2 | y | y | n | y | n | all | all |
| 7.Database architecture & rights | 2 | n | y | n | n | y | all | all |
| *Prohibit level (1-moderate, 2-high) | | | | | | | | |

### 3.3 Offline CL's behavior management

The client's computer usage log sequence pattern analysis gathered both CLs before and during computer access. Client's computer usage logs are always gathered from many logs-keeping processing mechanisms by DAS. The CL's computer usage patterns have accumulated collected since they first started working in this computer system. CL's today usage of the computer system will be captured and seek for current CL's computer system usage pattern, then detect whether the pattern is similar to his prior computer usage pattern. CL's computer system usage patterns are used to accurately identify the CL's authentication. Unfortunately, some CLs may have similar item sequence patterns as other CLs. Therefore, many interesting behavior topics about the client's computer usage log sequence patterns should be gently added later to reduce similar patterns for different clients. Many techniques could create the CL's behavior pattern [23], such as generalized sequential pattern (GSP), pattern-growth-based approaches, etc.

### 3.4 Authentication protocol

The suggested authentication protocol collaborates with CA to prevent a man-in-the-middle attacks. More ever, CA is responsible for generating the OTP randomly. This OTP is a conventional key that both Server and CL will use to encrypt all the messages that transfer between them during CL's access. If the session is finished and CLs would like to log on again, thus the CA will repeatedly generate the new OTP. Therefore, this protocol can preserve data privacy and man-in-the-middle attack problem. Meanwhile, the server does not need to be concerned about client login name and password table protection.

### 3.5 Client's computer usage log management

DAS is responsibility for analyzing the item's usage pattern from DAS_table lookup CL's behavior to present the "on accessing Client's usage behavior". DAS is an essential actor in monitoring and detecting the intruder, as explained in 2.1(B).

### 3.6 Sensitive data

The organizational information system section defines the sensitive data. This entity will define the detail of each entity and its attribute regarding the rights of entity activities, time, and where there can be manipulated. The details are shown in Table 5. Since many clients may have different authorized rights on the same sensitive data thus, this table could be joined to each specific CL (cid-1) authorized rights. However, this research chooses to separate into two tables since the authorized organizational right is not very complicated.

### 3.7 Decision-making on client's computer usage

DAS monitors the client's usage behavior consideration. For example, the CL's on accessing usage behavior will be compared to the cumulative computer usage pattern, LUW, then decide to do something on CL computer access.

### 3.8 Evaluation method

Research samples, and fifty clients (CL) observations were used to test the accuracy of intruder classification. Forty-two CLs are the truly typical usage client, and eight are not typical usage clients. Therefore, the evaluation will count the "true normal" and "not normal user" and then calculate the found classification result for classification accuracy.

## 4. TWO-STEP ITDP PROTOCOL RESULT AND EVALUATION

This section will explain the detail of the experiment and their results as following topics.

### 4.1 CL's computer usage (behavior) collection

The client's computer usage log sequence pattern analysis for prior and during computer access is conducted by DAS. The research context is about the education sector computer systems. The experiment project was created and tested during 2020-2021. There were 146 students and 28 staff participated in this experiment. The sample observations were fifty chosen students, the fourth-year undergraduate majoring in computer technology. The sample transaction will present only one client's behavior to quickly understand the CL's usage behavior analysis. The transaction of item access (or purchase) data were collected at least six weeks. Then it was set as the first usage sequence pattern. This firstly usage sequence will be implemented as the starting accumulate sequence pattern. The collection of the CL's item purchasing on each working day, the information system section actual purchasing items are shown in Table 7. There are two types of purchase items (1.routine works and 2.not routine works), and each type compose of five tasks (item code). These works, called items, are coded as {A, B, C, D, E, F, G, H, I, J}.

**Table 7.** The purchasing items definition

| Type | Detail | Item code |
|---|---|---|
| Routine work | Routine of business | |
| 1 | information system | A |
| 2 | Access on public information, under client's rights | B |
| 3 | Download file, under organization application | C |
| 4 | Upload file, under organization application | D |
| 5 | Organization e-mail using | E |
| Non-routine work | | |
| 1 | Social network using | F |
| 2 | Search engine | G |
| 3 | Portal website | H |
| 4 | Market place platform | I |
| 5 | Video sharing application | J |

For example, CL's sample computer usage was collected Monday through Friday. The time duration is divided into five durations. The actual working hours are code "2" and "3". This information should help DAS compare the current usage pattern with similar dates and times so that the analytic result will increase the classification accuracy. Time duration of data collection table look-up attributes are time durations, and their code;{'6.00-8.59 .am.':'1','9.00-12.00 am.':'2','12.01-12.59 am.': '3','13.00-17.00 am.': '4','17.01-18.00 am.': '5'}. The sample of experimental CL's computer usage (presenting just one week of data) transaction is shown in Table 8.

**Table 8.** The sample of one experimental CL's computer usages sequence

| Transaction | | Item usage | Client's usage sequence |
|---|---|---|---|
| Day | Time | | |
| Monday | 1 | E | |
| | 2 | A | |
| | 3 | G | <EAGB> |
| | 4 | B | |
| | 5 | - | |
| Tuesday | 1 | E | |
| | 2 | (AC) | |
| | 3 | G | <E(AC)G(BD)> |
| | 4 | (BD) | |
| | 5 | - | |
| Wednesday | 1 | E | |
| | 2 | (AC) | |
| | 3 | (GH) | <E(AC)(GH)(BD)(HJ)> |
| | 4 | (BD) | |
| | 5 | (HJ) | |
| Thursday | 1 | - | |
| | 2 | (AC) | |
| | 3 | (GH) | <(AC)(GH)(BD)(FH)> |
| | 4 | (BD) | |
| | 5 | (FH) | |
| Friday | 1 | - | |
| | 2 | (AC) | |
| | 3 | (GH) | <(AC)(GH)(BD)(EHI)> |
| | 4 | (BD) | |
| | 5 | (EH) | |

After 1-item, 2-item, 3-item, 4-item, and n-items-together extraction based on support criterion equal or greater than '4' (or about 80%) on six weeks of data collection, there are many derived data sequence patterns as shown in Table 9.

**Table 9.** The CL's computer usage sequence pattern (of fifty experiment samples) at the whole date and time (support≥4).

| 1-item | 2-item | 3-item | 4-item |
|---|---|---|---|
| A | AB | AGB | - |
| B | AD | AGD | |
| C | AG | CGB | |
| D | CB | CGD | |
| G | CD | | |
| | CG | | |
| | GB | | |
| | GD | | |
| | (AC) | | |
| | (BD) | | |

In another case, suppose the support value is set to "≥5", then it will derive the different generated data sequence pattern, as shown in Table 10. The met support criteria sequence pattern are reduced from nineteen to only five patterns. Since the transaction is classified as five working days, therefore at the support value "4", the support criteria are 80% (4/5) is the suitable criteria.

**Table 10.** The CL's computer usage sequence pattern at the whole date and time (support≥5)

| 1-item | 2-item | 3-item | 4-item |
|---|---|---|---|
| A | AB | - | - |
| B | AG | - | - |
| G | - | - | - |

The CL's computer usage sequence pattern (support ≥4) could illustrate the sequence architecture as shown in Figure 3.

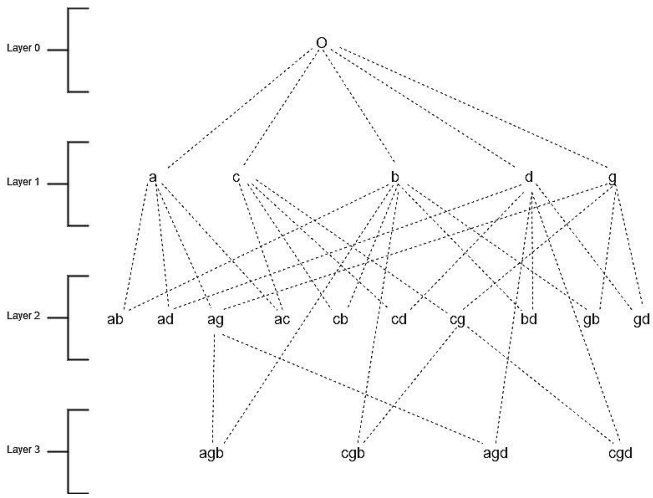


**Figure 3.** The computer usage sequence pattern architecture (support≥4)

For example, similarity calculation (consideration of the whole pattern of a week or five working days), suppose that the CL's , on accessing, usage sequence transaction at current is <C(AC)BJCGDBH>. The measurement will count all items before or after n-item together occurring and all the possible matching to the prior derived usage patterns.

**A**.Case: support ≥4 (Table 8)

1-item matching is "C", "A", "C", "B", "C", "B". (<C (A C) b J C G D B H>.

2-item matching is none.

2-items together matching is "(AC)". <C (AC) BJCGDBH>.

3-item matching is "CGD". <C(AC)BJ CGD BH>.

The LUW value is about 80% (8/(8+2)).

**B**.Case: support =>5 (Table 9)

1-item matching is “A”, “B”, “G”, “B”. (<C (A C) B G D B H>.

2-2-item matching is none.

The LUW value is about 50% (4/(4+4)).

This result indicates that the matching item pattern is a low percentage since there are few n-item patterns. The n-item pattern will occur more in the low value of “n” (1, 2). Therefore, the suitable support criteria should be repetitively train and tested on the extensive data sequence to choose the best support criteria value. The experimental result shows that a high support data sequence pattern will not ensure a high percentage of CL’s item usage pattern matching. On the other hand, the lower support should cause the type I and II classifications.

### 4.2 The suggested authentication log on protocol

There are four sequenced tasks in the authentication log-on protocol scenario, as shown in Figure 4.

***Step 1***: The client sent a message “CL's identity, ESV's k-pub (CL's CA cert ID, n=1)” to the server for computer system access permission. The message is encrypted with SV’s public key so no one else can reveal the “CL's CA cert ID, N=1”.

***Step 2*:** The server sent CL’s detail to CA to recheck whether CL is legitimate. The sending message is encrypted with CA’s public key to keep the message private. The message sequence is presented in the communication step by an integer number (‘N’), which will be increased in its value by step 1.

The N value is used to protect against replay attacks. SV's identity, ECA's k-pub(CL's CAcert-ID, N+1=1).

***Step 3-4****:* CA reveals the sending message with CA’s private key. If there is CL's CA cert ID presented in “CA’s secure database”, then CA sends a message to both server (step 3rd) and the client (step 4th). These messages, CA will prepare the conventional key (K-con) with 3-Triple DES default for SA, and CL to use to encrypt their communication message. However, SA and CL may change the received K-con (by 3.2.3), and encryption algorithm to the agreed one to prevent message privacy attacks from CA.

### 4.3 Decision-making on client’s computer usage

DAS is the one that has the responsibility to detect the client’s usage behavior and conduct the decision-making about the CL’s computer access.

### 4.4 Evaluation

The suggested “Two-step insider threat detection protocol (2S-ITDP)” is evaluated for the accuracy of internal threat classification. The accuracy of classification is depended on the number of items usage coverage, the thorough of sensitive data covering, and the employee job mobility rate. Therefore, the research will consider just the technical aspect of the number of the items usage (definite on ten items about computer usage). The support value of the sequence pattern is defined at eighty percent (support ≥4). The resulting pattern is nineteen generated data sequence patterns. The level of CL’s normal working (LUW) is set to passing if the CL’s LUW is equal to or greater than seventy percent similarity. The test result shows that the general sequence pattern at support criteria similar to “4” should give more accuracy in classification, as shown in Table 11.

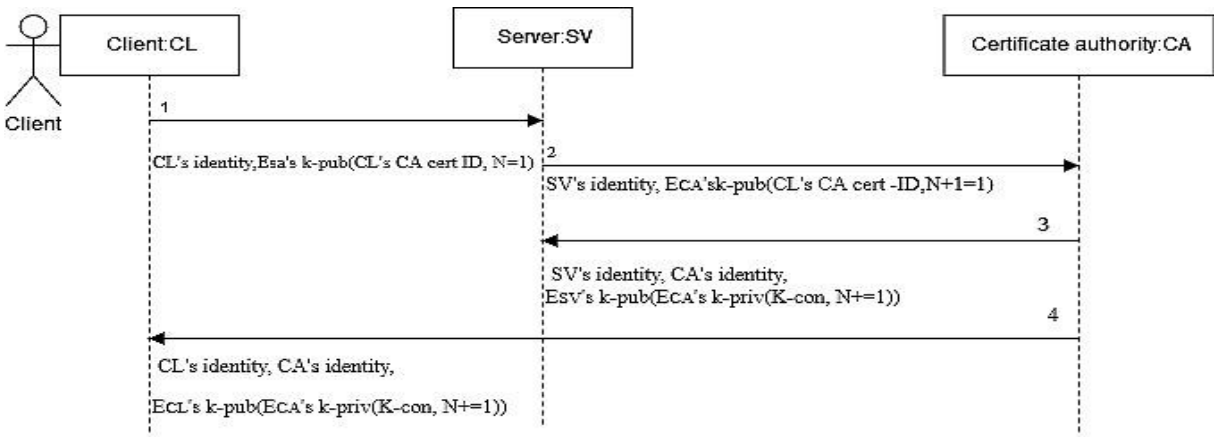


**Figure 4.** Suggested authentication log-on protocol

**Table 11.** The classification evaluation result

| | Support | Patt-rn | Real number usage use (42) | | Not normal usage user (18) | | Correct | Average correct |
|---|---|---|---|---|---|---|---|---|
| | | | LUW=>70% | LUW<70% | LUW=>70% | LUW<70% | Classification | Classification |
| 1 | =>4 | 19 | 37(88.10)% | 5 | 1 | 7(87.50)% | 44 | 88% |
| 2 | =>5 | 5 | 28(66.67%) | 14 | 2 | 6(75%) | 34 | 68% |

## 5. RESEARCH SUMMARY AND SUGGESTION

The experiment on two-step insider threat detection protocol is composed of the log-on under authentication support and client behavior analytics. The suggested log-on protocol (first step) can prevent password brute force attacks and the man-in-the-middle attack problem. More ever, by using data sequential pattern analysis (second step), client behavior analytics can monitor and detect fake client and untrusted clients with a good accuracy levels of classification.

The client's computer using (behavior) analytics considers ten attributes. There are five routine works and five non-routine works. These shared work attributes are the general activities for most clients; therefore, the proposed practical framework is helpful for insider threat detection. However, there are some suggested considerations for further research.

**5.1** The experiment was limited to ten items of computer usage, which may not cover all organization client's activities. The organization must collect and group the additional activities to include the important ones in the interest item list. The excellent data sequence pattern is possibly derived from the coverage, significant attributes, and large volume of transactions.

**5.2** The data sequence pattern ought to separate, presenting the specific item sequence pattern (on the individual clients, working day, and working periods) since patterns comparing the same situation (for example, time duration or which working day) should support DAS on identifying the CL's status with the best accuracy and precision. All coverage of situation data analytics will overcome the type 1 and type 2 errors of classification.

**5.3** The n large ($n > 4$) and be purchased together items pattern (e.g. (abcd)). These patterns rarely occur, but these patterns are also seldom found in other CL. Therefore, considering different weight sets on these rare patterns is a challenging topic for further research.

**5.4** There are many other behaviors type [24] that could present different points of view on insider threat. Therefore, these topics should be consider included in behavior analytic.

**5.5** There are some accidental tasks without the client's intention, which could cause the wrong decision in classification. Hence these kinds of activity should be advance informed to all clients to be aware of prohibiting tasks list.

## ACKNOWLEDGMENT

The experiment collected the experimental log data from the Department of information technology, Faculty of science, Chandrakasem Rajabhat University's server during 2020-2021 under the permission of the computer and network administrator. With all of their helping hands, the experiment was undertaken successfully.

## REFERENCES

[1] AL-Musawi, B.Q.M. (2012). Preventing brute force attack through the analyzing log. Iraqi Journal of Science, 53(3): 663-667.

[2] Mohammed, M.A., Degadzor, A.F., Effrim, B.F., Appiah, K.A. (2017). Brute force attack detection and prevention on a network using wireshark analysis. International Journal of Engineering Sciences & Research Technology.

[3] Abu-Shanab, E., Matalqa, S. (2015). Security and fraud issues of E-banking. International Journal of Computer Networks and Applications, 2(4): 179-188.

[4] Malik, M., Patel, T. (2016). Database security-attacks and control methods. International Journal of Information, 6(1/2): 175-183. https://doi.org/10.5121/ijist.2016.6118

[5] Kitone, M.K., Kelvin, O.K. (2013). Role of information systems in organizational business process. International Journal of Information Technology and Management, 8(1): 1214-1220.

[6] Kang, J.M., Seo, S.S., Hong, J.W.K. (2011). Usage pattern analysis of smartphones. In 2011 13th Asia-Pacific Network Operations and Management Symposium, Taipei, Taiwan, pp. 1-8. https://doi.org/10.1109/APNOMS.2011.6077030

[7] Oskouei, R.J., Chaudhary, B.D. (2010). Internet usage pattern by female students: A case study. In 2010 Seventh International Conference on Information Technology: New Generations, Las Vegas, NV, USA, pp. 1247-1250. https://doi.org/10.1109/ITNG.2010.76

[8] Alotaibi, S., Alharbi, K., Abaalkhail, B., Ibrahim, D.M. (2021). Sensitive data exposure: Data forwarding and storage on cloud environment. iJOE, 17(14): 4-18. https://doi.org/10.3991/ijoe.v17i14.27365

[9] Babkin, S., Epishkina, A. (2018). One-time passwords: Resistance to masquerade attack. Procedia Computer Science, 145: 199-203. https://doi.org/10.1016/j.procs.2018.11.040

[10] Steinfeld, R., Hawkes, P. (Eds.). (2010). Information Security and Privacy: 15th Australasian Conference, ACISP 2010, Sydney, Australia, July 5-7, 2010, Proceedings (Vol. 6168). Springer Science & Business Media.

[11] Polpong, J., Wuttidittachotti, P. (2020). Authentication and password storing improvement using SXR algorithm with a hash function. International Journal of Electrical and Computer Engineering (IJECE), 10(6): 6582-6591. https://doi.org/10.11591/ijece.v10i6.pp6582-6591

[12] Hijawi, H., Saheb, M. (2015). Sequence pattern mining in data streams. Computer and Information Science, 8(3). https://doi.org/10.5539/cis.v8n3p64

[13] Mooney, C.H., Roddick, J.F. (2013). Sequential pattern mining--approaches and algorithms. ACM Computing Surveys (CSUR), 45(2): 1-39. https://doi.org/10.1145/2431211.2431218

[14] Shou, Z., Di, X. (2018). Similarity analysis of frequent sequential activity pattern mining. Transportation Research Part C: Emerging Technologies, 96: 122-143. https://doi.org/10.1016/j.trc.2018.09.018

[15] Elser, K. (2020). It performance audit of city wide data classification and sensitive data encryption. Office of the City Auditor, San Diego, USA. https://www.sandiego.gov/sites/default/files/audit_of_citywide_sensitive_data_encryption_standards_and_data_classification_public.pdf

[16] Abbas, A.F., Qureshi, N.A., Khan, N., Chandio, R., Ali, J. (2022). The blockchain technologies in healthcare: Prospects, obstacles, and future recommendations; Lessons learned from digitalization. International Journal of Online & Biomedical Engineering, 18(9): 144-159. https://doi.org/10.3991/ijoe.v18i09.32253

[17] Funde, S., Swain, G. (2021). Security aware information classification in health care big data. International Journal of Electrical and Computer Engineering (IJECE), 11(5): 4439-4448. https://doi.org/10.11591/ijece.v11i5.pp4439-4448

[18] Nepal, S. (2022). User behavior analytics for insider threat detection using deep learning. Department of Electronics and Computer Engineering, Pulchowk Campus, IOE, Tribhuvan University, Nepal.

[19] Ghnemat, R., Jaser, E. (2015). Classification of mobile customers behavior and usage patterns using self-organizing neural networks. International Journal of

Interactive Mobile Technologies, 9(4): 4-11.
[20] Jiang, W., Tian, Y., Liu, W., Liu, W. (2018). 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, Springer International Publishing, pp. 421-429. https://doi.org/10.1007/978-3-030-00828-4_43
[21] Jacco, C. (2019). A smarter way to authenticate customers. Information protection and cybersecurity financial services, USA. https://assets.kpmg.com/content/dam/kpmg/xx/pdf/2019/06/smarter-way-to-authentication-customers.pdf
[22] Christiana, A.O., Oluwatobi, A.N., Victory, G.A., Oluwaseun, O.R. (2019). A secured one time password authentication technique using (3, 3) visual cryptography scheme. In Journal of Physics: Conference Series, 1299(1): 012059. https://doi.org/10.1088/1742-6596/1299/1/012059
[23] Doko, E., Bexheti, L.A., Hamiti, M., Etemi, B.P. (2018). Sequential pattern mining model to identify the most important or difficult learning topics via mobile technologies. International Journal of Interactive Mobile Technologies, 12(4): 109-122. https://doi.org/10.3991/ijim.v12i4.9223
[24] Aunger, R., Curtis, V. (2008). Kinds of behavior. Biology & Philosophy, 23(3): 317-345 https://doi.org/10.1007/s10539-007-9108-4