Future Generation Computer Systems 74 (2017) 417–429

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

CLAPP: A self constructing feature clustering approach for anomaly detection

Gunupudi Rajesh Kumar[a], Nimmala Mangathayaru[a,*], Gugulothu Narsimha[b], Gali Suresh Reddy[a]

[a] Faculty of Information Technology, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India

[b] Computer Science and Engineering, JNTUH College of Engineering Jagtial, Karimnagar, India

ARTICLE INFO

Article history:
Received 4 August 2016
Received in revised form
11 December 2016
Accepted 29 December 2016
Available online 4 January 2017

Keywords:
Anomaly
Intrusion
Outliers
Web of Things
Classifier

ABSTRACT

The term internet of things is a buzz word these days and as per Google survey conducted recently, it has even dominated the buzz word big data predominantly. However, IoT area is still not matured and is throwing light on lot of research issues towards the data mining researchers. Security in IoT throws several challenges because of limited resources. In this context, IoT gains importance once again from data miners towards anomaly mining or intrusion detection. Intrusion detection is classified as NP-class in the literature even today. Algorithms addressing privacy and security issues in IoT must consider the complexities involved and hence require re-attention from all researchers. One more problem faced when judging for intrusion is the use of high dimensionality, classifier choice, and distance measure. For example, the traditional distance measure, such as Euclidean misjudges the similarity. In this paper, the objective is to design a fuzzy membership function to address both dimensionality and anomaly mining so as reduce the computational complexity and increase computational accuracies of classifier algorithms. We validate the proposed measure using several experimentations on NSL-KDD and DARPA datasets using kNN, J48 and CANN using Gaussian measure. Improved accuracies of classifiers on U2R and R2L attacks have been recorded in the experimental results obtained for experiments conducted.

1. Introduction

It is becoming a significant movement that Internet of Things (IoT) shows the use of internet is growing on regular basis in everyday life. Consumption of news, maintenance of social communication and shopping pattern has changed drastically with internet services reaching almost everyone. As a result, internet occupies a predominant role in everyone's life and will be more in coming years. Consumers are, however not ready to use applications which are quantified to measure the health status and other medical data. The reason for this is that the consumers need not put effort for all these purposes with the introduction of the new types of the devices which are very much negligible in price, small in size and more over capable of connecting to internet. There are still some problems existing which are considered as a hurdle for IoT development, but this development will not stop because of their benefits. Internet has become a basic requirement for every day life. The facilities and services that are available through the internet is the primary reason for its gradual growth [1].

1.1. IoT and big data

Big data and Internet of Things (IoT) becomes the headlines of the press and occupies the top priority position in every organization. The combination of both technologies big data and IoT, is undoubtedly an excellent combination with innovative outcomes and vast scope for the business impact. Across the globe from industrial sensors to the sensors in WBAN that monitors health, from the vehicular sensors to the home automation enormous collection of sensors connect to the internet and share their data to provide useful information. In the mean while the financial requirement for storing data has gone drastically down and analysis made huge progress.

* Correspondence to: Department of Information Technology VNR Vignana Jyothi Institute of Engineering and Technology (AUTONOMOUS), Bachupally, Nizampet (SO), Hyderabad-500090, India.

E-mail addresses: gunupudirajesh@gmail.com (G. Rajesh Kumar), manga.surya@gmail.com, mangathayaru_n@vnrvjiet.in (N. Mangathayaru), narsimha06@gmail.com (G. Narsimha), gali.sureshreddy@gmail.com (G. Suresh Reddy).

http://dx.doi.org/10.1016/j.future.2016.12.040
0167-739X/© 2017 Elsevier B.V. All rights reserved.

It is customary that any technology brings new challenges, and projects in IoT are not an exception to the list. 96% of the IoT stakeholders report that they are facing new challenges every day. 58% of the challenges are generated around business and business policies and 51% of the population are adapting to new technologies. It is more appreciable that many of the participants take time to document other challenges they are facing on regular basis. Selection of appropriate platforms, identification of threats, change of pace, lack of proper understanding and stability, absence of global standards along with internal and external stakeholders including vendors form the population for this survey.

According to the IoT meet Big Data Analytics Survey [2] report, some of the interesting findings are discussed below.

1. **IoT projects face many challenges**: IoT projects are still in early stages, and many projects do not have hard measurements to track success. One third (33%) reported that they will track their success using quantifiable metrics. Slightly fewer (29%) do have documented goals for success, but these goals cannot be quantified. The most common IoT project evaluation criteria reported by participants is merely to gain experience (38%).
2. **Data is captured, but not used fully:**The majority of IoT projects do involve data collection, but very few are fully leveraging the opportunity that data provides. Only 17% of survey participants indicated that they do not capture data as part of their IoT projects. It has been observed that 83%, of people which is a major group are only collecting data, out of which only (8%) of the people reports very less usage of the data by capturing and analyzing data in a regular fashion. More than (58%) of the group are making an effort and are doing some techniques of analytics even though they know that, they can do better.
3. **Challenges are faced at all stages of IoT data collection and analysis**: Most IoT stakeholders, 94%, report challenges related to data capture and analysis. Any data project includes multiple steps: capturing the data, analyzing the data, and acting on that analysis. IoT projects are not being an exception. Other challenges with data collection and analysis reported by participants were varied. Issues included technical challenges such as dealing with unexpected data, data that is out of spec, data that is not in a usable format, and correlating with data from other sources. Some process centric challenges reported were including training users on how to use the data, dealing with information that belongs to siloes organizations, and developing new business models to leverage analysis.
4. **Excessive amount for providing Ease of use exceeds cost**: Surprisingly, the cost is always a limiting factor in many technology decisions, but especially for stakeholders of IoT, this ease of use appears to be a more demanding issue than cost. 76% of the participants say that it would be better to collect and save more data if it is easier than those, who always claim that they would if it available for free (68%).
5. **IoT data capture should be better will be beneficial**: stakeholders of IoT do always think that if data capture is easy, then definitely there would be a positive impact. 92% of the people that is a majority group claims that they would have been benefited a lot if the data collection and capture is faster and effective, so that the decision making process becomes better reaching expected benefits (70%).
6. **Use of better techniques for analytics would increase Return on Investment (ROI) of IoT projects**: If there is increase in techniques that are applied over IoT, obviously the data performance and accuracy will improve, in turn the RoI of the organization will improve [2].

2. Literature survey

With IoT, it is becoming possible to be virtually connected to the entire world. Smart environments are becoming real and possible which however are also not free from security and vulnerability threats. IoT makes it possible to have smart home environments and these could be well compromised by performing simple network traffic analysis [3]. A smart integrated solution is proposed which considers measuring normal traffic patterns generated by COTS smart devices to detect the vulnerabilities possible.

Authors in [4] discuss a future direction for the security and privacy of Internet of Things. The research [5] focuses on security considerations for IoT using perspectives in the context of IoT proliferation. With the rise of IoT, intelligent systems evolved and more powerful and efficient software for these intelligent systems are being developed. The complexities are also increasing with these systems.

Rivera [6] in his paper proposed a schema that unifies access control systems between IoT devices based intelligent systems and hybrid systems. A non-secure hardware makes the software built upon it non-secure. This happens when we fail to identify the vulnerabilities in hardware frameworks. With shortage of time or other marketing reasons the finished wearable devices are being released into the market without proper testing and validations activities.

IoT design flow practices with few case studies are discussed in [7] to ensure security and privacy. A novel security architecture of internet of things (IoT), proposed in [8] is based on the concept of software defined networking (SDN). They use the network access control global traffic monitoring to design this novel security architecture of internet of things.

In order to facilitate the research in the field of IoT, researchers need powerful mechanisms to embed and test a different set of privacy technologies. A new test bed for privacy research in IoT is provided by Sharad in [9].

A detailed study of general architecture, protocols, security, privacy issues, applications and extensive survey in context of IoT is discussed in [10]. The impact of real world implementation using arduino and various trends is also discussed.

Principles of artificial immune system [11] are applied to detect threats of security breach in IoT environment by defining libraries of attack information. A framework to detect attacks in IoT and smart environment is proposed in [11] where the attacks found and identified are added to information libraries incrementally and these are used to throw alarm in case of any security breach.

IoT is enabled with RFID, communication technologies, smart sensors and various internet protocols. Current structure of IoT is to have the smart sensors around us combined with human involvement generates intelligent decisions. The first phase of IoT includes internet revolution, mobile, machine to machine (M2M) technologies. In the second phase i.e. in near future the IoT is going to expand to multiple diversified field of technologies to bridge technological gaps so as to enable rapid decision making [12]. IoT is a combination of Identification, Sensing, Communication, Computation, Services and Semantics. Fuqaha [13] describes, IOT as a combination of horizontal markets which covers different domains and vertical markets which covers deeper domain level specifications and protocols.

Daisy [14] proposes a novel IP/MPLS (Internet protocol/ Multiprotocol label Switching) core which enables integration of cloud services and underlying devices that uses multiple protocols. Authors use Elliptic Curve Cryptography (ECC) to provide complete security and also to ensure integrity, confidentiality, privacy, and authentication, tested the framework with a test bed and published promised results. Even though the overall simulation is not yet done but the results were encouraging with this novel approach as it shows new directions to IoT researchers.