I.J. Computer Network and Information Security, **2014, 5, 9-20**
Published Online April 2014 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijcnis.2014.05.02

Visualization of Influencing Nodes in Online Social Networks

Prajit Limsaiprom, Prasong Praneetpolgrang, Pilastpongs Subsermsri
School of Information Technology, Sripatum University, Bangkok 10900, Thailand
crossprajit@yahoo.com, prasong.pr@spu.ac.th, pilastpongs@yahoo.com

Abstract—The rise of the Internet accelerates the creation of various large-scale online social networks. The online social networks have brought considerable attention as an important medium for the information diffusion model, which can be described the relationships and activities among human beings. The online social networks' relationships in the real world are too big to present with useful information to identify the criminal or cyber attacks. The methodology for information security analysis was proposed with the complementary of Cluster Algorithm and Social Network Analysis, which presented anomaly and cyber attack patterns in online social networks and visualized the influencing nodes of such anomaly and cyber attacks. The closet vertices of influencing nodes could not avoid from the harmfulness in social networking. The new proposed information security analysis methodology and results were significance analysis and could be applied as a guide for further investigate of social network behavior to improve the security model and notify the risk, computer viruses or cyber attacks for online social networks in advance.

Index Terms—visualization, influencing nodes, anomaly and cyber attacks, online social networks, clustering, social network analysis.

I. Introduction

New services will widen the Internet into an interactive medium for online social networks (Web 2.0) such as Facebook, MySpace, Twitter, Hi5, Google, etc. The online social network services increase the benefits of users, however these services may cause additional risks for users due to their trusted these online social networks. The attackers, spammers and scammers may have the opportunities to exploit their information easily or construct convincing social engineering attacks with all their data.

Many organizations have become vulnerable from the intrusion of cyber attackers that compromise the security of their networks. The security incidents comprise the violation of the given security policies. To be able to detect violations of the security, it must be able to observe all activities that could potentially be part of such violations [1]. Many computer security techniques have been intensively studied in the last decade to defend against various cyber attacks and computer viruses, namely cryptography, firewalls, anomaly and intrusion detection [2], [3], [4], [5], [6].

Online social networks can also play an important role as a medium for the spread of information. For example, innovation, hot topics and even malicious rumors can propagate through online social networks among individuals, and computer viruses can diffuse through email networks. Social Network Analysis (SNA) has become a powerful methodological tool alongside with statistics. The social network analysis is an approach and a set of techniques, which can use for studying the exchange of resources among actors (i.e., individuals, groups, or organizations). Regular patterns of information exchange reveal themselves as online social networks. The actors are nodes in the networks and information exchange relationships are connectors between nodes [7].

The previous research paper was studied about computer virus distribution in online social networks and found that the users who visited MySpace was the first order 40.58%, next was Hi5 35.12% and Facebook was the last one 24.30%. The virus behavior analysis was JS/PackRedir. A! tr. dldr 90.36% from private IP address 172.16.10.96. The result of correlation analysis between the usage of online social networks and virus distribution was significant at the 0.01 level (2-tailed) [8].

In this paper, we have carried out the anomaly and cyber attack patterns in online social networks, identified and visualized the influencing nodes of such anomaly and cyber attack patterns in online social networks.

The organization of this paper as follows: Section I demonstrates about the background. Section II reviews the related works. Section III introduces the proposed approach. Section IV explains about the researching methodologies. Section V presents the data source that used in this research. Section VI illustrates the result and analysis. Section VII presents the discussion. Finally, section VIII shows the conclusion and references.

II. Related Works

Data Mining has been interested in various research applications such as improved the performance of intrusion detection system, detected the anomaly and attack patterns, classified network behaviors, etc.

Recently, many researchers have interested to propose the frameworks or the methods in data mining to improve the performance of intrusion detection system:

C. Azad et al., studied many research papers to classify the area of interesting in Data Mining. Anomaly detection was interested in first priority that presented 67% [9]. R. Smith et al., introduced a new clustering algorithm based on the neural network architecture called an autoassociator (AA) [10]. D. Fu et al., presented the improved association analysis algorithm based on FP-Growth and FCM network intrusion detection technologies based on statistical binning [11]. The result of this research found the intrusion activity in time and solved the problem of data mining speed effectively; enhanced the detective ability of intrusion detection. J. Song et al., presented an approach that was different from the traditional detection models based on raw traffic data. The proposed method could extract unknown activities from IDS [12].

Data Mining was used to detect the anomaly and attack patterns: M. Bordie et al., developed a data-mining system with an integrated set of analysis methods that used to analyze a large amount of IDS log data to discover interesting, previous unknown and actionable information [13]. They presented an architecture and data mining process that used a set of integrated tools including visualization and data analysis and utilized several data-mining algorithms, including temporal association, event burst, and clustering, to discover valuable patterns. H. Yang et al., presented an anomaly detection approach based on clustering and classification for intrusion detection [4]. They used connections obtained from raw packet data of the audit trail and then performed clustering to group training data points into cluster, from which selected some clusters as normal and non-attack profile.

The influence-based classification algorithm was used to classify network behaviors. Z. Qi et al., proposed k-walks in (1+A) k matrix for studying the connective relationship of a network. They normalized the walker vector of each node into unit n-dimensional unit spherical surface, and then they computed geodesic lines between two nodes on the spherical surface, and got the best part of the community by k-means clustering method [14]. Z. Jiliu et al., used the criminal data mining by applying data mining technology and graph theory to Social Network Analysis (SNA) in crime groups or terrorist organizations [15]. They proposed a Crime Group Identification Model (CGIM) based on time step by calculating the group similarity and the group members aggregation under different time steps.

Besides, Social network is a phenomenon of the interaction among the people in a group, community or Internet world. It can visualize as a graph, where a vertex corresponds to a person in that group and an edge represents some kind of connections between the corresponding persons. Online social networks are interested in various research areas as well such as social network analysis, searching and discovering mechanisms in Social Networks, information diffusion, the problem of information dissemination, etc.

Some researchers have interested in the applications of social network analysis: A. Apolloni et al., used a probabilistic model to determine whether two people would converse about a particular topic based on their similarity and familiarity [16]. H. Xia et al., analyzed the key factors influencing knowledge diffusion and innovation through quantitative analysis about network density, centrality and the cohesive subgroup [17].

In addition, S.H. Sharif et al., reviewed on search and discovery mechanisms in Social Networks. They classified the existing methods into four categories: people search, job search, keyword search and web service discovery [18]. S. Sharma et al., presented a centrality measurement and analysis of the social networks for tracking online community by using betweenness, closeness and degree centrality measures [19].

The approaches and techniques were proposed in order to study of information diffusion: C. Haythornthwaite introduced social network analysis as an approach and a set of techniques for the study of information diffusion [7]. T. Fushimi et al., attempted to answer a question "What does an information diffusion model tell about social network structure?". They proposed a new scheme for an empirical study to explore the behavioral characteristics of information diffusion model [20]. C.T. Butts applied the notion of algorithmic complexity to the analysis of social network, based on theoretical motivation regarding constraints on the graph structure [21]. C. Lo Storto presented an approach useful to analyze the performance of the product development process [22].

Some researchers have concentrated on the problem of information diffusion in social networks: M. Kimura et al., addressed the problem of efficiently estimating the influence function of initially activated nodes in a social network under the susceptible/infected/susceptible (SIS) model [23]. A. Pablo et al., solved the influence maximization problem in social networks with greedy algorithm [24]. They considered two information diffusion models, Independent Cascade Model and Linear Threshold Model. Their proposed algorithm compared with traditional maximization algorithms such as simple greedy and degree centrality using three data sets.

III. Proposed Approach

Based on the objectives of this research, a new approach was proposed with the complementary of cluster algorithm and social network analysis.

The clustering technique was proposed to detect informative patterns of cyber attacks in online social networks. Consequently, the centroid of the cluster could use to represent the member within each cluster and discover attacks in the audit trail.

Visualization the influencing nodes of anomaly and cyber attacks in online social networks, active nodes would be discovered by social network analysis (SNA) into two classes-Influencing nodes (Influencing Nodes:

IN) and other nodes (Other Nodes: ON), IN was a class which indicated the influencing nodes and ON was a class which indicated the other nodes. The proposed architecture of this research presented as shown in Fig.1.

The software, namely WEKA was used to analyze the IDS data streams of online social networks and present anomaly and cyber attack patterns. The software, namely Pajek was used to analyze and visualize the influencing nodes with high degrees' nodes in online social networks' relationships with 7,035 clients and 25,778 edges of data in April 2012.

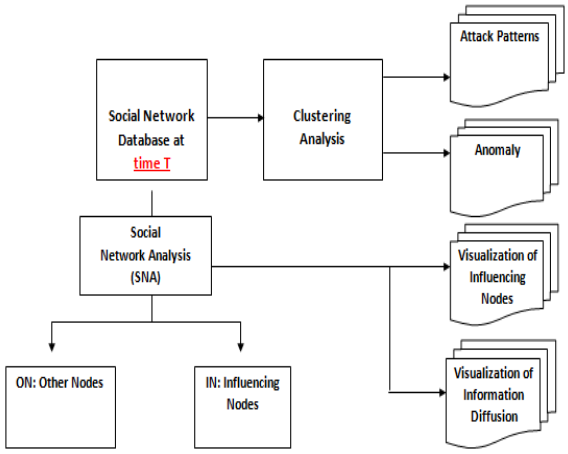

Fig 1: Architecture of the proposed system

IV. Methodologies

The new approach was proposed in this research, which developed with the complementary of cluster algorithm and Social Network Analysis. The cluster algorithm was used to analyze IDS logs data and discover anomaly and cyber attack patterns in online social networks. Social network analysis (SNA) was used to identify and visualize the influencing nodes of anomaly and cyber attacks in online social networks, which could be applied in many applications such as identifying the structure of a criminal network or monitoring and controlling information diffusion, or cyber attacks for secure computer systems in network applications.

A. Clustering Algorithm

Clustering is a useful technique for the intrusion detection of online social networks as malicious activities will cluster together, separate itself from non-malicious activities. The clustering algorithm will group the similar objects from different objects by considering between similarity measure and distance measure.

There are two most clustering analysis models, which are hierarchical cluster analysis and non-hierarchical cluster analysis. The hierarchical cluster analysis is included Agglomerative method and Division method while the method of non-hierarchical cluster analysis is K-Means clustering. The cluster analysis with distance measure and K-Means clustering method was used in this research.

Formally, a cluster analysis can be described as the partitioning of a number N of classifying objects-a number of patterns with an endless dimension P in K groups or clusters {Ck; k = 1,…, K}. Given N objects X = {xi, i = 1,…, N}, where xi, j denotes the jth element of xi, the grouping of all objects with index i = 1,…, N in clusters k = 1,…, K can be defined as follows:

$$W_{k,i} = \begin{cases} 1, \ if\ pattern\ X_i\ \epsilon\ cluster\ C_k \\ 0, \ otherwise \end{cases} \quad (1)$$

The association of each object to a cluster is unique by applying two conditions for the matrix.

$$W_{k,i}\ \epsilon\ \{0,1\}; \sum_{k=1}^{K} W_{k,i} = 1 \quad (2)$$

Furthermore, let the following definition denominate the number of objects belonging to a cluster Ck:

$$|C_k| = \sum_{i=1}^{N} W_{k,i} \quad (3)$$

The Euclidean distance was used as a similarity measure function to identify the similarity of two feature vectors. Whether a set of points is close enough to be considered a cluster. The distance measure D (x, y) was used to tell how far points x and y are. Often, the points may think to live in k-dimensional Euclidean space, and the distance between any two points, say as

$$x = [x_1, x_2, \ldots, x_k] \text{ and } y = [y_1, y_2, \ldots, yk] \quad (4)$$

$$D(x,y) = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \quad (5)$$

The correlation coefficient provides another possibility to measure similarities between two classification objects:

$$R(x,y) = \frac{\sum_{i=1}^{n} x_i, y_i}{\sqrt{\sum_{i=1}^{n}(x_i)^2}\sqrt{\sum_{i=1}^{n}(y_i)^2}} \quad (6)$$

B. Social Network Analysis (SNA)

Social network analysis is based on the principles of graph theory, which consists of a set of mathematical formulae and concepts in the study of patterns of lines. Actors are the points in the graph, and relationships are the lines between actors, Sociograms are graphs of social

networks. Centrality is the extent to which a person is in the center of a network. Central people have more influence in their network than people who are less central. Measures of centrality include degree, betweenness and closeness centrality.

1. A network is represented by a matrix that called the adjacency matrix A, which in the simplest case is a (n x n) symmetric matrix, where n is the number of nodes in the network. The adjacency matrix has elements.

$$A_{ij} = \begin{cases} 1 & if\ i\ and\ j\ are\ connected \\ 0 & otherwise \end{cases} \quad (7)$$

2. Let a graph G= (V, E) represents the graph of online social networks where vertices V corresponds to contact nodes (users) in the network, and edges E corresponds to the information sending events among users.
3. The degree of a vertex in a network is the number of edges attached to it. In mathematical terms, the degree Di of a vertex i is:

$$D_i = \sum_{j=1}^{n} A_{ij} \quad (8)$$

4. The betweenness Ba of a node a is defined as the number of geodesics (shortest paths between two nodes) passing through it:

$$B_a = \sum_{j}^{n} \sum_{a}^{n} g_{ij}(a) \quad (9)$$

Where $g_{ij}(a)$ indicates whether the shortest path between two other nodes i and j passes through node a.

5. The Closeness Ca is the sum of the length of geodesics between a particular node a and all the other nodes in a network. It actually measures how far away one node is from other nodes and sometimes called farness. Where l(i, a) is the length of the shortest path connecting nodes i and a:

$$C_a = \sum_{i=1}^{n} l(i, a) \quad (10)$$

6. For node i, we definitely InDg (I) as the unique number of edges are sent to node i.

$$InDg(i) = \frac{1}{|V|} \sum_{j=1}^{i} E_{ji} \quad (11)$$

7. For node i, we definitely OutDg (I) as the unique number of edges are received from node i.

$$OutDg = \frac{1}{|V|} \sum_{j=1}^{i} E_{ij} \quad (12)$$

8. The total unique number of edges is sent to the node i and the number of edges are received from node I definitely in TotalDg(i)

$$TotalDg(i) = \frac{1}{|V|} \sum_{j=1}^{|V|} \left\lceil \frac{E_{ij} + E_{ji}}{2} \right\rceil \quad (13)$$

V. Data Sources

The organization in this research is the public health sector, which concerns about high privacy and security of information system and needs to avoid from the harmful of network applications. The event log data from IDS (Intruder Detection System) of Health Care Organization with Head Office in Bangkok and 12 regional centers (RC1, RC2… RC12) had located in the different region cover 76 provinces of Thailand in April 2012 with high severity were used in this research. Sample record of raw IDS logs file presented as shown in Table I.

Table 1. Network IDS Log Record Data Defenition

Date	2012-11-13
Time	13:03:39
Attack_id	10873
Severity	High
Src	172.16.10.96
Dst	172.16.10.137
Src_port	2156
Dst_port	80
Msg	The remote attackers can gain control of vulnerable systems

In this example, besides date and time stamp, the attributes of IDS log file were: attacked ID (attack_id) for representing the type of intruder signature; severity presented the level of risk; source IP address, destination IP address, source port, destination port, and a detailed message for additional explanation of this signature. The meaning of the example alert above, the attack id was 10873; the source IP was 172.16.10.96 with port number 2156; the destination IP was 172.16.10.137 with port number 80; and activity was "The remote attackers can gain control of vulnerable systems".

As mention above, this research focused on log data of intrusion detection system of online social networks to

identify and discover unknown patterns, the data preprocessing as follows:

1. The attack logs data of online social networks were filtered from Intrusion Detection System log to analyze and discover anomaly and cyber attack patterns.
2. The raw events were enhanced by augmenting the attack log data with other relevant information not found in the raw event logs, such as the type of web (web 1.0, web 2.0), group number of online social networks in each real IP destination address, so on.
3. The attack logs data in April 2012 were filtered to IDS log data of online social networks (web 2.0) to analyze and discover unknown patterns with Clustering Technique and Social Network Analysis (SNA).
4. This research merged online social networks' relationships with members, who were attacked from the attackers of such online social networks. The online social networks' relationships in the real world are too big to present useful information as shown in Fig.2, which was the example of online social networks' relationships with 7,035 nodes and 25,778 edges in April 2012. They became very difficult to read as the member of actors' increases.

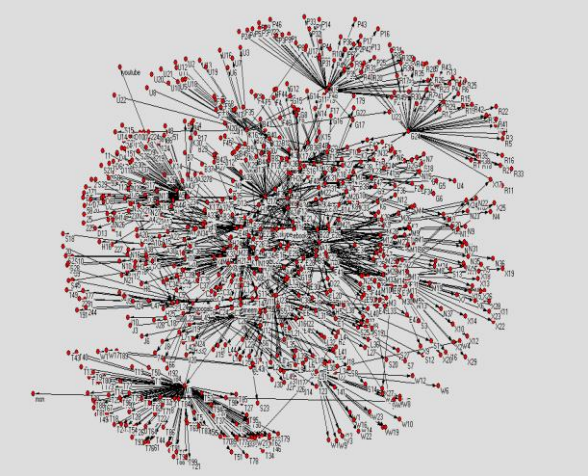

Fig 2: The relationships in online social networks

VI. Results and Analysis Results

The approach was proposed with the complementary of cluster algorithm and Social Network Analysis. The cluster algorithm was used to analyze IDS log data to discover anomaly and cyber attack patterns in online social networks. Social network analysis (SNA) was used to identify and visualize the influencing nodes of anomaly and cyber attacks in online social networks. The results of this research presented as follows;

A. Anomaly and Cyber attacks Patterns in Online Social Networks

The anomaly and cyber attacks patterns in online social networks were analyzed with a cluster algorithm by using source IP address. The model and evaluation of test set categorize signatures into four groups. The anomaly and cyber attacks patterns caused from attackers, which executed an arbitrary program on infected systems with 40% of the test set. The anomaly with HTTP with 37% of the test set. The remote attackers can gain control of vulnerable systems with 19% of the test set. Denial of Service (DOS) with 4% of test set. The result presented as shown in Fig. 3.

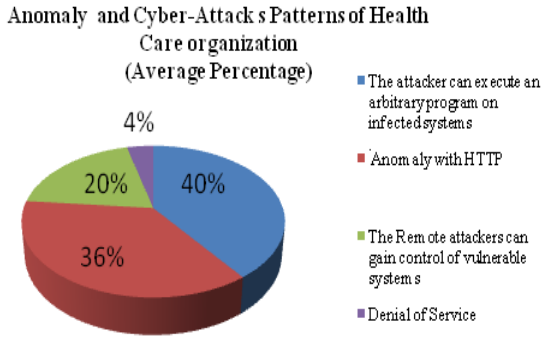

Fig 3: Social networks anomaly and cyber attacks patterns analysis result

B. The Proportion of Anomaly and Cyber attacks Patterns in Each Branch of the Organization

In this experiment, 60% of the data stream was training set and 40% of the data stream was test set. When the training process finished and system model was built, the association rules of anomaly and cyber attacks patterns in online social networks would be discovered. The association predictive model was evaluated and best rules presented with eight principles.

The anomaly and cyber attack patterns at head office and 12 regional centers were the same patterns. The attackers executed an arbitrary program on infected systems was the highest percentage of risk of RC6, located at Khon Khan. Anomaly with HTTP was the highest percentage of risk of RC2, located at Lopburi. The remote attackers could gain control of vulnerable systems was the highest percentage of risk of RC9and RC10, located at Pitsanuloke and Chiang Mai, respectively. DoS was the highest percentage of risk of Head office in Bangkok. The proportion of risk in each signature presented as shown in Table II and Fig. 4.

Table 2. Percentage of Anomaly and Attack Classify by Site and Signature

Sites	The attacker can execute an arbitrary program on infected systems	Anomaly with HTTP	The Remote attackers can gain control of vulnerable systems	Denial of Service
Head Office	44%	27%	19%	10%
RC1	37%	48%	14%	1%
RC2	29%	56%	12%	3%
RC3	41%	30%	24%	5%
RC4	36%	41%	21%	2%
RC5	42%	38%	15%	5%
RC6	47%	39%	13%	1%
RC7	42%	33%	19%	6%
RC8	43%	37%	19%	1%
RC9	38%	31%	27%	4%
RC10	37%	29%	27%	7%
RC11	42%	35%	19%	4%
RC12	41%	32%	26%	1%
Average	40%	36%	20%	4%

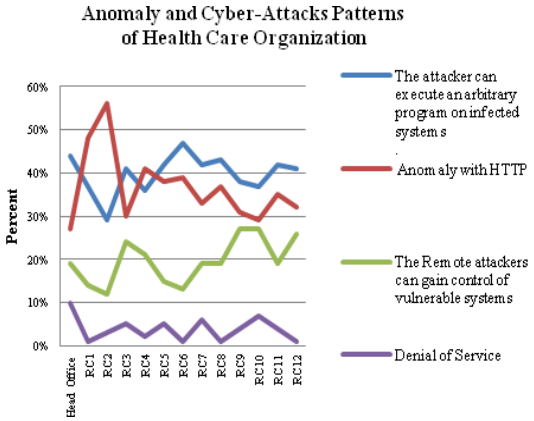

Figure 4: Percentage of anomaly and cyber attacks patterns classify by signature

C. The Proportion of Anomaly and Cyber attacks Patterns Classify by Online Social Networks

The proportion of anomaly and cyber attack patterns classified by online social networking presented as shown in Table III, Fig. 5 and Fig. 6.

1. The proportion of risk caused by the remote attackers could gain control of vulnerable systems when users access Facebook was 100%.
2. The proportion risk caused by the attacker can execute an arbitrary program on infected systems when users access googalz was 100%.
3. The proportion of risk caused by Denial of Service when users access Skype was 55.56% and MSN was 44.44% as shown in Fig. 5.
4. The proportion of risk caused by anomaly with HTTP when users access Facebook, imeem, YouTube, Hi5 and Twitter was 30%, 16%, 5 %, 47% and 2%, respectively as shown in Fig. 6.

Table 3. Percentage of Risks Classify by Online Social Networks

Social Networks	Signature	Percent
Facebook	The remote attackers can gain control of vulnerable systems	100.00
Googalz	The attacker can execute an arbitrary program on infected systems	100.00
Skype	Denial of Service	55.56
MSN	Denial of Service	44.44
Facebook	Anomaly with HTTP	29.73
imeem	Anomaly with HTTP	16.22
YouTube	Anomaly with HTTP	5.41
Hi5	Anomaly with HTTP	46.52
Twitter	Anomaly with HTTP	2.12

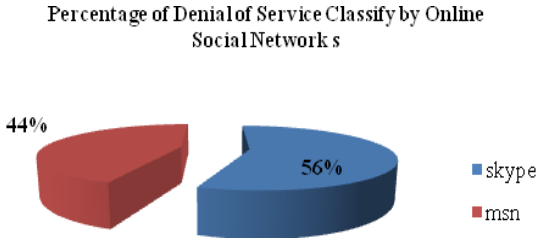

Fig 5: Percentage of Denial of Service classifies by online social networks

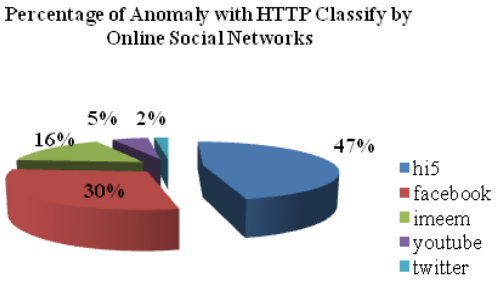

Fig 6: Percentage of Anomaly with HTTP classifies by online social networks

D. Identifying and Visualization the Influencing Nodes of Anomaly and Cyber attacks in Online Social Networks

The centrality is the extent to which a person is in the center of a whole network. Central people have more influence in their network. Measures of centrality include degree, betweenness and closeness centrality.

1. This research presented 42 nodes (node label 3, 60, 87, googalz, 224, 145, 119, B6, 216, 217, Facebook, Skype, MSN, Hi5, 234, 229, 30, 20, 179, 171, 96, 44, 229, A47, A12, B9, C26, B31, A24, C40, F48, J7, G24, E26, H21, M14, K16, L41, E29, E21, 235, and G19) these were the influencing nodes. A new network of 42 influencing nodes presented the cyber attack patterns, which called Egocentric network as shown in Fig. 7.
2. A network member with a higher degree could be the leader or "hub" in a network.
3. Top five In-degree nodes were node label 60, 87, 3, googalz and 224 with an In - degree equal 109, 109, 107, 72, and 60, respectively.
4. Top five Out-degree nodes were node label 234, J7, 217, 229 and 216 with an Out - degree equal 1451, 101, 75, 68 and 61, respectively.
5. Betweenness measures the extent to which a particular node lies between other nodes in a network. Top five Betweenness nodes were node label 216, 119, 234, C26 and J7 with Betweenness measure equal 1513.400, 1338.200, 1197.467, 785.233 and 695.000, respectively.
6. Top five Closeness nodes were node label 119, 217, 216, 171 and 234 with closeness measure equal 8.456, 8.436, 8.413, 8.301 and 8.293, respectively. This meant that node label 119, 217, 216, 171 and 234 were persons in the center of the network and could be the leader in the network because they were both influencing nodes and high degrees. If they were attacked from many threats or cyber attacks such as social engineering or malware, they would be influenced in their network.

A network member with a higher degree could be the leader or "hub" in the network. Betweenness measures the extent to which a particular node lies between other nodes in a network. Closeness is the sum of the length of geodesics between a particular node and all the other nodes in a network. It actually measures how far away one node is from other nodes. This research presented 42 influencing nodes, which could be partitioned to new network with high degree centrality.

E. Anomaly and Cyber attacks Diffusion by Influencing Nodes in Online Social Networks

Egocentric Network is the graphs in subgroups of the whole network, builds a picture of a typical actor in any particular environment and shows how many ties they maintain, and what kinds of information they send to and receive from others in their network. This research presented the sociogram of each representative node, for example node label 119, 216, 217, 234 and their linkages of relationships as shown in Fig. 8, Fig. 9, Fig. 10, Fig. 11, and Fig. 12 respectively.

Finally, the results of Egocentric Networks presented the closet vertices of node label 119 or node label 50 and node label 53. The closest vertices of node label 216 were node label 20 and node label 29. The closest vertices of node label 217 were node label 18 and node label 28. The closest vertices of node label 234 were node label 29 and node label 42. These could be implied that when node label 119, 216, 217, 234 were attacked from many threats such as social engineering or malware, they would send these attacks to node label 50, 53, 20, 29, 18, 28 29 and 42 with the highest proportion as shown in Fig. 13.

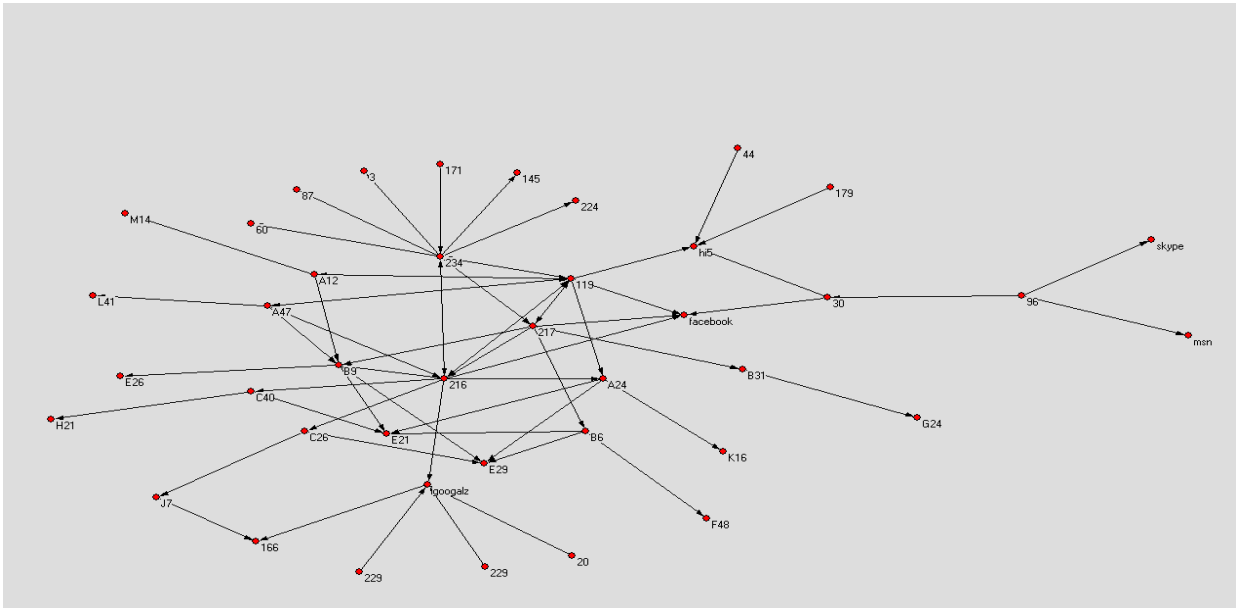

Fig 7: Subgroup Identification of influence nodes

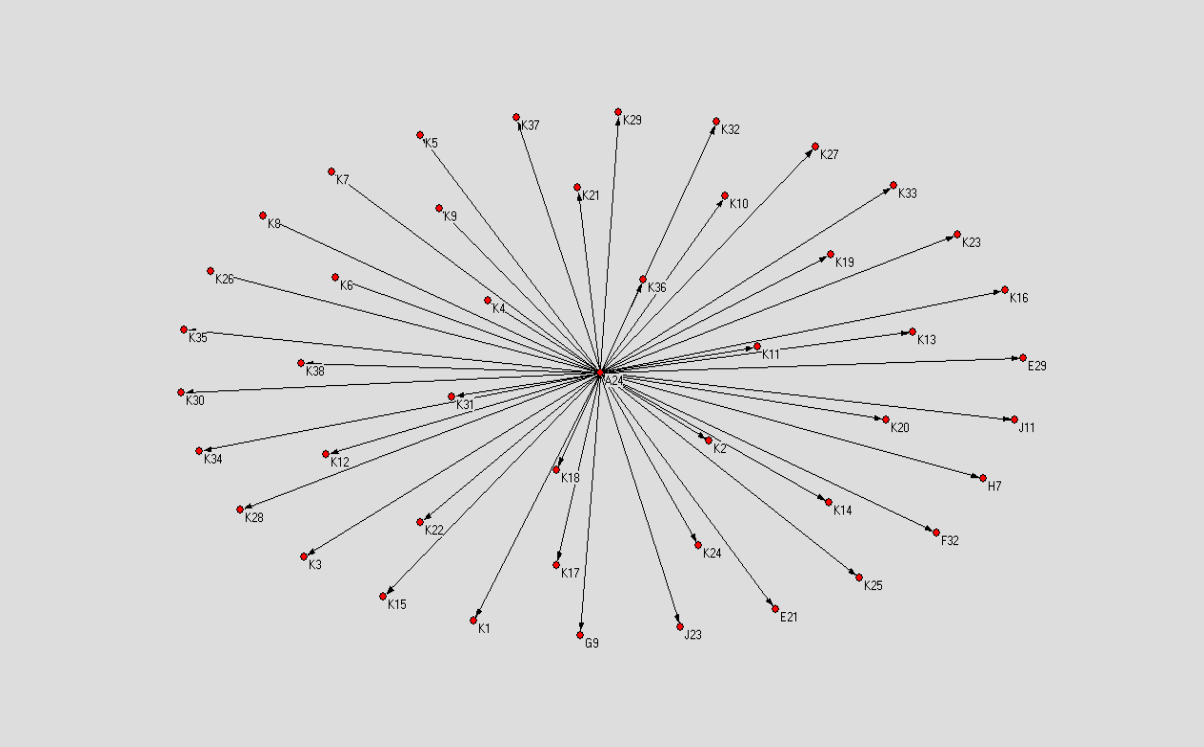

Fig 8: Egocentric network of representative node A24

Fig. 8 presented the characteristics of node label A24 as follows: Node label A24 had 93 edges and average degree was equal 1.9565217. The closest vertices were node label 32 and node label 36 with distance: 0.08033. The Closeness Centralization was equal 1.00000.

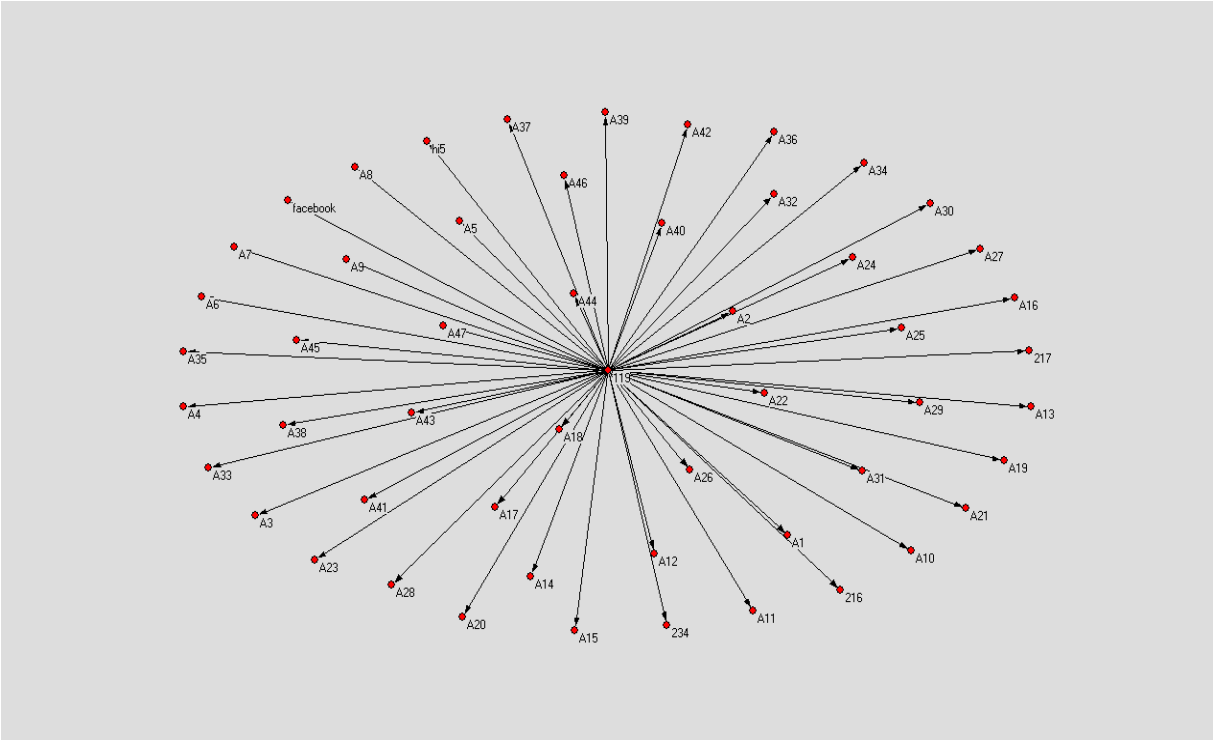

Fig 9: Egocentric network of representative node 119

Fig. 9 presented the characteristics of node label 119 as follows: Node label 119 had 107 edges and average degree was equal 1.9622642. The closest vertices were node label 50 and node label 53 with distance: 0.06899. The Closeness Centralization was equal 1.00000.

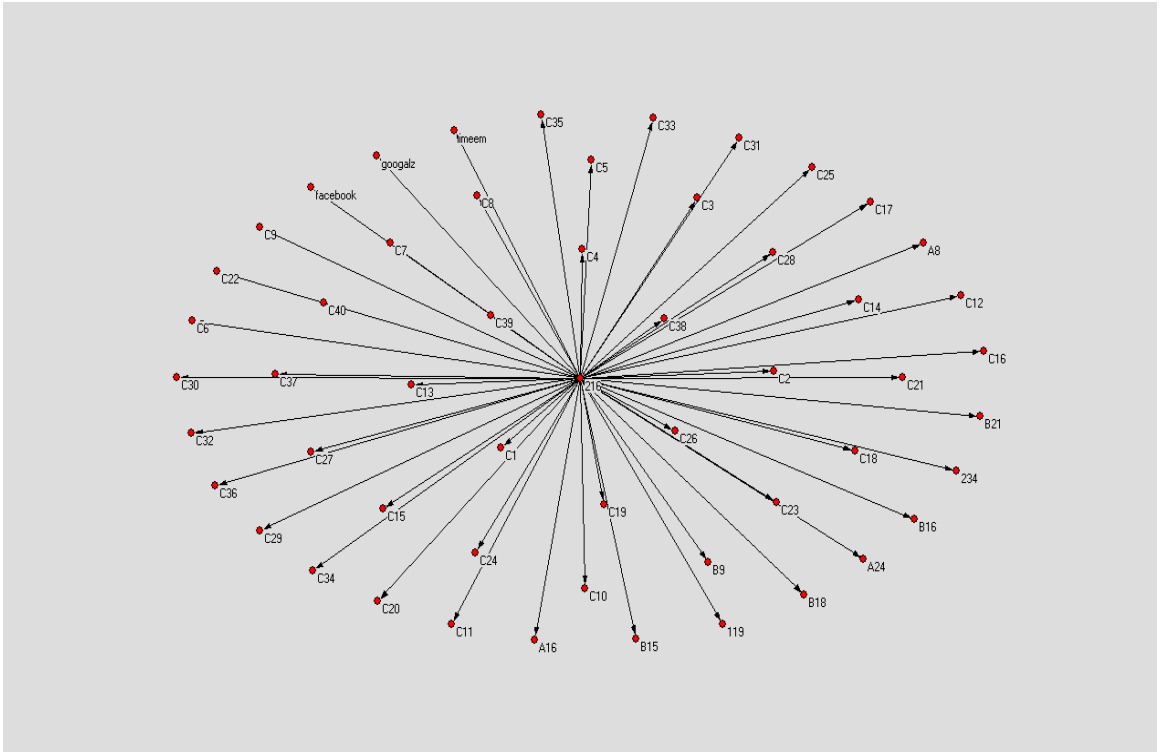

Fig 10: Egocentric network of representative node 216

Fig. 10 presented the characteristics of node label 216 as follows: Node label 216 had 109 edges and average degree was equal 1.9629630. The closest vertices were node label 20 and node label 29 with distance: 0.06895. The Closeness Centralization was equal 1.00000.

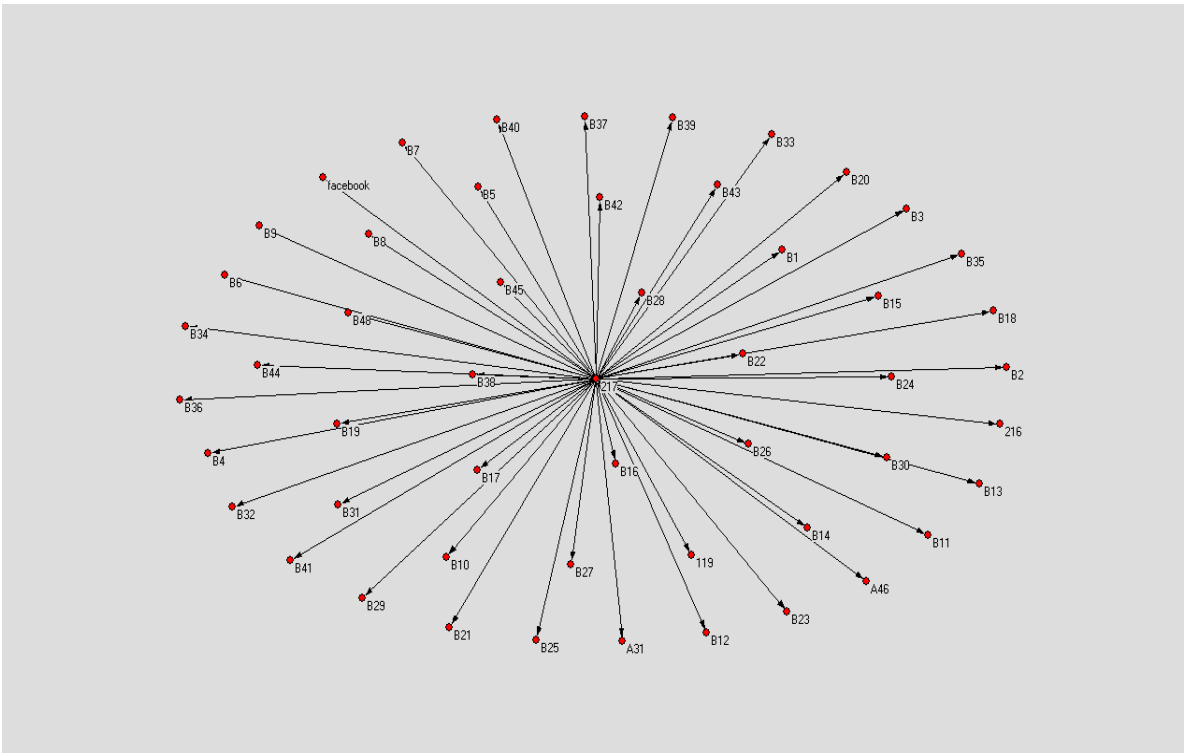

Fig 11: Egocentric network of representative node 217

Fig. 11 presented the characteristics of node label 217 as follows: Node label 217 had 105 edges and average degree was equal 1.9615385. The closest vertices were node label 18 and node label 28 with distance: 0.07135. The Closeness Centralization was equal 1.00000.

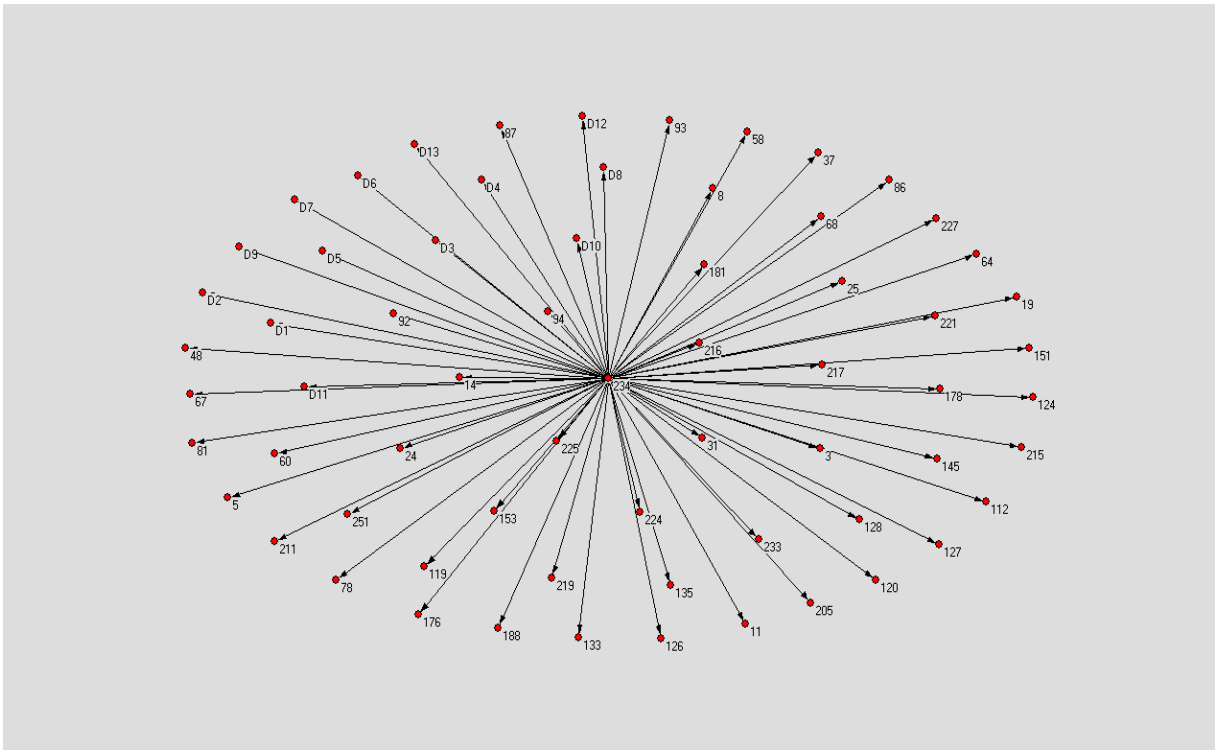

Fig 12: Egocentric network of representative node 234

Fig. 12 presented the characteristics of node label 234 as follows: Node label 234 had 131 edges and average degree was equal 1.9692308. The closest vertices were node label 29 and node label 42 with distance: 0.05801. The Closeness Centralization was equal 1.00000.

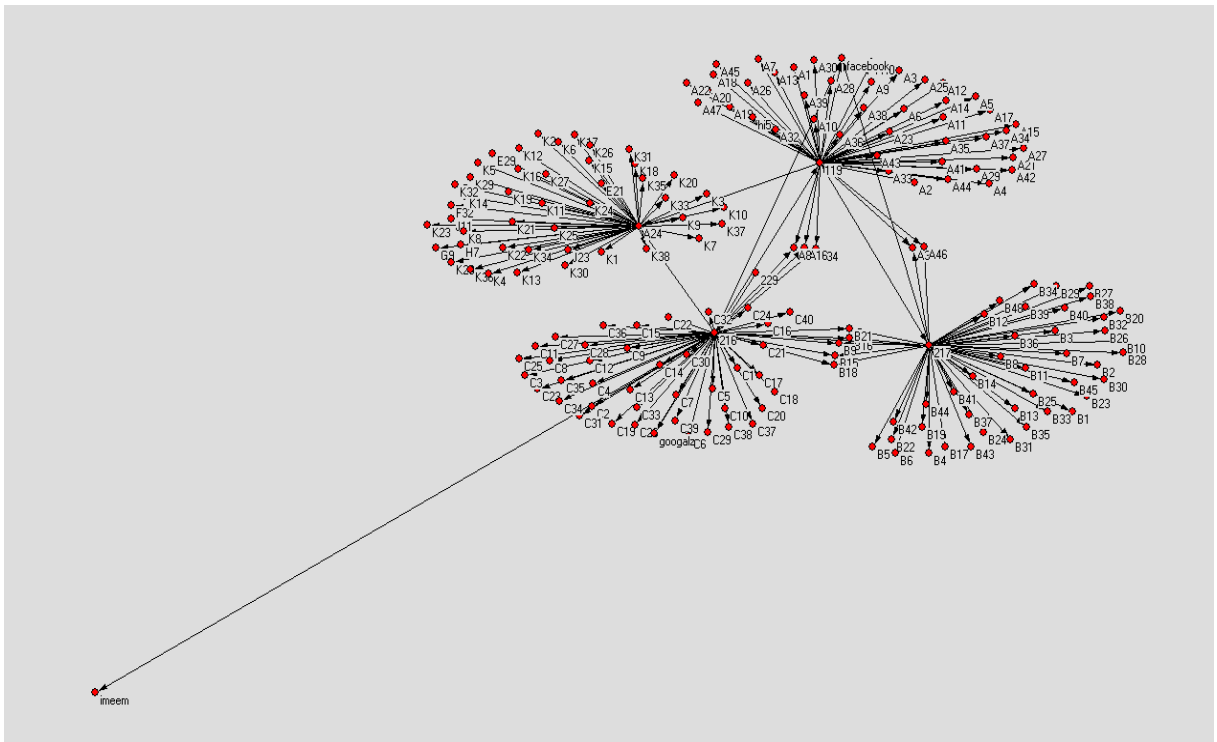

Fig 13: Information Diffusion of representative nodes 119, 216, 217 and 234

Fig. 13 presented the relationships of representative node label 119, 216, 217, and 234 as follows: Average degree was equal 2.1604278. The closest vertices were node label 17 and node label 21 with distance: 0.00331. The Closeness Centralization was equal 0.45397.

VII. Discussion

The IDS log data of online social networks was used to demonstrate the proposed method, which represented of the public health organizations in Thailand. They have concerned high privacy and security of information system and had to avoid from the harmful of network application such as an Internet application and other applications or services.

The results presented the anomaly and cyber attack patterns in online social networks, which differed in each region in Thailand. Visualization the influencing nodes of anomaly and cyber attack patterns in online social networks had very helpful in analyzing and understanding social groups.

The online social networks' relationships in the real world were too big to present with useful information to identify the criminal or cyber attacks. The proposed methodology could finalize the result with 42 influencing nodes of anomaly and cyber attacks in online social networks with key factors about density, centrality (degrees, betweenness and closeness) and the cohesive subgroup to show hidden information in each subgroup of whole networks, called Egocentric Networks.

VIII. Conclusion

This research proposed the methodology of information security analysis that was complement of Cluster Algorithm and Social Network Analysis to present anomaly and cyber attacks patterns in online social networks. The visualization of influencing nodes of such anomaly and cyber attacks patterns in social networking presented in this research as well.

The cluster algorithm analyzed the large amount of IDS data to identify the unknown online social networks anomalous and cyber attack patterns. This analysis was computed with a test set, presented attackers' patterns as (1) The attackers attempted to execute an arbitrary program on infected systems (2) An anomaly pattern (3) The remote could gain control of vulnerable systems (4) Denial of Service.

Social network analysis (SNA) found 42 influencing nodes of such anomaly and cyber attacks patterns and presented the visualization of the influencing nodes in whole network and closet vertices of influencing nodes, which could not avoid from the harmful of the network. The experiment and results illustrated the proposed information security analysis methodology was significance analysis.

The proposed approach could be applied as a guide for further investigate of social network behavior to improve the security model and notify the risk, computer viruses or cyber attacks for online social networks in advance.

Acknowledgements

I would like to thank National Research Council of Thailand (NRCT) for providing funding and material related to educational research.

References

[1] U. Flegel and J. Biskup, "Requirements of Intrusion Reductions for Cooperating Intrusion Detection Agents, in Emerging Trends in Information and Communication Security," vol. 3995, G. Muller, Eds. Berlin, Germany: Springer-Verlag, pp. 466-480, 2006.

[2] A. R. Kumar, P. and S. Selvakumar, "M_2KMIX: Identifying the Type of High Rate Flooding Attacks using a Mixture of Expert Systems," Journal of IJCNIS, vol. 4, no. 1, pp. 1-16, February 2012.

[3] C. Sheth, R. Thakker, "Performance Evaluation and Comparison of Network Firewalls under DDoS Attack," Journal of IJCNIS, vol. 5, no. 12, pp. 60-67, October 2013.

[4] H. Yang, F. Xie, and Y. Lu, "Clustering and Classification Based Anomaly Detection, in Fuzzy Systems and Knowledge Discovery," vol. 4223, L. Wang et al., Eds. Berlin, Germany: Springer-Verlag, September 2006, pp.1082-1091.

[5] I. S. I. Abuhaiba, H. B. Hubboub, "Reinforcement Swap Attacks against Directed Diffusion in Wireless Sensor Networks," Journal of IJCNIS, vol. 5, no. 3, pp. 13-24, March 2013.

[6] V. BiBhu, K. Roshan, K. Balwant Singh, D. Kumar Singh, "Performance Analysis of Black Hole Attack in Vanet," Journal of IJCNIS, vol.4, no.11, pp.47-54, October 2012.

[7] C. Haythornthwaite, "Social Network Analysis: An Approach and Technique for the Study of Information Exchange," Proc. The ALISE conference, San Antonio, Texas, Autumn 1996, pp. 323-342.

[8] P. Limsaiprom and P. Tantatsanawong, "Study of Computer Virus Distribution in Social Network: A case Study of National Blood Centre, Thai Red Cross Society," Proc. The National Conference on Computer Information Technologies, Thailand, March 2010, pp. 115-120.

[9] C. Azad and V. K. Jha, "Data Mining in Intrusion Detection: A Comparative Study of Methods, Type and Data Sets," Journal of IJITCS, vol. 5, no. 8, pp. 75-90, July 2013.

[10] R. Smith, N. Japkowicz, M. Dondo, and P. Mason, "Using Unsupervised Learning for Network Alert Correlation, Advances in Artificial Intelligence," vol. 5032, S. Bergler, Eds. Berlin, Germany: Springer-Verlag, Canada, May 2008, pp. 308-318.

[11] D. Fu, S. Zhou, and P. Guo, "The Design and Implementation of a distributed Network Intrusion Detection System based on Data Mining, World Congress on Software Engineering 2009," Proc. IEEE Xplore, Xiamen, May 2009, pp.446-450.

[12] J. Song, H. Ohba, H. Takakura, Y. Okabe, K. Ohira, and Y. Kwon, "A Comprehensive Approach to Detect Unknown Attacks Via Intrusion Detection Alerts," Proc. ASIAN 2007, Qatar, December 2007, pp. 247-253.

[13] M. Bordie, M. Mei, D. George, and S. Ma, "Next Generation of Data-mining Applications," edited by Kantardzic and Zurada, The institute of Electrical and electronics Engineerings, Inc., 2005, pp. 545-567.

[14] Z. Qi and M. Ying-Hong, "An Algorithm to Detect Community by Geodesic Line in Social Networks," Journal of AISS, vol. 3, no. 6, pp. 328-333, July 2011.

[15] Z. Jiliu, G. Wang, S. Qiao, C. Wang, Y.Lei, "A Crime Group Identification Model Based on Mobile Communication Vestige Records," Journal of JCIT, vol. 6, no. 5, pp. 69-77, July 2011.

[16] A. Apolloni, K. Channakeshava, L. Durbeck, "A Study of Information Diffusion over a Realistic Social Network Model," Proc. International Conference on Computational

Science and Engineering, Vancouver, BC, August 2009, pp. 675-682.
[17] H. Xia, R. Wang and S. Hu, "Social Network Analysis of the knowledge Diffusion among University Students," Proc. The Second International Symposium on Knowledge Acquisition and Modeling, Wuhan, December 2009, pp. 343-346.
[18] S.H. Sharif, S. Mahmazi, N. J. Navimipour, and B. F. Aghdam, "A Review on Search and Discovery Mechanisms in Social Networks," Journal of IJIEEB, vol. 5, no. 6, pp. 64-72, December 2013.
[19] S. Sharma and G. N. Purohit, "A New Centrality Measure for Tracking Online Community in Social Netowrks," Journal of IJITCS, vol. 4, no. 4, pp. 47-53, April 2012.
[20] T. Fushimi, T. Kawazoe, K. Saito, M. Kimura, and H. Motoda, "What does an Information Diffusion Model Tell about Social Network Structure," Proc. PKAW, D. Richards, and B-H. Kang, Eds. Berlin, Germany: Springer-Verlag, 2009, pp. 122-136.
[21] C.T. Butts, "The complexity of social networks: theoretical and empirical findings, In Social Networks," vol. 23, pp. 31-71, 2001.
[22] C. Lo Storto, "Investigating information flows across complex product development stages by using social network analysis," Proc. The Complexity in Engineering, Rome, February 2010, pp. 118-120.
[23] M. Kimura, K. Yamakawa, K. Saito, and H. Motoda, "Community Analysis of Influential Nodes for Information Diffusion on a Social Network," Proc. IEEE Xplore, Hong Kong, June 2008, pp. 1358-1363.
[24] A. Plabo, V. Pablo, and K. Saito, "Selecting the Most Influential Nodes on Social Networks," Proc. International Joint Conference on Neural Networks, Orlando, FL, August 2007, pp. 2397-2402.

Prajit Limsaiprom received the B. Sc in Applied Statistics from King Mongkut's Institute of Technology, Ladkrabang, Bangkok, THAILAND in 1989, the Master Degree in Information Technology in Business (Statistics) from Chulalongkorn University, Bangkok, THAILAND in 2001. She currently is Information Technology Manager, National Blood Centre, Thai Red Cross Society, Bangkok, THAILAND and presently a Ph.D. candidate in School of Information Technology, Sripatum University, Bangkok, THAILAND. Her research interests in the area of Data Mining Analysis, Information Security Visualization, Social Network Analysis and Social Networks Security. She has recorded in Who's Who in the world in Information Technology.

Gp. Capt. Assoc. Prof. Dr. Prasong Praneetpolgrang received the B.Sc. (1st Hons) in Electrical Engineering from the Royal Thai Air Force Academy, Bangkok, THAILAND, in 1987, the Master Degree in Computer Engineering, 1989, the Master Degree in Electrical Engineering, 1993, and the Ph.D degree in Computer Engineering from Florida Institute of Technology, Florida, USA, in 1994. He currently has the rank of associate professor at the Information science institute, Sripatum University, Bangkok, Thailand. His research interests are in the areas of Computer and Information Security, Trust Management and IT Governance, e-Commerce and Cloud Applications. Dr. Prasong Praneetpolgrang has more than 100 published articles in these areas. He has served on program committees of both international and national conferences on Computer Science and Engineering, Information Technology and e-Business. He is also a member of the IEEE, and ACM. He has recorded in Who's Who in the world in Information Technology.

Dr. Pilastpongs Subsermsri received Ph.D. in Computer Engineering, University of Houston, USA, M.S.EE. in Electrical Engineering, University of Houston, USA, and B.S. EE. in Electrical Engineering, Khonkaen University. He currently is the Director of the Master of Science Program in Computer Information Systems (CIS) at the Information Science Institute of Sripatum University (ISIS), School of Information Technology, Sripatum University, Thailand. His research interests are in the areas of Agricultural IT, IT Management, IT Strategic Planning and e-Logistics. Dr. Pilastpongs Subsermsri has many published articles in these areas.

How to cite this paper: Prajit Limsaiprom, Prasong Praneetpolgrang, Pilastpongs Subsermsri,"Visualization of Influencing Nodes in Online Social Networks", IJCNIS, vol.6, no.5, pp.9-20, 2014. DOI: 10.5815/ijcnis.2014.05.02