IADIS International Journal on Computer Science and Information Systems
Vol. 12, No. 1, pp. 75-87
ISSN: 1646-3692

PRIVACY AND DATA SECURITY IN INTERNET OF THINGS

Arun Nagaraja[1] and N. Rajasekhar[2]
[1]Faculty of Information Technology, VNR VJIET, Hyderabad-500090, India
[2]Department of Computer Science and Engg, Institute of Aeronautical Engineering, Hyderabad-500043, India

ABSTRACT

Privacy is the major concern in the present world today. Data is also playing the major role and how the data is secured in the network is the trivial task. When the data is transmitted in the network, it is initially encrypted and the information is secured with minimal crypto security features. The information is more secured with public and private keys and during retrieval the data is to be decrypted with the same. After gathering and transmitting the required information, how to provide privacy to the data in the network is concern. Data that is shared in the network are to be made private and more secured. By privacy we can preserve the data in any channel and transmit it with security and its standards. The paper discusses on how the recent trend is working on data and its security. The paper also tells about the various cryptographic algorithm used during data transmission.

KEYWORDS

Transmission, Privacy, Encrypt, Decrypt, Cryptography, Preserving, Data, Security

1. INTRODUCTION

Data mining is the emerging field in the present world with various features. Whenever the information is to be transmitted, the data is available. Large amount of data is gathered to gain the knowledge about various domains. The information is spread in the wide area network, such that every research area requires data. Present research is working on various fields where security is also becoming a key concern. Among the different emerging technologies, technologies like machine learning and deep learning are having the requirement of security and its features.

We generally receive coupons or offers from different stores on the products that we are planning to buy. Companies follow this method of data mining which is not a coincidence. One of the best examples is supermarkets, as a part of the Data mining program, the company developed rules to predict what the shoppers will buy in the future, by looking at the content of their customers shopping basket. This is the most famous technique that is globally being used. The used of Data mining is not solely reserved for corporate applications, it goes beyond that. Crime agencies use data mining to identify which areas are more prone for crime. The search based on time of occurrence, and based on their previous records. Now from all this we can define Data mining in our terminology as a process of extracting knowledge from massive volumes of data based on our requirements. The most common definition which has been stated is, Data mining is the non-trivial extraction of implicit, previously unseen, and potentially useful information using a large amount of data.

Through data mining, we can perform Characterization and Discrimination, Association, Classification and Prediction, Clustering, Outlier Analysis and other patterns for preserving data. In general, Data mining and KDD (Knowledge Discover from Data) are considered as synonyms by most people and are interchangeable. The stages in KDD process are as mentioned: Data Cleaning/Preprocessing and Data Integration Stage Data Selection and Data Transformation Stage Data Mining/Discovery Stage Data Analysis and Data Visualization Stage.

Rapid Miner is the open source system for data mining; it is a stand-alone application for data analysis. Rapid Analytics is built around Rapid Miner for analytical analysis. Weka is a collection of different machine learning algorithms for each task. PSPP is a program for statistical analysis of sample data, which has a GUI and a conventional Command-Line Interface. There are few others like KNIME, Orange, Apache Mahout, jHepWork, Rattle and many more.

The information gathered will be either processed or pre- processed. Depending on the user requirement on processing the data, the information is gathered in the form of matrix. When the rules are to be applied for the data, the data will associate with many techniques. The key factor in applying the techniques is on how to provide security and how to handle various issues related to the query.

The major problem in handling the data is security with various algorithms. Many algorithms like RSA or Diffie Hell- man are having many advantages in handling the data with more advanced security features. But the problem with the system is on how to identify the threats occurring with data. The threats that enter the data are in many forms, either in ".exe format or with system files", etc. This says data will be more pruned to obtain vulnerabilities in the network. This represents security is a major concern to handle data in different formats and how securely we can transmit in the communication channel is the task.

The attacker can target the data when data is in the communication medium. The data publisher does not require the knowledge of data mining. He can directly publish the data which may tend to many security issues. The data publisher can pass the information to various recipients having different security features for individuals. This helps the user to identify how information can be processed and how it can be secured and retrieved. The study tells about many security issues and its handling techniques (Buket Yksel, Alptekin Kp & Znur Zkasap, 2017).

This paper talks about various methodologies involved in making data secure and how to handle it without any vulnerability. Most of the data which is involved in data security is Dynamic data. The study also talks about moving of data and performing various transactions to make it much secure. Section 2 talks about the related work, Section 3 talks about the recent trends of data security and the final section conclude the paper.

2. RELATED STUDY

Data is the most happening thing which is used in wide areas. Any fields of research concentrate on various varieties of data, which is used either as raw form or in the processed form. When a query is passed in the secured channel, initially the data is identified with its format and then analyzed with association rules. The data is handling very delicate for its possible attributed to process the information. Data is stored either in the database or in the cloud.

Data which is stored in the database is handled with many rules and techniques and also security is also provided. Data is encrypted before processing. Various security features are added to the data to make it more secured. The data is not in the combined format in the database. To provide effective security features, the similar data is clustered together based on its size.

Stored data in the database are handled by different users. When various users transmit data to store in the location, the information is gathered together security features are provided. Before the data is transmitted to the destination, Data receiver will identify the type of transmission of data. If the data is static, only minimum association rules are applied. Most of the times, the data is dynamic. When it is dynamic, the dimensions of the data are identified.

Various security features are used for different forms of data. During transmitting from one channel to another, privacy is the key concern. Raw data is always less secured and possibility of intrusion of the data is more. The data is represented in the form of trees. When the security features are to be applied to the database, initially transmission will happen to the right leaf and then moves to the root data and then to the right leaf. Data that is not secured does not have more advanced features of transmission.

Privacy is playing the key role which is managing many security services. Before the data is made secured, low level encryptions and decryption is provided. To provide encryption and decryption various public and private keys are required. This says, information is secured from level one and then it is also made effective. Data sets are identified initially and then the attributes are also identified to provide levels of security. Depending on the attribute property, the information is gathered and security is also provided. The Composition of many rules and protocols are mapped together to fetch the information properly.

IoT is also having major issues in preserving the data. The Internet of Things (IoT) is nowadays becoming a buzzword everywhere. IoT covers all the devices but not limited to smart phones, wireless sensor nodes, embedded devices, WBAN equipment, VANET related devices and other communicating equipment such as routers, switches etc. Mistely Blowers, in his paper defines IoT is a collection of large number of sensors, nodes, devices communicating each other, coordinating them and sharing the data-analytics and actions. It is important to note that all these devices are of heterogeneous groups.

The property of IoT makes it easier to get imbibed into multiple fields like space, traffic-monitoring, health monitoring, medical research, pharma research, education, industrial monitoring and manufacturing, automation, security and surveillance, automobile, surveying, construction, robotics, control systems, information &technology, etc., As IoT has entered every domain, several complexities need to be addressed without fail. With collaboration of such different fields, in order to generate valid results, IoT devices have to communicate each other. This in turn becomes a complex task and need to be addressed by researchers. Each device has its ownformat, protocols and standards of data. It is difficult to establish the communication link between them without having a proper integration mechanism.

Many researchers are trying to classify the IoT devices based on different criteria, out of which recent classification by Alan Aronoff (2015) in his paper discussed, which is based on the types of data is classified as:

1. M2M data and smart sensor: consisting of data between mobile to mobile communications and other smart sensors.

2. Audio/Video Connected: consists of data in different formats of audio and video in different applications.

3. Different types of analytics, automotive: consisting of data that is being generated and used for analytics and automotive areas.

4. Computing Nodes in high-performance environment: consisting of data that is used in communication process among high performance computing systems.

Attempts made by Matt Web to classify IoT devices based on the data as wearable’s (includes sensors includes WBAN), the media (includes movies and music), automation (includes private and public such as traffic) and smart appliances (own network with tightly coupled services). This paper discusses on various challenges in IoT proposed by different researchers. The main intention of the author is that his contribution will be helpful for future researchers.

Before preserving the data, it should be encrypted such that levels of security can be provided. Various protocols can be used for different levels of data extraction and data security. This leads to more advanced security features to protect the data.

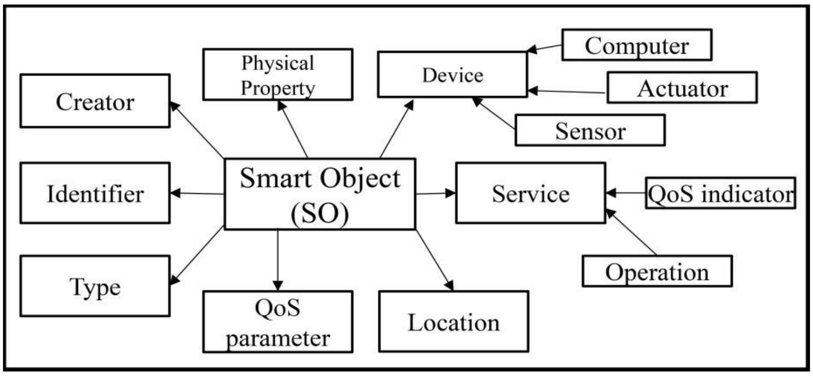

Figure 1. Metadata Associated Smart Objects used for Classification of IoT Devices

Giancarlo Fortino, Anna Rovella, Wilma Russo & Claudio Savaglio (2014) defines IoT as a collection of smart objects like autonomous, sensing and actuating capacity, processing networking and communicating capabilities. Figure 1 describes the metadata used for classification of IoT devices. It includes the creation of smart objects, devices, services and QoS parameters. The authors (Giancarlo Fortino, Anna Rovella, Wilma Russo & Claudio Savaglio, 2014) in their paper proposed a smart object description document (SODD) and profile description document (PDD) used for analysis of the IoT devices. An author (V. J. Jincy & Sudharsan Sundararajan, 2015) modifies the classification criteria as described in figure 2.

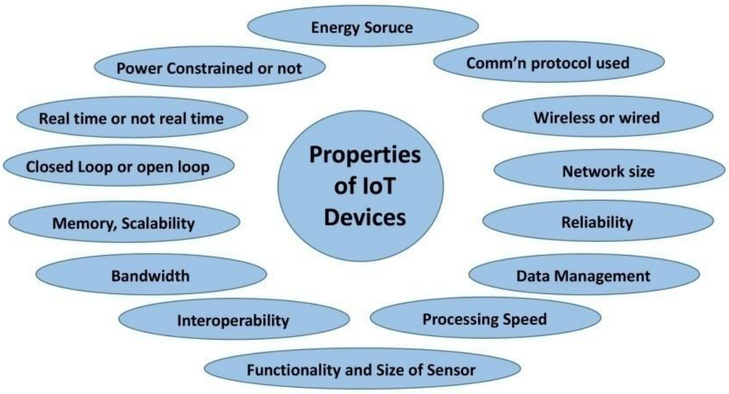

Figure 2. Modified Properties that are Used to Classify the IoT Devices

Many authors define IoT as devices that can be used anywhere any place, any device anything, any context anytime, anybody anytime, any business any service, in any network any path. In a paper the author explains the process of evolution of IoT how the transition from a network to IoT through the internet, mobile internet and people with the computer and mobile devices. The author also proposes the classification based on the type of node and system parameters. Various applications for IoT are as follows:

1. Industrial systems:e.g: laser distance meters, thermography camera modules.
2. Fire detection systems: fire detection sensor node, sprinkler node.
3. Home automation: motion detectors, smart meters, smart bulbs and switches
4. Health care: Pacemakers, WBAN systems, RFID, GSM, GIS based health monitoring tools.
5. Miscellaneous Monitoring Tools : Air traffic control node, earthquake-detection, weather monitoring, RFID smart conveyer belt, traffic control node, smart car sensors, sensor nodes for irrigation systems, personal safety systems, traffic surveillance equipment, tsunami detection nodes, etc.,

One of the key issues for the data mining issue is on social networking, but not on business or technology. It is a huge threat of providing security or individual privacy, as it gives us facts that are not obvious to human analysts of the data (E. Pavlov, J.S. Rosenschein & Z. Topol, 2004), (Yehuda Lindell & Benny Pinkas, 2000). The field of data mining is having significance to identify huge amounts of data, which are easily collected and stored with the help of computer systems. This large amount of data, gathered from various channels, contains much personal information. Consumers have become more assertive in demanding that their personal information be protected. The possible threats are prediction of information about the classified work from the correlation with unclassified works. Consider the example of budget estimation or staffing, where there was an increase in the number staff which was not updated and or where the records of the new staff members are not present. When predicting some information about the budget then it can lead to incorrect predictions. Another threat is detecting hidden information based on insufficient information. It is not only data which is to be protected but also the correlations among the data items.

We can use data mining to handle the security issues. Data mining tools can be used to examine for auditing and to recognize uneven behavior. Tools are being verified as a means to determine abnormal patterns and also to determine the type of problem (Filippo Maria Bianchi, Antonello Rizzi, Alireza Sadeghian & Corrado Mois, 2016). Some works in intrusion detection, are neural networks to detect abnormal patterns or SRI workingon Intrusion Detection Expert Systems which consists of a statistical subsystem that observes behavior on the computer system. If it is deviated significantly from the expected behavior then it is identified that, there has been an intrusion that has been taken place. It can be drawn heavily on fraud detection, where major work is going on in the telecommunication industries. There are different types of privacy preserving techniques such as Heuristic-based techniques; Cryptographic-based Techniques, and Reconstruction-based Techniques (Benjamin C. M. Fung, Ke Wang, Rui Chen & Philip S. Yu, 2010). The known information is that, to compute the detection model for every target that is being processed.

The authors (Srgio Moro, Paulo Rita & Bernardo Vala, 2016), (Gunupudi Rajesh Kumar, Mangathayaru Nimmala & G. Narasimha, 2015) talk about data security is playing the key role in finding intrusion in data sets. The authors also discuss more on how to find the intrusions using system calls by identifying the similarity for various system calls. The Data that is transmitted in the network is always not secure. To provide security, many algorithms can be used at various levels, such that to protect the information during transmission.

Many algorithms like RSA, DES etc.; are used to perform encryption and decryption. Author (Gunupudi Rajesh Kumar et al., 2016) explains about providing software security and identifying cloud security features. Improving the security aspects to provide various platforms leads to higher efficiency in transmitting the data. Cloud data is always not secure. Providing security is a key aspect to perform typical encryption and decryption. The data is encrypted using public / private keys for respective users. As the data reaches the destination, the decryption process is carried out and the analysis is performed to identify whether data has reached with proper order. This can be noticed by indexing the data.

To identify the intrusion and detection in the database, many methods can be used. One such method is single scanning of the database (Abdullah El-Haj & Shadi Aljawarneh, 2015). By performing this method as discussed in the reference, the space complexity can be reduced by minimizing most occurrences of system call patterns.

The new concept in the current technology is with cloud computing. The cloud is created virtually and it is also maintained with high security features. Confidentiality, authentication and integrity is maintained along with the security features, such that any access to the cloud can be made only with high security. The author (Aljawarneh, S. A., & Yassein, M. O., 2016) also explains about the providing scalability, availability and automatic backup. Most of the security techniques are always used for banking sectors.

Banking sectors have moved to online facilities to make the users more feasible in accessing the transactions and payments. But the drawback in using online banking system is with attacks. The attacks such as cross site scripting, DOS attacks, Eaves dropping, masquerading etc., make any web transactions more vulnerable. The author (Aljawarneh, S. A., 2017), (Alhaj, A., & Aljawarneh, S. A., 2017) discusses more on how to resolves the security threats and also on how to protect the data in cloud. The author also proposes the mechanism on secure data transmission with preemption algorithm to resolves the issues between quality of service and many security problems.

Study tells that, the security that is being provided then the requirements of data mining concerning to security. Primarily, physical database integrity is related to the power failure of the system. When the power fails then the intermediate records are not posted or retrieved correctly. Due to which the data mining leads to anomaly of the results. Another is, logical database integrity, where the modification of one field or record will not affect any other fields which leads to the logical integrity anomalies. Element Integrity is a must, every element should maintain integrity only then there will be no chance for change by human mistakes or any other programs. Audibility is another requirement, where the modification of the records in the database is considered as OLAP application. The details of the previous records shall be recorded in a log file which ensures proper modification. Access Control is the important requirement in which the system has the capability for the access control which gives the privileges such as the administration rights to the user for different records or files in the database. Finally, authentication is the at most requirement for the security issues.

Authors (M. B. Yassen, S. Aljawaerneh & R. Abdulraziq, 2016), (E. Aljarrah, M. B. Yassein and S. Aljawarneh, 2016) and (M. B. Yassein et al., 2016) discussed on providing security aspects to the data with the help of clustering and classification algorithms. They have proposed the many security features for data transmission in wireless and lossy networks so as to provide high security and lossless transmission. With the help of clustering and classification similar data can be identified and merged together and transmitted with security aspects in the network.

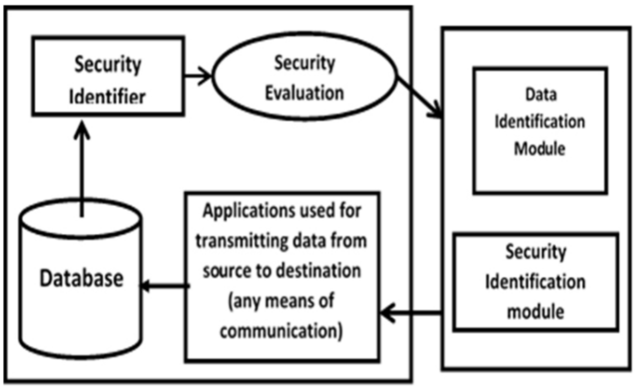
Security Identifier
Security Evaluation
Data Identification Module
Security Identification module
Database
Applications used for transmitting data from source to destination (any means of communication)

Figure 3. Database and Security Identification

3. STUDY ON DATA MINING AND SECURITY FEATURES

Figure3 talks about the database and security identification procedure where the database will identify the various applications which are accessing for the data. When the data is accessed from the application, initial the destination address is checked and what data to be accessed is identified. Once the data is identified, the index is maintained and identified based on the queries. As after indexing, data cannot be transmitted immediately. It has to be checked with, are there any security aspects present for the data. Once the security feature is identified, the evaluation procedure starts. Various cryptographic algorithms are used to provide data security. Security is the key concern in protecting the data and how to make the data transfer securely in any transmission channel is the proposed idea.

Security is the key role in any identification or processing of data in the network. How effectively the data can be preserved is based on the user requirement. If the user is having the confidential data, then security can be increased more by adding various encryption and decryption algorithms. Security algorithms like RSA, AES are more secure and provide high level security.

Most of the data transmission in the network is facing many security issues. Though the networks are less secure, the networks provide only WPA security with AES and TKIP encryption. To make the data secure enough, the wireless networks require providing WPA2 authentication with only TKIP keys. That makes the data secure in the wireless network.

When data moves from place to place, the possibility of occurring of attack is more. To prevent the attack from any worms or viruses, the privacy preserving algorithms to be used; such that data can be protected. Depending on the algorithm used, when the information is to be passed from place to place, the channel is to be protected. How to protect data and what encryption techniques are used to protect the data is the key concern. Any Encryption can be breached during transmission. This paper discusses survey, on how different encryptions are used, where are they used and how authors explained about the various security issues.

Novel Security algorithm and key management techniques are used to make the data more secure. Homomorphic encryption algorithm is used and also many key management algorithms are used to obtain more secured features. When the data is requested from the application, the database will look for the indexing. Depending on the index, the database will also look for the security identification. Once the security features are identified, the evaluation of data is performed. The security features are evaluated based onthe level ofauthentication and its effective utilization. The non-authenticated users who try to access the data cannot have access. The privacy is identified and how the data is prevented from security reasons will be the major concerns (Jaideep Vaidya & Chris Clifton, 2003), (Li, Qinghua & Cao Guohong, 2013).

The data mining concepts these days are routing towards the IOT concepts. Privacy and security is playing the key role. IoT explains that, data can be accessed anywhere and anyplace based on the user requirement. When the data is requested or transmitted from the source, the database will identify which module is concerned with and how the data is handled. Once the data is identified and user comes to know about the destination requirement, then the data can be processed with ease of transmission.

Many algorithms like, Homomorphism algorithm, RSA algorithm or EC Elgamal algorithms are used, when the encryption is to be provided. When these algorithms are looked upon, the EC Elgamal will provide the best efficiency with encrypting the data and giving more security to transmit the data from any source to the destination. Data is not stored in one level. It is stored in multi-level and based on this, when the data is accessed, it can be identified based on its weight. The weight of the data is calculated depending on the level of the tree. When the tree is to be accessed, the next root levels are identified and then the data is mapped based on data weight. The data level is depending on the nodes of the tree. When the root node is identified, based on the Binary search tree (BST) the data is placed at its position. The BST places the nodes those are having less weight than a root node to the left node and those having more weight than a root node, then that data is placed on the right side of the root. By using BST, data can be searched or identified easily.

Many algorithms like C4.5, K means, Apriori, etc., are used in data mining to perform classification. These algorithms are as known performs the best result in identification of the data and to predict which class the text belongs to. Based on different techniques, the data can be classified easily. The trendin IOT is, playing the major role in all the aspects. For example, when users need to access the house held devices remotely from being outside; this is possible with many sensors placed with internet connectivity. When data is requested from the application, the database identifies the authentication and based on that and access will be provided. There is a possibility that, once the intruder comes to know about any authentication keys, there is a possibility to get the data breach. This is the major study which the researchers are working on real time.

The algorithms which are used to perform authentication, must initially match with the requirements of data mining users. The information is then provided only when the user enters with proper authentication. To provide the unique authentication, the Fingerprint scanning can be made possible to perform high end security. By using IRIS chances of protecting the data is more. The data is becoming the major concern in all the fields. Without data no operations or tasks can be performed. Using K-means algorithm, the clustering of data is performed. Clustering is performed to identify similar items present in various groups. By using different clustering algorithms, data can be used with high levels of similarity and identification. This is possible only when there are similar data present in various groups of a database. Many algorithms are used to provide the clustering and classification. Clustering is performed on K clusters and also depending on the clustering members.

By performing clustering, the user will be able to identify data usage of other users and their methodology. This is possible only when there is a correct way of authentication and also with high security features. Encryption and decryption process is performed with random keys and private keys. This is shared by the database depending on the time of usage. Once the keys are shared, the keys will last only for one usage and again when the data are to be accessed, the new keys are to be fetched from the database. By doing this, security feature can be enhanced. Database is capable of identifying the security level for DB and how to use the data. Once the cipher texts are generated at the senders end with the private keys, it is said that, information is made secure and can be accessed only with proper authentication (Xun Yi & Yanchun Zhang, 2012).

While deciphering the text, many clusters will be present and identification of valid cluster to process the data securely is the major task. This is possible when the information is retrieved from end to end with high end security features. The data is distributed and distributed association rule mining can be used to compare the plain text over the cipher texts

(X. Yi & Y. Zhang, 2007). Studies performed by V. Radhakrishna et al. (2016) applied temporal pattern mining for intrusion detection. Similarity measures proposed in temporal context that include V. Radhakrishna et al. (2016), V. Radhakrishna et al., (2015), Aljawarneh, S. A., Moftah, R. A., & Maatuk, A. M. (2016), Shadi A. Aljawarneh et al. (2017) may be applied for intrusion detection as there is a wide scope for research in this direction. M.S.B. Phridviraj, Vangipuram RadhaKrishna, Chintakindi Srinivas & C.V. GuruRao (2015); Chintakindi Srinivas, Vangipuram Radhakrishna & C. V. Guru Rao (2013); Chintakindi et al. (2014); Chintakindi et al. (2015) propose similarity measures for data streams and software component clustering which may be adopted for anomaly identification.

Clustering and classification are the one major goal in data mining techniques. Many methodologies are used and handling techniques are identified base on the K-clusters algorithm. Many algorithms are used to identify the structure of the data. Whether the data is placed horizontally or vertically, accessing of the data is much easier when the data is placed in the vertical direction. Because it defines the tree like structure and dropping to the position where data is present is easier. When the horizontal structure is used, the data needs to search the entire database till the element is found. The complexity will be increased and sorting of the data will take more time. Heap Sort can be used to perform the sorting, as this has the complexity of O (nlogn).

4. CONCLUSION

This paper discusses on many techniques based on data and security features used by different algorithms. The detailed survey performed tells about how effectively data can be transmitted, how the data can be analyzed based on its identity and how data mining techniques. have been applied. Many algorithms like RSA, Homomorphic, EC Elgamal algorithms are used to securethe data. EC Elgamal algorithm gives the major accuracy and good efficiency in handling the data. Security is the always a key concern and handling data much more securely without any hassles is another key concern. Today’s world is very addicted to the technology and the data gathered is also getting breached with unknown transmission of data from end to end. To avoid the data breach, security algorithms are used with the help of private keys to perform encryption and decryption at various stages so that the user can handle the data with less risk and achieve high performance in handling data. The complexity is also identified and thus information can be gathered end-to-end and privacy can be preserved by saving the data from being attack.

REFERENCES

Abdullah Alhaj, Shadi Aljawarneh, Shadi R. Masadeh, Evon M. O. AbuTaieh. (2013). A secure data transmission mechanism for cloud outsourced data.*International Journal of Cloud Applications and Computing,3*(1), 34-43.

Abdullah El-Haj,Shadi Aljawarneh. (2015). A mechanism for securing hybrid cloud outsourced data: securing hybrid cloud.*Advanced Research on Cloud Computing Design and Applications, IGI Global,* 73-83.doi:10.4018/978-1-4666-8676-2.ch006

Alan Aronoff. (2015, August). Imagination technologies limited, a white paper on IoT - Opportunities for device differentiation v1.0. Retrieved August 11, 2015, from https://imgtec.com/blog/what-is-the-internet-of-things/.

Alhaj, A., & Aljawarneh, S. A. (2017). An algorithm for securing hybrid cloud outsourced data in the banking sector. *IGI Global*, 157-171. doi:10.4018/978-1-5225-0864-9.ch010

Aljawarneh, S. A. (2017). Online banking security measures and data protection. *IGI Global*, 1-312. doi:10.4018/978-1-5225-0864-9

Aljawarneh, S. A., Moftah, R. A., & Maatuk, A. M.(2016). Investigations of automatic methods for detecting the polymorphic worms signatures. *Future Generation Computer Systems, 60*, 67-77.

Aljawarneh, S. A., & Yassein, M. O. (2016). A conceptual security framework for cloud computing issues. *International Journal of Intelligent Information Technologies, 12*(2), 12-24. doi:10.4018/IJIIT.2016040102

Asif Imran, Shadi Aljawarneh, Kazi Sakib. (2016). Web data amalgamation for security engineering: Digital forensic investigation of open source cloud.*Journal of Universal Computer Science,22(*4),494-520.

Benjamin C. M. Fung, Ke Wang, Rui Chen, and Philip S. Yu. (2010). Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys. 42*(4), Article 14. doi:http://dx.doi.org/10.1145/1749603.1749605.

Buket Yksel, Alptekin Kp, znur zkasap. (2017). Research issues for privacy and security of electronic health services. *Future Generation Computer Systems, 68,* 1-13. doi:10.1016/j.future.2016.08.011

Chintakindi Srinivas, Vangipuram Radhakrishna, C.V. Guru Rao. (2013). Clustering software components for program restructuring and component reuse using hybrid XOR similarity function. *AASRI Procedia, 4,* 319-328.

Chintakindi Srinivas, Vangipuram Radhakrishna, C.V. Guru Rao. (2014). Clustering software components for program restructuring and component reuse using hybrid XNOR similarity function.*Procedia Technology, 12,* 246-254.

Chintakindi Srinivas, Vangipuram Radhakrishna, C.V. Guru Rao.(2015). Software component clustering and classification using novel similarity measure.*Procedia Technology,19*, 866-873.

Chintakindi Srinivas, Vangipuram Radhakrishna, C.V. Guru Rao.(2014). Clustering and classification ofsoftware component for efficient component retrieval and building component reuse libraries. *Procedia Computer Science,31*, 1044-1050.

E. Pavlov, J.S. Rosenschein, Z. Topol. (2004). Supporting privacy in decentralized additive reputation systems.*Proceedings of the Second International Conference on Trust Management*, 108-119. doi:10.1007/978-3-540-24747-0_9

E. Aljarrah, M. B. Yassein and S. Aljawarneh. (2016).Routing protocol of low-power and lossy network: Survey and open issues.*Proceedings of 2016 International Conference on Engineering & MIS,* 1-6. doi:10.1109/ICEMIS.2016.7745304

Filippo Maria Bianchi, Antonello Rizzi, Alireza Sadeghian and Corrado Moiso.(2016).Identifying user habits through data mining on call data records. *Engineering Applications of Artificial Intelligence, 54*, 49-61. doi:10.1016/j.engappai.2016.05.007

Giancarlo Fortino, Anna Rovella, Wilma Russo and Claudio Savaglio. (2014). Including cyber physical smart objects into digital libraries.*Proceedings of International Conference on Internet and Distributed Computing Systems,* 147-158. doi: 10.1007/978-3-319-11692-1_13

Gunupudi Rajesh Kumar, Mangathayaru Nimmala, G. Narasimha. (2015). A novel similarity measure for intrusion detection using gaussian function. *Technical Journal of the Faculty of Engineering, 39(* 2). 173-183.

Gunupudi Rajesh Kumar, N. Mangathayaru, and G. Narasimha. (2016). Intrusion detection a text mining based approach. *Special issue on Computing Applications and Data Mining International Journal of Computer Science and Information Security, 14 (*Special Issue 1). 76-88.

Gunupudi Rajesh Kumar, N. Mangathayaru, and G. Narasimha. (2015). An improved k-means clustering algorithm for intrusion detection using Gaussian function. *Proceedings of the International Conference on Engineering & MIS, Article 69*. doi: 10.1145/2832987.2833082.

Gunupudi Rajesh Kumar, N. Mangathayaru, and G. Narasimha. (2016). Intrusion detection a text mining based approach. *Special issue on Computing Applications and Data Mining International Journal of Computer Science and Information Security, 14 (*Speical Issue1). 76-88.

Gunupudi Rajesh Kumar, Nimmala Mangathayaru, Gugulothu Narsimha. (2016). An approach for intrusion detection using novel Gaussian based kernel function.*Journal of Universal Computer Science, 22*(4). 589-604. doi:10.3217/jucs-022-04-0589

Jaideep Vaidya, Chris Clifton.(2003). Privacy-preserving k-means clustering over vertically partitioned data. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining,* 206-215. doi: http://dx.doi.org/10.1145/956750.956776

Jun Tang, Yong Cui, Qi Li, Kui Ren, Jiangchuan Liu, and Rajkumar Buyya. (2016). Ensuring security and privacy preservation for cloud data services. *ACM Computing Surveys, 49* (1). Article 13. 39. doi:10.1145/2906153

Li, Qinghua and Cao, Guohong. (2013, July). Efficient privacy-preserving stream aggregation in mobile sensing with low aggregation error privacy enhancing technologies. In Emiliano De Cristofaro, Matthew Wright, *Privacy Enhancing Technologies 13th International Symposium*. Symposium conducted at Bloomington, IN, USA.

M.S.B. Phridviraj, Vangipuram RadhaKrishna, Chintakindi Srinivas, C.V. GuruRao. (2015). A novel Gaussian based similarity measure for clustering customer transactions using transaction sequence vector. *Procedia Technology, 19*, 880-887. doi: http://dx.doi.org/10.1016/j.protcy.2015.02.126

M. B. Yassen, S. Aljawaerneh and R. Abdulraziq. (2016).Secure low energy adaptive clustering hierarchal based on internet of things for wireless sensor network (WSN): Survey.Proceedings of 2016 International Conference on Engineering & MIS, 1-9. doi:10.1109/ICEMIS.2016.7745310.

M. B. Yassein, S. Aljawarneh, E. Masa'deh, B. Ghaleb and R. Masa'deh. (2016).A new dynamic trickle algorithm for low power and lossy networks. *Proceedings of 2016 International Conference on Engineering & MIS,* 1-6. doi:10.1109/ICEMIS.2016.7745314

S. Aljawarneh, V. Radhakrishna, P. V. Kumar and V. Janaki. (2016). A similarity measure for temporal pattern discovery in time series data generated by IoT. *Proceedings of 2016 International Conference on Engineering & MIS*, 1-4. doi:10.1109/ICEMIS.2016.7745355

Shadi A Aljawarneh, Raja A Moftah, Abdelsalam M Maatuk. (2016). Investigations of automatic methods for detecting the polymorphic worms signatures.*Future Generation Computer Systems, 60*,67-77.

Shadi A. Aljawarneh, Vangipuram Radhakrishna, Puligadda Veereswara Kumar, Vinjamuri Janaki. (2017). G-SPAMINE: An approach to discover temporal association patterns and trends in internet of things. *Future Generation Computer Systems*.doi: http://dx.doi.org/10.1016/j.future.2017.01.013

Srgio Moro, Paulo Rita, Bernardo Vala. (2016). Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach. *Journal of Business Research, 69*(9), 3341-3351.doi: http://dx.doi.org/10.1016/j.jbusres.2016.02.010

V. J. Jincy , Sudharsan Sundararajan. (2015).Classification mechanism for IoT devices towards creating a security framework. *In: Buyya R., Thampi S. (eds) Intelligent Distributed Computing, 321.* [Advances in Intelligent Systems and Computing]. doi:10.1007/978-3-319-11227-5_23

V. Radhakrishna, P. V. Kumar and V. Janaki. (2016). Looking into the possibility of novel dissimilarity measure to discover similarity profiled temporal association patterns in IoT. *Proceedings 2016 International Conference on Engineering & MIS,* 1-5. doi:10.1109/ICEMIS.2016.7745353

V. Radhakrishna, P. V. Kumar and V. Janaki. (2016). Mining of outlier temporal patterns. *Proceedings of 2016* Int*ernational Conference on Engineering & MIS,* 1-6. doi:10.1109/ICEMIS.2016.7745343

V. Radhakrishna, P. V. Kumar, V. Janaki and S. Aljawarneh. (2016). A computationally efficient approach for temporal pattern mining in IoT. *Proceedings of 2016 International Conference on Engineering & MIS*,1-4. doi:10.1109/ICEMIS.2016.7745354

V. Radhakrishna, P. V. Kumar, V. Janaki and S. Aljawarneh. (2016). A similarity measure for outlier detection in time stamped temporal databases.*International Conference on Engineering & MIS*,1-5. doi:10.1109/ICEMIS.2016.7745347

V.Radhakrishna, Puligadda Veereswara Kumar, Vinjamuri Janaki.(2016). A novel similar temporal system call pattern mining for efficient intrusion detection. *Journal of Universal Computer Science, 22(*4), 475-493. doi:10.3217/jucs-022-04-0475

Vangipuram Radhakrishna, P. V. Kumar, V. Janaki.(2015). An approach for mining similarity profiled temporal association patterns using Gaussian based dissimilarity measure. *Proceedings of The International Conference on Engineering & MIS 2015*, Article 57.doi:http:10.1145/2832987.2833069

X. Yi, Y. Zhang. (2007). Privacy-preserving distributed association rule mining via semi-trusted mixer. *Data Knowledge and Engineering, 63*(2), 550-567. doi:http://dx.doi.org/10.1016/j.datak.2007.04.001

Xun Yi, Yanchun Zhang. (2012). Equally contributory privacy-preserving k- means clustering over vertically partitioned data. *Information Systems, 38(*1), 97-107. doi:10.1016/j.is.2012.06.001

Yehuda Lindell, Benny Pinkas. (2000). Privacy preserving data mining. *Proceedings of the 20th Annual International Cryptology Conference on Advances in Cryptolog*, 36-54.