DISTRIBUTED PROVENANCE AS A SERVICE FOR TRUSTWORTHY COMPUTING IN THE DYNAMIC IOT CLOUD SYSTEMS

[1]M. Jeevan Kumar, [2]Dr. Vishal Bhatnagar

[1]Research scholar, Dept of C.S.E, Shri Venkateshwara University, Gajraula, Uttar Pradesh, India.

[2]Research Guide, Dept of C.S.E, Shri Venkateshwara University, Gajraula, Uttar Pradesh, India.

ABSTRACT: Through, allowing devices to utilize networked resources and make intelligent decisions, the Internet of Things (IoT) offers numerous advantages. Developing a reliable and secure network is necessary. They consider provenance is one of the crucial components of IoT-based solutions have not received the attention they require. Enormous computational and storage capabilities provided by the cloud provides an effective and easily accessible platform for processing and preserving IoT data. Distributed Provenance as a Service for Trustworthy Computing in the Dynamic IoT Cloud Systems is presented in this analysis. Data provenance, sometimes referred to as data lineage, offers a history of the changes made to a data item from the moment of creation to the present. To maintain independence from the underlying data processing system, Genoma separates the provenance collection and recording processes. The edge nodes can record these accesses and maintain the provenance information. Each specific edge gateway is connected to a specific fog Local Area Network (LAN). The following demonstrates that they mean by "provenance-as-a-service": Independent of the underlying data processing substrate; distributed provenance across IoT devices, the edge, and the cloud. In order to demonstrate the superiority of the described work, performance parameters such as data integrity, privacy, freshness, availability, and energy consumption were carried out in a real-world IoT ecosystem.

KEYWORDS: IoT cloud infrastructure, security, Distributed Provenance, data management.

I. INTRODUCTION

Data is often exchanged over the Internet by a complex network of smart devices collectively referred to as the Internet of Things (IoT) [1].

IoT is becoming essential for communication in the future. The IoT will provide the foundation for applications such as patient information for medical personnel, smart grids, smart meters, home automation, transportation, performance and statistics of automobiles. Data collected from sensors or Internet of Things nodes is transmitted to a cloud through the Internet, followed by acquisition by the necessary entities. The obtained data must be precise and include information about its source [2].

The management of IoT networks is distributed. In particular, the network is administrators by IoT network administrators from various trust domains and regulated through distributed protocols, such as Border Gateway Protocol (BGP) [3]. A global provenance diagnostics may collect provenance data from numerous network domains for global root cause analysis and performance optimization in a dispersed IoT network. Smart devices, Radio Frequency Identification (RFID) tags, and common sensor devices might all be integrated in a hospital setting. These devices keep track and gather different types of user data in a context [4].

The IoT healthcare application will utilize these data to inform choices and deliver common, automated services. This may be accomplished by combining, analyzing, aggregating, and mining all types of user data collected by IoT devices. Regarding

trust and data provenance, this raises a number of new challenges. The tracking of this data is referred to as provenance. To share information, generate knowledge, and conduct digital transactions, a trusted and secure IoT environment must be created.

Data provenance may be considered as additional data that supports in identifying a given data's history [5]. The data provenance metadata should allow revealing the true identity of the owner connected with the transferred IoT data when a privacy-preserving strategy is required, considering the inspection requirements are addressed. Additionally, the provenance information for the data should be permanently everywhere data is kept or shared, in transit or at rest, so that it can be traced and audited. For IoT applications, suitable privacy-preserving data provenance techniques are presently not being used in a way that satisfies these requirements [6].

The Internet of Things-based network has to be sufficiently secure to earn consumers trust. Due to IoT node's low energy consumption, the security method should be lightweight. It is also important to safe and reliable mutual authentication between IoT nodes and the server. In the IoT, accurate and secure data provenance is used to increase the level of trust. The data provenance may be used to track and describe the progression of data generation, starting with the initial sources. From the perspective of regulatory processes, the records can be utilized in order to protect intellectual property. Though, the integrity of the data provenance is an important issue. Whenever the information provenance is not adequately protected by following wasteful security rules, it may be generated or altered by an unauthorized party [7]. A security solution that is both lightweight and extremely secure should be developed in order to build IoT user trust.

Data provenance is used for a number of purposes. Data quality and reliability may be evaluated using the data provenance. The quality and trustworthiness of the data may be guaranteed if the source and nearby nodes that are creating or transporting it are reliable. It is also possible to use data provenance to determine the nodes and actions that contributed to problems in data collecting and processing, which can then be used in verified computation. On the other hand, detailed provenance information enables data recovery when data is no longer useable, preserving system availability and ensuring efficient data communications. The data flow brought on by smart devices has the potential to compromise privacy. At the structural and data/module level, privacy must be ensured for any provenance system that is aware of privacy. Data provenance may also define data citations and increase the readability of data.

The structure of this paper is as follows: The literature survey is explained in Section II, the provenance methodology is discussed in Section III, the result analysis is explained in Section IV, and the analysis is concluded with Section V.

II. LITERATURE SURVEY

Ebelechukwu Nwafor, David Hill, Andre Campbell, and Gedare Bloom et. al. [8] explains provenance is integrated into the Internet of Things. The Provenance Aware Internet of Things System (PAIoTS) is the name of the framework, the authors propose for collecting provenance data from IoT devices that is trace-based. In the fields of forensic analysis, business, scientific computing, and intrusion detection, data provenance has been analyzed. The provenance of data can support in the

detection and prevention of malicious cyberattacks. They create a proof-of-concept prototype system to evaluate the efficiency of described framework. Sabah Suhail, Zuhaib Uddin Ahmad, Faheem Zafar, Choong Seon Hong, Abid Khan, et. al. [9] discussed the difficulties associated with the technical infrastructure of IoT based on RFID, IP (Internet Protocol), and WSN (Wireless Sensor Network). Provenance may be extremely important in maintaining the data trails of IoT devices since it addresses several problems with data replication, decision-making, and trustworthiness. Based on security concerns and other IoT resource limitations, they have identified some potential approaches to integrating safe provenance into the IoT.

Compton M., Corsar D., and Taylor K., et al. [10] presents an approach for the alignment of Provenance Ontology (PROV-O), using the Semantic Sensor Network (SSN), the semantic ontology is used to express sensor observation data. This model offers information on the provenance ontology representation of the SSN subcomponents. The model does not handle additional sensor semantic representations and is suited appropriate for SSN. The main goal of author effort is to provide a generic Provenance-Sensor alignment that is unrelated to any one semantic ontology.

Mohd Izuan Mohd Saad, Kamarularifin Abd Jalil, Mazani Manaf, et. al. [11] presents the description of the provenance as well as the difficulties of providing security assurances in the cloud. The research additionally indicates an innovative data provenance trust model for cloud computing. The model's flow mechanism for achieving a high level of cloud service trust is thoroughly described in the paper. The purpose of this research is to develop a complete trust model that incorporates all relevant security components in order to increase the degree of trust among cloud services. P. Buneman and S. Davidson et al. [12], state the importance of provenance in understanding data quality. Additionally, it has been noted that database users have an interest about the provenance of a particular piece of data in the context of data provenance and the field of databases. They specify that the majority of the data in databases are chosen subsets (views), with source data constituting a very small fraction of the total. The accuracy and timeliness of the data are thus determined by knowing provenance.

E. Pignotti and P. Edwards, et. al. [13] provides a mobile-enabled software prototype with an integrated semantic model for the purpose of capturing, storing, and using metadata linked to the device, origin, functionality, and usage. They said that despite the benefits of IoT apps running on devices, customers challenge with understanding the capabilities, uses, and workings of the devices. Jie Yin, Jun Li and Peichao Ke, et. al. [14] provides an algorithm based on provenance data to improve service chain performance by improving the dispatch probability of businesses with excellent reputations as determined by user ratings. In the logistics chain process, RFID tags are used. A number of businesses that offer comparable services are included in each service provider pool. This method expresses user evaluations and enterprise reputation using fuzzy linguistic values, considering the various variables utilized for evaluating business reputation as well as the uncertainty of user assessments on service chain results. The outcomes of simulations show effectively this strategy performs.

Fenye Bao, Ing-Ray Chen, et. al. [15] the development of a scalable trust management protocol for the Internet of Things

maintaining a focus on interpersonal connections. In order to account for social interaction, they take into consider a number of trust qualities, such as honesty, cooperation, and communal interest. These results demonstrate that the case of false proposal attacks carried out by malicious nodes, there is a trade-off between trust assessment accuracy against trust convergence time.

III. DISTRIBUTED PROVENANCE AS A SERVICE FOR TRUSTWORTHY COMPUTING

The architecture of Distributed Provenance as a service for Trustworthy computing in the Dynamic IoT Cloud systems is represented in below Fig. 1.

The inputs, systems, entities, and processes that have an effect on significant data is recorded in an IoT-based system's provenance. The user receives the data, performs inference on it to understand the provenance data and derives significant conclusions. The data is collected, recorded, and put into the storage system. Furthermore, since provenance data is a permanent record of usage, it should be expected that it will be read-only; it can only be saved, archived, or removed; it cannot be rewritten.

Genoma separates the recording and capturing of provenance to provide independence from the underlying data processing technology. In accordance with the policies established by the Policy Modeller, provenance data is sent to the cloud after being stored on the edges. The two methods for gathering provenance data in the cloud are the provenance collector at the edge and the data stream processing system directly. All provenance data that is gathered subsequently maintained in the cloud by Apache Atlas.

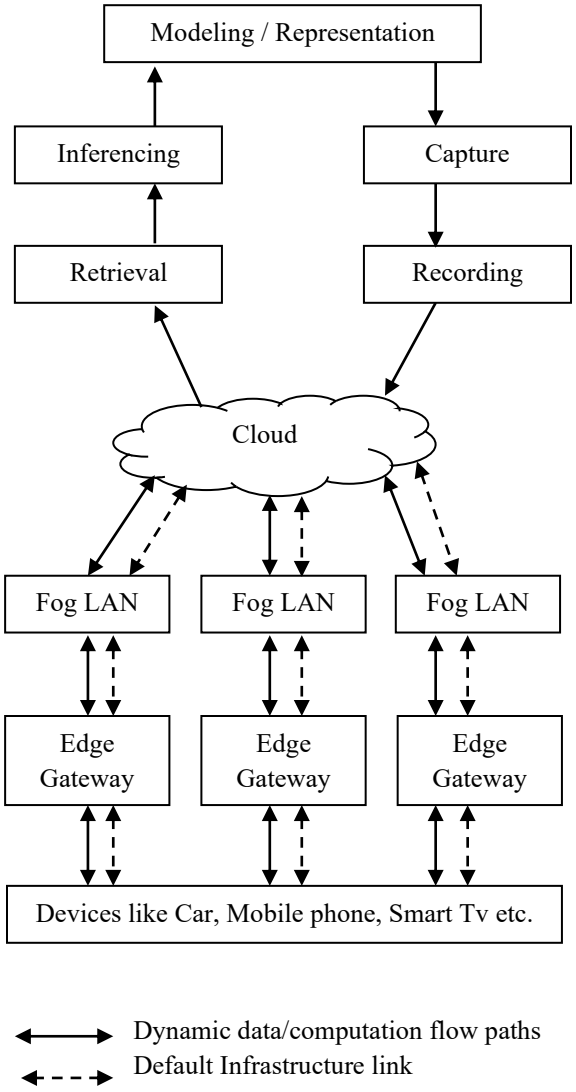

Fig. 1: ARCHITECTURE OF PROVENANCE MODEL

This Provenance -As-A-Service (PAAS) has been one of the most important works on provenance in the cloud. It automatically gathers, maintains, stores, and offers provenance searches. It serves as a provenance solution as well as a foundation for further provenance systems. Therefore, it is strongly connected to the basic cloud data platform on which it was constructed.

Considering that many IoT systems gather data for quickly decision-making, the decisions may be meaningless due to the

significant latency. Solutions for the IoT cloud's network latency issue have been proposed through edge and fog computing. Fog and edge computing move the computation from the central cloud to the edge. Edge nodes can carry out some early calculations and often serve as the gateway for IoT devices. Thus, the layered infrastructure, from IoT devices, to the edge gateways, to the fog LANs, and to the cloud should be integrated to provide real time and powerful computation for the IoT systems.

Decomposition of computations enables preliminary decisions to be made regarding fog LANs and nearby edge nodes. The cloud can handle intensive computations and globalized data processing. According to application-specific designs each layers, data processing tasks are decomposed. Some major applications such as IoT systems for fault diagnosis, supply chain tracking, autonomous vehicle control, etc., can make use of this computing structure.

Each specific edge gateway is connected to a specific fog Local Area Network (LAN). The fog LAN then connects to a specific cloud. However, a static configuration could acquire resource underutilization, have unbalanced load, and be prone to single-point failures. Thus, it is frequently desirable to have dynamic IoT cloud configurations. For example, when a default edge gateway is overloaded or failed, a nearby alternate can be selected dynamically to serve. In Fig. 1 the black doted lines are the default connections and the red lines represent the dynamically selected connections for non-dedicated data and computation flow paths.

This dynamic computing structure may not be feasible for some IoT systems. For example, in a smart home, if the edge gateway is owned by the home owner, then it will be infeasible to share it. But if the gateway is owned by the Internet provider and the provider offers “software defined edge capability” and proper “isolation”, then gateways from multiple nearby homes can be shared when needed. Such dynamic configuration can be more conveniently realized by public or semi-open IoT systems such as surveillance camera networks and road side gateways for autonomous vehicles.

Provenance data may be temporarily kept and processed on the edge due to Genoma's distributed storage method, which allows for the resolution of network connection challenges that will probably grow more prevalent in large-scale distributed IoT systems. However, all obtained provenance data is intended to be stored on the cloud.

The provenance information is formatted for our data model by a connection called provenance recording, which receives it from the provenance capture connector. The file system storage is then used to store this data on the edge device. Therefore, Genoma may operate as a genuine "provenance as a service" as it would be independent of the underlying data stream processing system due to the separation of provenance recording and storage. The user may view the provenance data that has been stored due to the provenance visualization component.

IV. RESULT ANALYSIS

The presence of a low-end device, such as a Raspberry Pi or low-end laptop, for our implementation of the edge, a minimum of 80 GB of storage and 2 GB of RAM are needed. They are now developing a miniprovenance engine on the edge because there are no provenance tools available that can be used with such devices. Due to the

initial time needed to converge the topology, simulations run for 10 minutes before reporting starts. Five times during each simulation, the average results are provided.

Essentially, to recreate it using the system in a resource-constrained setting, the Data Integrity, Data Privacy, Freshness, and Availability are taken into consideration as performance indicators. However, security and service quality assurance mechanisms in charge of ensuring these characteristics. Following are explanations of the performance matrices taken into consideration for evaluation:

Integrity: Integrity involves preventing unauthorized persons from tampering or modifying the data to ensure its integrity, correctness, and validity at the source, across communication networks, and at the destination.

Privacy: Privacy is one of the most important aspects of protecting personal information in the Internet of Things (IoT), where virtually every physical or logical object has a special identification and is capable of independent communication across the Internet or another comparable network. The provenance created as a result of the data overflow caused by smart things might be concerning for privacy. Privacy must be protected at the data/module level and structural level for any provenance system that is privacy-aware.

Freshness: An adversary must be unable to replay recorded data and provenance without being seen in order to meet the freshness requirement. The implication is an attacker cannot replay data to authorized nodes without auditors recognizing access, even they have captured some of the data and provenance information.

Availability: The provenance chain prevents nodes from selectively removing certain nodes, according to the availability requirement. This means that records in a valid provenance chain between whether they collaborate or not, two nodes cannot be removed by either node.

The comparative performance of Distributed Provenance as a service for Trustworthy computing model and Normal provenance model is graphically represented in below Fig. 2 in terms of Integrity and Privacy. Similarly performance of these two models in terms of Freshness and Availability are represented in below Fig. 3.

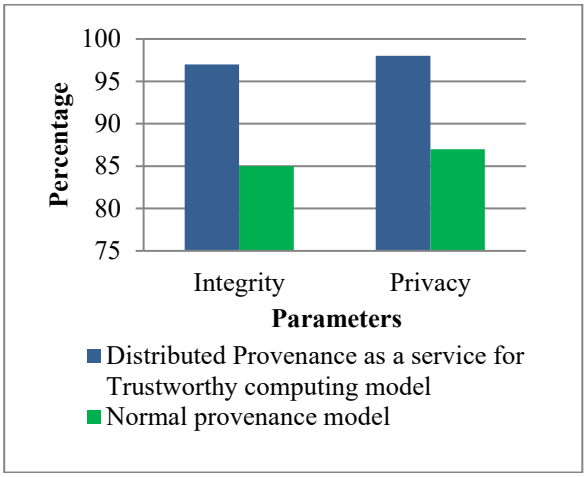

Fig. 2: COMPARATIVE PERFORMANCE IN TERMS OF INTEGRITY AND PRIVACY

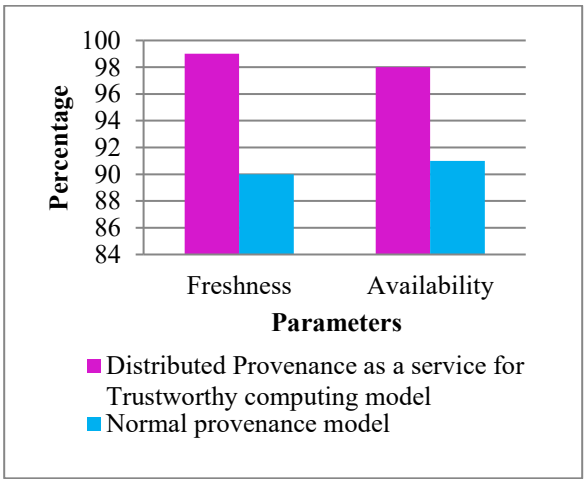

Fig. 3: COMPARATIVE PERFORMANCE IN TERMS OF FRESHNESS AND AVAILABILITY

Energy Consumption: Energy is a limited resource because IoT nodes are often battery-powered. Therefore, using the nominal values of the Edge IoT things, for both the distributed Provenance as a service for Trustworthy computing model and the model without distributed provenance, and they measure energy consumption at the system level. In contrast to the Routing Protocol for low-power and Lossy networks (RPL) without the Distributed Provenance as a Service model, Fig. 4 demonstrates that the energy usage in the case of the Distributed Provenance as a Service model is minimal.

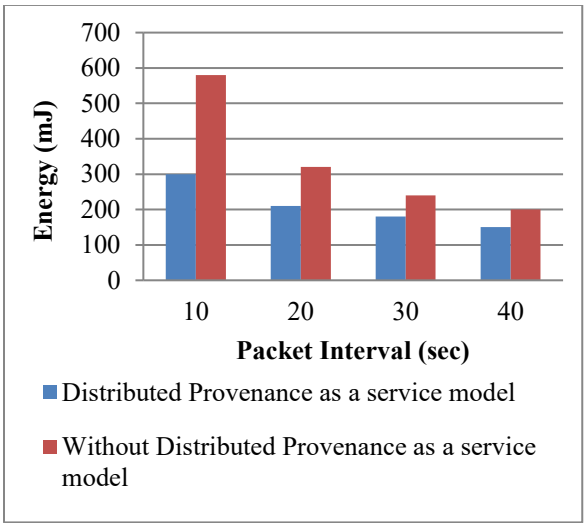

Fig. 4: COMPARISON OF ENERGY CONSUMPTION

From results it is clear that, performance parameters as Data Integrity, Data Privacy, Freshness and Availability are high for described Distributed Provenance as a service for Trustworthy computing model. Similarly, energy consumption is very low compared with without Distributed Provenance as a service model. Therefore, the effectiveness of described model is enhanced.

V. CONCLUSION

In this paper, Distributed Provenance as a service for Trustworthy computing in the Dynamic IoT Cloud systems is described. In the layered IoT cloud, the accesses are validated through the edge nodes. To maintain independence from the basic foundation for information management, Genoma isolates the provenance assortment and recording processes. The edge nodes can record these accesses and maintain the provenance information. The overall provenance information is essentially stored in the IoT cloud infrastructure. Each specific edge gateway is connected to a specific fog Local Area Network (LAN). The fog LAN then connects to a specific cloud. The reason they intend is the following "provenance-as-a-service": along with independence from the underlying data processing substrate, distributed provenance across IoT devices, the edge, and the cloud are also important. The Data Integrity, Data Privacy, Freshness, Availability and energy consumption are considered as performance metrics used in this study. From results it is clear that, High Data Integrity, Data Privacy, Freshness, Availability and Low energy consumption for described Distributed Provenance as a service for Trustworthy computing model. Therefore, the effectiveness of described model is enhanced.

VI. REFERENCES

[1] Joseph Henry Anajemba, Tang Yue, Celestine Iwendi, Pushpita Chatterjee, Desire Ngabo,Waleed S. Alnumay, “A Secure Multiuser privacy Technique for Wireless IoT Networks Using Stochastic privacy Optimization”, IEEE Internet of Things Journal, Volume: 9, Issue: 4, Year: 2022

[2] Dawei Wei, Huansheng Ning, Feifei Shi, Yueliang Wan, Jiabo Xu, Shunkun Yang, Li Zhu, “Dataflow Management in the Internet of Things: Sensing, Control, and Security”, Tsinghua Science and Technology, Year: 2021

[3] Teimuraz Matcharashvili, Archil Prangishvili, "Quantifying regularity of the Internet Interdomain Routing based on Border Gateway Protocol (BGP) data bases", 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), Year: 2020

[4] Usman A. Haider, M. Noman, Hidayat Ullah, Farooq A. Tahir, "A Compact Chip-less RFID Tags for IoT Applications", 2020 IEEE International Symposium on Antennas and Propagation and North American Radio Science Meeting, Year: 2020

[5] Hala Hamadeh, Akhilesh Tyagi, "Privacy Preserving Data provenance Model Based on PUF for Secure Internet of Things", 2019 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS), Year: 2019

[6] Sabah Suhail, Choong Seon Hong, M. Ali Lodhi, Faheem Zafar, Abid Khan, Faisal Bashir, "Data trustworthiness in IoT", 2018 International Conference on Information Networking (ICOIN), Year: 2018

[7] Mohsin Kamal, sMuhammad Tariq, "Light-Weight Security and Data provenance for Multi-Hop Internet of Things", IEEE Access, Volume: 6, Year: 2018

[8] Ebelechukwu Nwafor, Andre Campbell, David Hill, and Gedare Bloom "Towards a provenance collection framework for Internet of Things devices," in Proc. IEEE SmartWorld, San Francisco, CA, 2017, pp. 1-6

[9] Sabah Suhail, Choong Seon Hong, Zuhaib Uddin Ahmad, Faheem Zafar, Abid Khan, "Introducing Secure Provenance in IoT: Requirements and Challenges", 2016 International Workshop on Secure Internet of Things, 2016

[10] M. Compton, D. Corsar, and K. Taylor, "Sensor data provenance: Ssno and prov-o together at last." 2014

[11] Mohd Izuan Mohd Saad, Kamarularifin Abd Jalil, Mazani Manaf, "Achieving trust in cloud computing using s ecure data provenance", 2014 IEEE Conference on Open Systems (ICOS), Year: 2014

[12] P. Buneman and S. Davidson, "Data provenance--the foundation of data quality," in Book Data provenance--the foundation of data quality', 2013, vol. 4, no. 1, pp. 1–8

[13] E. Pignotti and P. Edwards, "Trusted tiny things: making the internet of things more transparent to users," ASPI '13 Proc. Int. Work. Adapt. Secur., 2013

[14] Jie Yin, Jun Li and Peichao Ke, "A Provenance Based Scheduling Algorithm for Logistics Chain in IOT", 2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering, 2013

[15] Fenye Bao, Ing-Ray Chen, "Trust management for the internet of things and its application to service composi tion", 2012 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), Year: 2012