Anonymization Framework for IoT Resource Discovery based on Edge Centric Privacy Model

Santosh Pattar, Lakshmi K N, Rajkumar Buyya, Venugopal K R, S S Iyengar, L M Patnaik

Abstract**: *As the efficacy of Internet of Things is expeditiously growing, maintaining privacy with respect users and applications has become a significant aspect. Since the data is getting generated at tremendous rate that includes Sensitive data (any data considered as private by the Data-owner) which has to be hidden, especially the data collected from the Crowd-Source. Due to resource-constrained sensing devices, IoT infrastructures use Edge devices for real-time data processing. Protecting sensitive data from malicious activity becomes a key factor, as all the communication flows through insecure channels. To develop security infrastructures for IoT and distributed Edge networks, this article proposes a user-centric security solution. The proposed security solution shifts from a network-centric approach to a user-centric security approach by authenticating users and devices before communication is established. The method presented herein is applied to an amusement park scenario, which is modeled as a typical smart IoT network. Here, data from sensors and social networks can boost smart lighting to provide citizens with an elegant and safe environment. However, it is challenging and infeasible to transfer and process zillions bytes of data using the current cloud-device architecture due to bandwidth constraints of networks, potentially uncontrollable latency of cloud services, and privacy concerns while collecting data from IoT devices. Firstly, a standalone IoT-edge system is developed, and later, an integrated IoT-based edge-cloud system is designed to compare the systems' effectiveness. The implementation results show a close correlation between the standalone edge and dual mode edge system. However, the edge-cloud system provides more flexibility and capability to counter the sensitive data streaming and analytics services within the constrained IoT framework. In this paper we have developed a system that uses fog computing approach to perform various tasks and filters the sensitive data, thus helps in preserving privacy.***

Keywords: Cloudlet, Data Dissemination, Edge Computing, IoT Ecosystem, Privacy

I. INTRODUCTION

Internet of Things (IoT) is a vast collection of things, human beings, sensors, and other physical objects that are connected over a network which accredits exchange of information among objects and also enables for some action to be taken by these things based on the decision or knowledge obtained from the data collected through the devices [1]. Due to the gigantic size of the data originated by gross sensors across the globe, it is not viable to gather and compute on all the data that is generated. Hence, it is advisable to fetch only from nominated sensors, selected based on the application domain or the requirements of the application, that helps to overcome the challenges posed by the large size of the data space. However, with the evolution of diverse IoT applications (e.g., smart city, industrial automation, and connected car), it becomes challenging for edge computing to deal with these heterogeneous IoT environments. In an IoT ecosystem, data is assembled in any-form, anywhere, and at any time from the sensors. There are possibilities that the fetched data might include confidential content owned by the people, organization *etc* and thus leads to the contravention of the *privacy in IoT*. Thus, bearing privacy has turned out to be a vital part of the IoT [2]. Data collected from an IoT ecosystem contains sensitive data that includes any identity related information like name, address, age, salary, location etc. that has to be accessed based on the considerations of accountable, fair and lawful processing, security safeguards, limited and controlled disclosures, transparency, choice and individual participation, collection and purpose limitations. Constraints for maintaining privacy are transparency, choice and individual participation [1]. IoT is the future of the internet where objects are connected to each other and having capabilities to sense, actuate and process the collected data using which numerous IoT applications can be built across various domains. viz., social, medical, logistics etc. The data collected by the various sensors that are embedded into numerous objects includes personal data which users does not want to share among the IoT devices, applications and other users [2]. It is very challenging to maintain privacy due to issues with heterogeneity, massive scale. Data collection process should clearly state "why the data is being collected" and "*how it is collected*".

Revised Manuscript Received on December 25, 2020.
* Correspondence Author

Santosh Pattar*, Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bangalore, India 560-001. (e-mail: santoshpattar01@gmail.com, lakshmikyadav4@gmail.com)

Lakshmi K N, Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bangalore, India 560-001. (e-mail: santoshpattar01@gmail.com, lakshmikyadav4@gmail.com)

Rajkumar Buyya, Cloud Computing and Distributed Systems (CLOUDS) Lab, School of Computing and Information Systems, The University of Melbourne, Melbourne, Victoria, Australia - VIC 3053. (email: rbuyya@unimelb.edu.au)

Venugopal K R, Bangalore University, Bangalore, India 560-001. (e-mail: venugopalkr@gmail.com)

S S Iyengar, Department of Computer Science and Engineering, Florida International University, Miami, Florida USA 33199. (e-mail: iyengar@csc.lsu.edu)

L M Patnaik, Consciousness Studies Program, National Institute of Advanced Studies, Indian Institute of Science, Bangalore, India 560-012. (e-mail: lalitblr@gmail.com)

Retrieval Number: 100.1/ijeat.B21091210220
DOI:10.35940/ijeat.B2109.1210220
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication
© Copyright: All rights reserved.

It should precisely mention the node of this collected data. "*Privacy Notices*" helps to deal with these issues and bring transparency to the data collection process, thereby increasing the acceptability of an IoT application by wider community [3]. Denaturing algorithm is used to hide the sensitive data where the content captured by camera is denatured with techniques like blurring, blanking, etc. so that the privacy of user is preserved. However, the privacy solution is applicable only for the video data and in IoT the data is in variety of format collected by the sensors like location, time, temperature, body movements etc. and the solution fails to consider this myriad range of formats [4].

An approach to solve the privacy problem (Methodology): The proposed framework is comprised of two subsystems: user portfolio and virtual machine portfolio (as shown in Fig. 4 and 5). The physical subsystem further comprises of an IoT sensor layer and various stakeholders viz. Data manager, cloud, user and privacy policies, and the cyber subsystem comprises of two layers: edge and Cloud. Each layer is formed of closely related functionality so that every other layer can function in an independent and efficient manner. The physical subsystem facilitates the data acquisition from the stranded individuals and disaster-affected environment, and the provision of various data analytics in the form of information services to the respective stakeholders. Whereas the cyber subsystem employs the data acquired by the IoT sensors and edge devices from physical subsystem and facilitates the various data analytics processes through the local cloudlets.

Contributions: Develop and implement a novel IoT framework integrated edge computing assisted system in leveraging sensitive data from edge networks.

1) To develop and implement an ideal IoT-based edge-cloud system to provide real-time sensitive data streaming and analytics service.

2) A novel end-to-end secure model for IoT and cloud-based edge networks.

3) To discuss and evaluate the proposed implementations in the purview of obtained results.

4) A privacy-based edge centric security model avoiding attackers from the first step of communications.

Organization: The rest of the paper is organized as follows. In the second section, we discuss state of the art research works that address search and discovery problem in in the IoT and also the privacy concerns in IoT by considering an use case scenario. In the next section, the proposed framework is discussed in detail. In the fourth section we propose our anonymization framework and denaturing algorithm. Experimental setup and result evaluation of the proposed framework's various components are compared in the sixth section. Finally, the sixth section concludes the paper.

II. LITERATURE SURVEY

A. Search and Discovery Techniques in IoT

Perera et al. [4] developed a context aware sensor search, selection and ranking model to overcome the demand of effective subset selection from relevant sensors among number of sensors in the IoT Network. The advantage is that many user preferences are taken into consideration like reliability, battery life etc. and also solutions for effective distributed sensor search are discussed. Tianqi et al. [5] designed a physical topology at the cloud layer for large scale IoT systems using unmanned aerial vehicles to monitor large scale environments. however, due to the use of predefined network topology the applicability of the proposed system for a generic application is of limited use. Amir et al. [6] introduced an indexing mechanism for IoT resources to retrieve their data in an efficient manner using data discovery services. Although it is impervious to dynamic and frequently updating data, the discovery time significantly increases due to the large size of indices [7]. Tanganelli et al. [8] implemented an edge centric distributed architecture that provides resource discovery and access service to IoT application. due to the use of distributed architecture the proposed system is scalable when compared to traditional model. However, the use of edge devices prevents the wide scale application of the system for larger environments [9]. Zhipeng et al. [10] conceptualized a method to bridge the technical gap for itemized product management in M2M interactions. customized protocol for machine discovery is developed that takes into account presence and messaging bottlenecks. However, entire system is disturbed if M2M interaction gets interrupted and thus leads to single point failure. Jeon et al. [11] gave forth a general model for performance analysis of neighbour discovery process in bluetooth low energy networks [12]. But, it is not always possible to maintain advertising interval and scan window size equal and thus leads to communication overhead.

B. Amusement Park and IoT

Bitar et al. [13] demonstrated the use of drones in amusement park for surveillance and monitoring of user with respect to medical emergency. Yet, the authors fail to provide a detailed user study of the designed protocol and thus its efficiency cannot be ganged. Kurkovsky et al. [14] applied Radio-Frequency Identification (RFID) to track and monitor people with broad objective of increasing safety and productivity of the deployed environment. proposed approach has applications to locate lost children, track military law enforcement and medical personnel. Despite that the use of RFID tags for identification hinders the deployment of encryption and security algorithms due to their low computational power and hence suffers from privacy concerns.

C. Privacy in IoT

Zhou et al. [15] scrutinized the Iot concerns, security threats, existing solutions, investigation demands, the advancement direction of present day IoT is also presented. Nonetheless, this w does not account for any specific solution to overcome security issues in the IoT. Xiong et al. [16] adduced a privacy and availability data clustering scheme based on k-means algorithm and differential privacy. A method is implemented to pick initial center point and calculate its distance from other points which eliminates outliers during the cluster formation. Although it heads to the formation of noise free clusters, the proposed scheme overlooks the data points that are dissimilar and merges into the same group.

Table 1: Comparison of The Literature Survey

Authors	IoT-domain	Description	Objective	Proposed Solution	Drawbacks
Zhou et. al. [10]	Generic	Effect of IoT contemporary features on security and privacy.	Illustrates developing trend of IoT security considering eight features of IoT, its solutions and drawbacks.	Dynamic analysis simulation platform, homomorphic encryption, anonymous protocols.	Paper does not provide any specific solution to overcome the drawbacks mentioned.
Xiong et. al. [16]	Energy Engage-ment	Enhancing privacy and viability for data clustering in intelligent electrical service of IoT.	Paper proposes a privacy and availability data clustering scheme (PADC) which enhances the selection of initial center points and the distance calculation method from other points to center point.	K-means algorithm and differential privacy algorithms.	The proposed scheme might sometimes overlook at the data points that are dissimilar and merges into the same group.
Lou et. al. [17]	Healthcare	Privacy protected data collection in IoT based health care systems.	Preventing some sophisticated attacks and threats such as collusion and data leakage has been presented in this paper.	Secure signature and encryption strategy, threshold secret sharing	Paper lacks in providing in-depth implementation and evaluation.
Zhou et. al. [18]	Cloud based IoT	Security and privacy for cloud-based IoT, challenges, counter measures and future directions.	Addresses challenging issues of secure packet forwarding and efficient privacy preserving authentication by proposing efficient preserving authentication.	Public key infrastructure using verification algorithm and anonymous public-secret key pair.	It is easy for the hackers to attack.
Yu et. al. [19]	Industrial IoT	Assured Data Deletion with Fine-grained Access Control for Fog-based Industrial Applications.	Proposed an assured data deletion scheme for secure data deletion.	Linear secret sharing schemes, attribute based encryption.	Even a minute error leads to the data loss and authorized users cannot access.
Sun et. al. [20]	Industrial IoT	Location Privacy Protection based on Differential Privacy Strategyfor Big Data in Industrial Internet-of Things.	Proposed a privacy protection method that satisfying differential privacy constraint to protect location data privacy and maximize the utility of data.	Laplace mechanism, privacy protection algorithm.	It can be applied only to the location based data.

Lou et al. [17] conceptualized a protection scheme for healthcare data produced from an IoT ecosystem. Tangled medical devices are liable to security violation. To protect them, the authors correlated slepian-wolf coding based secret sharing protector that interpolates the notion of secret sharing and reformation. It overlays on abundant concerns namely data leakage and destruction, flam attacks, insider attacks, the amount of data handling and the amount of data storage. Yet sharing mechanism, in-depth implementation and evaluation is lacking here. Zhou et al. [18] originated the architecture, distinctive security and privacy requirements for the next generation mobile technologies on the cloud. The approach labels the issues of secure packet forwarding and efficient privacy preserving authentication by proposing a privacy preserving data aggregation despite of any public key homomorphic encryption.

Yu et al. [19] set-forth framework that advocates data from industries and users by concatenating the cloud, fog and things of the physical environment.

It embeds data deletion scheme to verify the data and regulates sensorial data. Since, smart objects verify the data, even a minute error leads to failure and authorized users cannot decrypt which might also give rise to data loss. A solution to maintain privacy through a model ”*Giga-Sight-Architecture*” that includes cloudlets interconnected with each other and the cloud through internet was proposed [20] [21]. Yin et al. [22] proposed a procedure to protect location privacy that satisfies various challenges associated with location data in industrial IoT by constructing a multilevel location information tree model. In Table 1, we compare the recent works.

III. PROPOSED PRIVACY MODEL

A. Problem Description and Use Case Scenario

To ensure anonymity to smart devices and end users of search applications in IoT by concealing their sensitive information through “*fog-computing approaches*”. Amusement park is taken as the use case in the proffered work. An amusement park comprises of many astonishing rides of discrepant themes like water theme, thrilling games, dry games etc. The park is lodged with many sensors to collect the data based on its specification which is make to bestow startling befitting experience to the visitors and to supervise the park serenely. Sensors used are of two types which are static and dynamic. Static sensors are commonly coupled with the belongings of park and dynamic sensors are attached to the visitors by embedding with the wearables like watch, belt etc. a user who himself carries a GPS (determines location) sensor by holding a cell phone through which he can set the rules and pass queries (for search). An user is provided with an application (interface) through which he can set the policies to notify which data relevant to him can be shared and which has be to be kept private, the data which has to be kept private is denoted as a sensitive data that undergoes filtering process at the cloudlet level shown in Figure 1. With the help of smart phones, software applications and wearable devices, a person visiting the amusement park is surrounded with various types of sensors that records and emits monitored signals. once the user goes on an adventurous ride, the sensors connected to him via belt, watch etc. states his level of excitement and he can also review on it which can in turn be viewed by other visitors. This might lead to the disposal of personal details of the user which has to be maintained private. An user can set the policies based on his interest of disclosing information.

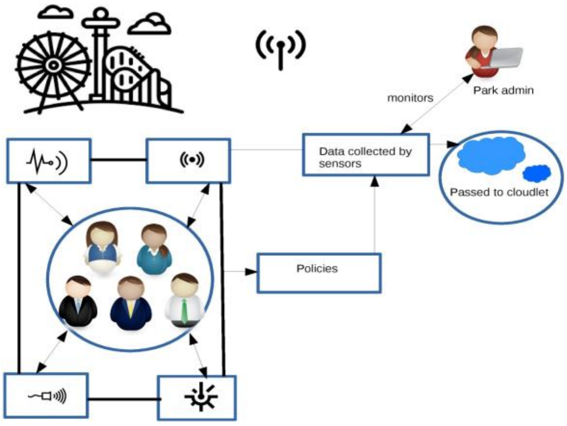


Figure 1: Amusement Park as a Use Case.

B. Data Acquisition and Privacy Policies

IoT acts as a backbone of sensing infrastructure to several mission-critical applications such as smart health, disaster management and smart cities [23]. The devices in regard of IoT generates mountainous data which can be anesthetized and assorted into five categories, which includes volume, velocity, veracity, variety, value and certified as five V’s of data generated in IoT. Since there are variant sensors attached to/around the users throughout an amusement park (which is taken as an use case in the proposed work), each sensor depicts contrastively in terms of anticipating data according to its characteristic, the amount of data it generates, the rate in which data is produced, accuracy in the obtained data, forms of data reproduced, value which can be derived. Employing five V’s as the basis, each sensor is being assigned with different identification tags. Table 2 depicts five V’s with various properties described below:

Volume: Volume is a bulk of data originated by sensing object which is described as the amount of space consumed. It accredits to the ridiculous amounts of data engendered each instant from the sensors. Volume can be in turn classified into three categories based on the amount of data generated by the sensors. If sensors generate data in a large volume it is categorized under huge volume and the device generating huge volume of data is assigned with “*High Volume Flag [HVF*]”. If the sensors produce data of moderate volume or the generated data requires moderate space then it falls under medium volume, the sensor generating medium volume of data will be assigned with “*Medium Volume Flag [MVF]”* [24]. The sensor or the devices attached with sensors producing data of less magnitude is called as profound volume as the data captured need less space and the sensor is assigned with “*Profound Volume Flag [PVF]”.*

Velocity: Velocity is the rate of data produced by the sensor which is described on the amount of speed. It imputes to the acceleration at which giant flock of data being procreated. Since the speed of induced data varies from sensor to sensor it is categorized under three modes of speed. Sensor producing data continuously at a high speed is termed as swift velocity sensor and labeled as “*Swift Velocity Label [SWVL]*”. Sensor producing data in medium intervals of time is termed as sluggish velocity and labeled as "*Sluggish Velocity Label [SLVL]*". If the velocity of data produced by the sensor is at a slow rate then it is termed as slow velocity and labeled as "*Slow Velocity Label [SLVL]*".

Veracity: Veracity represents the quality and truthfulness of data which can be interpreted in terms of validity. Accumulating mass of data has no point if the data lacks in quality and truthfulness. Thus it is categorized under two types, if the data generated and processed by sensor is always valid then it is coined as eternal and enrolled with an "*Eternal Tag [ET]*". Conceding that the data generated by sensor is valid only in an intrinsic span i.e. the identity of data which is considered as valid only if it is collected in a particular period or in an appropriate region based on locality, date, time etc. then it is coined as span based and enrolled with a "*Span Based Tag [ST]".*

Variety: Data existent looks merely distinct than data from the precedent. Data can be generated in various forms, it can be structured, unstructured, text format, audio, video etc.

Variety signifies heterogeneity of the data which can be represented as the type of data produced by the sensor devices. The sensor generates integer data, floating point, data in the form of a string, variable data and it is referenced by "*Integer Valued Pin [IPIN]*", "Floating Point Pin [FPIN]", "*String Valued Pin [SPIN]*", "*Variable Valued Pin [VPIN]*" (having both numerals and characters) respectively.

Value: Value interpolates the virtue concerned with the data being elicited. It refers to the stature of data being excerpted. Value implies usefulness of data based on domain requirements, value of generated data is portrayed by accessibility. The sensor generating data which can be converted into a useful form is forenamed as conventional form and is casted as "*Conventional Form Key [CK]*". The data which cannot be converted into a useful form is forenamed as a non-conventional form and is casted as "*Non-Conventional Form Key [NCK]*". The data which is obtained in an useful form and does not require to be converted into any forms is termed as eternal and the corresponding sensor is casted with "*Strict Key [SK]*".

i) <u>*Rule (Inferring number of visitors entering):*</u>
park(?p), visitedBy(?v), visitor(?u)
→count(?v, ?p)

ii) <u>*Rule (Inferring an attraction of amusement park):*</u>
park(?p), hasAttraction(?a)
→attraction(?a)

iii) <u>*Rule (Inferring an attraction with a given ID):*</u>
park(?p), hasAttraction(?a), hasID(?id)
→attraction(?a)

iv) <u>*Rule (Inferring heart rate and excitement rate of an user):*</u>
visiter(?u), hasHeartRate(?hr), hasExcitementRate(?er)
áheartRate(?hr)
→excitementRate(?hr)

v) <u>*Rule (Inferring speed of a ride):*</u>
park(?p), hasAttraction(?a), hasSpeed(?s)
→speed(?s)

vi) <u>*Rule (Inferring attraction having maximum speed):*</u>
park(?p), hasAttraction(?a), hasSpeed(?s)
→maxSpeed(?m)

vii) <u>*Rule (Inferring age of a visitor):*</u>
park(?p), visitedBy(?v), visitor(?u), hasAge(?y)
→age(?y)

Encryption: Encryption is a routine of striving an information in the interest of making only affianced users to procure it, since privacy has become an ethical challenge in IoT, it is obligatory to impinge privacy concerns of IoT. Table 3 illustrates various encryption algorithms used in the affianced work. The encryption algorithm has to be casted in accordance with precise lineaments. Data generated by the defined sensors in the use case is encrypted by different algorithms based on the categories entitled under five V's. The algorithms are distinguished and correlated to encrypt data coming from specific sensor hinge on length of the block size, key size, speed, efficiency.

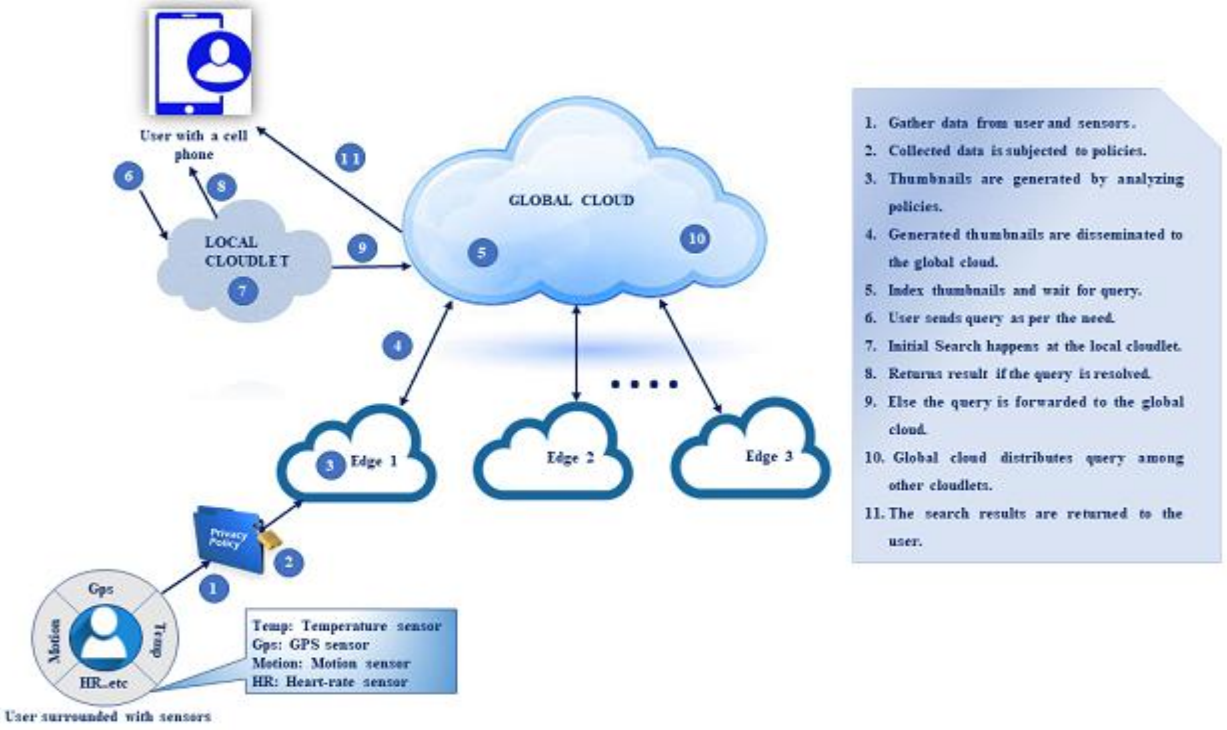
- *AES:* AES algorithm was originated by Vincent Rijmen, Joan Daemen in 2001 which uses a key length of 128 bits, 192 bits or 256 bits and it takes 10, 12, 14 rounds for 128 bit, 192 bit and 256 bit key respectively. The block size is of 64 bits and the symmetric type of an AES algorithm is symmetric block cypher. It is the fastest running algorithm and provides excellent security. Based on the above characteristics mentioned this algorithm can be used for the data which is generating at high speed, data of large volume. Since the algorithm is efficient in terms of speed, input size and security the algorithm is used to encrypt data coming from sensors labeled with HVF (high Volume Flag), SWVL (Swift Velocity Label), ET (Eternal Tag).
- *MD5:* MD5 algorithm was envisioned by Ronald Rivest, it has a key length of 128, 192 or 256 bits and takes 4 rounds, the block size is 512 bits, the cypher type of an algorithm is symmetric block cypher, slow compared to AES hence it is used to encrypt the data getting generated by the sensors assigned with MVF (Medium Volume Flag), SGVL (Sluggish Velocity Label) and ST (Span based Tag).
- *RSA:* RSA algorithm was bring into being by Ron Rivest, Adi Shamir and Leonard Adleman in 1978. The key length confides on number of bits in the modulus and where n=p*q, it takes a single round, the block size is not constant, the cypher type of an algorithm is asymmetric block cypher. This algorithm is slowest among all and least secure. By ruminating the all signed aspects,RSA is accounted to encrypt data generated by the sensors which are enrolled with PVF (Profound Volume Flag), SLVL (Sluggish Velocity Label), NCK (Non Conventional Form Key).
- *RC6:* RC6 algorithm was designed by Ron Revist, Matt Robshaw, Ray Sidney, Yiquen Lisa Yin in the year 1998. The block size is 128 bits, and the key sizes are 128, 192 and 256 bits, it is a symmetric key block cypher, it has high speed and minimal code memory. This algorithm works taking varieties of data as an input, it is fast and flexible thus it can be used to encrypt various forms of data which can be an integer, floating point, string format, or variable

Table 2: List of Key Definition for Five V's

Characteristic	Definition	Unit	Category	Definition To Category	Identification
Volume	The bulk of data originated.	Space Consumed	Huge Medium Profound	Data is generated at large volume. Data generated require moderate space. Data is captured in a less magnitude.	HVF-High Volume Flag MVF-Medium Volume Flag PVF-Profound Volume Flag
Velocity	The rateof data production.	Speed	Swift Sluggish Slow	Data produced continuously at high speed. Data produced in medium intervals of time. Data produced at slow rate.	SWVL-Swift Velocity Label SGVL-Sluggish Velocity Label SLVL-Slow Velocity Label
Veracity	Represents the truthfulness of data.	Validity	Eternal Span Based	Identity of processed data which is always valid. Identity of processed data which is valid only in a particular span.	ET-Eternal Tag ST-Span Based Tag
Variety	Signifies heterogeneity of the data.	Type	Integer Float String Variable	Collected data is of integer type. Data collected is of type float. Data is collected in the form of a string. Data collected is a variable type.	IPIN-Integer Valued Pin FPIN-Floating Point Pin SPIN-String Valued Pin VPIN-Variable Valued Pin
Value	Implies usefulness of data based on domain requirements.	Accessibility	Conventional form Non conventional form Strict form	Data which can be converted into its useful form. Data which cannot be converted into its useful form. Data which is obtained in useful form.	CK-Conventional Form Key NCK-Non Conventional Form Key SK-Strict Key

Table 3: Various Encryption Algorithms used in the Proposed Work.

Tag	Algorithm	Tag	Algorithm
HVF-High volume flag	AES	IPIN- Integer valued pin	RC6
MVF- Medium volume flag	MD5	FPIN- Floating point pin	RC6
PVF- Profound volume flag	RSA	SPIN- String valued pin	RC6
SWVL- Swift velocity label	AES	VPIN- Variable valued pin	RC6
SGVL- Sluggish velocity	MD5	CK- Conventional form key	Triple DES
SLVL- Slow velocity label	RSA	NCK- Non conventional form	RSA
ET- Eternal tag	AES	SK- Strict key	DES



1. Gather data from user and sensors.
2. Collected data is subjected to policies.
3. Thumbnails are generated by analyzing policies.
4. Generated thumbnails are disseminated to the global cloud.
5. Index thumbnails and wait for query.
6. User sends query as per the need.
7. Initial Search happens at the local cloudlet.
8. Returns result if the query is resolved.
9. Else the query is forwarded to the global cloud.
10. Global cloud distributes query among other cloudlets.
11. The search results are returned to the user.

Figure 2: Architecture Module of the Proposed Work.

form data. Henceforth, the algorithm can be casted to encrypt data assembled by sensors which are slotted with IPIN, FPIN, SPIN, VPIN.

- *Triple DES:* Triple DES algorithm was founded by IBM in the year 1978, it uses a key length of 168 bits, and takes 48 number of rounds, the block size is 64 bits, it belongs to a symmetric block cypher, it is used to encrypt data coming from the sensor assigned with CK (Convention Form Key) because the conventional form data does not require any conversion and thus triple DES endure the vogue of conventional form data encryption.

C. Ontological Model for Privacy in the IoT

Ontologies epitomize a collection of views associated with a specific domain/field/concern including the relationships coupled between them. It is the representation of knowledge derived from the relationships and attributes bonded with the several views. Using ontologies it is possible to obtain relationships among objects based on different properties in diversified ways. The vital elements of ontologies are that it guards the better interpretation of information as they form obvious speculation on the field/domain, it has a definite interconnection among the objects and etymology of metadata is efficient and thus ontologies strengthen the data quality in terms of understanding and representation. OWL (Web Ontology Language) is a semantic web computational language used in designing ontologies which offers precise, persistent and eloquent discrepancy between classes, attributes and relationships.

D. Architectural Design

The key challenge is the high rate of incoming data streams from devices with the various rules set by the users. Working of the entire process is shown in Figure 2. The system includes various components mentioned below.

1) Client: Smart phone is used as a client which has an app for human user and embedded firmware for IoT devices. Using the app user is able to set rules on the data being routed to the cloudlet for processing. The user can define various policies based on the privacy measures he is willing to take on the data he publishes, it can be based on time, location etc.

2) Cloudlet: The rules set by the user leads to the filtering of data which is done at the level of cloudlet which is responsible for hiding sensitive information. The task of denaturing algorithm is performed here and is implemented as a personal VM on cloudlet. The data passed by the user is denatured inside personal VM before being stored on the cloudlet.

3) Data Storage: By facilitating computing, data storage, and controls closer to the edge network, there is a necessity for developing a promising solution to meet the requirements of low latency, high scalability, and a privacy-aware edge centric model. [25]. Data manager runs in a separate VM on cloudlet. It manages the data repository which holds the data streams filtered by the cloudlet. It is a source of storage which also includes the database of associated metadata. Data is represented in the form of segments where each segment contains a set of data streams. The stream contains data from various sensors used in the use case like temperature sensor, accelerometer, GPS etc [26].

4) Cloud: Cloud is responsible for storing the metadata. Data is indexed and stored in the cloud, necessary tags are then annotated to it for the global search workflow.

E. Execution Flow

As delineated in Figure 2, the flow of execution includes several steps which are given below:

1) Get data: Data is collected from the user (entails user information like name, age, address, annual income etc.) and passed for the further process.

2) Filter: The collected data is subjected for the filtering process, where the data is filtered out based on the policies set by the user which includes sensitive and non-sensitive data. Any data considered as private should not be shared with others and thusly will be filtered out.

3) Process and generate thumbnails: The data is filtered and processed by applying certain methods (various encryption algorithms) to hide sensitive data and is further processed to generate thumbnails which holds the data obtained after processing and filtering.

4) Disseminate thumbnails to cloud: Since the data is collected and processed in the local edge which has to be sent to the global cloud. Therefore, generated thumbnails will be disseminated to the global cloud.

5) Indexing: Thumbnails will be indexed and converted into meta data which will be stored in the cloud.

6) Sending query: In this step users can send query stating few properties, the query is initially sent to the local cloud where the search is performed locally and sends back the result to the respective user if resolved or else the query will be forwarded to the global cloud.

7) Searching at Local cloudlet: Query sent by the user will be first checked in the local cloudlet and the results are returned by resolving the given query.

8) searching at Main cloud: If the relevant information is not found, the query will be directed and it is further searched in the global cloud.

F. Flow of Data

The below given Data Flow Diagram (as shown in the Figure 3) illustrates the flow of data:

1) Policy and Data Acquisition: The data provided is filtered based on policies set by the user and rest of the data is is acquired and passed for the next step.

2) Data Dissemination: Data is collected from user along with rules set by user which is then passed for processing and analysis.

3) Data Processing and Analysis: Filtered data from the previous step is passed for further processing.

4) Data Storage: Data is stored in the data manager of virtual machine.

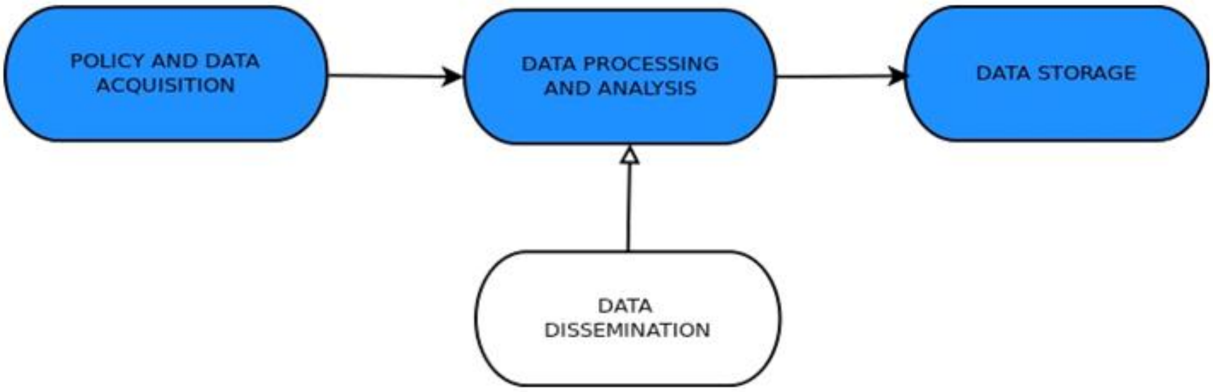
IV. ANONYMIZATION FRAMEWORK AND DENATURING ALGORITHM

A. Edge Computing

Recent years have witnessed the proliferation of edge computing and fog computing, in which billions of IoT devices are connected to IoT generate zillions bytes of data at the network edge [26]. Computation performed by cloudlets is similar to cloud computing since both approaches employ idle resources embedded with it in order to unfold the tasks. Yet, the incongruity among the two is that resources of the network are used in cloud computing, while in edge computing cloudlets only avail resources stationed at the edge of the network. Edge computing is planted on the notion of the utility computing which endures the view of sharing resources similar to network computing, distributed computing, elastic computing etc (due to several short comings in cloud computing) [27].

Edge computing has come into tableau with regards to impinge shortcomings in cloud computing like it extremely hampered the client encounter with the use of remote networking in cloud computing centers, it could not give full help to portable situations (mobile scenarios) particularly for the fast vehicle mounted system conditions, in which driver should rapidly find out about the street conditions and traffic stream progressively, it failed to meet the real time prerequisites of the non-cognitive conditions allied with geographical distribution [28]. On extensive scale sensor systems, for instance the sensor hubs must repeatedly forward their recently received information to other nodes. Countless devices linked with the cloud and system transmission capacity and network bandwidth ended up inadequate thus the cloudlets have been introduced as a new computing model having indistinguishable features of the cloud but nearer to clients in order to balance speed and other criteria.

Cloudlet is the flexibility reinforced information processing center assisting asset poignancy and synergistic nomadic supplications by offering effectual computational resources to the devices associated with it is the main purpose of cloudlet. It can be presumptively glanced as a information processing center in a corner whose intent is to get the cloud nearer. The goal of a cloudlet is to collar several applications like mobile use cases which are resource radical and inter mutual. A valid flap is found in the essentials for cloud and cloudlet [29].



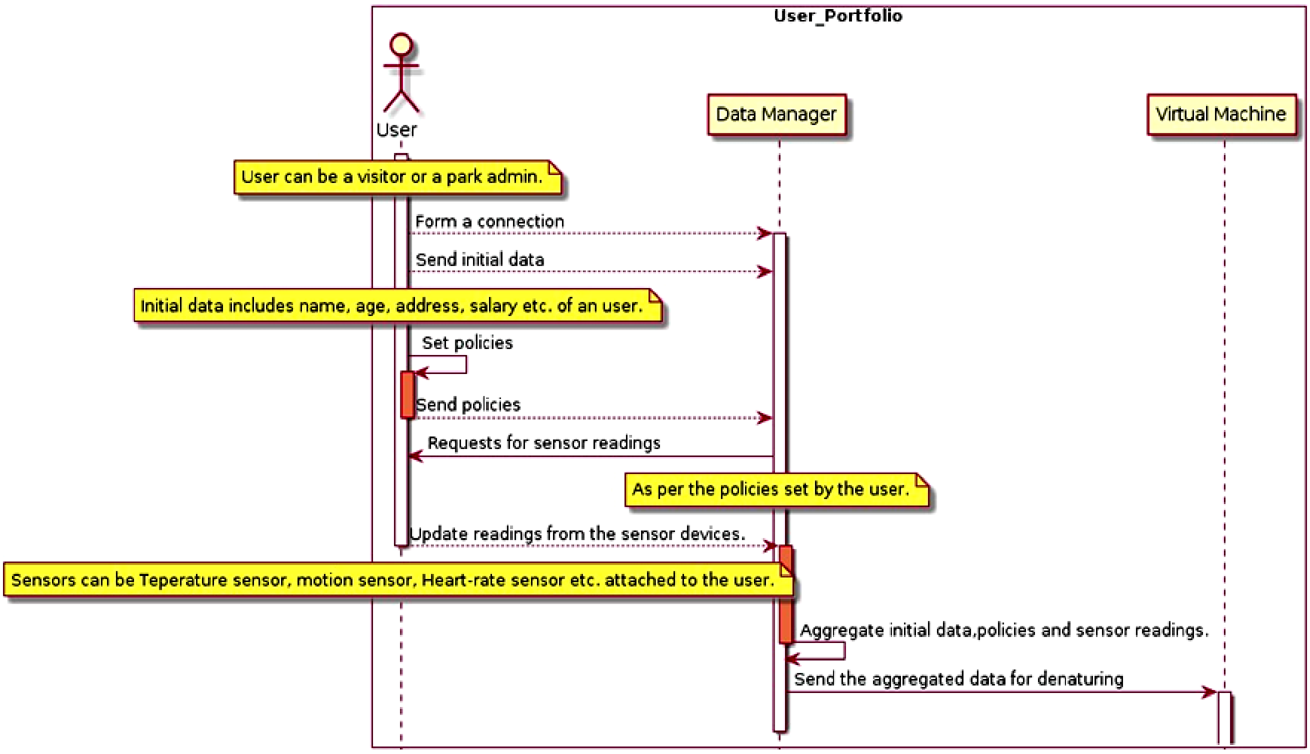

User_Portfolio
User
Data Manager
Virtual Machine
User can be a visitor or a park admin.
Form a connection
Send initial data
Initial data includes name, age, address, salary etc. of an user.
Set policies
Send policies
Requests for sensor readings
As per the policies set by the user.
Update readings from the sensor devices.
Sensors can be Teperature sensor, motion sensor, Heart-rate sensor etc. attached to the user.
Aggregate initial data,policies and sensor readings.
Send the aggregated data for denaturing

Figure 4: User Portfolio Sequence Diagram.

Both the levels require firm exile amongst uncertain user level data processing substantiation, exercise control, computing mechanisms. Since the computation of data coming from various devices take place in the cloudlet and only the meta data is passed into the cloud. The user can set policies and filtering is done at the nearest cloudlet level which helps in maintaining privacy even more efficient [30].

B. User-Portfolio

The amusement park incorporates several entities which are concerned and themselves included in gorging the data from sensors to the next level for processing and consecutively persevere data solitude. In accordance with the user portfolio, data manager requests for the user (user can be a park visitor or a park admin.) related information which includes visitor ID, name, age, address, date of birth, phone number, salary etc. Data manager also gathers sensor readings from the respected user where the sensor can be dynamic which is attached to the wearing of visitor like wrist watch, belt etc. or static sensors fixed in the rides or any other locales of an amusement park. Later the policies are received by the user whose sensitive data has to be hidden. The user information along with the sensor readings and policies set by the user is be aggregated and the aggregated data is sent for further denaturing process [31].

C. Virtual Machine (VM) Portfolio

The undeniable filtering process occur in the virtual machine. Virtual machine portfolio holds VM, data manager and an indexer to make an outfit. VM collects data from data manager having user information, sensor readings alongside user set policies. VM undergoes denaturing by employing various encryption algorithms. An appropriate algorithm is decided by considering the constraints like type of data, speed at which it is generated, restraints framed by user in policies etc. The private data is filtered out in the process of denaturing and the outcome is directed to the next level called indexer where the data is converted into meta data and sent to the cloud [32].

V. IMPLEMENTATIONS AND PERFORMANCE ANALYSIS

The proposed anonymization framework satisfies various user requirements with an assured privacy at a high level. The searching mechanism is based on user customized policies and filtering, where the system is embedded with several algorithms to handle different forms of data. As there are many considerations used collectively in one system and still works efficiently when compared with other resembling exertions with respect to the below mentioned metrics.

A. Query Distribution atop Number of Attempts

The current work has a design which addresses differently configurable data. As discussed in the proposed model section it handles velocity, veracity, volume, variety and value. The query processing efficiently works and demonstrates efficacious performance when compared with Novel indexing method. Figure 6 illustrates the comparison between anonymization framework and novel indexing method for scalable IoT [6] on correlation among number of queries and process of attempts made to execute the query considering user constraints.

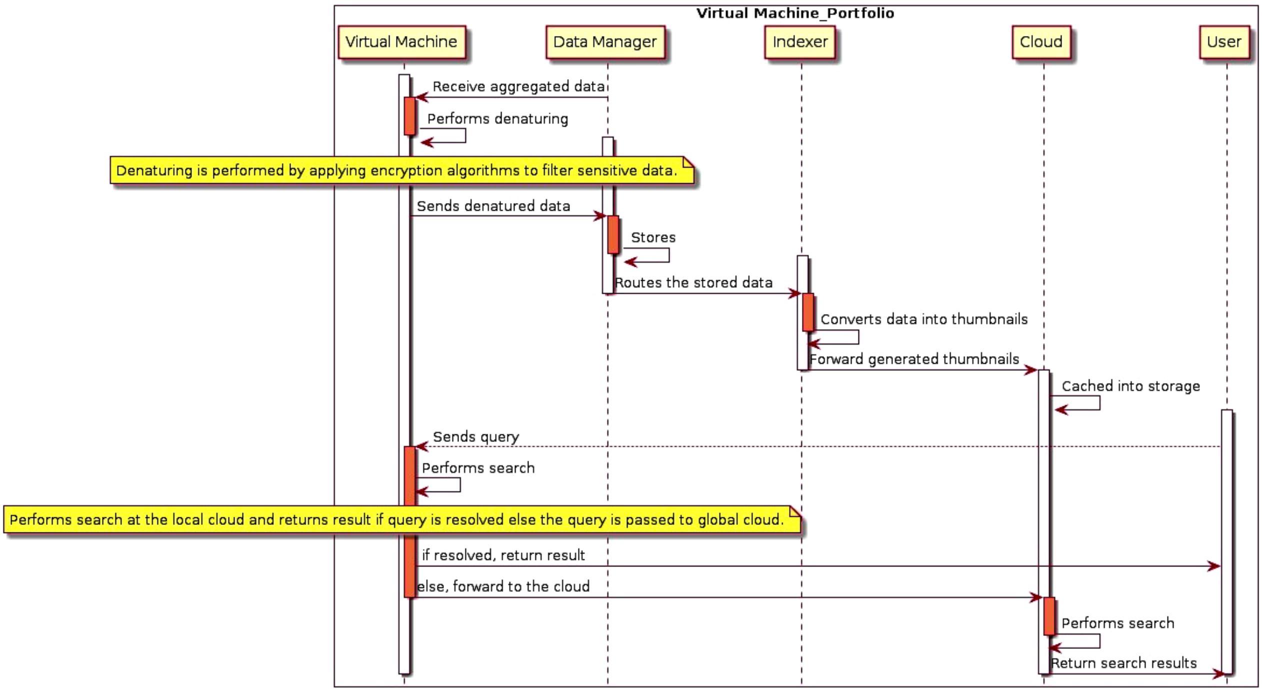

Virtual Machine_Portfolio
Virtual Machine
Data Manager
Indexer
Cloud
User
Receive aggregated data
Performs denaturing
Denaturing is performed by applying encryption algorithms to filter sensitive data.
Sends denatured data
Stores
Routes the stored data
Converts data into thumbnails
Forward generated thumbnails
Cached into storage
Sends query
Performs search
Performs search at the local cloud and returns result if query is resolved else the query is passed to global cloud.
if resolved, return result
else, forward to the cloud
Performs search
Return search results

Figure 5: Virtual Machine Portfolio Sequence Diagram.

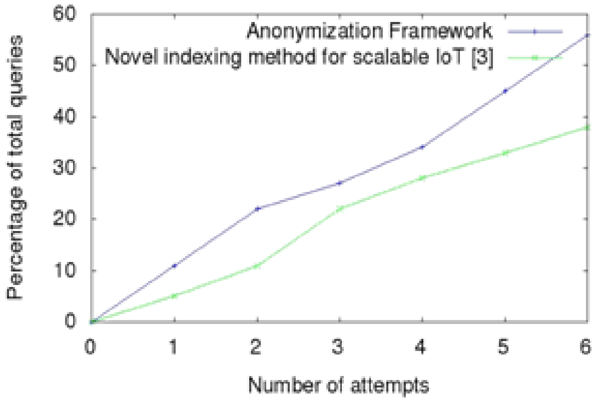


Figure 6: Comparison between distribution of queries atop number of attempts between anonymization framework and Novel Indexing method of scalable IoT [6].

B. Computation Overhead

The differentiation amongst static and dynamic devices has reduced the first level of filtering as there is a comprehension on which type of data is carried by which sensor inscribing "*variety*" data. Thus, while computing any operation the number of times devices visited was less when compared with any other related work. Figure 7 simulates the comparison between computation overhead of Game-theoretic greedy approximation offloading scheme [GT-GAOA] [8] and anonymization framework.

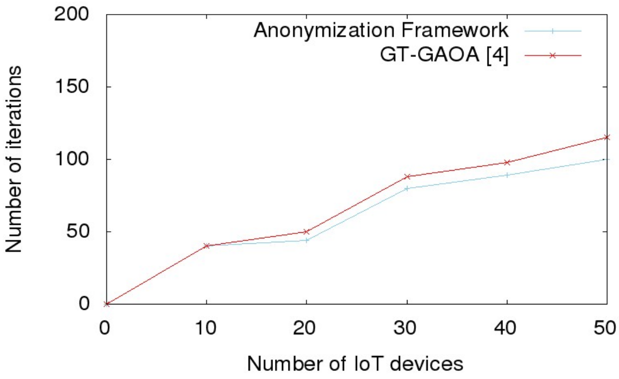


Figure 7: Comparison between computation overhead of anonymization framework and GT-GAOA [8].

C. Verification Time for Increased Policies

The current system is designed implying various algorithms that provides speed of access to the data stored in cloud and takes minimal time to communicate with fog devices and nearby cloudlets. The addition of algorithms like Triple DES handles lengthy data which in turn gives rise to the increase in speed of processing.

Figure 8 shows the verification time of Fine-grained access control [19] and anonymization framework.

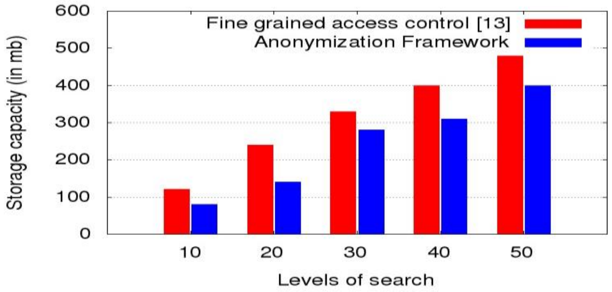


Figure 9: Comparison between verification time for increased policies of anonymization framework and fine-grained access control [13].

D. Accuracy and Scaling

The proposed model is planned for a crowd centric IoT ecosystem. The sensors are placed all over and can be scaled up maintaining the accuracy without affecting the performance. The system cannot fail in giving accurate results and the sensors can be scaled up among several sub-regions which gives more accurate with maintained privacy. Figure 9 displays the comparison of accuracy and scaling between anonymization framework and Topological multi-dimensional scaling method [Topo-MDS] [5].

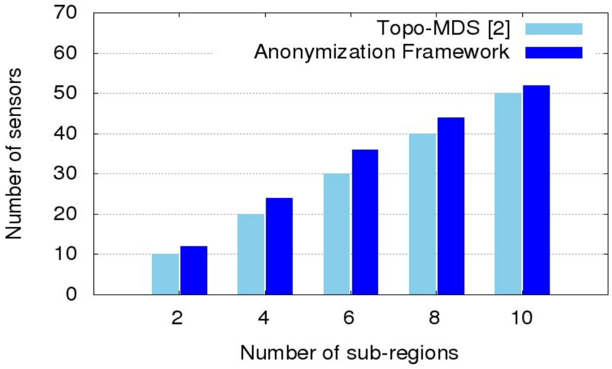


Figure 8: Comparison between accuracy scaling of anonymization framework and Topo-MDS [5].

E. Encryption Time

The introduced work encloses implementation of different algorithms in order to handle divergent forms of data. The system supports many constraints where variety and volume is not a limitation. Most of the relevant works include one level of encryption and also does not support variety of data having volume a restraint. Figure 10 demonstrates the comparison of encryption time between anonymization framework and privacy protector [17].

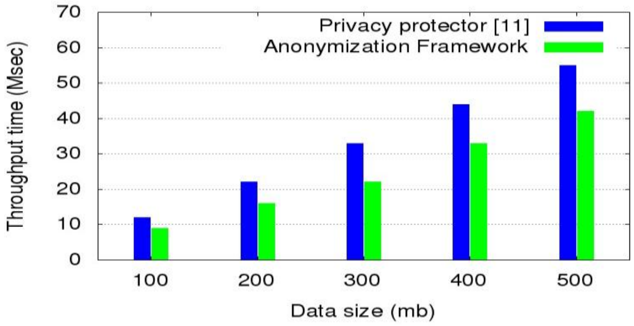


Figure 9: Comparison between encryption time of anonymization framework and privacy protector [17].

VI. CONCLUSIONS

This article proposed a user-centric security solution for IoT and Edge networks, where the security approach to secure complete systems is shifting from network-centric to edge centric. The proposed security model uses a centralized cloud server and edge controller to authenticate the IoT devices. IoT devices always initiate the security process to establish secure channels with Edge controller for further data processing. The proposed framework can potentially improve future improvements like the energy efficiency of the sensors, device failure management, and Fog layer computation offloading. Even the framework can be utilized for tracing the panic health of individuals in various distressing scenarios. The framework can be extended to social networking and social internetworking strategies, where humans and objects simultaneously interoperate.

REFERENCES

1. S. Pattar, S. K. Dwaraka, V. Darshill, R. Buyya, K. R. Venugopal, S. S. Iyengar, and L. M. Patnaik, "Progressive Search Algorithm for Service Discovery in an IoT Ecosystem," *in Proceedings of the 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, pp. 1041–1048, 2019.
2. S. Pattar, C. R. Sandhya, V. Darshill, D. Chouhan, R. Buyya, K. R. Venugopal, S. S. Iyengar, and L. M. Patnaik, "Context-oriented User-centric Search System for the IoT based on Fuzzy Clustering," *in Proceedings of the International Conference on Computational Intelligence, Security & IoT (ICCISIoT)*, pp. 1–14, 2019.
3. S. Pattar, C. R. Sandhya, V. Darshill, R. Buyya, K. R. Venugopal, S. S. Iyengar, and L. M. Patnaik, "SoCo-ITS: Service Oriented Context Ontology for Intelligent Transport System," *in Proceedings of the 2019 The 7th International Conference on Information Technology: IoT and Smart City (ICIT-2019)*, pp. 1–6, 2019.
4. C. Perera, A. Zaslavsky, C. H. Liu, M. Compton, P. Christen, and D. Georgakopoulos, "Sensor Search Techniques for Sensing as A Service Architecture for the Internet of Things," *IEEE Sensors Journal*, vol. 14, no. 2, pp. 406–420, 2014.
5. T. Yu, X. Wang, J. Jin, and K.McIsaac, "Cloud-Orchestrated Physical Topology Discovery of Large-Scale IoT Systems Using UAVs," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 5, pp. 2261–2270, 2018.
6. S. A. Hoseinitabatabaei, Y. Fathy, P. Barnaghi, C. Wang, and R. Tafazolli, "A Novel Indexing Method for Scalable IoT Source Lookup," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 2037-2054, 2018.
7. S. Pattar, R. Buyya, K. R. Venugopal, S. S. Iyengar, and L. M. Patnaik, "Searching for the IoT Resources: Fundamentals, Requirements, Comprehensive Review, and Future Directions," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2101–2132, 2018.
8. G. Tanganelli, C. Vallati, and E. Mingozzi, "Edge-Centric Distributed Discovery and Access in The Internet of Things," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 425–438, 2018.
9. M. S. Roopa, S. Pattar, R. Buyya, K. R. Venugopal, S. S. Iyengar, and L. M. Patnaik, "Social Internet of Things (SIoT): Foundations, Thrust Areas, Systematic Review and Future Directions," *Computer Communications*, vol. 139, pp. 32–57, 2019.
10. Z. Wu, Z. Meng, and J. Gray, "IoT-based Techniques for Online M2M Interactive Itemized Data Registration and Offline Information Traceability in a Digital Manufacturing System," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2397–2405, 2017.
11. W. S. Jeon, M. H. Dwijaksara, and D. G. Jeong, "Performance Analysis of Neighbor Discovery Process in Bluetooth Low-Energy Networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 2, pp. 1865–1871, 2017.

12. K. R. Venugopal and R. Buyya, “Mastering C++,” *2nd Edition, McGraw Hill Education*, 2013.
13. A. Bitar, A. Jamal, H. Sultan, N. Alkandari, and M. El-Abd, “Medical Drones System for Amusement Parks,” in *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*. IEEE, 2017, pp. 19–20.
14. S. Kurkovsky, E. Syta, and B. Casano, “Continuous RFID-Enabled Authentication: Privacy Implications,” *IEEE Technology and Society Magazine*, vol. 30, no. 3, pp. 34–41, 2011.
15. W. Zhou, Y. Jia, A. Peng, Y. Zhang, and P. Liu, “The Effect of IoT New Features on Security and Privacy: New Threats, Existing Solutions, and Challenges Yet to be Solved,” *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1606-1616, 2018.
16. J. Xiong, J. Ren, L. Chen, Z. Yao, M. Lin, D. Wu, and B. Niu, “Enhancing Privacy and Availability for Data Clustering in Intelligent Electrical Service of IoT,” *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1530-1540, 2018.
17. E. Luo, M. Z. A. Bhuiyan, G. Wang, M. A. Rahman, J. Wu, and
18. M. Atiquzzaman, “Privacy Protector: Privacy-Protected Patient Data Collection In IoT-based Healthcare Systems,” *IEEE Communications Magazine*, vol. 56, no. 2, pp. 163–168, 2018.
19. J. Zhou, Z. Cao, X. Dong, and A. V. Vasilakos, “Security and Privacy for Cloud-Based IoT: Challenges,” *IEEE Communications Magazine*, vol. 55, no. 1, pp. 26–33, 2017.
20. Y. Yu, L. Xue, Y. Li, X. Du, M. Guizani, and B. Yang, “Assured Data Deletion with Fine-Grained Access Control for Fog-Based Industrial Applications,” *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, pp. 4538-4547, 2018.
21. M. Satyanarayanan, P. Simoens, Y. Xiao, P. Pillai, Z. Chen, K. Ha, W. Hu, and B. Amos, “Edge Analytics in The Internet of Things,” *IEEE Pervasive Computing*, vol. 14, no. 2, pp. 24–31, 2015.
22. M. Satyanarayanan, “The Emergence of Edge Computing,” *Computer*, vol. 50, no. 1, pp. 30–39, 2017.
23. C. Yin, J. Xi, R. Sun, and J. Wang, “Location Privacy Protection Based On Differential Privacy Strategy for Big Data in Industrial Internet-of-Things,” *IEEE Transactions on Industrial Informatics*, 2017.
24. D. Puthal, L. T. Yang, S. Dustdar, Z. Wen, S. Jun, A. v. Moorsel, and R. Ranjan, “A User-Centric Security Solution for Internet of Things and Edge Convergence,” *ACM Transactions on Cyber-Physical Systems*, vol. 4, no. 3, pp. 1–19, 2020.
25. N. Vinutha, S. Pattar, P. Deepa Shenoy, and K. R. Venugopal, “SliceNetAD: Slice Selection based Convolution Neural Network Model for Classification of Alzheimer’s Disease,” *International Journal of Image Mining*, vol. XX, p. XX, 2020.
26. E. Grande and M. Beltran, “Edge-Centric Delegation of Authorization for´ Constrained Devices In The Internet of Things,” *Computer Communications*, vol. 160, pp. 464–474, 2020.
27. L. Yang, X. Chen, S. M. Perlaza, and J. Zhang, “Special Issue On Artificial-Intelligence-Powered Edge Computing for Internet of Things,” *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9224–9226, 2020.
28. KR, Venugopal, and L. M. Patnaik. "Moving Vehicle Identification Using Background Registration Technique for Traffic Surveillance." *Proceedings of the International MultiConference of Engineers and Computer Scientists*. Vol. 1. 2008.
29. Tarannum, S., Aravinda, B., Nalini, L., Venugopal, K.R. and Patnaik, L.M.,*“*Routing Protocol For Lifetime Maximization of Wireless Sensor Networks*”, In 2006 International Conference on Advanced Computing and Communications*, pp. 401-406, 2006.
30. Venugopal K R, Rajan EE, Kumar PS “Performance Analysis of Wavelength Converters in WDM Wavelength Routed Optical Networks”, *In Proceedings. Fifth International Conference on High Performance Computing*, pp. 239-246, 2000
31. Venugopal, K. R., E. Ezhil Rajan, and P. Sreenivasa Kumar. "Impact of Wavelength Converters in Wavelength Routed All-Optical Networks." *Computer communications* vol. 22, no. 3 pp. 244-257, 1999
32. Kanavalli, A., Sserubiri, D., Shenoy, P. D., Venugopal, K. R., & Patnaik, L. M, “A flat routing protocol for sensor networks,” *In 2009 Proceeding of International Conference on Methods and Models in Computer Science (ICM2CS)*, pp. 1-5, 2009
33. Tarannum, S., S. Srividya, D. S. Asha, R. Padmini, L. Nalini, K. R. Venugopal, and L. M. Patnaik. "Dynamic Hierarchical Communication Paradigm for Wireless Sensor Networks: A Centralized, Energy Efficient Approach." *In 2008 11th IEEE Singapore International Conference on Communication Systems*, pp. 959-963, 2008.

AUTHORS PROFILE

Santosh Pattar received the Bachelor of Engineering degree in computer science from the RNS Institute of Technology, Visvesvaraya Technological University, Bengaluru, India, in 2012 and the Master of Engineering degree in bioinformatics from the University Visvesvaraya College of Engineering, Bangalore University, Bengaluru, in 2015, where he is currently pursuing the Ph.D. degree in computer science. His current research interests include the Internet of Things, with a focus on device search and discovery techniques, cloud computing, data mining, and bioinformatics.

Rajkumar Buyya is a Redmond Barry Distinguished Professor and the Director of the Cloud Computing and Distributed Systems Laboratory, University of Melbourne, Australia. He has authored over 900 publications and seven text books, including the book entitled *Mastering Cloud Computing* published by McGraw Hill, China Machine Press, and Morgan Kaufmann for Indian, Chinese, and international markets, respectively. He is one of the highly cited authors in computer science and software engineering worldwide with an H-index of 138 and over 104,0000 citations. Software technologies for cloud computing developed under his leadership have gained rapid acceptance and are in use at several academic institutions and commercial enterprises in 40 countries around the world. He was a recipient of the Web of Science Highly Cited Researcher Award by Thomson Reuters in 2016 and 2017, and Scopus Researcher of the Year 2017 with Excellence in Innovative Research Award by Elsevier for his outstanding contributions to Cloud computing. He is a Fellow of IEEE.

K. R. Venugopal received the Bachelor of Engineering degree from the University Visvesvaraya College of Engineering (UVCE), the master’s degree in computer science and automation from the Indian Institute of Science Bangalore, and the Ph.D. degree in economics from Bangalore University and in computer science from the Indian Institute of Technology Madras. He is currently the Vice Chancellor, Bangalore University, Bengaluru. He has a distinguished academic career and has degrees in Electronics, Economics, Law, Business Finance, Public Relations, Communications, Industrial Relations, Computer Science, and Journalism. He has authored and edited 76 books on Computer Science and Economics, which include Petrodollar and the World Economy, C Aptitude, Mastering C, Microprocessor Programming, Mastering C++, and Digital Circuits and Systems. He has been granted 101 patents. During his four decades of service at UVCE he has over 900 research papers to his credit. His research interests include computer networks, wireless sensor networks, parallel and distributed systems, digital signal processing, and data mining. He is a ACM Distinguished Educator and an IEEE Fellow.

S. S. Iyengar received the Ph.D. degree from MSU, USA in 1974. He is currently a Ryder Professor with Florida International University, USA. He was a Roy Paul Daniels Professor and the Chairman of the Computer Science Department, Louisiana State University. He heads the Wireless Sensor Networks Laboratory and the Robotics Research Laboratory with USA. He has been involved with research in High Performance Algorithms, Data Structures, Sensor Fusion, and Intelligent Systems. He has directed over 40 Ph.D. students and 100 post graduate students, many of whom are faculty of Major Universities worldwide or Scientists or Engineers at National Labs/Industries around the world. He has published over 900 research papers and has authored/co-authored six books and edited seven books. His books are published by Wiley, CRC Press, Prentice Hall, Springer-Verlag, IEEE Computer Society Press. One of his books entitled *Introduction to Parallel Algorithms* has been translated to Chinese. He is a fellow of ACM and an IEEE Fellow.

L. M. Patnaik is currently a Senior Scientist with Consciousness Studies Program, National Institute of Advanced Studies, Indian Institute of Science, India. He was the Vice Chancellor with Defense Institute of Advanced Technology, Pune, India, and has been a Professor since 1986 with the Department of CSA, Indian Institute of Science, Bengaluru. During the past 35 years of his service at the Institute he has over 1150 research publications in refereed international journals and conference proceedings. His areas of research interest have been parallel and distributed computing, mobile computing, CAD, soft computing, and computational neuroscience. He was a recipient of 20 national and international awards; notable among them is the IEEE Technical Achievement Award for his significant contributions to High Performance Computing and Soft Computing. He is a fellow of all the four leading Science and Engineering Academies in India and a fellow the Academy of Science for the Developing World.