Original Article

Comprehensive Event-Time-Action Threat Intelligence To Prevent Advanced Persistent Threats

Abdul Khadar A[1*], Dr. Shrishail Math[2], Dr. Brahmananda SH[3], Dr. Shivamurthy G[4]

[1*]SJC Institute of Technology, Chickballapur
[2]SKIT, Bangalore
[3]GITAM University, Bangalore
[4]VIAT, Muddenahalli, Chickballapur

***Corresponding Author:** Abdul Khadar A

*SJC Institute of Technology, Chickballapur
DOI: 10.47750/pnr.2023.14.S02.277

Abstract

The overall efforts and expenses put across by the organizations and institutions to protect their sensitive information from the larceners has drastically elevated in the recent past with the enhancement and sophistication of advanced persistent threats (APT) attacks, yet obfuscation that these APTs create while in business are proving these organization's and institution's security systems feeble. To prevent the APT's proliferation, the computer security systems shall get out-of-the-box solutions that can challenge the invaders. The proposed comprehensive event-time-action (CETA) threat intelligence has come up with such a model. The basic cons of the existing security systems is to go with the flow which they have found in forensic investigation of the attacks that APTs have made. The CETA model uses the pattern and procedures that the APTs follow, as proposed by Lockheed Martin's intrusion kill chain (IKC), with a new shade every-time. CETA with the consolidation of shuffle-selective-search module, octa-secured entry module, ensnare the invader module and machine learning tools has successfully detected and with better efficiency has prevented the APT's attacks trails. The rigorous experiments with penetration tests and vulnerability exploitations the CETA model has found that sensitive information system could not be breached with success rate of 97.43% of prevention and 98.67% accuracy of detection of the attacks by the APT's spear phishing and social engineering attack vectors.

Keywords: IKC, shuffle-selective-search, octa-secured entry, spear phishing, social engineering

INTRODUCTION:

The entire world from house hold to health, from manufacturing to military and from education to entertainment every sector is trying to connect itself with the rest of the world to communicate and to grow. This communication and connection have become easy and accessible by the huge and ever developing computer network. And it is not hidden from any of the sector that their communication and connection is under threat, always. Its obvious that when something is so popular and used enormously there is always a negative factor that tries to hinder its growth.

When an organization or an individual is sharing or transferring some data or secure or personal data from its or his perspective the organization or the individual will try to make sure that his data never being tampered or lost or manipulated while being transferred over the computer network or stored in the secured system. In this view a lot of safeguard or security measured are taken on different levels.

As the data or information being transferred or being stored at the desired stage, the threats of this data or information becoming vulnerable would be more prominent because in this data era the data or information is a tool, is a weapon, is a treasure and is an investment and sometimes asset to the both the owner al well as the thief of this data. As data or information travels over the computer network it comes across many security threats.

Computer network security threats have found a more sophisticated enemy the Advanced Persistent Threats. All the threats discussed are launched or planted on the target and the purpose is achieved within a fraction of time and there are many identification prevention methods for their attacks that can to some extent avoid the calamity to happen. But the Advanced Persistent Threats-APTs quite different from regular attacks. These are the threats planed and implanted by an organization or they are generally state sponsored attacks. These attacks are organized not executed instantaneously, and are organized with an intension of not to harm the hardware or defame the target organization or not to gain some financial advantage. These state sponsored attacks are made to steal or gain information, sensitive information, like Intellectual Proprietary

Rights, Nuclear Reactor plant data, access credentials of a server that stores the complete business details and business strategies, etc., The life span of the execution of these APTs is minimum of 6 months to 2-3 years.
Advanced Persistent Threats are the threats to the network security software. These threats though follow similar work procedure the IKC but cause different effects. There are classified based on the attack vectors.

- Phishing
- Social Engineering
- DNS modifications
- Zero-day attacks
- Exploiting the Vulnerabilities
- Internal Attacks
- Pirated Software
- Ransom ware

The APTs are broadly categorized based on four major parameters like APT group name, sectors that these groups target, associated malware they used to attack and the attack vectors employed. As the research goal is to detection and prevention of the APTs the most suited criteria for categorize is Attack Vectors.

After first planning for who, how, or why the attack is going to happen and then building or acquiring the attack tool which can be named as Study phase. The next step is delivering it to the infiltrated device which can be named as Spear phase. This can be done remotely, locally, or even during the manufacturing of the device itself. Finally, there is the deployment, wherein the malicious [7] payload is looking for the location in which to become persistent, i.e., to be saved or deployed in the Flash / Non-Volatile Memory in order to survive a restart or power loss which can be named as Search phase. These attacks can be carried out through different vectors: by outsiders or hackers, by insiders, which are often deceived technicians or disgruntled employees, or through the supply chain (i.e., deceived contractor), where malware is injected directly into a device during manufacturing or delivery. The final phase is Sabotage when the attacker gains the required plunder from the target.

Related work:

Y Zhang, Research on strong-association rule based web application vulnerability detection [1], this work is on web vulnerabilities, the approach of resolving the web vulnerabilities is found to be the most effective way of preventing the web applications from threats. This is achieved by initially traversing through the web site to collect all the pages visited and a regression test is performed on association between pages traversed later make a collection of pages. The intrusion detection method like Intrusion Kill Chain are utilized to trace the way the web sites are exposed to vulnerabilities [6] and the intruder's behavior is also traced to identify the stage of the attack underway. It proposes the strict methods of controlling and monitoring the web sites by the organizations users as part of their regular day to today activities. [2] A graph analytic metric for mitigating advanced persistent threat. This work proposes a metric which is calculated from the considered cyber network's processed sub graph's selected edges belonging to those vertices for which the given property is defined. And the every node's metric is processed as probability that a given node can be reachable from this node within the graph. This metric is possible to process from the layers like auditing and authorization dynamically while authorization of the page is done. And can effectively avoidance of threats like pass the hash. This metric can be used to measure the vulnerability of the cyber network to a specific persistent threat and back door exit from the targeted secure data system. The work analyses the results of its processing the graph and each node as performance of star reachability graph and as well the performance of the single path reachability graph and plots the maximum, the average and minimum reachability of the nodes from a specific node of the graph by considering the order of the graph. A Theoretical Framework to Counter Advanced Persistent Threats [3], this work states that deceit of the attempt of an advanced persistent threat needs in depth knowledge of the attack vector as well the process of the attack, though this has proved as the most effective method of defending the attack of the APTs. As it is well known fact that attack of advanced persistent threats, which are usually state or government sponsored, are very severe and generally sophisticated and unique and can surpass all traditional and signature based identification kind of intrusion detection systems. Which leads a way that generally traces the browsing and usage behavior of the user's secure system. This work proposes a hidden Markov model that uses the indicators of compromises are utilized as key factors to detect [8] and identify the traces of advanced persistent threats attacks on the target. Any attack that resemble the attack vector of the APTs [9] could be also be traced with this framework. This framework uses ransom ware as the main attack vector to identify and detect the APT attacks on the targeted system. Anomaly detection of malicious users' behaviors for web applications based on web logs [4], mentioned work states that as the digital communication is being updated and improvised the threat of security issues are too scaling at the same pace. The available intrusion detection systems instead of recording user's behavior are analyzing each and every request coming to the server and they could detect merely the known threats which have been recognized clearly but there is emergency need to detect the zero day attacks that before the software vendor come to know the existence of which, the invader identifies and uses them to attack the target. In this regard the work meticulously analyzes the web resources like web servers and web logs and the set of user behavior into regular and vulnerable links. The results of the work show that higher accuracy and lesser false alarm rate can be achieved using web user behaviors.

Proposed methodology:

Through observations till time it is quite clear that the most frequent and the most destructive threat category are Phishing and Social Engineering. The proposed research work has found detection and prevention methods for each of these types of threats. The Phishing attack and Social engineering attacks are detected with the ETA method. Phishing attack can be prevented using the Octa-Secured-Entry and EI method. The Social Engineering attack is prevented using the Shuffle-Selective-Search.

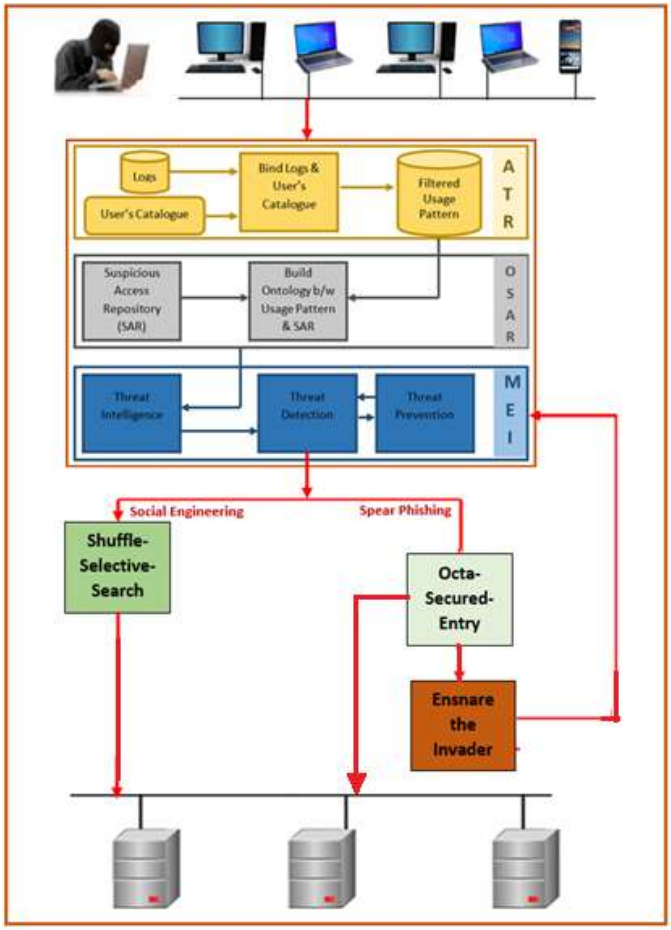

Figure 1: CETA Architecture

Through observations till time it is quite clear that the most frequent and the most destructive threat category are Phishing and Social Engineering. The proposed research work has found detection and prevention methods for each of these types of threats. The Phishing attack and Social engineering attacks are detected with the ETA method. Phishing attack can be prevented using the Octa-Secured-Entry and EI method. The Social Engineering attack is prevented using the Shuffle-Selective-Search.

The proposed methodology uses the Correlation between Event of any system data related activity, the Time at which the system's sensitive data was tried to access and the Action that determines what was the action that ATP tried to do on the sensitive data of the system.

The Access Track Record (ATR) module of the proposed methodology maintains and tracks this system usage information at the layer one (L1) of the threat intelligence system which falls between the resourceful system and the regular users of the system. The complete architecture of the proposed methodology is as shown on the figure 1. The user's catalogue is bound with system logs to create a unified access portfolio about the resourceful system. The intelligence system maintains a suspicious system access repository, which is s dynamic repository of the blacklisted Ips, web pages and links that are engaged in the APT or any ingress activities and are collected through various web search engines on regular basis (depending upon the type of organization). This repository along with the filtered system usage pattern is used as two major components to build the ontology between time-event-action of each and every access to the resourceful system. This is executed at the level two (L2) at every external access to the system.

The EI (Ensnare the Invader) perceives the ontology generated by the OSAR module. With the machine learning text classification algorithm SVM (Support Vector Machine) it classifies the OSAR dataset as either suspicious or not based on the attributes of ontology O and attributes *t, aip* and *atype* from LU set. From the dataset available it predicts the current activity within the resourceful system to be TP then an alarm is set to the resourceful system admin and the threat intelligent system updates its dataset with this new data as suspicious. This way the threat intelligence system dynamically

goes on training the system with new fishy data so as to keep track of all the misbehaving activities from either the regular users of the resourceful system and as well the external users. The system admin has to take care of the safeguarding the system and further nullifying attempt of APT to ingress the system. APTs are so sophisticated that if found to be created doubt on the target system or after successful attack they will efface all the logs within the system and even change the mode of hiding as well become passive for over a period of time to just not get caught and be persistent in the system. To ensnare the ATPs agent the MEI module creates a virtual environment for the accessing ATP so as to make it feel that it is in the original environment and be in the same mode of ingress. The intelligence system continues to monitor to prevent the resourceful system from this user for its every access and action in the system. If in the future time period the same user or the client system found to be behaving fishy then it will be blocked and the system admin will be instructed to take a survey of the entire system and terminate any future access of the user or the client.

The ordered set O (LU, R) that is generated from the OSAR module is fed as input to the machine learning module. This set would have all the fields that are related to the log files, user catalog and data from the suspicious website repository. As all these attributes reach the ML module the data is analyzed by the Support Vector Machine module that analyzes this data for any true positive prediction.

Machine learning involves predicting and classifying data and to do so we employ various machine learning algorithms according to the dataset. SVM or Support Vector Machine is a linear model for classification and regression problems. It can solve linear and non-linear problems and work well for many practical problems. The SVM creates a line or a hyper plane which separates the data into classes.

The Shuffle-Selective-Search [8] dissection is a process of identifying the intrusion of the invaders into the sensitive data environment through IKC method. As the APTs invade over the sensitive data repeatedly and continuously for a prolonged period of time to gain the access and control over the intrusion-point environment the SSS dissection suits well to mitigate the APTs that follow IKC to succeed the attack. It works on the computer networks of the organization or sensitive data environment where each computer of the network could be identified by a unique number like IP address. These IP addresses need to be taken as an array of elements in an order.

This dissections receives the behavior of the network from the MEI (Model and Ensnare the Invader) model of the proposed methodology and the if it receives true positive alerts then the system admin is informed of the presence of APTs activity in the network else it allows the access to the sensitive data

The proposed SSS is tested at the laboratory with eight systems and a sensitive data holding system whole as a network of computers. All the computers were setup with internet connectivity and had specific IP addresses. From all the eight systems access was provided to the system having the sensitive data. The DVM that is Damn Vulnerable Machine was setup with the DVM website to allow free access to the sensitive data holding machine. And all possible attacks were made on the central system with spear phishing and social engineering attacks as main ones.

Table1: Result and Analysis

Network Size		No of Shuffles	Key Position		Time Taken =n*log(n)
			Bucket 'A'	Bucket 'B	
n=4	$2^n=2^4=16$	n-1=3	$1^2+1=2$	$3^2+1=10$	2.408 sec
n=5	$2^n=2^5=32$	n-1=4	$3^2+1=2$	$5^2+1=26$	3.494 sec
n=6	$2^n=2^6=64$	n-1=5	$2^6/2-22=10$	$2^6-22=42$	4.668 sec
n=7	$2^n=2^7=128$	n-1=6	$2^7/2-22=42$	$2^7-22=106$	5.915 sec

The above analysis of the result could be plotted as graph as shown in following figures. The graph in Figure4.1 depicts the 4 different scenarios of different size of network which is under threat and if after the n^{th} shuffle the bucket 'A' is found to have the intrusion-point then the point of intrusion is given by the 'Key Position' in each case.

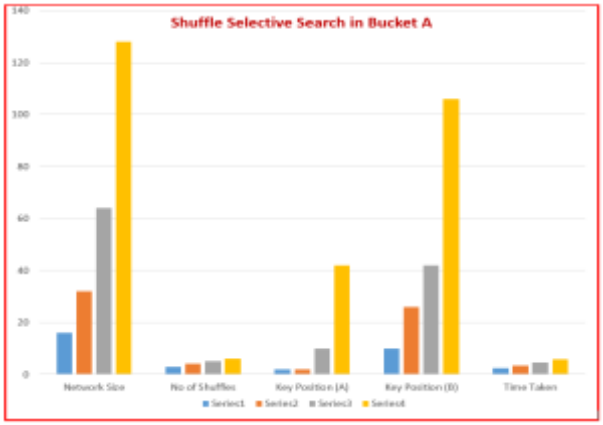

Figure 2. Graph of intrusion-point in Bucket 'A' after last shuffle and computation time.

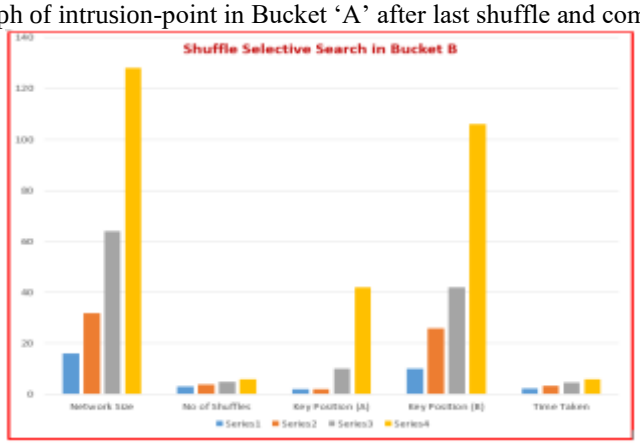

Figure 3. Graph of intrusion-point in Bucket 'B' after last shuffle and computation time.

Octa secured entry uses a two dimensional matrix that has the scope of scalability too, initially each cell of the 4x4 matrix will have at least one as its contents but to max the row or column pertaining to the specific cell shall have total eight as contents. And the cells (2,1) to (2,3), (3,2) to (3,3) these would not contain any value so will not be considered for processing and treated as NULL values. With initial move one can alter the contents of the cell with one shift to any corresponding cell of the same row or column by retaining the total row or column value to not exceed eight. Proceeding in such way one can make any number of moves but one should know does not let any value more than eight in it.

The row wise and column wise sum of the data of the matrix shall not be exceeding eight and if this happens the OSE algorithm gives three chances to manage the total to eight by all means and then at failure of the third try the OSE throws true positive ingress into the system by third party.

The OSE works better than the captcha of any category and even better that than the OPT lead entry into the system. The security is so hard that unless the row and column wise total is not matches then the entry will be not allowed.

Octa Secured Entry Methodology

Algorithm:

1. Initialize row and column to zero
2. add (i, i+3)
   Initialize SRi, SRi+4, SCi, SCi+4 to zero
   For columns i & i+3
   Increment its sum value by one
   For rows i & i+3
   Increment its sum by one
   Do not alter the other cells
3. Get the random position for the cell (random position(i,i+3)
   Repeat the following steps till selected cell is within range
   Get random position for the row between i and i+3
   Get random position for the column between i and i+3
   Check if the chosen cell lie within the specified range
   If range meets continues
   Else
   Break to main code by returning the row and column position

4. Update the matrix by randomly selected cell
5. Add to the selected cell
6. Update each cell value for the new position to have eight, value to a total
7. Add to the next cell a value one.
8. Repeat the above steps through step 3 until the proper move is made.

The above is the algorithm is the octa secured methodology.

Table 2: Octa Secured Entry Performance Analysis

Sl. No	Source of Access	Attempts Made	Success Rate	Failure Rate	Reason for Failure
1	DVW	25	97%	3%	The result from MEI module was a true negative one
2	Malicious Mails	36	98%	2%	Mails had zero day attack code
3	Embedded URLs	18	95%	5%	The used URLs had links to dark web

The above table2 depicts the result of the experiments carried out on the octa secure entry at the laboratory. The research used three various kinds of spear phishing attacks on the sensitive information in the network. The module has successfully blocked the ingress by all the means of invades. There were few instances where the module could not tackle the ingress the reasons for the same are mentioned in the table.

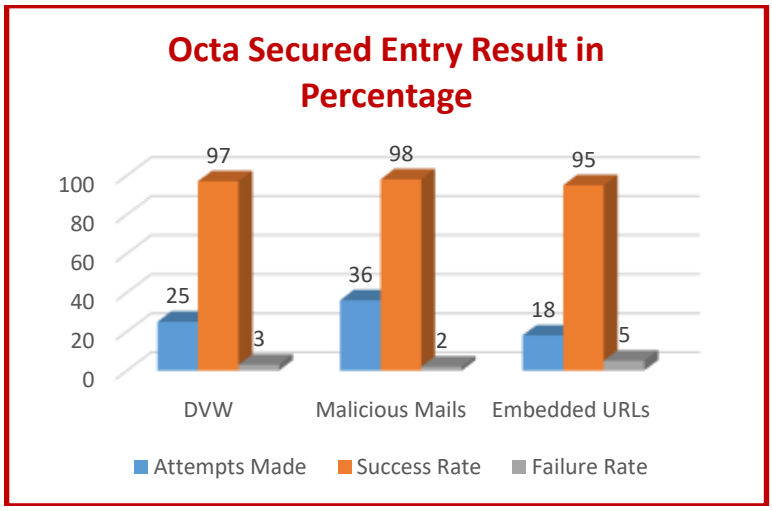

Figure 4: Result of Attack on Octa Secure Entry

Comparative analysis of CETA model with other threat intelligence tools:

Table 3: CETA Performance Analysis

	False Positive	***Human involvement***	***Complexity***	***Initial setup***	***Rest of the alerts***
Spectral Ops	High	Low	Moderate	Moderate	High
Echosec	High	High	High	Moderate	Low
IntSights (ETP) Suite	Moderate	Low	High	Moderate	Moderate
ThreatConnect	Low	Low	High	High	Moderate
ZeroFOX	High	Low	High	Moderate	High
CETA	Low	Low	Moderate	Low	Low

Note: High (3 = Above 75%), Moderate (2 = 74% to 21%), Low (1 = 20% & Below)

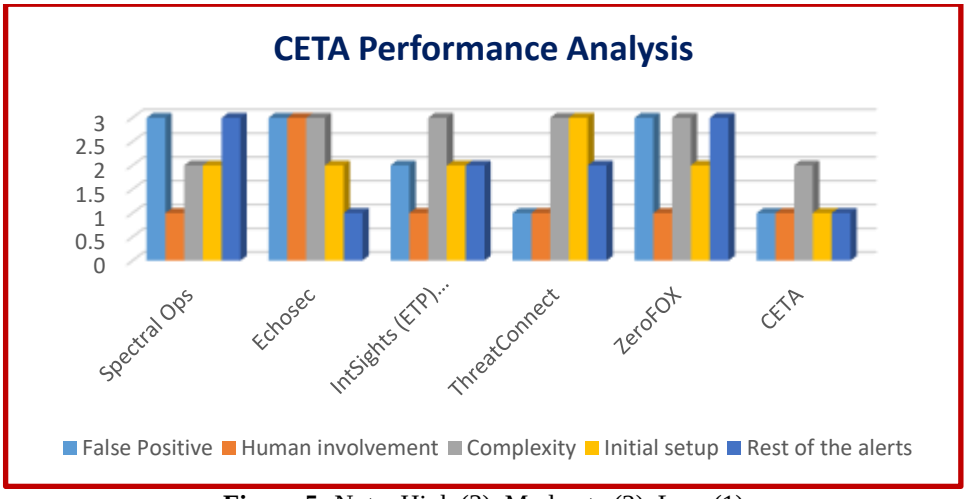

Figure 5: Note: High (3), Moderate (2), Low (1)

The table 3 lists the performance comparison of CETA with standard threat intelligence tools (TIT) where CETA has performed exceedingly well in all the key parameters like generating lowest false positive alarms, better rate of true positive alerts, lowest human interference for the processing the threats and lowest complexity of methodology used. The same is depicted in the figure 5 as a graph.

Future enhancements and conclusion:

The CETA model is well versed in threat intelligence to detect and prevent the advanced persistent threats but the model concentrates on two specific threat vectors spear phishing and social engineering. There is scope for developing threat intelligence for all the possible APTs listed in the introduction section. CETA has proved its efficiency in detection and prevention of APTs through extensive experiments.

REFERENCES:

1. H Tian, Jing X, K Lian, Y Zhang, Research on strong-association rule based web application vulnerability detection, International Conference on Computer Science and Information Technology, IEEE, August 2009, DOI: 10.1109/ICCSIT.2009.5234394, INSPEC Accession Number: 10867977
2. John R. Johnson; Emilie A. Hogan, A graph analytic metric for mitigating advanced persistent threat, IEEE International Conference on Intelligence and Security Informatics, IEEE, June 2013, ISBN:978-1-4673-6213-9.
3. Rudra P. Baksi and Shambhu J. Upadhyaya, Decepticon: a Theoretical Framework to Counter Advanced Persistent Threats, Information Systems Frontiers, Springer, November 2020, 10796-020-10087-4.
4. Yang Gao, Yan Ma and Dandan Li, Anomaly detection of malicious users' behaviors for web applications based on web logs, International Conference on Communication Technology (ICCT), IEEE, October 2017, DOI: 10.1109/ICCT.2017.8359854, 978-1-5090-3944-9.
5. Ingo Paenke, Jürgen Branke, Member, IEEE, and Yaochu Jin, Senior Member, IEEE, Efficient Search for Robust Solutions by Means of Evolutionary Algorithms and Fitness Approximation, Ieee Transactions On Evolutionary Computation, IEEE, August 2006, vol. 10, no. 4, issn 1089-778X
6. Miao Liu And Bin Wang, A Web Second-Order Vulnerabilities Detection Method, IEEE. Translations, IEEE, October 18, 2018, Volume 6, Electronic ISSN: 2169-3536
7. Tarique Mustafa, Malicious Data Leak Prevention and Purposeful Evasion Attacks: An Approach to advanced Persistent Threat (APT) Management, Saudi International Electronics, Communications and Photonics Conference, IEEE, July 2013, INSPEC Accession Number: 13641016.
8. Parth Bhatt, Edgar Toshiro Yano, Dr. Per M. Gustavsson, Towards a Framework to Detect Multi-Stage Advanced Persistent Threats Attacks, International Symposium, IEEE Computer Society, 2014, 978-1-4799-2504-9
9. Bernard Lee Jin Chuan, Manmeet Mahinderjit Singh & Azizul Rahman Mohd Shariff, APT Guard : Advanced Persistent Threat (APT) Detections and Predictions using Android Smartphone, Computational Science and Technology, Springer, pp 545–555, August 2019, 978-981-13-2622-6_53