Indian Journal of Science and Technology, Vol 8(34), DOI: 10.17485/ijst/2015/v8i34/74505, December 2015
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Latent Relation Analysis based Discovering Fraudulent Ranking Identification on Mobile Web Apps

L. Velmurugan*

Department of Computer Science, AMBO University, Ethiopia; velmuruganphd15@gmail.com

Abstract

Objective: The main objective of this work is finding a fraudulent ranking behaviour of mobile apps where mobile app developers may generate fraudulent evidences for providing a top ranking for them. The primary goal of this work is to find out the fraudulent evidences present in the ranked mobile apps. This work attempts to improve the accuracy of detection of fraudulent ranking behaviour of mobile apps by performing concept vector based review evidence analysis. **Method:** Mobile app ranking fraudulent behaviour is the biggest issue in the mobile app development environment due to the degradation of mobile app's important level. In the existing work, Leading Session Methodology based evidence aggregation (LSMEA) is introduced to leverage the fraudulent ranking activities. This LSM analysis the three types of evidences such as ranking based, rating based and review based and aggregates their output finally for detecting the fraudulent ranking behaviour of mobile apps. Among the above mentioned evidences, review based evidence is based on user opinion about the corresponding mobile app. LSM analysis the users review comments by using latent semantic approach which will find the important semantic terms from the user review comments. However this method failed to identify the concepts of semantic terms accurately which might lead to wrong assumption of fraudulent ranking behaviour. This problem is overcome in this work by introducing the Concept Vector based Review Evidence Analysis (CVREA) which is done by using WordNet tool. Word Net tool will retrieve the most important concepts present in each sentence of user review comments based on which fraud signature would be computed. Finally, result of these three evidences would be combined together to detect the fraudulent ranking behaviour of mobile apps. **Application/ Improvements:** This proposed research methodology would be more helpful in the mobile app markets where the number of apps developed for the specific purpose has been increased considerably. In this situation, it is required to provide truthful and most popular mobile apps to the users to increase the reputation level. This proposed research methodology provides a way for increasing the reputation level of the mobile owners by detecting and eliminating the fraudulent ranking behaviour of mobile apps.

Keywords: Mobile Apps, Fraudulent Behaviour, Ranking Evidences, Sematic Relation

1. Introduction

Mobile apps usages are increased in number in today real world environment due to the increased number of smart phones. The mobile apps are released by various industries and in many forms. There are lots of mobile apps are released which are doing the same process. The apps need to be ranked to provide the flexible way for users to select their most wanted apps. For example, there are mobile apps are present in real world for chatting purpose like whatsApp, hike, chat on and so on.

As with the increased usage of mobile apps, the fraudulent behaviours are also increased in number. The mobile apps can be ranked in terms of rating and the usage of that particular app by the users. This ranking would be changed periodically due to arrival many new software. Among these apps, fraudulent mobile web app detection plays a critical role in many scenarios[1,2].

The main contribution of this work is finding a fraudulent behaviour of a mobile apps where mobile app developers may generate an fraudulent evidences for providing an top ranking for them. The primary goal of

* Author for correspondence

this work is to find out the fraudulent evidences present in the ranked mobile apps. And also this work aims to filter the mobile fraudulent ranking behaviour based on the semantic relation present among the evidences of mobile apps.

The organization of this work is given as follows: In this section brief introduction about the mobile web app behaviour is given. In section 2, various previous researches that have been conducted for detecting the fraudulent behaviour in most application are discussed briefly. In section 3, proposed methodology of our work is discussed in the detailed manner for detecting the fraudulent web app behaviour. In section 4, experimental tests that have been conducted are discussed deeply to know the improvement of the proposed methodology. Finally in section 5, the overall research work has been concluded to indicate the improved methodology.

The techniques used in misuse detection and anomaly detection are described as follows:

1.1 Expert Systems

An skilled system is outlined as a computer system capable of representing and reasoning concerning some knowledge-rich domain with a read to finding issues and giving recommendation[3,4]. Skilled system detectors encrypt data concerning attacks as if-then rules. NIDES developed by SRI uses the skilled system approach to implement intrusion detection system that performs time period observation of user activity[5]. NIDES consist of applied mathematics analysis part for anomaly detection and rule based mostly analysis part for misuse detection.

1.2 Neural Networks

"ID (Neural Network Intrusion Detector) is an anomaly intrusion detection system enforced by a back propagation neural network beneath OS surroundings[6]. It's trained to spot users supported what commands and the way typically they used throughout on a daily basis. It's simple to coach and cheap as a result of it operates off-line on daily log information. ANN (Artificial Neural Networks) provides the power to generalize from antecedently's discovered behaviour (normal or malicious) to acknowledge similar future unseen behaviour for each anomaly detection and misuse detection[7]. It's enforced by a hack propagation neural network.

1.3 Model-based Reasoning

Model-based detection may be a misuse detection technique that detects attacks through noticeable activities that infer AN attack signature. There's information of attack eventualities containing a sequence of behaviours creating up the attack. Garvey and role player combined models of misuse with evidentiary reasoning[8]. The system accumulates additional and additional proof for an intrusion try till a threshold is crossed; at now, it signals AN intrusion try. A pattern matching approach supported coloured Petri Nets to find misuse intrusion is projected by Kumar and Spafford[9]. It uses audit trails as input below UNIX operating system setting.

1.4 Data Mining

Data processing approaches is applied for intrusion detection. A crucial advantage of knowledge mining approach is that it will develop a replacement category of models to find new attacks before they need been seen by human consultants. Classification model with association rules rule and frequent episodes is developed for anomaly intrusion detection. This approach will mechanically generate apothegmatic and correct detection models from great amount of audit information. However, it needs an oversized quantity of audit information so as to figure the profile rule sets. Moreover, this learning method is associate integral associated continuous part of an intrusion detection system as a result of the rule sets employed by the detection module might not he static over an extended amount of your time. A team of researchers at Columbia projected the detection models exploitation cost-sensitive machine learning algorithms[10]. Audit information is analysed by association rules rule so as to see static options of attack information.

1.5 State Transition Analysis

State Transition Analysis could be a misuse detection technique, that attacks are painted as a sequence of state transitions of the monitored system. Actions that contribute to intrusion situations are outlined as transitions between states. Intrusion situations are outlined within the variety of state transition diagrams. Nodes represent system states and arcs represent relevant actions. If a compromised (final) state is ever reached, an intrusion is claimed to own occurred. STAT (State Transition Analysis Tool) could be a rule-based skilled system designed to hunt out better-known penetrations within the audit trails of multi-user laptop systems[11].

USTAT (UNIX State Transition Analysis Tool) could be a UNIX-specific paradigm of STAT[12].

1.6 Other Techniques

A genetic rule[13] is applied to notice malicious intrusions and separate them from traditional use. A genetic rulemay be a technique of artificial intelligence downside resolution supported the idea of Darwinian evolution applied to mathematical models. This genetic rule was designed in order that every individual depicted a attainable behavioural model. This approach provides a high detection rate and a coffee warning rate. Dokas and Ertoz projected building rare category prophetical models for characteristic illustrious intrusions. This technique will address the lack of normal data processing techniques once addressing inclined category distribution.

Iterative mechanism[14] is introduced for detecting the fraudulent behaviour existing in the distributed and parallel system in terms of improved privacy and security violation resides in the patterns of transaction. This is done to prove the various tolerating mechanisms in terms of market abased analysis where the multiple extraction techniques are used in the data retrieval mechanism. Different iteration mechanism that are induced to provide an efficient and flexible way of deriving the fraudulent patterns resides in the mobile app behaviour.

2. Fradulent Mobile App Behaviour Detection

Mobile app fraudulent becomes most critical issue in the real world environment where the number of mobile app users is increased in number. These mobile apps need to be ranked honestly for providing the better services to the mobile app users. The app leader board is responsible ranking the apps based on their usage and the reputation level. The mobile apps can be ranked based on characteristics called the number of users, percentage of rating, popularity level of application and so on. App leader board would perform analysis over the available mobile apps based on these characteristics to give prioritization for the mobile apps.

Fraudulent activities are increased due to this ranking scenario where the users will prefer the most popular apps only. Many fraudulent companies attempt to increase the ranking of their newly developed apps in the shortest period of time by doing many malicious activities. Detection of fraudulent behaviour is most critical task where the characteristics of the mobile app would be available in the better manner. The efficient prediction of fraudulent behaviour is introduced in this work which attempts to detect the fraudulently ranked mobile apps that are available online to prevent the mobile app users to install the worst app.

The fraudulent behaviour of mobile app detection is done by gathering the various evidences from the mobile app and finding fraudulent signature of apps. This is done by analysing the mobile app evidences. The evidences that are considered in this work are

- Ranking based evidence
- Rating based evidence
- Review based evidence

These evidences are analysed and the fraudulent ranking behaviour of mobile apps are found. Among the above mentioned evidences, review based evidences are based on the user review comments where the user opinion about the mobile apps would be present. In the proposed methodology called Concept Vector based Review Evidence Analysis (CVREA) is introduced. This work makes use of word net tool for extracting the concepts from the user review comments based on semantic meaning. The fraudulent ranking behaviour is predicted by processing the following steps

- Mining leading session
- Gathering Evidences
- Concept Vector based Review Evidence Analysis

2.1 Process Flow

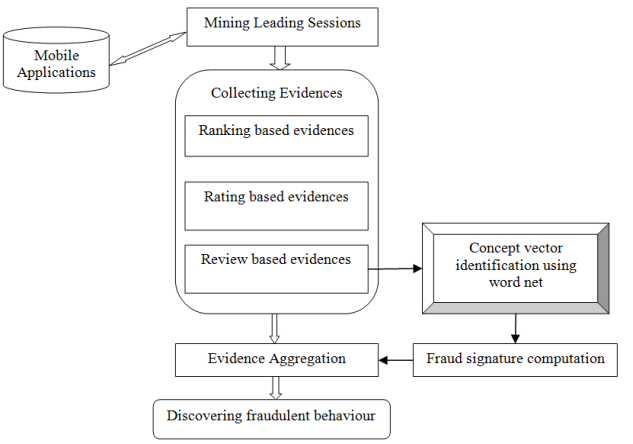

Figure 1. Flow of process.

The proposed research methodology of this work is represented in the Figure 1. In the figure, overall flow of this proposed work is depicted where it will mine the leading session from the mobile behaviour where the leading session is evaluated based on the timing value. After gathering the leading session, based on that evidences for mobile app fraudulent behaviour would be extracted. This fraudulent app behaviour evidences would be aggregated together to know the overall fraudulent behaviour. Finally, fraudulent behaviour would be predicted by evaluating the latent semantic relationship present between the evidences that are gathered. These processing flow discussed in the detailed manner in the following sections.

2.2 Mining Leading Session

Leasing session is defined as the sequence of leading events. Leading events are the time sequence which is spent in the particular mobile app. This leading session are extracted with the help of analysing the historical data of the mobile app which are available online. The main steps that are followed to mine the leading sessions are

- Find the leading events
- Build leading session by merging leading event

In the first step, time period spent for the each and every task of mobile app would be extracted. In the second step, those leading event would combined together to create the leading session. The algorithm for mining the leading session is depicted in algorithm1.

ALGORITHM 1: Mining Leading Session

Input 1: a's historical ranking records R_a;
Input 2: the ranking threshold K*;
Input 3: the merging threshold Φ;
Output: The set of a's leading session S_a;
Initialization: $S_a = \Phi$
1. $E_s = \phi; e = \phi; s = \phi; t^e_{start} = 0$
2. for each i ∈ [1, $|R_a|$] do
3. if $r_i^a \le K*$ and $t^e_{start} = 0$ then
4. $t^e_{start} = t_i$;
5. else $r_i^a \le K*$ and $t^e_{start} \neq 0$ then
6. // found one event
7. $t^e_{end} = t_{i-1}; e =< t^e_{start}, t^c_{end} >$;
8. if Es == Φ then
9. $Es \cup e; t^e_{start} = t^s_{start}; t^c_{end} = t^s_{end}$
10. Else If $(=< t^e_{start}, t^c_{end} >) < \phi$ then
11. $Es \cup = e; t^c_{end} = t^s_{end}$
12. Else then
13.// Found one session
14. $s = \langle t^s_{start}, t^s_{end}, E_s \rangle$
15. S_a U = S; S = Φ is a new session
16. $E_s = \{e\}; t^s_{start} = t^e_{start}; t^s_{end} = t^e_{end}$
17. $t^c_{start} = 0; e = \phi$
18. Return S_a;

2.3 Gathering Evidences

After mining of leading sessions, the evidences that are related to the fraudulent mobile app behaviour would be predicted to find the apps that are ranked wrongly. Evidences are gathered based on three behaviours of mobile apps. Those are ranking based evidences, rating based evidences, review based evidences. The value of these evidences would be gathered with the consideration of the various time sessions, mainly based on the leading sessions. Ranking based evidences are the one which is done by the app leader board to give the better review of apps to the users who using smart phones.

Ranking of apps would consist of three phases. Those are rising phase, maintenance phase, and the recession phase. In the rising phase, ranking value of the mobile app would be increased suddenly whereas in the maintenance phase, the ranking value of mobile would be maintained without degradation by providing valuable services to the users. In the recession phase, the ranking value would be degraded suddenly from higher level to the lower level. These ranking phases of mobile apps would help to detect the fraudulent behaviour which may vary in different time sessions. From this ranking analysis, we can predict the fraudulent by finding the unexpected ranking rising or recession phase.

Rating based evidences are other important evidences which can be done anonymously in order to increase the reputation of the mobile apps. Rating of mobile apps which are done by anonymously need to be detected to prevent the web apps from the fraudulent ranking. This is done by analysing the leading session that is extracted.

Review based evidences are the one where the product comments would be left by the users about the mobile apps. The comment may consist of both positive and negative comments where the fraudulent companies may leave many positive comments to increase the usage of mobile apps. In the existing work, important terms

present in the user review comments are extracted by analysing the more repeated verbs based on which fraud signature would be identified. However this cannot identify the fraudulent ranking behaviour accurately in presence of less knowledge about the concepts of user review comments. This done by using the Concept Vector based Review Evidence Analysis (CVREA) which is discussed detailed in the following sub section.

2.4 Decision Based Latent Semantic Relationship Extraction

User review comments about the mobile apps are one of the most important things that can be used for understand about the popularity about the mobile apps. Mobile app developers may leave more positive comments about the mobile apps to increase the popularity of the mobile apps in the considerable manner. This fraudulent behaviour of app developers who leaves the wrong comments needs to be identified. This can be done by finding the more important terms present in the user review comments and computing the fraudulent signature of those terms for different mobile apps. In this proposed research methodology Concept Vector based Review Evidence Analysis is introduced which will identify the concepts of user comments based on which fraudulent signature would be computed. The concept vector identification is done in this work by using the word net tool.

WordNet is a large lexical data base of English language. WordNet tool expresses the group of nouns, verbs, adjectives and adverbs along with their syntactic meaning. These terms would be interlinked with each other based on their semantic meaning. Concept Vector based Review Evidence Analysis performs sentence based concept mining, where importance of concepts that are present in the user review comments would be calculated for both sentence and group of sentences. After finding the different concepts of user review comments, the important concepts are filtered by finding the conceptual term frequency (ctf). The overall flow of this work is given as follows:

ALGORITHM 2. Fraudulent ranking behaviour detection with Concept Vector based Review Evidence Analysis

Input: Mobile apps

Outputs: Fraudulent ranking behaviour of apps

1. Gather the user review comments of different mobile apps
2. Mine the leading sessions as given in algorithm 1
3. For every leading sessions Si ∈ Sa
4. Find the ranking based evidences
5. Find the rating based evidences
6. Find the review based evidences
7. CVREA ()
8. End for
9. Aggregate the evidences based on unsupervised approach
10. Output fraudulent ranking behaviour of mobile apps
11. CVREA ()

Begin

12. Load the user review comments
13. Divide the review comments into sentences
14. Parse sentences into WordNet for identifying concepts Ci
15. For each concepts Ci ∈ C
16. Find the ctf of concept c in sentence s
17. Find the ctf of concept c in document d

$$ctf = \frac{\sum_{n=1}^{sn} ctf_n}{sn}$$

Table 1. The comparison analysis graph

Number of Web Apps	Time Complexity		Precision		Recall	
	LSMEA	CVREA	LSMEA	CVREA	LSMEA	CVREA
5	20	34	0.12	0.35	0.35	0.41
10	40	50	0.19	0.57	0.5	0.58
15	50	71	0.34	0.65	0.55	0.63
20	65	82	0.59	0.83	0.67	0.72
25	70	89	0.71	0.87	0.79	0.82
30	78	95	0.82	0.95	0.84	0.96

18. End for
19. Store concept with more ctf value in concept vector
20. Compute fraud signature of each user review in terms of concepts using cosine similarity

$$Sim(s) = \frac{2x\sum_{1\le i<j\le N_s} Cos(\overrightarrow{w_{ci}}, \overrightarrow{w_{cj}})}{N_s x(N_s - 1)}$$

22. Return Sim (s)

End

Where sn → total number of sentence in user review

$\overrightarrow{w_{ci}}$ → Concept vector review 1

$\overrightarrow{w_{cj}}$ → Concept vector of review 2

The above pseudo code provides improved detection fraudulent ranking behaviour of mobile apps using the proposed methodology called the Concept Vector based Review Evidence Analysis. This method improves the accuracy of detection of fraudulent ranking behaviour of mobile apps by finding the important concepts that are present in the user review comments based on which fraudulent signature is identified. The performance evaluation of the proposed research methodologies in terms of detection of fraudulent ranking behaviour is given and discussed detailed in the following sections.

3. Experimental Results

In this section performance evaluation is done to show the improvement in the proposed methodology. The experimental tests conducted were proving the effectiveness of the proposed methodology by comparing it with the existing approach. The comparison is done against the parameters called the time complexity, precision, recall. In our work, varying number of apps is taken for analysis to predict the malicious behavioural based apps. The performance measure values are given in the Table 1.

The performance evaluation is shown in the following Figures 2 to 4.

3.1 Time Complexity

Time complexity is the measure which is consumed by the application to find the fraudulent behaviour present in the mobile app ranking. The time taken to find the fraudulent behaviour is measured in the unit called millisecond. The comparison graph is shown in Figure 2:

In that graph, the time taken to detect the fraudulent behaviour present in the mobile app rank is compared against the existing work and the proposed methodology. In X axis number of web apps taken and in Y axis time complexity in millisecond is taken. This graph proves that the proposed methodology provides better result than the existing methodology with better improvement.

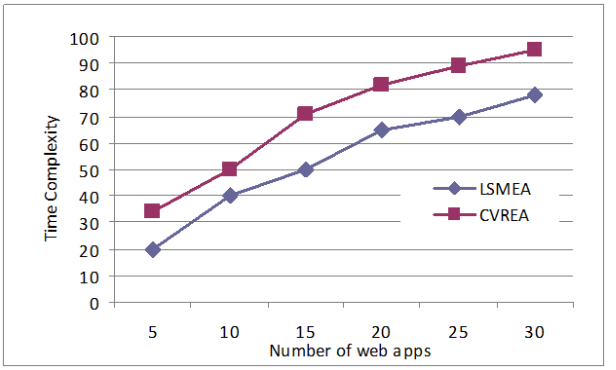

Figure 2. Time complexity comparison.

3.2 Precision

Precision is used to predict the number of accurate fraudulent behaviour detection among the set of all possible solution in which better solution can be obtained. That is precision or positive predictive value is defined as the proportion of the true positives against all the positive results (both true positives and false positives). The precision is calculated as follows:

Precision = No of TP / (No of TP + FP)

The comparison graph is shown in the Figure 3. In that graph, the precision value of detecting the fraudulent behaviour present in the mobile app rank is compared against the existing work and the proposed methodology. In X axis number of web apps is taken and in y axis precision value is taken. This graph proves that the proposed methodology provides better result than the existing methodology with better improvement.

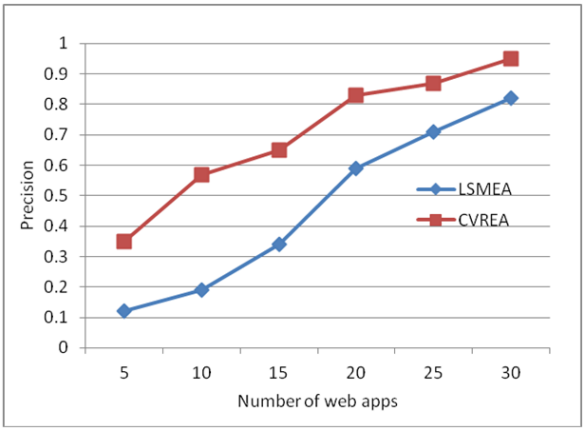

Figure 3. Precision comparison.

3.3 Recall

Recall is used to measure the whether the retrieved result of fraudulent behaviour detection is done correctly or not. Recall in information retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved.

Recall = |{relevant documents} ^ {retrieved documents}|/ |{relevant document}|

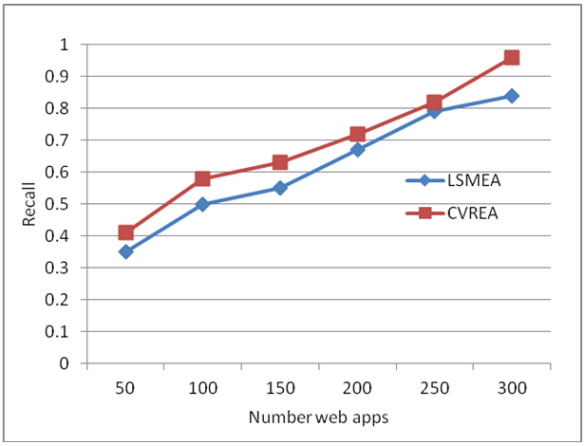

Figure 4. Recall Comparison.

The comparison is shown in the Figure 4: In that graph, the recall value of detecting the fraudulent behaviour present in the mobile app rank is compared against the existing work and the proposed methodology. In X axis number of web apps is taken and in y axis recall value is taken. This graph proves that the proposed methodology provides better result than the existing methodology with better improvement.

4. Conclusion

Mobile apps become the most popular technology among the people which leads to an development of many apps with similar features. However the ranking of mobile is done in the fraudulent manner which needs to be avoided for filtering the unwanted apps from the set retrieved apps. In this work, we developed a ranking fraud detection system for mobile Apps. Specifically, we first showed that ranking fraud happened in leading sessions and provided a method for mining leading sessions for each App from its historical ranking records. Then, we identified ranking based evidences, rating based evidences and review based evidences for detecting ranking fraud. Moreover, we proposed an optimization based aggregation method to integrate all the evidences for evaluating the credibility of leading sessions from mobile Apps.

5. References

1. Available from: http://venturebeat.com/2012/07/03/ apples-crackdown-on-app-ranking-manipulation/, 2012
2. Available from: https://developer.apple.com/news/index.php?id=02062012a, 2012
3. Lunt TF, Tamaru A, Gilham F, Jagannathan R, Neumann PG, Javitz HS, Valdes A, Garvey TD. A real-time lntrusion Detection Expert System (IDES) -Final Technical Report. SRl Computer Science Laboratory, SRl International, Menlo Park, CA. 1992. p. 273–85.
4. Sehring M, Shellhouse E, Hanna M, Whitehurst R. Expert system in intrusion detection: A case study. Proceedings of the 11th National Computer Security Conference. 1988 Oct. p. 85–91.
5. Anderson D, Frivold T, Tamaru A, Valdes A. Next generation intrusion detection expert system (nides). Computer Science Laboratory. 1994.
6. Ryan J, Lin MJ, Miikkulainen R. Intrusion detection with neural networks. In: Jordan MI, Keams MJ, Solla SA, editors. Advances in Neural Information Proceeding Systems. The MIT Press. 1998; 10:72–7.
7. Ghosh AK, Schwartzbard A. A study in using neural networks for anomaly and misuse detection. Proceedings of the 8th USENIX Security Symposium, D. C., 1999; 141–52.
8. Lee W, Stolfo S. Data mining approaches for intrusion detection. Proceedings of the 7th USENIX Security Symposium, San Antonio, TX. 1998; 6.
9. Kumar S, Spaford EH. A pattern matching model for misuse intrusion detection. Proceedings of the 17th Notional Computer Security Conference. 1994. p. 11–21.
10. Stolfo SI, Lee W, Chan PL, Fan W, Eskin W. Data mining-based intrusion detectors: An overview of the Columbia ids project. In ACM SIGMOD Record. 2001; 30(4):5–14.

11. Ilgun K, Kemmerer RA, Porras PA. State transition analysis: A rule-based intrusion detection approach. Software Engineering. 1995; 21(3):181–99.
12. Ilgun K. USTAT A real-time intrusion detection system for UNIX. Proceedings of the I993 IEEE Symposium on Research in Security and Privacy, Oakland, CA. 1993; 16–28.
13. Chittur A. Model generation for an intrusion detection system using genetic algorithms. In Ossining High school Honors, 2001; 1–17.
14. Sherly KK, Nedunchezhian R. A improved incremental and interactive frequent pattern mining techniques for market basket analysis and fraud detection in distributed and parallel systems. Indian Journal of Science and Technology. 2015 Aug; 8(18):55109.