Comparing Structure Learning Algorithms of Bayesian Network in Authentication via Short Free Text

Charoon Chantan
Technopreneurship and Innovation Management Program, Graduate School, Chulalongkorn University, Bangkok, Thailand.

Sukree Sinthupinyo
Department of Computer Engineering, Faculty of Engineering Chulalongkorn University, Bangkok, Thailand.

Tippakorn Rungkasiri
Faculty of Commerce and Accountancy, Chulalongkorn University, Bangkok, Thailand.

ABSTRACT
In this paper, we empirically evaluate effectiveness of structure learning of Bayesian Network when applying such networks to the domain of Keystroke Dynamics authentication. We compare four structure learning methods of Bayesian Network Classifier – Genetic, TAN, K_2, and Hill Climbing algorithms, on our authentication model, namely Classify User via Short-text and IP Model (CUSIM). The results show that Genetic algorithm was best suited to our model. The findings from the study also indicate that the Accuracy, FAR, and FRR rate of Genetic algorithm are better than other algorithms tested in this work. Moreover, we found that TAN algorithm gives better results in some scenario than other algorithms.

Keywords
Authentication, Classification, Short Free Text, Keystroke Dynamics, Bayesian Network, Internet security.

1. INTRODUCTION
Nowadays, Internet security is an important mechanism to detect and prevent unauthorized users. There are numerous activities, especially Internet Banking in which we cannot identify the users and the intruders, referring to hackers, attackers, and crackers access. An Internet criminal would be the insider or external attacker, who can be the spurious user or examine personal information or forge transactions. Therefore, to prevent important information on the Internet systems, we need a reliable security system to detect the intruders. The current Internet security system uses multi-component authentication which is composed of (1) known information, (2) stored information, and (3) information of each person. However, the security system still has a problem where a weak point can be found in these components [1]. The above-mentioned problems are critical points to find the effective solutions to identify users and prevent the intruders.

Keystroke Dynamics is a biometric method which intends to identify a user from the habitual typing rhythm as a form of personal behavior on a computer keyboard. Various researchers have applied the keystroke dynamics on the fixed text and used some tools to analyze and recognize the typing pattern for authentication. Gaines et al. [2] studied keystroke dynamics by verifying fixed password. The study of Gunetti and Picardi [3] found that the accuracy of research results required long length of free text. Roadrungwasinkul and Sinthupinyo [4] studied and proposed a method of identification by short free text, using the combination of the average/standard deviation score and the Artificial Neural Network. The values of FRR and FAR obtained from the research are better than the existing method based on the short-free-text fashion. In addition, the work of [4] also found that the short free text of 100 characters achieved higher accuracy than text with other length. Also, this mechanism can be applied to detect and classify the user along with IP address and location. The combination of this method with the information of IP address and location can help increase the accuracy without requires addition cost. Aldridge et al. [5] proposed the verification method to verify user using IP address.

Bayesian Network (BNs), known as Belief Networks, has been widely used in the field of pattern recognition. The advantage of learning by Bayesian Network is to find the network structure and parameters that best fit for the training data, according to the scoring function [6].

Thus, the objective of this paper is to explore the role of structure learning algorithm of Bayesian Network used to find the best performance compared to the result of [7]. In addition, Genetic Algorithm, Tree Augmented Naïve Bayes (TAN), K2, and Hill Climbing Algorithm were used to compare and explore the best accuracy in our model. The remaining of this paper was organized as follows: Section 2 provided a literature review of our research. In Section 3, described our research methodology, and proposed the model derived from prior knowledge. Section 4 presented the experimental results. Finally, the conclusions were presented in Section 5.

2. LITERATURE REVIEW
2.1 Biometrics (Keystroke Dynamics)
As we stated before, the problems of Internet security have become more serious in present. Keystroke Dynamics has been studied for a period of time in applying to identify and classify users. The input text of KD used in the existing study could be categorized into fixed and free text. Fixed text could be directly represented as duration of each character or latency of each digraph. However, there were some researches that concerned with free text. The feature vectors of free text could be extracted from any different text input and length of text. Monrose and Rubin [8] proposed the experiment using both fixed text and free text input in the authentication process. A user profile was created based on the statistics (average and standard deviation) of both the keystroke duration and the digraph latency. The experiments were composed of three methods. The first method was the Euclidean distance. Second method was the probability and the last method was the weighted probability score. Hu et al. [9] proposed the method to classify user profile using *k*-nearest neighbor. The result showed the great accuracy which is less than 0.05 of FAR and

less than 0.005 of FRR. In 2005, Gunetti and Picardi [3] proposed the method to support free text in authentication and classification process. The result showed a good performance of FAR and FRR. The study of Gunetti and Picardi could be regarded as one of the best free text classification methods. In this paper, we applied the research result of [4] used in our experimental as the prior knowledge of the probability of Keystroke Dynamics of the model.

2.2 IP Address and Location

Many past researches had used IP Address to authentication or identification the user. Aldridge et al. [5] proposed the method to verify user using an IP address along with other component to better authenticate a user. Park et al. [10] proposed the method of intrusions detection which monitored and verified the intrusion from IP information. In addition, we can use the location and device to verify the intrusion on the Internet as in [11], [12]. In consequence, we can use the information form IP address and location to be the prior knowledge as another variable of the model.

2.3 Bayesian Networks

Bayesian Networks (BNs) (also known as Bayesian Belief Network, Causal Probabilistic Network, Probabilistic, and Cause-Effect Model) is a directed acyclic graph (DAG) with a conditional probability distribution (CP table) for each node. The link between nodes represents probabilistic dependencies among the corresponding random variables. These conditional dependencies in the graph are often estimated using known statistical and computational methods. Hence, BNs combine principles from graph theory, probability theory, computer science, and statistics [13]. BNs have been *popular* in *statistics, machine learning and artificial intelligence.* They enable an effective representation and computation of the joint probability distribution (JPD) over a set of random variables [14], as shown in equation 1,

$$P(y_i, \ldots\ldots, y_n) = \prod_{i=1}^{n} P(y_i | Parent(Y_i)) \qquad (1)$$

Where $\prod$ is the multiplication of $P(y_i|Parent(Y_i))$

The decomposition of joint probability distributions can enable Bayesian Networks to analyze data and extract useful information for decision making, controlling, predicting, and reasoning. In order to use Bayesian inference, prior probabilities and posterior probabilities are required. Various researches use machine learning techniques to test the efficiency of the model, possibly by different algorithms. In addition, there are many researches using Bayesian network for user classification and authentication [15], [16], [17], [18]. To increase accuracy of user classification and authentication, we propose a Bayesian network model which consists of KD, IP, and Location to improve efficiency.

2.4 Bayesian Network Learning

The learning of BNs could be classified into two important tasks: 1) learning of the graphical structure model, and 2) learning of the parameters for that structure model. It is trivial to learn the parameters for the structure that best fits complete data [19]. We will focus on learning the BN structure. There are two methods to construct Bayesian Network: 1) top-down modeling methods, and 2) reverse-engineering methods. Top-down modeling methods focus on finding the direct solutions of Bayesian network structure and parameter assignments from any prior knowledge. On the contrary, reverse-engineering approaches try to utilize learning algorithms to train Bayesian network structure and parameters from a collection of past observations. This study uses the top-down modeling, constraint-based method, and probabilistic relations using Markov boundary. However, our model does not deal with constraint-based algorithms. The method of search-and-score comprises two elements: a search procedure for a network structure and a score (metric) evaluating each structure found in the search [20]. The best-fit data leads to the scoring based algorithms that seek for a structure that maximizes the scoring function [21].

2.5 Bayesian Network Classifiers

There are several algorithms in BNs based on the search-and-score method. This study focuses on four algorithms, which have the potential to learn the dependencies and causal relationship among the variables. They are composed of Tree Augmented Naïve Bayes (TAN), K2, Hill Climbing [22], [23], and Genetic algorithms (GA) [24]. Figure 1 illustrates the structure of algorithms in the Bayesian classifiers considered in this paper. TAN is an extension of the Naive Bayes classifier. The Naive Bayes' assumption that all the features are independent is removed. The dependencies between variables are also taken into account. K2 is a score-based greedy search algorithm for learning Bayesian networks from data. It maximizes the probability of an optimal graph topology, given a dataset, using a Bayesian score to rank different graphs. The algorithm is restricted by an order on the variables. Hill Climbing (HC) is an algorithm used for adding, deleting, and reversing arcs. The search is not restricted by an order on the variables. HC will follow the graph from node to node to increase the value of the solution, until a local maximizing. The concept of Genetic Algorithm is a principle of Charles Darwinian theory of evolution to natural biology. The working of genetic algorithm starts with a population of random chromosomes. The algorithm evaluates these structures and allocates reproductive opportunities. This provides a better solution for the problem and gives a better chance to reproduce. GA operation basically depends on the Schema theorem. GA is recognized as the best optimize function and widely used in pattern discovery, image processing, signal processing, and training Neural Networks [24].

The objective of this study is to compare the performance among these four algorithms of BN classifiers in order to find the best accuracy of our model.

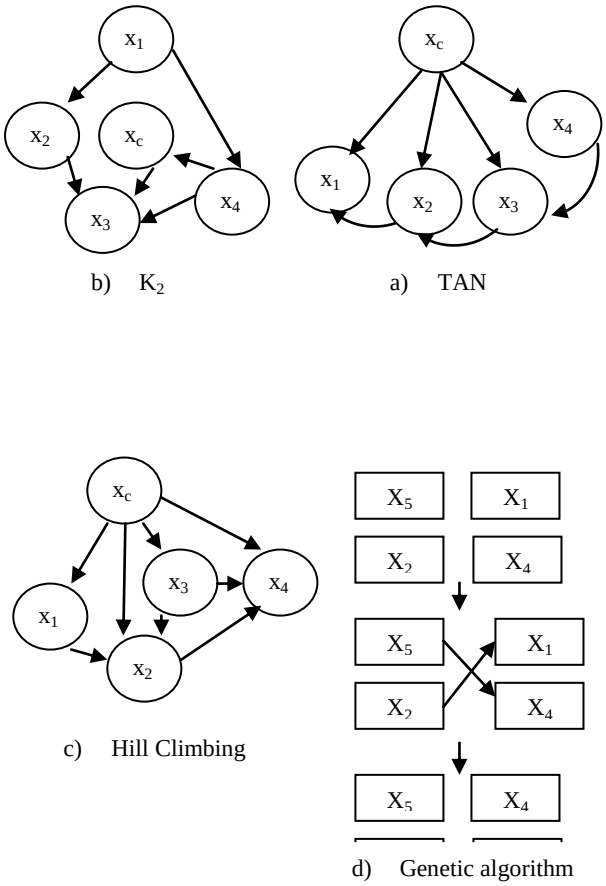

Fig 1: Different Bayesian Network structure of Genetic algorithm [24], TAN, K_2, and Hill climbing [25].

3. RESEARCH METHODOLOGY

3.1 Construction of the Bayesian Network model

Refer to [7], the model constructed from Bayesian network is composed of User of Intruder node, KD node, Location node, IP node, and Output node (CUSIM) (see Figure. 2). Each node uses the prior knowledge for conditional probability. The knowledge derives from results of [4] and statistic survey of [26]. The equations and conditions of CUSIM model depend upon the concept of Bayesian network. (See more details in [7]).

We use the rules of CU-SIM for setting condition and calculation in model shown as below [7].

1). CUSIM answer "Yes", If KD score is high and IP is same.

2). CUSIM answer "Yes", If KD score is high but IP is not same.

3). CUSIM answer "Yes", If KD score is moderate and IP is same.

4). CUSIM answer "No", If KD score is moderate and IP is not same.

5). CUSIM answer "No", If KD score is low and IP is same.

6). CUSIM answer "No", If KD score is low and IP is not same.

3.2 Determine the condition Probability distribution

We use the historical, prior knowledge data to determine the CPT table. The table is composed of the probability distribution of "User/Intruder", "Keystroke Dynamics", "location", and "IP address". Also, the same scenarios to test the model are applied. (See more details in [7]).

3.3 Learning CUSIM network

This study aims to learn the structure and parameter through algorithms, consisting of K2, TAN, Hill Climbing, and Genetics algorithm, and to compare the performance of CUSIM, based on the training data. Summary of the output model of each algorithm consists of scenario analysis and causal analysis to assess the accuracy of the CUSIM. We use WEKA [25] to test and train data set to evaluate the performance of the CUSIM learning by four algorithms. The standard 10-fold cross validation and resample technique is used in the test.

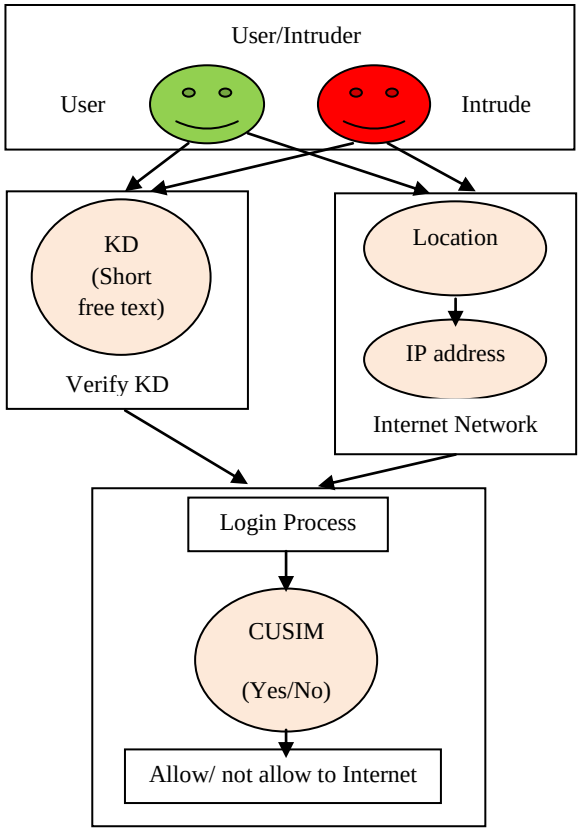

Fig 2: Bayesian network of CUSIM

4. EXPERIMENTAL RESULTS

The results of four algorithms of BNs classifiers were evaluated by the standard metrics of accuracy, precision, recall, F-measure, FRR, FAR and ROC area for CUSIM. Those values were calculated by the Confusion Matrix. In addition, another statistical analysis was used to assess the performance of the different algorithms for comparisons. The kappa statistic measured the agreement of prediction. Mean absolute error is used to range of possible values in terms of

the unit of measurement. The weighted average of all the absolute errors was found from cross validations and was relative to absolute error. Absolute error is a ratio of the mean absolute error of the learning algorithm over the mean absolute error found by predicting the mean of the training data. The lower of the percentage is, the better performance of the classifier would be.

The test results of the four BN's algorithms were presented in two situations and three scenarios. Each situation is shown in Table 8, and Table 9.

In table 8, Scenario 1 is the situation that the conditional probability of P (KD | Real User = High), P (KD | Real User = Medium), and P (KD | Real User = Low) of the user and the intruder were the same value. The conditional probability of P (Location | Real User = Same-Location), P (Location | Real User = Change-Location) of the user and the intruder were the same value. The conditional probability of P (IP| Location = Same-IP), P (IP | Location = Change-IP) of the same-Location and the Change-Location were the same value. The value of probability was sorted in a descending order from P(High) to P(low) value. The results of accuracy, FRR, FAR, Precision, Recall, F-Measure, and ROC of Genetics algorithm are better than TAN, K_2, and Hill climbing algorithms.

In Scenario 2, the conditional probability of P (KD | Real User = High), P (KD | Real User = Medium), and P (KD | Real User = Low) of the user and the intruder were in disorder direction to user. The conditional probability of P (Location | Real User = Same-Location), P (Location | Real User = Change-Location) of the user and the intruder were opposite direction to user. The conditional probability of P (IP| Location = Same-IP), P (IP | Location = Change-IP) of the same-Location and the Change-Location were opposite direction to same-location. The values of probability were not sorted in the same way values but the probability of user was higher than scenario 1. The results of accuracy, FRR, FAR, Precision, Recall, F-Measure, and ROC of Genetics algorithm are better than TAN, K_2, and Hill climbing algorithms.

In Scenario 3, the conditional probability of P (KD | Real User = High), P (KD | Real User = Medium), and P (KD | Real User = Low) of the user and the intruder were opposite direction to user. The conditional probability of P (Location | Real User = Same-Location), P (Location | Real User = Change-Location) of the user and the intruder were opposite direction to user. The conditional probability of P (IP| Location = Same-IP), P (IP | Location = Change-IP) of the same-Location and the Change-Location were opposite direction to same-location. The values of probability were not sorted in the same way but the probability of user was less than scenario 1 and 2. The results of accuracy, FRR, FAR, Precision, Recall, F-Measure, and ROC of TAN algorithm are better than Genetic, K_2, and Hill climbing algorithms. The results were similar to situation 2 "P(User/Intruder) = 90:10".

The results in Table 8 show that the Genetic algorithm yielded to be the best results. Thus, we selected the Genetic algorithm used in CUSIM model for representing the authentication process via short-Free-Text, Location, and IP address. Moreover, the results of CUSIM in Scenario 3 of Genetic algorithm show that the accuracy and FAR values decreased more than those in scenarios 1 and 2. It was the values of Accuracy, FAR, and FRR depending on the conditional probability of Keystroke Dynamics when we know whether it was the real user or the intruder.

Table 9 shows the statistics Kappa, Mean absolute error, and Relative absolute error. The statistic values of Genetic algorithm achieved the better results than TAN, K_2, and Hill climbing. It shows that there was consistency in Genetic Algorithm.

Table 8: The Experiment of Performance results of classifiers

Classifiers/ Algorithms	**P (User/Intruder) = 70:30**						
	Scenario 1						
	Accuracy	**FRR**	**FAR**	**Precision**	**Recall**	**F-Measure**	**ROC Area**
Genetic Algorithm	1	0.000	0.000	1	1	1	1
TAN	0.99	0.000	0.050	0.994	0.975	0.984	1
K_2	0.95	0.038	0.100	0.916	0.931	0.923	0.994
Hill Climbing	0.94	0.038	0.150	0.906	0.906	0.906	0.906
	Scenario 2						
	Accuracy	**FRR**	**FAR**	**Precision**	**Recall**	**F-Measure**	**ROC Area**
Genetic Algorithm	0.99	0.000	0.029	0.99	0.99	0.99	0.989
TAN	0.99	0.000	0.029	0.993	0.985	0.989	0.996
K_2	0.86	0.152	0.118	0.842	0.865	0.850	0.98
Hill Climbing	0.86	0.152	0.118	0.842	0.865	0.850	0.98
	Scenario 3						
	Accuracy	**FRR**	**FAR**	**Precision**	**Recall**	**F-Measure**	**ROC Area**
Genetic Algorithm	0.96	0.000	0.103	0.96	0.95	0.96	0.995
TAN	0.98	0.000	0.051	0.984	0.974	0.979	0.995
K_2	0.96	0.033	0.051	0.958	0.958	0.958	0.994
Hill Climbing	0.96	0.033	0.051	0.958	0.958	0.958	0.994
Classifiers/ Algorithms	**P (User/Intruder) = 90:10**						
	Scenario 1						
	Accuracy	**FRR**	**FAR**	**Precision**	**Recall**	**F-Measure**	**ROC Area**
Genetic Algorithm	1	0.000	0.000	1	1	1	1

TAN	0.98	0.000	0.100	0.988	0.950	0.968	1
K_2	0.93	0.038	0.200	0.896	0.881	0.889	0.993
Hill Climbing	0.93	0.038	0.200	0.896	0.881	0.889	0.993
	Scenario 2						
	Accuracy	**FRR**	**FAR**	**Precision**	**Recall**	**F-Measure**	**ROC Area**
Genetic Algorithm	0.98	0.000	0.080	0.987	0.960	0.973	0.980
TAN	0.98	0.000	0.080	0.987	0.960	0.973	0.980
K_2	0.930	0.053	0.120	0.903	0.913	0.908	0.989
Hill Climbing	0.930	0.053	0.120	0.903	0.913	0.908	0.989
	Scenario 3						
	Accuracy	**FRR**	**FAR**	**Precision**	**Recall**	**F-Measure**	**ROC Area**
Genetic Algorithm	0.93	0.057	0.100	0.914	0.921	0.917	0.994
TAN	0.94	0.043	0.100	0.929	0.929	0.929	0.991
K_2	0.90	0.100	0.100	0.874	0.900	0.885	0.986
Hill Climbing	0.90	0.100	0.100	0.874	0.900	0.885	0.986

Table 9: The Statistical Analysis of classifiers

Classifiers/ Algorithms	**P (User/Intruder) = 70:30**		
	Scenario 1		
	Kappa statistic	**Mean absolute error**	**Relative absolute error**
Genetic Algorithm	1	0.0234	7.2315%
TAN	0.9682	0.0414	12.7691%
K_2	0.8466	0.0764	23.5834%
Hill Climbing	0.8125	0.0734	22.6451%
	Scenario 2		
	Kappa statistic	**Mean absolute error**	**Relative absolute error**
Genetic Algorithm	0.9776	0.0464	10.3028%
TAN	0.9776	0.0529	11.7455%
K_2	0.7009	0.1132	25.1354%
Hill Climbing	0.7009	0.1132	25.1354%
	Scenario 3		
	Kappa statistic	**Mean absolute error**	**Relative absolute error**
Genetic Algorithm	0.9143	0.0535	11.2343%
TAN	0.9576	0.0518	10.866%
K_2	0.9159	0.0412	8.6369%
Hill Climbing	0.9159	0.0412	8.6369%
Classifiers/ Algorithms	**P (User/Intruder) = 90:10**		
	Scenario 1		
	Kappa statistic	**Mean absolute error**	**Relative absolute error**
Genetic Algorithm	1	0.0247	7.633%
TAN	0.9351	0.046	14.2129%
K_2	0.7771	0.0804	24.828%
Hill Climbing	0.7771	0.075	23.1583%
	Scenario 2		
	Kappa statistic	**Mean absolute error**	**Relative absolute error**
Genetic Algorithm	0.9452	0.0458	12.1065%
TAN	0.9452	0.0526	13.9077%
K_2	0.8158	0.0799	21.1245%
Hill Climbing	0.8158	0.0799	21.1245%
	Scenario 3		
	Kappa statistic	**Mean absolute error**	**Relative absolute error**
Genetic Algorithm	0.8349	0.0635	15.0681%
TAN	0.8571	0.079	18.7252%
K_2	0.7706	0.0869	20.6063%
Hill Climbing	0.7706	0.0869	20.6063%

5. CONCLUSIONS

In this paper, we propose the empirical evaluation of the effectiveness of structure learning of Bayesian Network in the model CUSIM [7] which is a method of user classification and authentication using the verification by keystroke dynamics; Location, and IP address. A comparison results in our work show that the results of genetic algorithm are better than the other algorithms. However, in Scenario 3, in which the conditional probabilities of P (KD | Real User = High), P (KD | Real User = Medium), and P (KD | Real User = Low) of the user and the intruder were opposite direction to user, the results of TAN algorithm are better than Genetic Algorithm. The Genetic Algorithm is thus selected for CUSIM. The empirical study also shows that the proposed model is accurate in the authentication and user classification process. This is a newly effective way to prevent the intrusion in Internet system.

6. REFERENCES

[1] M. Pearce, S. Zeadally, and R. Hunt, "Assessing and improving authentication confidence management", Information Management & Computer Security, Vol.8, No. 2, 2010, pp. 124-139.

[2] R.S. Gaines, W. Lisowski, S.J. Press, and N. Shapiro, "Authentication by Keystroke Timing: some Preliminary Results", Technical Report R-2526-NSF: Rand Corporation, 1980..

[3] D. Gunetti, and C. Picardi, "Keystroke analysis of free text", ACM Trans. Inf. Syst. Secure, 2005, Vol. 8, No. 3, pp.312–347.

[4] W. Roadrungwasinkul, and S. Sinthupinyo, "A Combination of Statistical Features and Neural Networks to Classify Users Based on Free Text", IRAST International Congress on Computer Applications and Computational Science (CACS 2010)., 2010.Information Management & Computer Security, Vol.8, No. 2, 2010, pp. 124-139.

[5] A. Aldridge, M. White, and K. Forcht, "Security considerations of doing business via the Internet: cautions to be considered", Internet Research: Electronic Networking Applications and Policy, Vol. 7, No. 1, 1997, pp. 9-15.

[6] M.C. Alexandra, L. O. Arlindo, and F.S. Marie, " Efficient learning of Bayesian network classifiers: An extension to the TAN classifier", Advances in Artificial Intelligence Lecture Note in Computer Science, Vol. 4830, 2007, pp. 16-25.

[7] C. Chantan, S. Sinthupinyo, and T. Rungkasiri, "Improving Accuracy of Authentication Process via Short Free Text using Bayesian Network", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No.3,2012, pp. 10-16.

[8] F. Monrose, and A. Rubin, "Authentication via keystroke dynamics", Proceedings of the 4th ACM conference on Computer and communications security, 1997, pp. 48–56.

[9] J. Hu, D. Gingrich, and A. Sentosa, "A k-Nearest Neighbor Approach for User Authentication through Biometric Keystroke Dynamics", Communications 2008 ICC '08 IEEE International Conference, 2008, pp. 1556-1560.

[10]J. Park, B. Ahnl, and H. Cho, "Intrusion Detection Using a PCB and IP address", Communications, Computer and Signal Processing, PacRim 2007 IEEE Pacific Rim Conference, 2007, pp. 223-226.

[11] J. Winterbottom, and C. Bryce, "The Internet Location Service", Intelligence in Next Generation Networks (ICIN), 14th International Conference, 2010, pp. 1-7.

[12] D. Jaros, and R. Kuchta, "New Location Based Authentication Techniques in the Access Management", Wireless and Mobile Communications (ICWMC) 6th International Conference, 2010, pp. 426 – 430.

[13] I. Ben-Gal, *"Bayesian Networks*", In: F., Ruggeri., F. Faltin, & R. Kenett (Eds.), "*Encyclopedia of Statistics in Quality and Reliability"*, John Wiley & Sons,2007.

[14] J. Pearl, "*Probabilistic Reasoning in Intelligent Systems*", Morgan Kaufmann, San Francisco.1988.

[15] A. Cufoglu, M. Lohi, and K. Madani, "A Comparative Study of Selected Classifiers with Classification Accuracy in User Profiling", Computer Science and Information Engineering 2009 WRI World Congress, 2009, Vol. 3, pp.708 – 712.

[16] L. Huijuan, C. Kejie, and L. Bai, "Bayesian Network Based Behavior Prediction Model for Intelligent Location Based Services", Industrial Electronics and Applications ICIEA 2007, 2nd IEEE Conference, 2007, pp. 703 – 708.

[17] A. Cufoglu, M. Lohi, and K. Madani, "Classification Accuracy performance of Naïve Bayesian(NB) Bayesian Networks(BN) Lazy
Learning of Bayesian Rules (LBR) and Instance-Based Learner (IB1) comparative study", Computer Engineering & Systems ICCES 2008, International Conference, 2008, pp.210 – 215.

[18] M. Bartlett, I. Bate, and J. Cussens, "Learning Bayesian Networks for Improved Instruction Cache Analysis", Machine Learning and Applications (ICMLA) Ninth International Conference, 2010, pp.417 – 423.

[19] G.F. Cooper, and E. Herskovits,(1992). "A Bayesian Method for the induction of probabilistic networks from data". *Machine Learning*, 9,1992, pp. 309-347.

[20] L. Boaz, and M. Roy, " Investigation of the K2 algorithm in learning Bayesian network classifiers", Applied Artificial Intelligence, Vol. 25, No. 1, 2011, pp. 74-96.

[21] C.O. Hong, "Improving classification in Bayesian networks using structural learning", World Academy of Science, Engineering and Technology,Vol. 75,2011, pp. 1407-1411.

[22] M. Stuart, H. Yulan, and L. Kecheng, "Choosing the best Bayesian classifier: An empirical study", IAENG International Journal of Computer Science, Vol. 36, No. 4, 2009, pp. 322-331.

[23] G.M. Michael, "The performance of Bayesian network classifiers constructed using differebt techniques", in working Notes of the 14th European Conference on Machine Learning (ECML), 2003, pp. 59-70.

[24] J.Holland, "*Adaptation in Natural and Artificial Systems*", University of Michigan Press. 1975.

[25] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1,2009.

[26] Website Ranking", at:
http://www.stats.in.th/?cmd=stats_global&list=y,
(accessed December 2011).