User Behavioral Analysis Using Markov Chain and Steady-State in Tracer and Checker Model

V. Arun* and R. Sudhakar

Department of CSE, Madanapalle Institute of Technology and Science, Madanapalle, Andhra Pradesh, India
E-mail: drarunv@mits.ac.in; drsudhakarr@mits.ac.in
**Corresponding Author*

Received 07 April 2018; Accepted 21 May 2018;
Publication 25 January 2019

Abstract

Tracer and checker model is an intrusion detection technique that uses mobile agent to track the user behaviour in ad-hoc network. Mobile agent can migrate to host and execute tasks parallelly. We enhanced TCM model to identify the intrusion in a host by analysing user behaviour during authentication process. Markov chain is a random process that transit from one state to another which depends only on the current state but not the sequence of events. Mobile agent is used to analyse the user input behaviour during authentication process which helps to predict intrusion in the system. In this paper, a behavioural approach is handled to identify the intrusion process. Markov-chain is used with the proposed behaviour approach and Mobile agents are used to distribute this functionality. Behavioural analysis is illustrated and simulation are experimented.

Keywords: Biometric authentication, mobile agent, intrusion detection, Markov chain process, TCM server, HIDS.

Journal of Cyber Security and Mobility, Vol. 8_2, 277–294.
doi: 10.13052/jcsm2245-1439.826

1 Introduction

Authentication plays vital role in any application or services and requires high security measures to protect the resources. It allows user to interact with the system with password and allow access to the resources. Authentication passwords can be a group of characters that are known only to the user. Choice of identifying the password is the key to enable security to the system. If the user choose a weak passwords, hackers can easily able to access to the resources without any trace in the system. Weak passwords are considered as the phone number, birth date or any other relevant information attached to the user. Hacker can able to identify the password with a guess or by using brute force attack. Continuous injection of the password with various characters may explore the password to the attacker. Intrusion of the system is most important while choosing the password by the user. Same password for multiple application or services may affect the system if any one of the service is compromised. Most of the services in the network rely on the authentication with characters rather than using any biometric measures [14]. Biometric authentication such as iris scanner or finger print scanner may require additional hardware support at the user end. And also any damage in the finger or eye problem may cause the system inaccessible by the legitimate user. Since ad-hoc wireless network does not rely on any pre-existing infrastructure and have access points in the wireless network, it is hard to maintain the hardware for biometric analysis on user end.

Various task in the network are performed in the host but in some cases we need the executable knowledge should be transferred from server to the host. Mobile agent is widely used in the ad-hoc wireless network to these knowledge in the host remotely. It can parallelly perform tasks in the hosts and update the server with essential details. Tracer and Checker model (TCM) is used as an intrusion identification model in the wireless network [1]. TCM delivers two mobile agent to the host to identify the intrusion in the network. Host-based intrusion detection system (HIDS) is an intrusion mechanism which monitors the host internal computing and logs the event to the system [2]. It identifies the intrusion by identifying the malicious activity and logs the event. Dynamic network packets are examined by HIDS and detect what program access which resource. TCM analysis these activities by triggering the mobile agent to the host and analysing the behaviour of the program. Even though the model analysis the behaviour of the program, it is necessary to analyse the user interaction during the authentication process.

Known-password attack checks for a certain password and passes the authentication check unknowingly in the system [4]. Brute force attack is widely used by hackers to analyse the password and access the resource without any trace in the HIDS. To avoid the password attack, we proposed a behavioural model during the authentication process. When user enters the password, the behaviour of the user is analysed and are used by the mobile agent by TCM to identify intrusion. TCM has internal alarming technique to alert other servers.

Dictionary attack is the common attack targeted to predict the user password. Commonly used words are used to identify the password to identify the password of the user. Dictionary attack along with Brute force attacks are joined together to iterate the possible password combination and the user password is hacked.

Mobile agent [6] analyse the behaviour of the user by identifying the timestamp of each user input. Timestamp range is determined during the analysis phase which is explained in Section 2. User input are analysed with the timestamp intervals and the behaviour is analysed with the time ranges. A threshold value for the time range is determined with the Markov chain process. Markov chain is a random process that transit from one state to another [7]. Markov chain has "memoryless" state in which each state does not depend on sequence of events. Steady-state is determined for each range of value with their probability. These threshold values are used to identify the intrusion in the system by checking the timestamp against the possibility of time stamp value.

Figure 1 shows the example or Markov-chain which independently has future, present and past state with a sequence of random variables such as X1, X2, X3 . . . Xn.

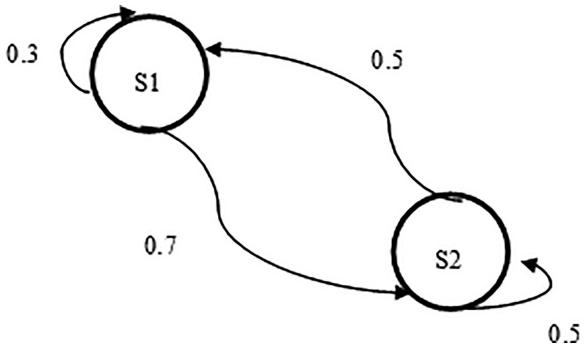

Figure 1 Markov-chain example.

Representative user technique is used to identify the behaviour of the user and is considered as the existing model in this paper. Alon Schclar [3] proposed the representative user technique which analyse the user behaviour with the user inputs. The system is trained with the series of input from the user and are used to analyses the behaviour of the user. During authentication process, the user behaviour is compared with the behaviour of the stored user and accepts the user if the behaviour matches the user input. But existing system uses the basic timestamp feature to identify the user behaviour and will fail when the user inputs while training session is limited. It is hard to gather the user training inputs when the user count is high. The major drawback in the existing model is the lack of training set which does not prove the analysis of large set of users. Proposed model does not require the training set and does not depend on the training inputs. Efficiency of the existing model rely on the number of training set of the user. Proposed model does not require more training sets to prepare the behavioural nature of the user. The behavioural pattern is determined when the user signup or set the password for the first time. Usually user are asked to set the password with two time confirmation and it is used to build the behavioural pattern. This means that only two training input is enough to identify the user behaviour. The behaviour is updated frequently when the user authenticate to the system, this feature will update the behaviour of the user.

Michael Fagan [13] proposed a new approach to handle user behaviour using password managers. It is the most recommended by many security experts, but are still not used by many users. An online survey is distributed to a total of 137 users and 111 non-users of the tool that asked about their experiences with password managers. Analysis of the differences in emotions between "users" and "non-users" reveals that the participants who never use a password manager are more likely to feel suspicious compared to "users," which could be due to misunderstandings about the tool. But "users" of password manager noted convenience and usefulness as the main factors for using a password manager, "non-users" noted security issues as the chief reason for not using a password manager.

We will be analysing the behaviour of the user input during authentication phase under Section 2. In Section 3, probability mass function is derived for each user input value. Using Section 3, a steady state function is derived from the user input in Section 4. We used Markov-chain process to analyse the user behaviour in Section 5 and we showcased the results in Section 6.

2 Authentication Analysis Phase

Authentication process should be performed to analyse the user behaviour. It is an enrol process to learn the user behaviour to the mobile agent. Analysis of the user password is performed when the user signup or create a new password to the account. The input is given to the mobile agent and range of possible input timestamp is analysed. User may have different pattern to make key stroke [9] during authentication process. Timestamp required for each input values and holding the key are unique for different users. The range of timestamp are used to analyse the user behaviour and may vary for different mode of input devices. Each type of device have different range values and are analysed by the mobile agent.

The process is illustrated with an example password as shown in Figure 2. Consider a user has a password 'Hello123#!' and assuming that no modifier key are used to enter special characters.

N denotes to number of characters in the password and (N–1) is the time range for the password. So, for given example the number of character is 10. So, it contains (N–1) timestamp.

(10–1) = 9 timestamps

A series of input is given to the mobile agent as shown in Table 1. It shows the user inputs with timestamp required to learn for mobile agent identifying ranges.

H	e	l	l	o	1	2	3	#	!

Figure 2 Characters of sample password.

Table 1 Timestamps for the given password

Input	Phase 1 (ms)	Phase 2 (ms)
H E	125	132
e 1	225	224
l 1	154	145
l o	321	355

Input	Phase 1 (ms)	Phase 2 (ms)
O 1	754	757
1 2	211	211
2 3	211	211
3 #	877	855
# !	966	988

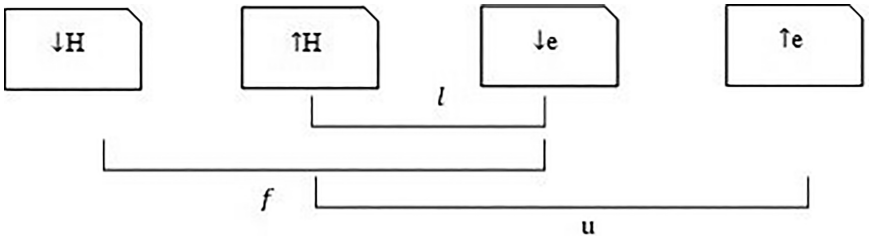

Figure 3 Latency time, flight time and up time.

With the range calculation, the other key factors are studied from the user input. Latency time (l) denotes the time taken for the user to release one key and press the other key. Flight time (f) denotes the time between two keys are pressed and the up time (u) represent the first key and second key is released. The latency time, flight time and up time are calculated during the two phase of input from the user. Figure 3 shows l, f and u for the given example keys "H-e" in which ↓ denotes key pressed down and ↑ denotes the key pressed up.

3 Probability Mass Function

Probability mass function specifies that X can be specific value of x if $p(x)$

$$P(X = x) = p(x) = p_x \textit{ is non-negative for all } x. \quad (1)$$

$$P(a \leq X \leq a_n) = P(x = a) + P(X = a_1) + P(X = a_2) + \cdots + P(X = a_n) \quad (2)$$

For input value 'He', the probability mass function is tabulated below and same way other input value should be determined. The overall time required for a key press is determined by two ways such as $(f - l)$ or $(u - l)$. User will have different values for two different approaches, so we apply the probability mass function with less probability for the determined values as 0.1 [10]. The other probability value will be distributed to the input time stamps as shown below in Table 2. Splitting the ranges with n/2 and arranging the ranges as shown below in Table 3.

It shows that for the input 'He' 80% possibility are there that it will fall in 124–132 time range 10% for range (133–150). Table 4 represent the probability mass function for the input password during the signup or password assignment phase.

Table 2 Probability mass function for input timestamps

Value	Probability Mass Function
125 (R_{min})	0.4
132 (R_{max})	0.4
148 (f–l)	0.1
150 (u–l)	0.1

Table 3 Probability mass function for time stamps range

Value	Probability Mass Function
124–132	0.8
133–150	0.2

Table 4 Timestamp range for input

Input	Range Value	Percentage
H	124–132	80%
e	133–150	20%
e	220–330	90%
l	331–340	10%
l	420–430	90%
l	431–440	10%
l	111–130	80%
o	131–140	20%
o	165–179	90%
1	180–211	10%
1	322–345	90%
2	346–366	10%
2	119–130	80%
3	131–140	20%
3	366–455	90%
#	456–510	10%
#	211–266	80%
!	267–287	20%

Since it has two training input values, the probability is assigned roughly separated. These values are used to identify the user behaviour. Table 5 shows the confined table which is updated after further login by the user. This proves that the change in user behaviour will also change the user behaviour analysis. Existing model requires more training input from the user and cannot be used when the user count increases. It is difficult to gather many input from the user and store the user behaviour.

Table 5 Updated Timestamp range

Input	Range Value	Percentage
H	121–135	66%
e	136–150	34%
e	211–310	78%
l	311–399	22%
l	410–422	55%
l	423–426	45%
l	110–130	78%
o	131–140	22%
o	120–130	64%
1	131–213	36%
1	320–354	45%
2	355–146	55%
2	100–110	75%
3	111–140	25%
3	360–411	47%
#	412–512	53%
#	120–221	65%
!	222–321	35%

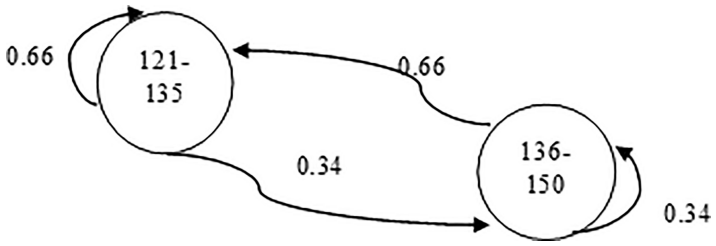

Figure 4 State diagram.

4 Steady-State Threshold Value

Steady-state determines the equilibrium condition of the process in which the effect of momentary variations. The ranges determined with probability mass function is used to identify the steady-state function. The steady state values are used as the threshold for the input. When the user enters password beyond the range of the steady-state determines that the user is not genuine. TCM model possess mobile agent to trace the user input and compare with the steady-state value [11]. Steady-state value is determined during the initial phase of analysis as shown in Figure 4 and it shows that there is 0.66 probability weight is assigned for 121–135 state. Whereas balance 0.34 probability is assigned

for the next state to transfer. But the steady state 136–150 has 0.66 probability for next state whereas 0.34 probability for the self-probability weight.

$$P = \begin{bmatrix} 0.66 & 0.34 \\ 0.34 & 0.66 \end{bmatrix} \tag{3}$$

$$qP = q \tag{4}$$

$$q(P - 1) = 0 \tag{5}$$

$$[q1 \;\; q2] = \begin{bmatrix} 0.66 & 0.34 \\ 0.34 & 0.66 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{6}$$

$$= \begin{bmatrix} -0.34 & 0.34 \\ 0.34 & -0.34 \end{bmatrix}$$

$-0.45q1 + 0.45q2 = 0$ *and we know that* $q1 + q2 = 1$.

Solving the simultaneous equation gives the steady state distribution as Steady-state vector $= [0.5 \;\; 0.5]$.

5 Behavioral Analysis Using Markov-Chain

The following steps illustrates the overall implementation of the proposed Markov-chain analysis of the user behaviour. During analysis phase, proposed model identifies the user behaviour and update the details to the TCM server using mobile agent. The following steps are performed during the analysis phase of the user behaviour during signup of the user.

***Pseudocode*:** *UserBehavior()*
Begin
Identify the timestamp for latency time, flight time and up time
Identify the minimum range, R_min
Identify the maximum range, R_max
Arrange the timestamp in ascending order
Identify the probability mass function
Identify the timestamp range and probability mass function for the time stamp range
Identify the threshold value using Steady-state function
Return user behaviour.
End

Read the user input during signup process

call UserBehavior()
Update the user behaviour in TCM server.

When the user involves in the authentication process, the user behaviour is used to compare with the user input. Mobile agent in TCM model calculate the user behaviour and compares with the user input. The following steps are performed during the authentication process.

***Pseudocode*:** *AuthenticateUser()*
Begin
Read the user input during authentication process
UserBehavior= call UserBehavior()
If(UserBehavior.IsLegitimateUser)
Begin
Update the user behaviour in the TCM server.
Allow user to access the resource.
Update HIDS in the host.
End
Else
Begin
Deny access to the user.
Initiate alarm in the TCM model.
End
End

6 Result and Discussion

User behaviours are stored in the TCM server which is a cloud based approach. The mobile agent migrates from server to host to read the user behaviour during authentication process. When user enters the authenticating password, the mobile agent reads the user input and calculates the user behaviour. Mobile agent migrates to the server and analyse the user behaviour with the stored user database. If there is a match in the user input behaviour, the mobile agent perform an update in the user behaviour and return to the host to allow the user. Table 6 shows sample inputs during authentication process in which two inputs are authenticated with different user. Our proposed model compares the result and authenticate the user successfully. Due to the lack of less training data, the existing model identifies only one input whereas accuracy is deviated on the other one.

Table 6 Analysed Input ranges with three inputs

Input Range	Input 1	Range Percentage	Input 2	Percentage	Input 3	Range Percentage
H e	135	67%	245	45%	135	76%
e 1	145	65%	235	34%	148	74%
l 1	355	55%	125	55%	896	32%
l o	655	78%	452	45%	984	53%
o 1	545	65%	441	64%	135	35%
1 2	125	54%	524	33%	541	64%
2 3	155	57%	321	22%	654	56%
3 #	954	87%	451	52%	135	55%
# !	985	12%	235	76%	562	76%

Figure 5 shows that the input 1 failed to match with the threshold value while entering key pair (#,-!) and input 2 failed in many analysis. Existing model suggest that input 1 and input 3 are legitimate user while proposed model analyse that input 3 is from real user [12]. Proposed model determine the threshold value from the steady state vector which helps the model to identify the real input. We have carried out cluster of inputs from different users, Next section shows the comparative result of the existing model and proposed model.

6.1 Comparative Study

The underlying concept of the proposed algorithm uses the behavioural approach with Markov chain process to identify the steady state. We have used cluster of users to test the proposed model and analysed the result with existing model.

6.1.1 Experiment to identify the attacker

We have trained the model with two set of user input and experimented with the group of user to the existing model and proposed model. 98.34% of users

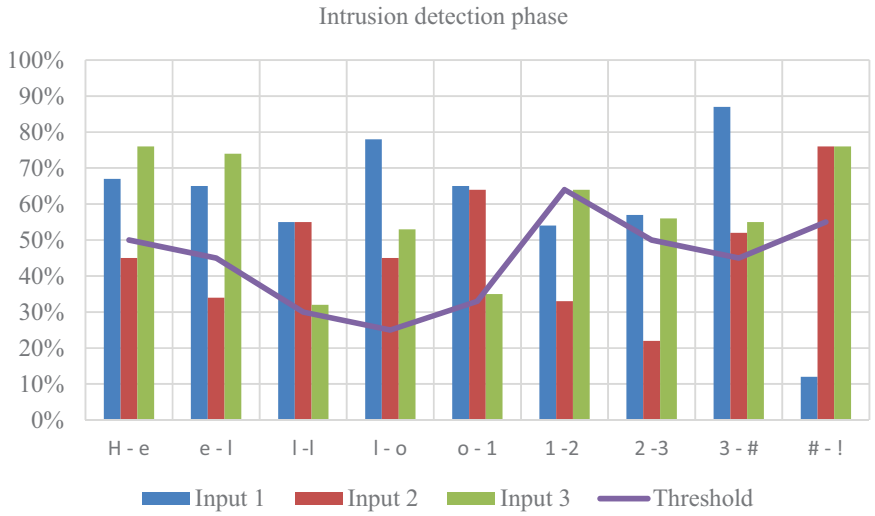

Figure 5 Input ranges with threshold value.

Table 7 ANOVA result for non-legitimate (Representative model Vs Markov chain model)

ANOVA Result						
Source of Variation	Sum of Squ	df	Mean Square	F	P-value	F Crit
Between Groups	2131.6	1	2131.6	78.94815	2.04E-05	5.317655
Within Groups	216	8	27			
Total	2347.6	9				

are identified as false users in the proposed model where as 63.12% of users are identified as false user in the existing model. The effect of less training set over the existing model result in breaking the algorithm.

Analysis of variance (ANOVA) [8] is determined on the group of users to check the hypothesis in Table 7. Below table shows the hypothesis result. P-value > 1 which determines that there is null hypothesis is not satisfied. This proves that the proposed algorithm have efficiency over the existing model.

Figure 6 shows the experimental result of cluster of users in which all users are not the legitimate users. There are 324 users are grouped into 10 different cluster and are tested again the existing representative user model and proposed Markov chain model. The result proves that the proposed model identifies the intruder effectively compared to the existing model.

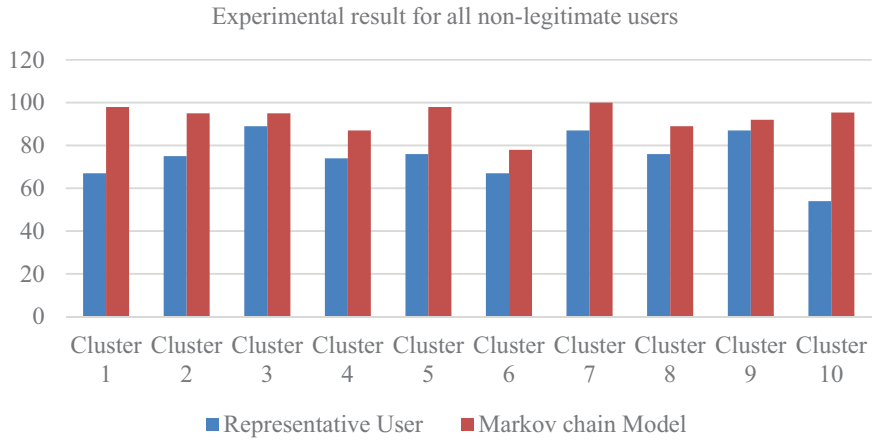

Figure 6 Percentage of success rate Representative user model vs Markov Chain Model.

6.1.2 Experiment to identify legitimate user

We carried out an experiment with all legitimate users to accept the user authentication. The determination result in 87.11% of users are identified as real users in the proposed model where as 23.12% of users are identified as user in the existing model. With limited input training values, proposed model effectively identified the legitimate user.

Analysis of variance (ANOVA) is determined for the group and Table 8. shows the hypothesis result. P-value $>$ 1 which shows that the proposed algorithm have efficiency over the existing model.

Figure 7 shows the successful rate of the cluster of user among representative model with Markov-chain model. The successful rate is higher in proposed model due to the effective analysis of the behaviour of the user with minimal training values.

Table 8 ANOVA result for legitimate user (Representative model vs Markov chain model)

ANOVA						
Source of Variation	Sum of Squ	df	Mean Square	F	P-value	F Crit
Between Groups	6132.0714	1	6132.07142	63.812438	3.8168E-06	4.74722534
Within Groups	1153.1428	12	96.0952381			
Total	7285.2142	13				

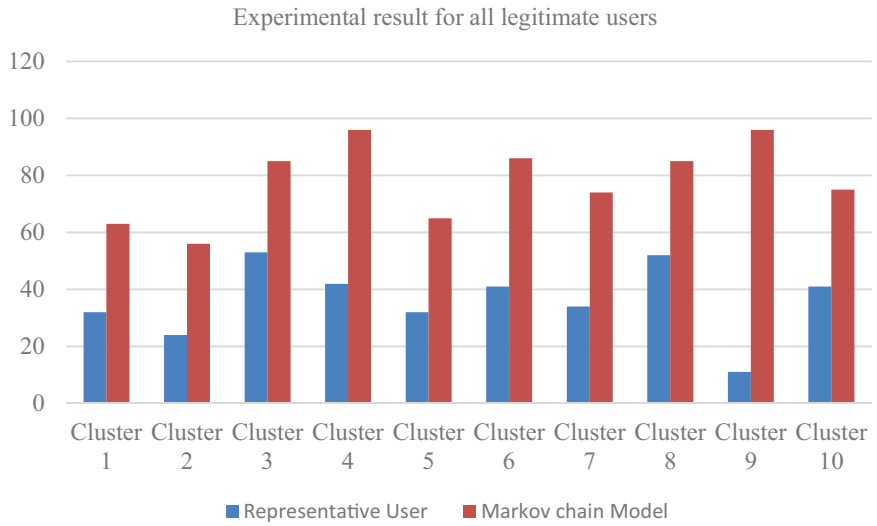

Figure 7 Percentage of success rate for Representative user model vs Markov chain model.

7 Conclusion

No separate training subsets are needed. Limited training inputs are assigned when the user signup with the new password. Algorithm is used in the mobile agent of TCM model which detects intrusion in the network. User behaviour is analysed in the paper with behavioural approach to enhance the user security. This approach is proposed specially for ad-hoc network where especial hardware for biometric mechanism cannot b1 adopted. Since proposed model does not require many training input subsets, it is easy to implement in the ad-hoc network. Steady state is a proved solution which is used in the proposed model to determine the threshold value. This helps the proposed algorithm to effectively analyse the user input with limited training data. The proposed model updates the behaviour of the user during authentication process. TCM integration alerts other servers to avoid the user when intrusion is detected.

References

[1] V. Arun, K. L. Shunmuganathan, 'Encrypted Tracer and Checker Model', Journal of Emerging Technologies – Image Processing and Networking, Vol. 6, No. 2, pp. 23–27, 2011.

[2] O. Al-Jarrah, 'Network Intrusion Detection System using attack behaviour classification' International Conference on Information and Communication Systems (ICICS), pp. 1–6, 2014.

[3] Alon Schclar, LiorRokach, Adi Abramson, and Yuval Elovici, 'User Authentication Based on Representative Users', IEEE transactions on systems, man, and cybernetics—part c: applications and reviews, Vol. 42, No. 6, 2012.

[4] R. Kirushnaamoni, Mepco Schlenk, 'Defenses to curb online password guessing attacks', International Conferene on Information Communication and Embedded Systems, 2013.

[5] Na Zeng, Wuhan China, Xiaolong Zhang, Hong Zhang, 'Intramural Network Intrusion Detection by Monitoring User Behavior', International Symposium on Knowledge Acquisition and Modeling, pp. 178–181, 2009.

[6] M. B. Nirmala, A. S. Manjunath, 'Mobile agent based secure code update in wireless sensor networks', International Conference on Information Networking pp. 75–80, 2015.

[7] G. Ioannou, P. Louvieris, N. Clewley, G. Powell, 'A Markov multi-phase transferable belief model: An application for predicting data exfiltration APTs', International Conference on Information Fusion (FUSION), pp. 842–849, 2013.

[8] Zheng Zhang, Xiu Yang, Oseledets, G. E. Karniadakis, 'Enabling High-Dimensional Hierarchical Uncertainty Quantification by ANOVA and Tensor-Train Decomposition', Computer-Aided Design of Integrated Circuits and Systems, Vol. 34, pp. 63–76, 2015.

[9] S. Bleha, C. Slivinsky, and B. Hussein, 'Computer-access security systems using keystroke dynamics', IEEE Trans. Pattern Analysisand Machine Intelligence., Vol. 12, No. 12, pp. 1217–1222, 1990.

[10] A. El-Saddik, M. Orozco, Y. Asfaw, S. Shirmohammadi, and A. Adler, 'A novel biometric system for identification and verification of haptic users', IEEE Transaction on Instrumentation and Measurement, Vol. 56, pp. 895–906, 2007.

[11] E. Frank and I. H. Witten, 'WEKA: A machine learning workbench for data mining', in Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers, O. Maimon and L. Rokach, Eds. New York: Springer, pp. 1305–1314, 2005.

[12] N. J. Grabham and N. M. White, 'Validation of keypad user identity using a novel biometric technique', Journal of Physics, Vol. 76, pp. 012023-1–012023-6, 2007.

[13] Yusuf Albayram, MhammadMaifi Hasan Khan and Ross Buck, 'An investigation into users' considerations towards using password managers', Human-centric Computing and Information Sciences, 2017.
[14] T. Subburaj and K. Suthendran, 'Detection and Trace Back of DDoS Attack Based on Statistical Approach', Journal of Advanced Research in Dynamical and Control Systems, Vol. 13, pp. 66–74, 2017.

Biographies

V. Arun recieved his Ph.D. degree from the University of Sathyabama at Chennai in 2011. He attended the University of Bharathiar, Coimbatore where he received his M.S in Computer Science in 2004. He received his B.E in Electronics and Instrumentation from Annamalai University, Chidambaram in 2002, His current areas of interest include intrusion detection, network security, and image processing. He has gained experience has Software Engineer in Chennai. In specialized in DotNet Course in the software working experience field. Then he entered into the teaching field and elaborated his software experience to the college students in the academic field. He has participated many seminars and conference.

R. Sudhakar recieved his B.E degree in Computer Science & Engineering from Anna University, Chennai; M.Tech. degree in Computer Science from Dr. M.G.R University, Chennai and Ph.D. degree in Computer Science Engineering from the Anna University, Chennai. His current areas of interest include wireless communication, network security, and image processing. He has taught various subjects in the Computer Science and Engineering Department over a period of 11 years. He now serves as Senior Assistant Professor of the Department of Computer Science and Engineering at Madanapalle Institute of Technology & Science, Andhra Pradesh, India.