Journal of ISMAC (2020)
Vol.02/ No.03
Pages: 147-153
http://irojournals.com/iroismac/
DOI: https://doi.org/10.36548/jismac.2020.3.002

Process Mining Error Detection for Securing the IoT System

Dr. Subarna Shakya
Professor, Department of Electronics and Computer Engineering,
Central Campus, Institute of Engineering, Pulchowk,
Tribhuvan University,
Pulchowk, Lalitpur, Nepal.
Email: drss@ioe.edu.np.

Abstract: As the use of Internet-of-Things in day to lives increases, its connection with objects and use of sensors has increased in number largely. These objects are also integrated with the internet, enabling its application in many more complex systems. Though efforts have been implemented to protect the security management, there are some major challenges faced by system because of the limited resources, heterogeneity and complexity of the system. This gives way to detecting the various attacks by characterizing the IoT system. Using a novel architecture with appropriate components, we have proposed a prototype of our concept that is used to determine the performance of the system by means of real-time input from the industries by extensive experimentation.

Keywords: Anomaly Detection, Data Mining, Process Mining, Internet of Things, Security Management;

1. Introduction

Over the years, the era of Internet-of-Things has grown greatly and evolved in variety of domains with applications in various fields like industrial infrastructure and smart home networks with advancement in industrial evolution. The complexity involved in the IoT system is often not properly estimated in terms of system complexity, which has resulted in new challenges arising from security measures. However the discrepancies that the IoT systems face will further impact other devices that do not depend on Internet-of-Things. This means that even if a device is not dependent on IoT, it is still vulnerable to the attacks it receives from the ones that are already infected. The botnet which is used in Distributed DoS attacks [1] is an example of an IoT device which is compromised. It was the Mirai botnet [2] which caused the IoT devices to be vulnerable to attacks, making them unavailable for a long period of time on the internet platforms. Similarly, in the year 2019, over 4,00,000 devices were exploited by a botnet attack through many applications that functioned using online streaming. However, due to naïve implementation of the system, it is possible that all the devices that are connected with face a major threat when there is an attack. In fact, a single weak IoT device is all it takes to manipulate a wider network to which it is connected. Antiviruses [3], firewalls and intrusion detection systems [4] prove to be inefficient when it comes to attacks on IoT. Moreover these traditional protection methods are more specific to particular protocols and systems and as such will not be able to protect against the complex security attacks. The proposed work introduces a data mining approach which is used to detect the misbehaviours and other attacks and provide appropriate security measures for the IoT device. This methodology is suitable for heterogeneous protocols and platforms giving a solution that can be used for unsupervised datasets. Data is collected in a passive manner and will not cause any additional overloads when information is added to the network. The main aim of the proposed methodology is to detect false alerts and also give a proper analysis of security measures with a contextual environment. Data pre-processing methods and data mining techniques form the base for the proposed work. First the states that characterize the IoT system are identified by data pre-processing. When there is concrete way to define these states, it is then possible to have a complete analysis of IoT system without disturbing its pre-defined protocols. The proposed work is used to find attacks of security by exploiting the behavioural model. The outputs are recorded with ProM library and are analysed using a database of experiments.

ISSN: 2582-1369 (online)
Submitted: 20.05.2020
Accepted: 01.07.2020
Published: 08.07.2020

2. Related Works

Though the aspect of using IoT devices has become a common agenda with growing interest, as their applications begin to expand, they become more vulnerable to major attacks. Many analytics methods infer the potential attacks based on the characterisation of data. In [5] the authors study the datasets generated from the smart cities which using many machine learning methodologies. Similarly in [6] authors have also designed methods to detect botnet behaviour based on analysis of IoT based devices that were attacked. The collected data is standardised as the first step followed by dividing it into two clusters [7]. The characterization aspect is done based on the clusters with highest data points. On the other hand, the cluster with smallest points is used for misbehaviour characterization [8]. Though this method will enable automation to a large level, it will not be able to give concrete conceptual data to aid security. Each and every data point is taken into consideration, without taking into account the previous states. In [9] the author has also experimented with the use of neural networks to address this criteria. The results obtained are complicated in nature and will not aid exploitation of security. Net models are generated by process mining method which has proven to have advantages in various fields, in comparison with other methods of mining. The output can be characterized for attempts at attacks and system failure based on the abnormalities of the sequence of the logs. It is also possible to collaborate these methods with machine learning for better results. In [10] the author proposes a PM based approach that can identify the patterns in the output and Machine Learning is used to assign the resources accordingly. Our proposed work exploits PM in order to secure [11] the IoT system which are limited with respect to resources and are heterogeneous and distributed in nature [12].

3. Proposed Method

The architecture of the proposed method is represented in Fig.1. There are three main blocks of the mining: a misbehaviour detection block, model building block and a data pre-processing block. Here the misbehaviour and model building blocks are dependent on the data pre-processing block. The input to the system is the runtime monitoring datasets as well as the training datasets fed from the previous block. The first step of the process would be to process the raw data and transform it into a usable refined data with the help of the data pre-processing block. This data is then transformed into behavioural models with the help of process mining algorithms. Based on the observed system, the petri nets are used to represent the behavioural models, resulting in two independent and disjoint sets. The first set represents the system's states while the next set represents the event that brings about a change in the state. Boxes and Circles are used to represent the transitions and places.

ISSN: 2582-1369 (online)
Submitted: 20.05.2020
Accepted: 01.07.2020
Published: 08.07.2020

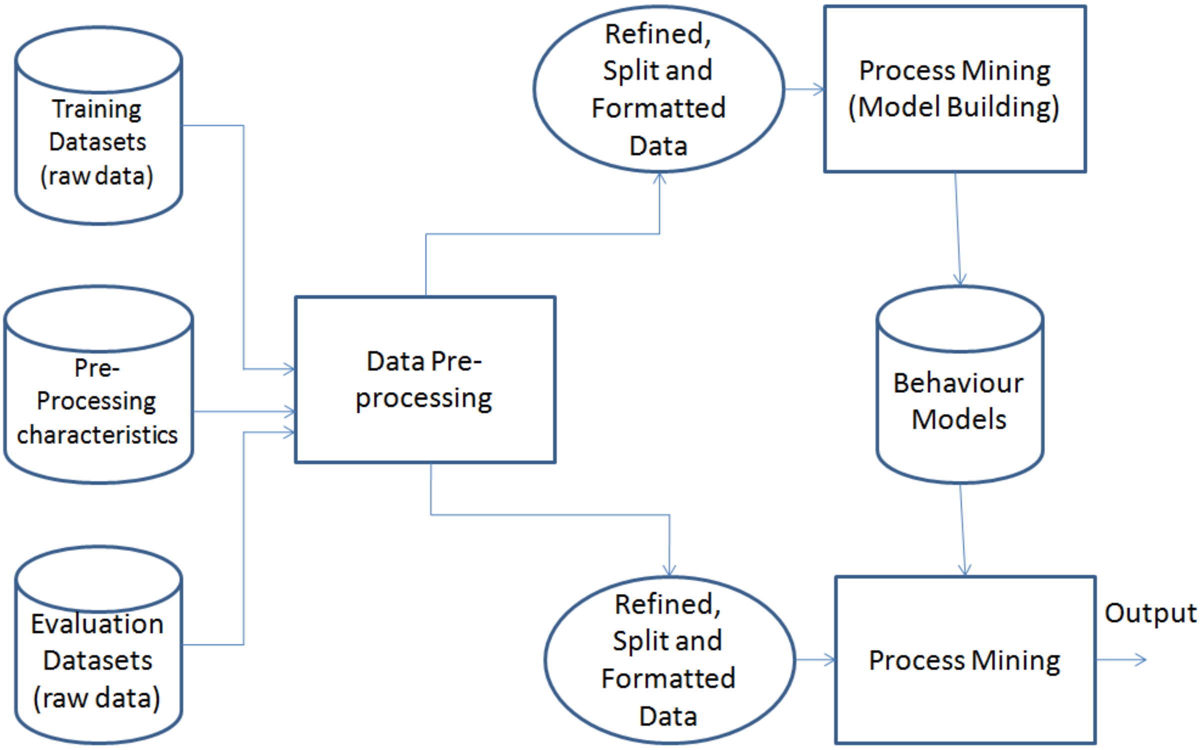
Training Datasets (raw data)
Pre-Processing characteristics
Evaluation Datasets (raw data)
Data Pre-processing
Refined, Split and Formatted Data
Process Mining (Model Building)
Behaviour Models
Refined, Split and Formatted Data
Process Mining
Output

Fig.1. Process Mining Architecture

3.1 Building of behavioral models

From the raw data, behavioural models are generated in the model building phase and this output is used as the refined data for the system. The proposed system is defined in such a way that it can adapt to the heterogeneity of platforms and protocols used by a typical IoT system. The phase initiates with the transformation of the raw data with the help of data preprocessing such that it can be used bythe process mining algorithm. Data pre-processing block and Process Mining block are the two crucial blocks of this category.

3.1.1. Data Pre-processing block

Data normalization, clustering and splitting together form the composition of data pre-processing block. This block gives freedom to observe and analyse the various states of the system. Each state is denoted by a ‘tuple’ of features such that in order for the tuples to belong to the same state, they must be equal. But it was found that this methodology was not apt enough to handle non-boolean and non-categorical features. Taking this into consideration, this data can be used directly to generate refined datasets while the other data can be process using clustering sub blocks and data normalization. At this junction, the data is re-scaled and integrated such that it can be compared.

ISSN: 2582-1369 (online)
Submitted: 20.05.2020
Accepted: 01.07.2020
Published: 08.07.2020

- **Data normalization sub-block:** In this block, the different features of the datasets are re-scaled and integrated in order to ensure that they can be compared in the steps that follow. Based on the collected data, predictive security is analyzed for the given datasets. This can be expanded to showcase the distance between the data points. But this type of quantification will not be enough to normalize or scale the data in a proper and efficient manner. Hence the normalized parameters are saved in the database of the output such that they can be further used during the detection phase, ensuring that consistency is maintained.
- **Clustering Sub-block:** In this block the data processed is first classified into cluster based on the system state. Clustering is a common methodology employed in the field of mining and can be used to decrease the total states that are used in the IoT system. Every tuple will be linked to a single cluster that will serve as the head. During detection phase, barycenter within the cluster, distance between normalized data and the barycenters are stored and exploited.
- **Splitting block:** The traces of the clustering sub-block is split into n data subsets such that each subset corresponds to a particular duration of time. This type of process will aid in decreasing behavioural complexity of the system and also stops high characterization which might prevent detection attacks and misbehaviours.

3.1.2 Process Mining Block

Various subsets of the process mining block is used to develop the behavioural model. This is later used to find the deviations in the detection phase. A IoT system is said to be misbehaving when the models built with the subsets do not come close to the runtime data. In this work, we have taken into consideration inductive mining algorithm which has the ability to support the behavioural model such that they are a perfect match to the logs. There are three steps involved in organizing this algorithm:

- For the refined data developed, a directly follow graph is formulated based on the logs of events.
- Based on this graph, a process tree is formed using enough operators to disseminate the graph.
- The events are further split along with the sublogs such that sub log S_1 holds the elements of the event E_1 while sub log S_2 holds the elements of the event E_2.
- Till the elements in the event set is reduced to 1, the cut is executed in all the sub-logs.

3.2 Detection of misbehaviors

Detection of misbehavior in an IoT system involves observing and analyzing the data gathered during the runtime operation and comparing it with that of the proposed behavioural model that is built based on the proposal. Let us first take into consideration the metric which is used to quantify the discrepancies from the proposed model.

- **Deviation Quantification:** The refined data is sent to the next block which quantifies the misalignment on comparison with the behavioral model. This process will help to determine the performance of the model and will also point out potential misbehaviours in the system. In this proposed work, we have used a refined dataset and behavior model to determine the output. The first step in this process would be synchronization of the events. This is possible when the movements of the dataset match that of the behavioural model. When log and model are not synchronized, the movement cost is 1 and if they are synchronized, then the movement cost is 1.
- **Detection Mechanism:** The output of data pre-processing is sent to this block which is typically used to create sub-logs that hold information on state identifiers and timestamps. A constraint is imposed on the

ISSN: 2582-1369 (online)
Submitted: 20.05.2020
Accepted: 01.07.2020
Published: 08.07.2020

sub-blocks to confine the mechanism through normalization such that the distance between data point and cluster shouldn't be more than the distance maintained at the building phase.

4. Results and Discussion

Based on the observation, we have taken into consideration different time splitting, normalization and clustering (birch, k-means). By changing the k-means clustering k parameter between the numbers 0.5 and 9.5 we can choose the value of epsilon to be between 0 and 1 while the birch value can be between 2 and 999.

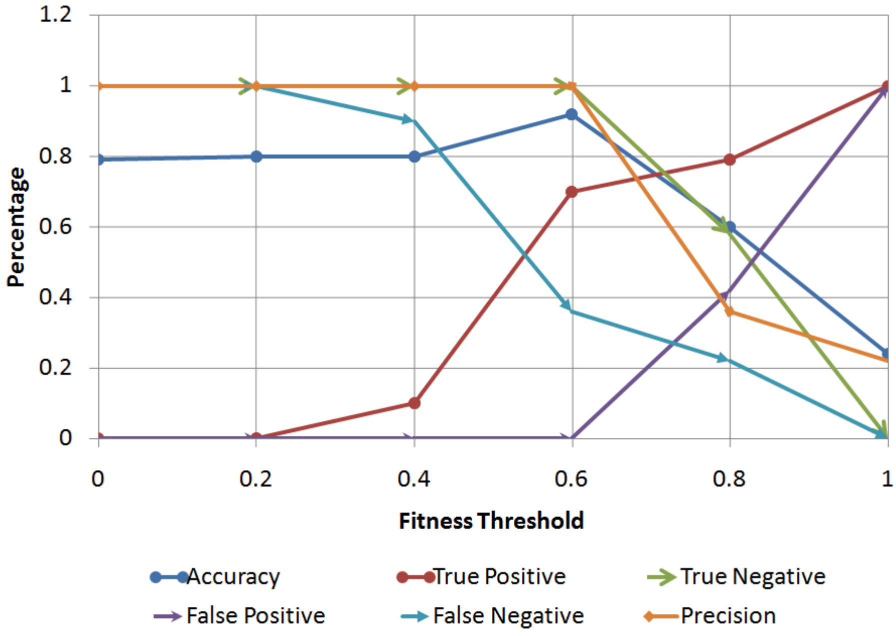

Fig.2 Detection Performance for Timesplit (25 seconds), Normalization, Clustering (k=8)

The normalization techniques that we have used in this proposed work will not need specific parameterization. Time splitting when parameterized in a likewise pattern, the results obtained are very accurate during the misbehaviour detection and model building phases. Based on the observation and experimentation, we have found that at k=8 of k-means clustering, it is possible to obtain the best detection performances as shown in Fig.2. Based on the performance metrics, we can see that the detection threshold can be observed between 0 and 1. Similarly Fig.3 depicts how IoT security can be characterized using the ROC curve.

ISSN: 2582-1369 (online)
Submitted: 20.05.2020
Accepted: 01.07.2020
Published: 08.07.2020

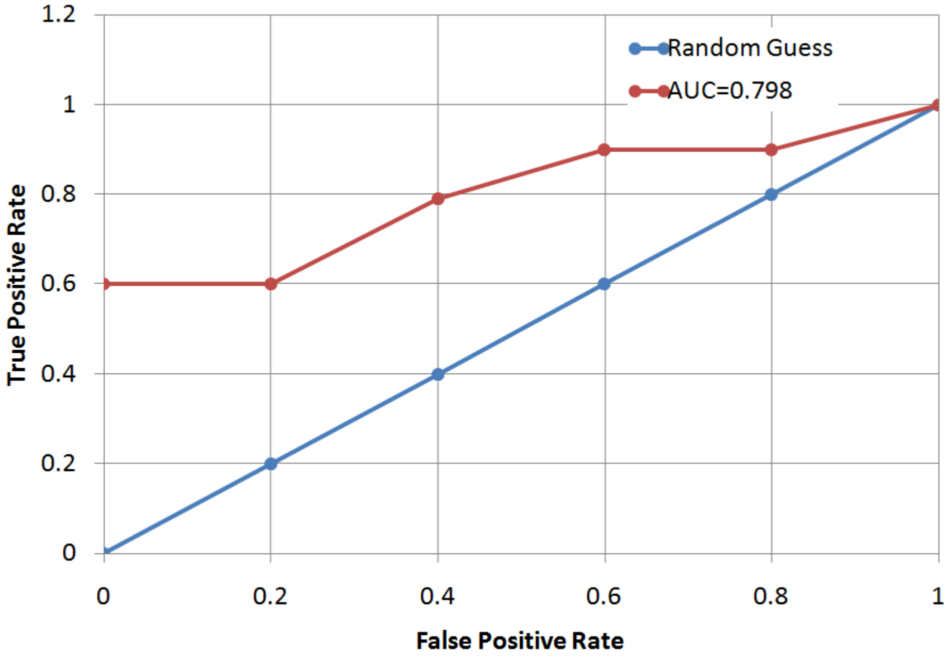

Fig.3. ROC for Timesplit (25 seconds), Normalization, Clustering (k=8)

5. Conclusion

One of the crucial challenges in security is the means of protecting the IoT devices despite the limited resources that are available using their characterization through heterogeneous platforms and protocols. To overcome this challenge, the proposed work explores the use of process mining approach which can manage many protocols and devices in order to support the security force of the IoT system. This technique is combined with data pre-processing methodology which can be used to characterize the IoT based system. Similarly behavioural models are developed with the aid of process mining such that they are adaptable to the heterogeneity of the devices as well as the protocols. They operate in such a way that they can observe the data at run time and identify potential attacks as well as misbehaviours. This paper presents a prototype which gives a solution to enhancing security of the IoT devices by exploiting ProM library. An elaborate experimentation is carried out and the observed output is recorded to evaluate the performance of our methodology.

References

[1] Duraipandian, M. (2019). Performance Evaluation of Routing Algorithm for Manet Based on the Machine.

[2] K. Delaney and E. Levy, “Connected Futures Cisco Research : IoT Value : Challenges, Breakthroughs, and Best Practices.” Cisco System Report, May 2017.

[3] M. Zaman and C. Lung, “Evaluation of Machine Learning Techniques for Network Intrusion Detection,” in Proceedings of the IEEE/IFIP International Network Operations and Management Symposium (NOMS 2018), April 2018, pp. 1–5.

ISSN: 2582-1369 (online)
Submitted: 20.05.2020
Accepted: 01.07.2020
Published: 08.07.2020

[4] M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis et al., "Understanding the Mirai Botnet," in Proceedings of the USENIX Security Symposium, 2017, pp. 1092–1110.

[5] Shakya, S. (2020). Performance Analysis of Wind Turbine Monitoring Mechanism Using Integrated Classification and Optimization Techniques. Journal of Artificial Intelligence, 2(01), 31-41.

[6] E. Bertino and N. Islam, "Botnets and Internet of Things Security," Computer, vol. 50, no. 02, pp. 76–79, feb 2017.

[7] He, Z., Wu, Q., Wen, L., & Fu, G. (2019). A process mining approach to improve emergency rescue processes of fatal gas explosion accidents in Chinese coal mines. *Safety science*, *111*, 154-166.

[8] C. Kolias, G. Kambourakis, A. Stavrou, and J. Voas, "DDoS in the IoT: Mirai and Other Botnets," Computer, vol. 50, no. 7, pp. 80–84, 2017.

[9] Haoxiang, W., & Smys, S. (2020). Soft Computing Strategies for Optimized Route Selection in Wireless Sensor Network. Journal of Soft Computing Paradigm (JSCP), 2(01), 1-12.

[10] Ghasemi, M., & Amyot, D. (2020). From event logs to goals: a systematic literature review of goal-oriented process mining. Requirements Engineering, 25(1), 67-93.

[11] Adithya, M., Scholar, P. G., & Shanthini, B. (2020). Security Analysis and Preserving Block-Level Data DE-duplication in Cloud Storage Services. *Journal of trends in Computer Science and Smart technology (TCSST)*, *2*(02), 120-126.

[12] L. Rouch, J. François, F. Beck, and A. Lahmadi, "A Universal Controller to Take Over a Z-Wave Network," in Proceedings of Black Hat Europe, 2017, pp. 1–9.

ISSN: 2582-1369 (online)
Submitted: 20.05.2020
Accepted: 01.07.2020
Published: 08.07.2020