Contents lists available at Article URL

KKU Journal of Basic and Applied Sciences

Journal homepage: jbas.kku.edu.sa

Real-time Defense System against Insider Attacks Using User Behavior Analytics and Evolving Cluster

Mohammed Yahya Alzahrani[a]

[a] *Information Technology Dept., College of Computer Science & IT, Albaha University, Albaha, Saudi Arabia.*.

Received: 02 December 2018 / Revised: 02 February 2019 / Accepted: 27 February 2019 / Published: 06 March 2019

Abstract: recent significance accidents happened. i.e.: millions email account were stolen, private information are leaking from social media portal and crucial data from institutions are held for ransom. Within years, system administrators were not aware that there are intruder inside the network. The accidents happened due to the lack of intelligent tools to monitor user behavior in internal network. In addition, accuracy is still a challenge in the existing detection systems. This paper presents an intelligent detection system of user behavior anomaly/malware and adopts user behavior analytics and evolving clustering method to improve the accuracy of the detection. A prototype of the proposed system has been built and experiments were conducted on real traffic at College of Computer Science and Information Technology, Albaha University network. The experimental results show that the proposed system better clustering results and high percentage of accuracy

Keywords: Cyber-Attack, Anomaly Detection, User Behavior, Evolving Cluster

Correspondence:
E-mail: auther@mail.com
Tel: o*******

1. INTRODUCTION

Access control, firewalls and anti-virus software in information security have successfully helped us to stop people from doing wrong things in a computer network system. However, we still have limitation in providing insights into human behavior and advanced malware because the rules are set up statically in our traditional reliance. Thus, the rules are hard to update dynamically when needed due to the changes in the network. Security information and event management (SIEM), an approach to security management, provides us with a better insight into patterns of user behavior.

Insider threat is a generic term for a threat to an organization's security or data that comes from within. Such threats are usually attributed to employees or former employees, but may also arise from third parties, including contractors, temporary workers or customers.

A technique called User Behavior Analytics has started to become particularly useful in giving solutions that have a level of flexible pattern recognition, which rules are unsuitable, and which humans simply cannot achieve due to the excruciatingly large amount of data involved. The user behavior is considered a hard problem due to the fact that human behavior is erratic and hard to predict.

In fact, System Administrators are often based only on ad hoc methods developed from years of experience in spotting and finding anomalous behaviors in their networks. Many commercial and open source tools have been developed to assist the System Administrators to guard their networks from the attacks. Nevertheless, these tools are not dynamic to adapt any traffic changing that requires the System Administrators to redefine policies from time to time with the aim to trigger alerts. The accuracy of the detection depends on how good the description of the anomalous behavior.

Therefore, this paper presents an intelligent detection system that is able to adapt to any changes of user behavior with a minimum involvement of network administrator. The system applies evolving cluster method to perform on-the-fly learning.

Many research works on user behavior analytics, evolving cluster methods in the area of cyber security have been carried out. Authors in [1] proposed a framework for analysis-based learning the user behavior. The authors introduce a cluster-based outlier detection algorithm to detect anomalous data and also describe the set of factor of user behavior profile, program profile and system resource usage. The proposed framework is sufficient to analyze behavior in host-based intrusion without explaining about how to capturing data in stream network and result of the alerts to be forwarded to response mechanism after identifying attack. Lim & Jones [2], present taxonomy of anomaly detection techniques in network-based intrusion. The authors confirmed the extent of the use of misuse-based detection system has become the dominant strategy for counter-measure from suspicious attacks and identified two approaches for building the behavior model; learning-based and specification-based models. Koch [3] uses fast-learning method with neural network based on pre-processed component. Prior to this work, authors in [4] through series of experiments investigate the possibility of automatically constructing a user profile from the logs users' activities, and introduce a standard formalism for regular expressions adopted for describing episodes and profiles. The most recent works on user behavior analytics on malware detections include Nitesh et al. [5] that presents malware classification using static, dynamic and hybrid approaches; Hansen et al. [6] use Random Forest classifier; Daku [7] use machine learning method, and Shashanka et al. [8] work on classification of malwares in enterprise security.

The existing network policies are not adequate to adapt to the continually changing network conditions arising from the explosive traffic growth. Thus, we need to rethink how the network traffic control can be improved. The incorporating intelligence into systems of network traffic control can play a significant role in guaranteeing security and Quality of Service (QoS) in Internet Protocol (IP)-based networks [9].

Works on evolving cluster method/algorithm have been carried out [10-17], more specifically some of them focus on dynamic approaches [18-19]. The works consider a micro clustering technique to deal with how to cluster evolving data stream. Recent works on evolving cluster method/algorithm in the area of malware detection include the following.

- Zhang et al. [20] use Convolutional Neural Network combined with Back Propagation Neural Network.
- Kamaruddin et al. [21] use Apache Spark and applied to Banking network.
- Bezerra et al [22] propose a new ECM for online data streams. The proposed ECM uses a statistical method based on the concepts of typicality and eccentricity able to group similar data observations.
- Mustofa et al [23] propose adaptive memetic differential evolution optimization algorithms for data clustering problems.
- Carnein and Trautmann [24] present a comprehensive survey on stream clustering algorithms.
- Škrjanc et al [25] present a survey on evolving fuzzy and neuro-fuzzy approaches in clustering, regression, identification, and classification.

- Bodyanskiy et al [26] propose a new class of evolving fuzzy networks, namely the evolving GMDH-neuro-fuzzy systems (Group Method of Data Handling).

Works on real time anomaly/malware detection systems also get serious attention from researchers. A study by Wan and Horng [27], presents an intelligent monitoring system for local-area network traffic to capture, parse and analyze the captured packets in real time fashion. Alharthi et al [16] present an intelligent real time Internet of Thing monitoring. Based on the proposed works by Alharthi et al. [16] and Pasha et al [17], this paper presents a system for detecting user behavior anomalies thru analyzing their network traffic behavior that involves user profiling, user behavior analytics. Furthermore, this paper utilizes an evolving cluster method that can react to sudden emerging of specific clusters such as attacks within a continuing data stream.

2. MATERIALS AND METHODS

The proposed system consists of three main components: Data acquisition, baselining/profiling, and detection modules. Figure 3 depicts the architecture of the proposed system.

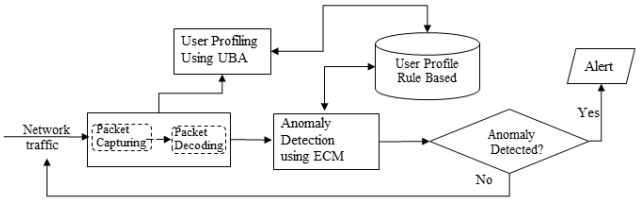


Fig. 3. User profiling process flow.

2.1 Data Acquisition Module

The module captures the traffic packet in passive mode, using pcap library, then the captured data is converted into text format and sent to Decoding sub-module. The decoded packets are then sent to the profiling module to produce the dataset. The Detection module also uses this data acquisition module during the training and testing experiment in real time fashion.

Traffic Dataset

The PacketCapturing module captures College of Computer Science and IT (CCSIT) network traffic of 96 different users of eight segments for period of one month. Figure 1 shows the snapshot of the packet capturing during the observation. At the bottom-right of the figure, it shows computers/nodes attached to the network. At the bottom-left, it displays the decoded process of the capturing packets. Figure 2 shows an example of captured data in text format. Then the system also captures the traffic of the same network for another one month duration to create dataset for training/learning and testing the ECM engine.

As general knowledge, network traffic packet has a header. The header consists of fields such as time stamp, MAC addresses, IP addresses, Protocol use in the upper layer, packet ID, packet size, etc. To identify a suspicious packet, the detection system checks the value of the certain fields and compare to values resulted from the profiling phase. Those fields then are considered as features of the packet. This paper considers packet header fields in Table 1 as the features of the packets in the created dataset.

TABLE I: FEATURES OF DATASET

Features	**Domain (value)**
Time stamp	Time
Source, destination MAC	Hexadecimal
Soure, destination IP	Integer
Protocol	String char
ID	String char
Packet number	Integer
Packet size	Integer

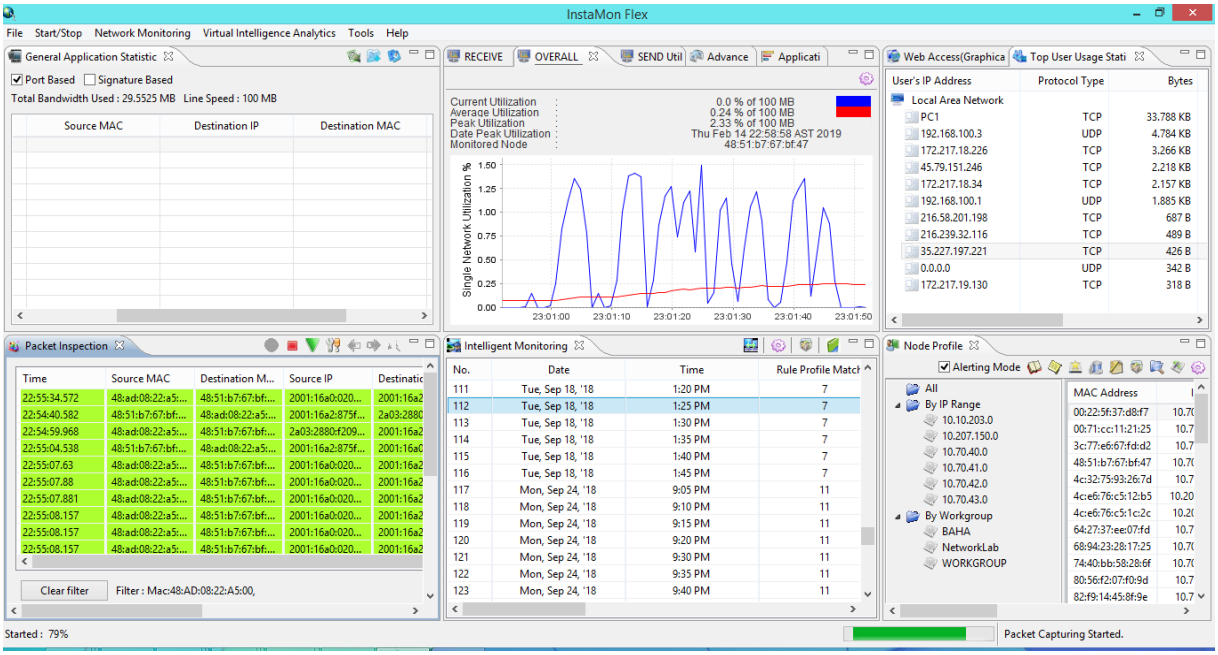


Fig. 1. A snapshot of data capturing activity using Instamon®.

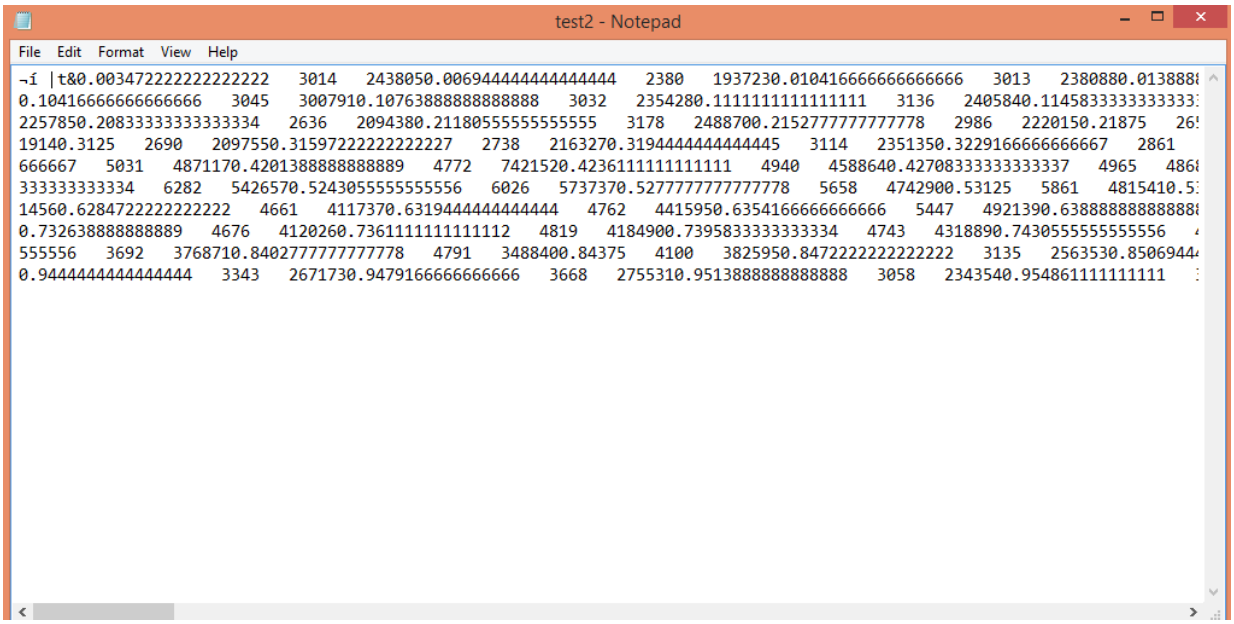


Fig. 2. A snapshot of captured data in text format.

2.2 User Baselining/Profiling Module

Users' attributes include User-ID, Name, Department, Authorized files, Authorized applications, Working-hours type (flexible or fixed time), IP/MAC address. The baselining/profiling considers only three type of applications that use protocols: HTTP, ICMP and overall traffic. The module uses User Behavior Analytics, combining the user's attributes and user's traffic profile to determine the user behavior.

User Behaviour Analytics

User behavior analytics (UBA), also known as user and entity behavior analytics (UEBA), is a process of gathering insight into the network events that users generate every day. Once collected and analyzed, it can be used to detect the use of compromised credentials, lateral movement, and other malicious behavior. UBA deviates from traditional consumer behavioral analytics to focus on the behavior of systems and the user accounts on them. UBA exposes stealthy, attacker activities by uncovering patterns in user behavior to identify what's "normal" behavior, and what may be evidence of intruder compromise, insider threats, or risky behavior on a network.

UBA connects activity on the network to a specific user as

opposed to an IP address or an asset. This means that if a user starts to behave in a way that is unusual or unlikely, even if it is not flagged by traditional perimeter monitoring tools, the UBA will be able to spot the behavior quickly, determine whether it is anomalous, and start an investigation if needed. As an example, a user that use HTTP protocol has a daily quota of bandwidth and access to the university system, suddenly show intensive use of the university system with high bandwidth. So, we may suspect this kind of behavior and system will produce an alert.

This paper uses a neuro-fuzzy system as shown in Figure 4. The input to the fuzzy interface are a set of data on users' attributes and other information as profile producing from the baselining process. The Fuzzy interface will then process this information to produce users' current profile to be fed into the neural network.

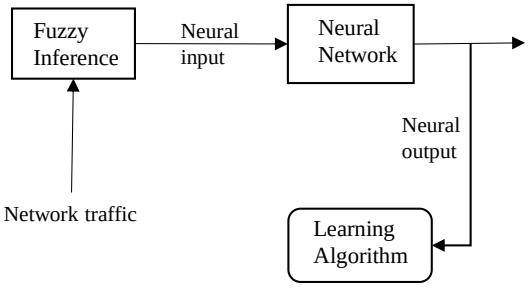


Fig. 4. Neuro-Fuzzy Engine for User Behavior Analytics.

Evolving Cluster

The Evolving Clustering Method (ECM) is an online clustering method that performs well on one-pass partitioning of an input space through partitioning scarce input. ECM_m is a clustering method for clustering streams of network traffic data that combines the ECM algorithm and its extension ECM_c so that it can use the number of cluster created in previous process and optimize its cluster center in online mode.

The input for ECM_m is fed with cumulative information every 5 minutes, which contains the time when the traffic is captured. This information is extracted from the collected network traffic data streams. As for the output, information about the cluster center and its radius, a matrix that maps the inputs index into the cluster where it belongs, the objective value and the input itself after being normalized is included. Figure 5 illustrates the ECM process.

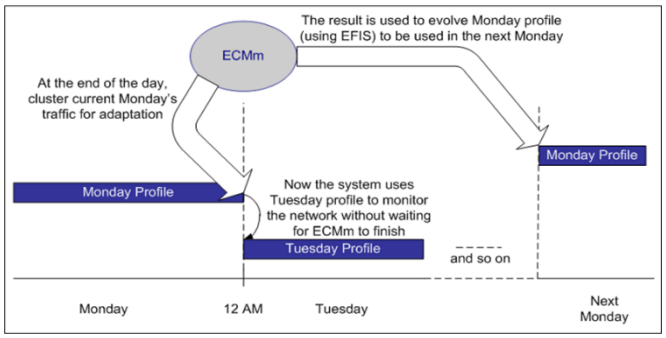


Fig. 5. The ECM workflow diagram [16], [17].

2.3 Detection Engine

The real-time network traffic is captured, and then the user behavior analytics is performed with the reference to the user profile rule-based. Figure 6 depicts the anomaly detection mechanism.

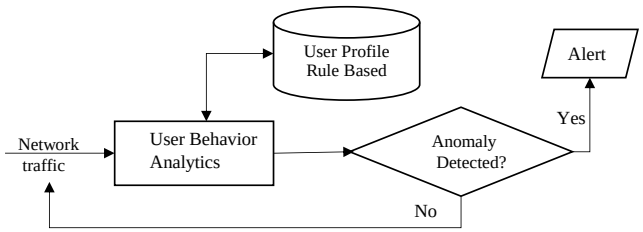


Fig. 6. The anomaly detection mechanism.

2.4 Implementation

The baselining/profiling strategy, the mechanism of user behavior analytics, and the anomaly detection system have been implemented as a prototype using Java on Eclipse platform. The prototype is incorporated into Instamon® as a defense system against insider attacks. The implementation is run on high-end server machine with the following specifications. VPS 2 cores processor, 16GB RAM, and 1TB SSD storage.

3. RESULTS AND DISCUSSIONS

3.1 Baselining/profiling Results

This paper considers three applications: HTTP, ICMP and overall traffic, of 96 users in 8 network segments. The module captures the 96 users traffic from Sunday to Thursday, every two hours starts from 00.00 to 24.00 o'clock. The experiments are conducted during Mid of August until Mid of October 2018. Table 2 displays results of dataset creation.

TABLE II: CREATED DATASET

No.	Dataset	Size	Period
1.	For baselining	127,611,398 KB	Aug.-Sept. 2018
2.	For clustering	179,554,324 KB	Sept.-Oct. 2018

Table 3 shows an example of results of a user traffic profiling on a Sunday observation, specifically on HTTP packets. For each time interval, the number of packets are accumulated and converted into Byte size.

TABLE III: EXAMPLE OF A USER TRAFFIC PROFILING

Observation time	Total packet	Total size (in Byte)
00.00-02.00	1247	2291992
02.00-04.00	1382	2738281
04.00-06.00	1293	2283928
06.00-08.00	1829	3828388
08.00-10.00	3920	6939198
10.00-12.00	5020	10238328
12.00-14.00	6950	15388288
14.00-16.00	5829	10028372
16.00-18.00	7382	17488392
18.00-20.00	4923	8382872
20.00-22.00	2381	4828188
22.00-24.00	1832	2038281

3.2 Clustering Results

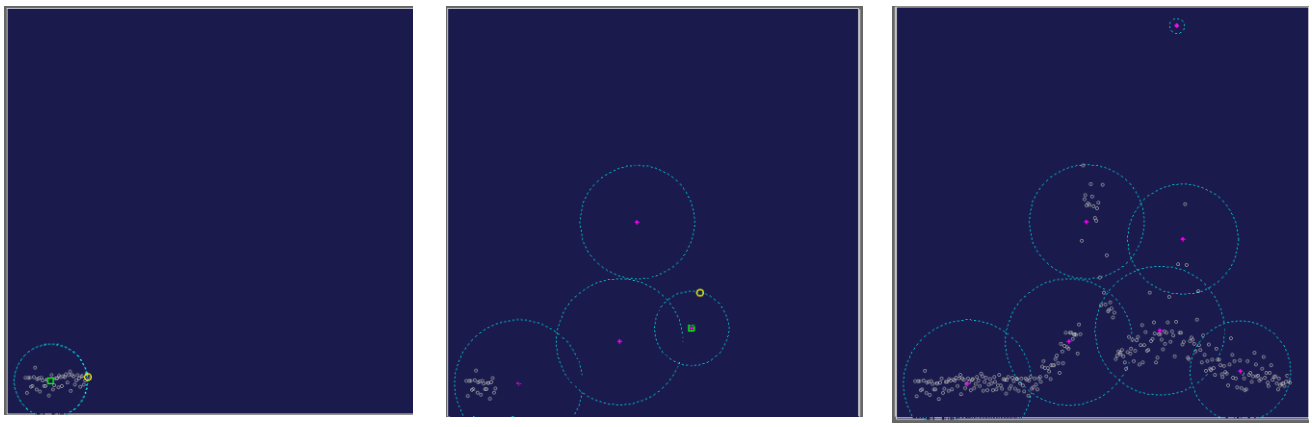
The implemented prototype performs the clustering by utilizing the created dataset that includes data of each day (Monday to Sunday). Then, the clustering is performed every end of the day using the data captured during the day. If there is no significance change in the clustering results, the detection procedure may use the old thresholds/rules.

Table 4 shows the information on the clustering results of a user. The clustering utilizes 287 datasets and involves 3 dimensions (Time, Total packet number, and Total packet size) The clustering process achieves objective value of 13.1993 and consumes 1.85 seconds of CPU. The processing time can be scale up by using more sophisticated machine. Thus, it can be considered as a real time system. The results is 7 clusters: 6 different normal traffic categories and one outliers.

TABLE IV: A USER TRAFFIC CLUSTERING RESULT

Parameters	results
Datasets #	278
Dimension	3
Obj. Value	13.1993
CPU time	1.85 sec.
Cluster #	7

Figure 7 illustrates how the ECM produces the clusters. Figure 7(a) shows the clustering process starts, Figure 7 (b) in the middle shows ongoing process and Figure 7(c) shows the result of the clustering. Small circle in the figure represents a traffic data, '+' represents a cluster center and the dot-line circle represents a cluster.

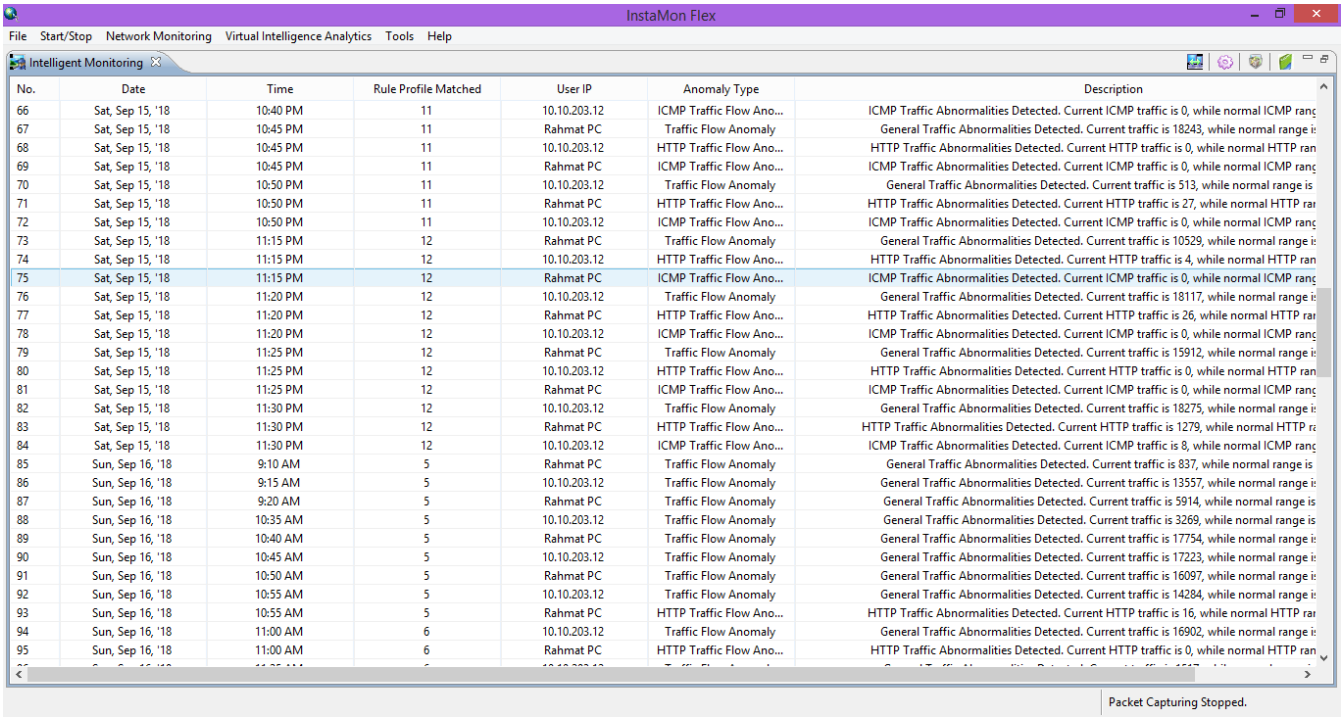


(a) (b) (c)

Fig. 7. Illustration of the ECM result.

3.3 Anomaly Detection using Created Dataset

Having done the baselining/profiling and clustering each user's traffic, the prototype is tested on the created dataset. The proposed system has capability to run captured traffic for forensics analysis purpose. Figure 8 shows the alerts from the prototype system representing the detected anomalies. The prototype system is able to detect anomaly on HTTP and ICMP traffic that happened during the experiment. Overall traffic anomaly is also detected and categorized as general traffic anomaly.



No.	Date	Time	Rule Profile Matched	User IP	Anomaly Type	Description
66	Sat, Sep 15, '18	10:40 PM	11	10.10.203.12	ICMP Traffic Flow Ano...	ICMP Traffic Abnormalities Detected. Current ICMP traffic is 0, while normal ICMP ran
67	Sat, Sep 15, '18	10:45 PM	11	Rahmat PC	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 18243, while normal range i
68	Sat, Sep 15, '18	10:45 PM	11	10.10.203.12	HTTP Traffic Flow Ano...	HTTP Traffic Abnormalities Detected. Current HTTP traffic is 0, while normal HTTP ran
69	Sat, Sep 15, '18	10:45 PM	11	Rahmat PC	ICMP Traffic Flow Ano...	ICMP Traffic Abnormalities Detected. Current ICMP traffic is 0, while normal ICMP ran
70	Sat, Sep 15, '18	10:50 PM	11	10.10.203.12	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 513, while normal range is
71	Sat, Sep 15, '18	10:50 PM	11	Rahmat PC	HTTP Traffic Flow Ano...	HTTP Traffic Abnormalities Detected. Current HTTP traffic is 27, while normal HTTP ra
72	Sat, Sep 15, '18	10:50 PM	11	10.10.203.12	ICMP Traffic Flow Ano...	ICMP Traffic Abnormalities Detected. Current ICMP traffic is 0, while normal ICMP ran
73	Sat, Sep 15, '18	11:15 PM	12	Rahmat PC	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 10529, while normal range i
74	Sat, Sep 15, '18	11:15 PM	12	10.10.203.12	HTTP Traffic Flow Ano...	HTTP Traffic Abnormalities Detected. Current HTTP traffic is 4, while normal HTTP ran
75	Sat, Sep 15, '18	11:15 PM	12	Rahmat PC	ICMP Traffic Flow Ano...	ICMP Traffic Abnormalities Detected. Current ICMP traffic is 0, while normal ICMP ran
76	Sat, Sep 15, '18	11:20 PM	12	10.10.203.12	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 18117, while normal range i
77	Sat, Sep 15, '18	11:20 PM	12	Rahmat PC	HTTP Traffic Flow Ano...	HTTP Traffic Abnormalities Detected. Current HTTP traffic is 26, while normal HTTP ra
78	Sat, Sep 15, '18	11:20 PM	12	10.10.203.12	ICMP Traffic Flow Ano...	ICMP Traffic Abnormalities Detected. Current ICMP traffic is 0, while normal ICMP ran
79	Sat, Sep 15, '18	11:25 PM	12	Rahmat PC	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 15912, while normal range i
80	Sat, Sep 15, '18	11:25 PM	12	10.10.203.12	HTTP Traffic Flow Ano...	HTTP Traffic Abnormalities Detected. Current HTTP traffic is 0, while normal HTTP ran
81	Sat, Sep 15, '18	11:25 PM	12	Rahmat PC	ICMP Traffic Flow Ano...	ICMP Traffic Abnormalities Detected. Current ICMP traffic is 0, while normal ICMP ran
82	Sat, Sep 15, '18	11:30 PM	12	10.10.203.12	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 18275, while normal range i
83	Sat, Sep 15, '18	11:30 PM	12	Rahmat PC	HTTP Traffic Flow Ano...	HTTP Traffic Abnormalities Detected. Current HTTP traffic is 1279, while normal HTTP ra
84	Sat, Sep 15, '18	11:30 PM	12	10.10.203.12	ICMP Traffic Flow Ano...	ICMP Traffic Abnormalities Detected. Current ICMP traffic is 8, while normal ICMP ran
85	Sun, Sep 16, '18	9:10 AM	5	Rahmat PC	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 837, while normal range is
86	Sun, Sep 16, '18	9:15 AM	5	10.10.203.12	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 13557, while normal range i
87	Sun, Sep 16, '18	9:20 AM	5	Rahmat PC	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 5914, while normal range is
88	Sun, Sep 16, '18	10:35 AM	5	10.10.203.12	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 3269, while normal range is
89	Sun, Sep 16, '18	10:40 AM	5	Rahmat PC	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 17754, while normal range i
90	Sun, Sep 16, '18	10:45 AM	5	10.10.203.12	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 17223, while normal range i
91	Sun, Sep 16, '18	10:50 AM	5	Rahmat PC	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 16097, while normal range i
92	Sun, Sep 16, '18	10:55 AM	5	10.10.203.12	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 14284, while normal range i
93	Sun, Sep 16, '18	10:55 AM	5	Rahmat PC	HTTP Traffic Flow Ano...	HTTP Traffic Abnormalities Detected. Current HTTP traffic is 16, while normal HTTP ra
94	Sun, Sep 16, '18	11:00 AM	6	10.10.203.12	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 16902, while normal range i
95	Sun, Sep 16, '18	11:00 AM	6	Rahmat PC	HTTP Traffic Flow Ano...	HTTP Traffic Abnormalities Detected. Current HTTP traffic is 0, while normal HTTP ran

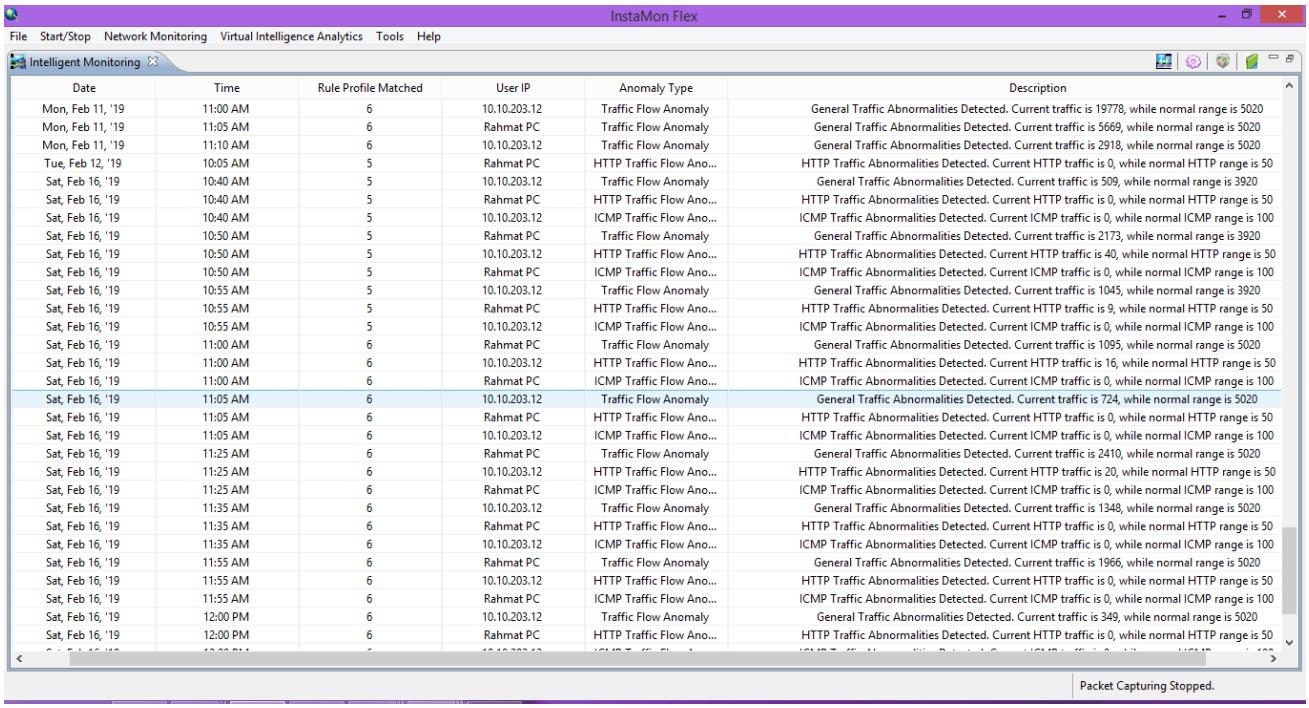
Fig. 8. Alerts produced by the prototype using the captured dataset.

Experimental results also show that the prototype system is able to perform on-the-fly clustering every midnight and then update the rules set up automatically.

3.4 Anomaly Detection on Real-time Traffic Results

Finally, the prototype is run to monitor and detect any anomaly on the CCSIT network. The experiment is conducted during the first and second week of February 2019. Figure 9 shows the alerts on anomaly from the prototype system. It is observed that the alerts are less compared to the alerts on the dataset.

This improvement is because the system is able to adapt thru the online clustering and adjust the profile of users' traffic as such the detection system works more efficient.



Date	Time	Rule Profile Matched	User IP	Anomaly Type	Description
Mon, Feb 11, '19	11:00 AM	6	10.10.203.12	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 19778, while normal range is 5020
Mon, Feb 11, '19	11:05 AM	6	Rahmat PC	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 5669, while normal range is 5020
Mon, Feb 11, '19	11:10 AM	6	10.10.203.12	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 2918, while normal range is 5020
Tue, Feb 12, '19	10:05 AM	5	Rahmat PC	HTTP Traffic Flow Ano...	HTTP Traffic Abnormalities Detected. Current HTTP traffic is 0, while normal HTTP range is 50
Sat, Feb 16, '19	10:40 AM	5	10.10.203.12	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 509, while normal range is 3920
Sat, Feb 16, '19	10:40 AM	5	Rahmat PC	HTTP Traffic Flow Ano...	HTTP Traffic Abnormalities Detected. Current HTTP traffic is 0, while normal HTTP range is 50
Sat, Feb 16, '19	10:40 AM	5	10.10.203.12	ICMP Traffic Flow Ano...	ICMP Traffic Abnormalities Detected. Current ICMP traffic is 0, while normal ICMP range is 100
Sat, Feb 16, '19	10:50 AM	5	Rahmat PC	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 2173, while normal range is 3920
Sat, Feb 16, '19	10:50 AM	5	10.10.203.12	HTTP Traffic Flow Ano...	HTTP Traffic Abnormalities Detected. Current HTTP traffic is 40, while normal HTTP range is 50
Sat, Feb 16, '19	10:50 AM	5	Rahmat PC	ICMP Traffic Flow Ano...	ICMP Traffic Abnormalities Detected. Current ICMP traffic is 0, while normal ICMP range is 100
Sat, Feb 16, '19	10:55 AM	5	10.10.203.12	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 1045, while normal range is 3920
Sat, Feb 16, '19	10:55 AM	5	Rahmat PC	HTTP Traffic Flow Ano...	HTTP Traffic Abnormalities Detected. Current HTTP traffic is 9, while normal HTTP range is 50
Sat, Feb 16, '19	10:55 AM	5	10.10.203.12	ICMP Traffic Flow Ano...	ICMP Traffic Abnormalities Detected. Current ICMP traffic is 0, while normal ICMP range is 100
Sat, Feb 16, '19	11:00 AM	6	Rahmat PC	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 1095, while normal range is 5020
Sat, Feb 16, '19	11:00 AM	6	10.10.203.12	HTTP Traffic Flow Ano...	HTTP Traffic Abnormalities Detected. Current HTTP traffic is 16, while normal HTTP range is 50
Sat, Feb 16, '19	11:00 AM	6	Rahmat PC	ICMP Traffic Flow Ano...	ICMP Traffic Abnormalities Detected. Current ICMP traffic is 0, while normal ICMP range is 100
Sat, Feb 16, '19	11:05 AM	6	10.10.203.12	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 724, while normal range is 5020
Sat, Feb 16, '19	11:05 AM	6	Rahmat PC	HTTP Traffic Flow Ano...	HTTP Traffic Abnormalities Detected. Current HTTP traffic is 0, while normal HTTP range is 50
Sat, Feb 16, '19	11:05 AM	6	10.10.203.12	ICMP Traffic Flow Ano...	ICMP Traffic Abnormalities Detected. Current ICMP traffic is 0, while normal ICMP range is 100
Sat, Feb 16, '19	11:25 AM	6	Rahmat PC	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 2410, while normal range is 5020
Sat, Feb 16, '19	11:25 AM	6	10.10.203.12	HTTP Traffic Flow Ano...	HTTP Traffic Abnormalities Detected. Current HTTP traffic is 20, while normal HTTP range is 50
Sat, Feb 16, '19	11:25 AM	6	Rahmat PC	ICMP Traffic Flow Ano...	ICMP Traffic Abnormalities Detected. Current ICMP traffic is 0, while normal ICMP range is 100
Sat, Feb 16, '19	11:35 AM	6	10.10.203.12	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 1348, while normal range is 5020
Sat, Feb 16, '19	11:35 AM	6	Rahmat PC	HTTP Traffic Flow Ano...	HTTP Traffic Abnormalities Detected. Current HTTP traffic is 0, while normal HTTP range is 50
Sat, Feb 16, '19	11:35 AM	6	10.10.203.12	ICMP Traffic Flow Ano...	ICMP Traffic Abnormalities Detected. Current ICMP traffic is 0, while normal ICMP range is 100
Sat, Feb 16, '19	11:55 AM	6	Rahmat PC	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 1966, while normal range is 5020
Sat, Feb 16, '19	11:55 AM	6	10.10.203.12	HTTP Traffic Flow Ano...	HTTP Traffic Abnormalities Detected. Current HTTP traffic is 0, while normal HTTP range is 50
Sat, Feb 16, '19	11:55 AM	6	Rahmat PC	ICMP Traffic Flow Ano...	ICMP Traffic Abnormalities Detected. Current ICMP traffic is 0, while normal ICMP range is 100
Sat, Feb 16, '19	12:00 PM	6	10.10.203.12	Traffic Flow Anomaly	General Traffic Abnormalities Detected. Current traffic is 349, while normal range is 5020
Sat, Feb 16, '19	12:00 PM	6	Rahmat PC	HTTP Traffic Flow Ano...	HTTP Traffic Abnormalities Detected. Current HTTP traffic is 0, while normal HTTP range is 50

Fig. 9. Alerts produced by the prototype during the real-time experiment.

3.5 Evaluation

With the aim to evaluate the performance of the anomaly detection system in term of accuracy, this work performs a measurement of clustering process of four existing evolving cluster methods on the same dataset. The author believes that the better the clustering result the better the detection accuracy. Table 5 shows the comparison results. EC algorithm [11], Local Mean clustering [12] and Vector Quantification [15] provide fast clustering process, however, not so accurate. Whereas Parallel ECM, ESOM and Adapted ECM_m provide a better clustering. Parallel ECM achieves the best performance because it uses parallel machines. Compare to ESOM, the adapted ECM_m performs faster.

Method	Author(s)	Cluster #	CPU Time
EC algorithm	Bazerra et al [11]	6	1.38 sec.
ESOM	Kasbov, Deng [14]	7	1.62 sec.
Local. means	Baruah, Angelov [12]	6	1.36 sec.
Vector quantif.	Lughofer [15]	6	1.41 sec.
Paralell ECM	Komaruddin [21]	7	1.02 sec.
Adapted ECM_m	This paper	7	1.58 sec.

TABLE V: CLUSTERING RESULTS COMPARISON

5. CONCLUSION

A mechanism for detecting anomaly of user traffic behavior has been developed in this research work. The mechanism adopts the evolving clustering method as the main tool to adapt against the changes in users' traffic behavior. The proposed mechanism was tested in a dataset as well as in real-time network. The experiment results showed that the detection mechanism is able to alert the system administrator on the anomalies with alerts that are more accurate. This results show that the ECM performs a good learn during the detection. The author plans to adopt the dynamic ECM for the clustering engine as a future work.

ACKNOWLEDGMENT

This research is partially funded by Scientific Research Deanship, Albaha University under grant no.: 020//1439.

REFERENCES

[1] Qiao H, Peng J, Feng C, Rozenblit J.W, **"Behavior analysis-based learning framework for host level intrusion detection"** 14th Annual IEEE International Conference and Workshops on the Engineering of Computer-Based Systems (ECBS'07), Tucson, AZ, USA, 26-29 March, 2007, DOI: 10.1109/ECBS.2007.23 (2007).

[2] Lim S.Y. & Jones A. **"Network anomaly detection system: the state of art of network behaviour analysis"** International Conference on Convergence and Hybrid Information Technology Proceedings, 459-465, (2008)

[3] Koch R Changing **"network behavior"**, Third International Conference on Network and System Security, 60-66, (2009)

[4] Galassi U., Giordana A. & Mendola D. **"Learning user profile from traces"**, Symposium on Applications and the Internet Workshops (SAINT 2005 Workshops), pp.166-169, (2005)

[5] Nitesh K, Subhasis M, Mugdha G, Anand H, Sandeep K.S, **"Malware classification using early stage behavioral analysis"**, 14th Asia Joint Conference on Information Security, Kobe, Japan, 1-2 Aug, 2019, DOI: 10.1109/AsiaJCIS.2019.00-10.

[6] Hansen, S.S. Larsen, T.M.T, Stevanovic, M, Pedersen, J.M, **"An approach for detection and family classification of malware based on behavioral analysis"**, International Conference on Computing, Networking and Communications (ICNC), Kauai, Hawaii, USA, 15-18 Feb, (2016). DOI: 10.1109/ICCNC.2016.7440587 .

[7] Daku H, Zavarsky, P, Malik, Y, **"Behavioral-based classification and identification of ransomware variants using machine learning"**, 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, New York, USA, 1-3 Aug. (2018). DOI: 10.1109/TrustCom/BigDataSE.2018.00224 .

[8] Shashanka M, Shen M-Y & Wang, J., **"User and entity behavior analytics for enterprise security"**. IEEE International Conference on Big Data, Washington DC, USA, 5-8 Dec. 2016, 1867-1874, (2016). DOI: 10.1109/BigData.2016.7840805.

[9] Barabas M., Boanea G., and Dobrota V. **"Multipath routing management using neural networks-based traffic prediction"**, 3rd International Conference on Emerging Network Intelligence Proceedings. 118-124, (2011).

[10] Tseng F.F. **"Evolving Clustering algorithms and their application for condition monitoring, diagnostics, & prognostics"**, PhD Thesis Wayne State University. 1750, (2017)

[11] Bezerra C. G., Costa B. S. J., Guedes L. A. And Angelov P. P. **"A new evolving clustering algorithm for online data streams"**, EAIS Proceedings, 162-168, (2016). doi: 10.1109/EAIS.2016.7502508 .

[12] Baruah R.D. and Angelov P. **"Evolving local means method for clustering of streaming data"** IEEE World Congress on Computational Intelligence Proceedings, 2-8, (2012).

[13] Serir L., Ramasso E., Nectoux P., Bauer O., Zerhouni N. **"Evidential evolving Gustafson-Kessel algorithm (E2GK) and its application to PRONOSTIA's data streams partitioning"**, 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC) Proceedings, 8273-8278, (2011).

[14] Kasabov N. and Deng D. **"ESOM: An Algorithm to Evolve Self-Organizing Maps from on-line data streams"**, IEEE International Joint Conference on (IJCNN), 6003-6009, (2000). doi:10.1109/IJCNN.2000.859364 .

[15] Lughofer E. **"Evolving vector quantization for classification of on-line data streams"**, International Conference on Computational Intelligence for Modelling Control & Automation Proceedings, 779-784, (2008).

[16] Alharthi A.F., Alzahrani M., Aldmour I., Pasha M.F., Stiawan D., Budiarto R. **"Smart Real-Time Internet of Thing Network Monitoring"**, **Powering the Internet of Things With 5G Networks**, Editors: Mohana, V, Budiarto R, Aldmour, IGI Publisher, pp. 202-225, (2018). DOI: 10.4018/978-1-5225-2799-2.ch008 .

[17] Pasha M.F., Budiarto R., Syukur M., Yamada M. **"Adaptive real-time network monitoring system: detecting anomalous activity with evolving connectionist system"**, In Joaquim Filipe, Helder Coelhas, Communications in Computer and Information Sciences (CCIS), Springer-Verlag. Vol. 3, 113-125, (2008).

[17] Lughofer E. "**Dynamic evolving cluster models using On-line split-and-merge operations"**, 10th International Conference on Machine Learning and Applications Proceedings, 20-26, (2011).

[17] Baruah R.D. Angelov P., Baruah D. **"Dynamically Evolving Fuzzy Classifier for Real-time Classification of Data Streams"**, Conference on Evolving and Adaptive Intelligent Systems (EAIS) Proceedings, 383-389, (2014).

[17] Zhang J, Qin Z, Yin H, Ou, L, Zhang K, **"A feature-hybrid malware variants detection using CNN based opcode embedding and BPNN based API embedding"**, Computers & Security July 2019 84:376-392, Elsevier, (2019). DOI: 10.1016/j.cose.2019.04.005 .

[17] Kamaruddin Sk, Vadlamani R, & Mayank P. **"Parallel evolving clustering method for big data analytics using apache spark: Applications to banking and physics"**, International Conference on Big Data Analytics, 278-292, Nov. (2017) DOI: 10.1007/978-3-319-72413-3_19 .

[17] Bezerra C.G, Costa B. S. J, Guedes L. A Angelov and P. P, **"A new evolving clustering algorithm for online data streams"**, 2016 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS), Natal, pp. 162-168, (2016). DOI: 10.1109/EAIS.2016.7502508.

[17] Mustafa, H., Ayob, M., Nazri, M., & Kendall, G. **"An improved adaptive memetic differential evolution optimization algorithm for data clustering problems"**, PloS one, 14(5), e0216906, (2019). doi:10.1371/journal.pone.0216906.

[17] Carnein M & Trautmann H, **"Optimizing data stream representation: An extensive survey on stream clustering algorithms"**, Business & Information Systems Engineering, The International Journal of WIRTSCHAFTSINFORMATIK, Springer;Gesellschaft für Informatik e.V. (GI), vol. 61(3):277-29, (2019).

[17] Škrjanc I, Iglesias J.A, Sanchis A, Lughofer E, Gomide F, **"Evolving fuzzy and neuro-fuzzy approaches in clustering, regression, identification, and classification: A Survey"**, Information Sciences, Elsevier, July 2019, 490:344-368, (2019).

[17] Bodyanskiy Y, Boiko O, Zaychenko Y and Hamidov G, **"Evolving GMDH-neuro-fuzzy system with small number of tuning parameters"**, 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Guilin, pp. 1321-1326, (2017). doi: 10.1109/FSKD.2017.8392957.

[17] Wan M-H, Horng M.F, **"An intelligent monitoring system for local-area network traffic"**, 8th IEEE International Conference on Intelligent Systems Design and Applications, (ISDA '08). Kaohsiung, Taiwan, 26-28 Nov, (2008). DOI: 10.1109/ISDA.2008.366.