TMP UNIVERSAL JOURNAL OF RESEARCH AND REVIEW ARCHIVES

VOLUME 3 | ISSUE 2 | YEAR 2024 | APR-JUN 2024

RECEIVED DATE	REVISED DATE	ACCEPTED DATE
20/03/2024	23/04/2024	17/05/2024

Article Type: Research Article

Available online: www.tmp.twistingmemoirs.com

ISSN 2583-7214

PRESERVATION OF INTERNET OF THINGS PRIVACY IN THE COLLECTION PHASE (CASE STUDY: AGRICULTURAL PRODUCTS)

[1]Iman Naderi[*], [2]Setare Bazargan, [3]Dr Hossein Yarahmadi

[*] 1 Master of science computer, Department of Industrial Engineering,Master of science Engineering, Boroujerd Branch, Islamic Azad University,Boroujerd, Iran

[2] Supervisor, Department of Computer, Master of Science Computer Engineering, Boroujerd Branch, Islamic Azad University, Boroujerd, Iran

[3] Department of Computer, Master of Science Computer Engineering, Boroujerd Branch, Islamic Azad University, Boroujerd, Iran

ABSTRACT

Objective and Background: The Internet of Things (IoT) devices store a great volume of data. Therefore, with the advancements in technology, some new protection guidelines have to be developed to preserve data privacy and collection. The present study aimed to determine a secure protective framework using the BTD structure in a specified zone, and provide the idea of hiding the sensory data of an operator device to preserve data privacy without losing the data integrity.

Methodology: Two simple and enhanced methods were used. The BTD framework and estimated ground truth were used in both modes. The enhanced method allows the users to use a random weight (variance) when combining their sensory data with the estimated ground truth provided by the agent. We ran the BTD framework in the CloudSim (network simulation framework). Each simulation is run for 500 virtual minutes with 12 iterations. The target zone is set to 50×50 meters. The test results were collected on real-world data traces from sensory systems. In addition to the BTD structure, the Voting method has been also used, and to do so, majority voting and CRH truth discovery method (Conflict Resolution on Heterogeneous Data) which does not take any measures to break the security of the sensor during the process is used.

Findings: The comparative results show a 0.70-0.71 error level between the BTD and other functions, at least. As long as there is at least one reliable source, the BTD will have an error level of 0. Simulations show that BTD has a worst-case error of 0.05 and a weighted variance of +/- 5%. Based on previous frameworks, the BTD is improved by three requirements of a population identified in the IoT: Preservation of privacy for the device user, data integrity for the data collection group, and low-cost computation on the user device.

Keywords: Internet of Things, privacy, data collection, data volume

CORRESPONDING AUTHOR

Name: Iman Naderi

Affiliation: Master of science computer, Department of Industrial Engineering, Master of science Engineering, Boroujerd Branch, Islamic Azad University, Boroujerd, Iran

INTRODUCTION

IoT is a new technology which is becoming ubiquitous and will affect the human life. IoT is connecting all things to each other and to human beings as well as their identification and discovery under an integrated network. Naturally, creating such a network is associated with many risks. The World Wide Web, which has been around for years, is still full of security flaws that endanger the property, and even the lives of people. Thus, it can be said that the security of the IoT is a key discussion in the implementation of such technology and requires intensive research to preserve people's privacy. The concept of the IoT, which has been introduced as a new concept in recent years, is only at the starting point. Various issues such as standardization, technical problems, IoT costs, and privacy preservation require further discussions by researchers and testing companies [1].

A set of standards, protocols, devices, and technologies required to connect and transmit data between smart devices (to each other and human beings) at a global level is called "IoT" [2]. In fact, the IoT is a concept in which smart things are equipped with small sensors, actuators, and microprocessors and are capable of performing multiple processes and communicating with each other [1]. A technology called 'RFID' is used in the structure of IoT. RFID technology has actually revolutionized the embedded communication model, which helps configure the microprocessors used in wireless communication. There are two RFID tags named active and passive tags. The active RFID tags use an internal power source (battery) which is inside the tag to provide the energy to tags and the circuits related to them, while a passive tag requires a reader to provide the energy to its components.

The hardware of the IoT includes the RFID tags, Zig Bee, Bluetooth, and sensor nodes. Among the specifications of the RFID tags, authentication and a unique specification that implements the exchange of information between tags and readers in wireless communications can be named. The Zig Bee includes a radio, microcontrollers, and simple rules that are small in size and reliable, with a low and cheap energy consumption. The Bluetooth includes a frequency spectrum that allows the two devices to connect wirelessly.

The Internet of Things, by collecting sensors and different objects, can create communication between them without human intervention. With the increase in communication level and resolving the requirements, security concerns are also increasing rapidly. However, the research in the field of IoT is still at the initial levels and requires broader discussions in the field of security threats and related vulnerabilities [3]. The most important challenge in the field of IoT is the provision and acceptance of a comprehensive architecture for it that, in addition to covering the functional and communicative issues, also addresses the problems related to security, privacy, and reliability.

Privacy is a subjective term whose degree and definition often differ from one stakeholder to another, from one context to another, and from one domain to another. Therefore, in a highly dynamic and large-scale IoT ecosystem, the framework of privacy should be completed by a legal framework [4]. It requires solutions that can iterate the privacy protection in a smart configuration, which can be otherwise permitted by the users on real occasions. In such solutions, the principles of privacy protection can be met using recommendations to minimize the data and target specifications during data collection. In addition, other security threats during the combination and collection of the data can be used by at least a set of protective mechanisms

that have been proven in various settings [5].

Regarding the fact that the use of IoTs is increasing in all parts of daily life, e.g., in the health sector, the confidentiality of a large volume of personal data of people which are stored and preserved is of great importance [6]. In the cellular design, several devices provide an access point for the devices that have lower technology for data exchange with the network, as a single unit. Among the important challenges in the IoT, the privacy, authentication, reliability, access control, data storage, and system configuration and management can be named [7].

To preserve data security and privacy, Yoe et al. [2015] have introduced a method in which none of the data pairs are the same in terms of the ABE (Attributed-Based Encryption), and will not have the same address based on a design based on the ECC. Bose et al. (2015) and Raza et al. (2013) proposed a light authentication scheme to create a security policy to control the level of reliability, assess the rank of security obtained from the sensors, and preserve the content in the transmission security. Ziegeldorf et al. [8] provided solutions such as distributed access control and identification all over the streams, however, they do not provide details on how these solutions should be produced. The present study is derived from this study. It is a description of "informed consent" concerning data integration as a solution to fill the gap between the "context" of data collection and use.

Information agreements are an important element to protect the data in the ICTs, because the consent of an individual-data (e.g., a citizen) is usually necessary for a third-person, to legally process personal information [9]. To provide informed consent for personal information, citizens should have a clear understanding of how their system data is being used.

While data connectivity is considered a security threat, it also increases the intensity of other privacy threats such as user profiles, locating, tracing, and identification [10]. The gravity of the information link directly affects the risk of the user profile. More precise details in which the data can be linked increase the risk of user profile as a feature dependency. Therefore, user profiling has become a more pressing concern due to the granular detail at which IoT collects data for big data business models. With the increase in the number of various types of devices that connect to the IoT network, independent interactions between the number of privacy violations have increased. Article No. [11] deals with the inclusive nature of personal information with the emergence of the IoT and obviating the identification concerns arising from linking an identifier to a user's identity. In addition, privacy threats such as location and tracking are high due to the extensive use of GPS, internet traffic, and smartphones. For example, the devices' metadata usually contain the device location which, with added details to users in different social network systems (even if anonymous), can reveal sensitive information about users.

To identify the threatening risk factors when integrating the data that cause data link, the tasks on the "unlikability" [12], the best rules and methods for "analysis of large data" [13], and legal concerns about the cyber security in IoT [14] have been considered.

The PPTD framework is focused on the preservation of users' privacy to participate in the sensory systems. This structure hides a high share of sensory data available in the users' device and requires a large sample space to correct the ground truth. The PPTD is especially used to obtain weight calculations and is a part of the revealing process of random ground truth. The calculations are small in isolation but can incur huge overhead costs to users' devices on large scales. The real concern is the possibility of invading the collected data. Yet, invasions by the third party that have access to the collected data address are more important [15]. Based on the above-mentioned, the present study aimed to preserve the privacy of the IoT in the collection phase.

METHODS

The Balanced Truth Discovery (BTD) has been proposed to obviate the serious limitations of previous structures used for user privacy and integrity of the collected data in IoT. By creating limitations for users' devices in the calculation of random ground truth, optimizations are made in both calculation costs and allocation of space for the devices. All BTD structure objectives are increasing to create security in sensory systems.

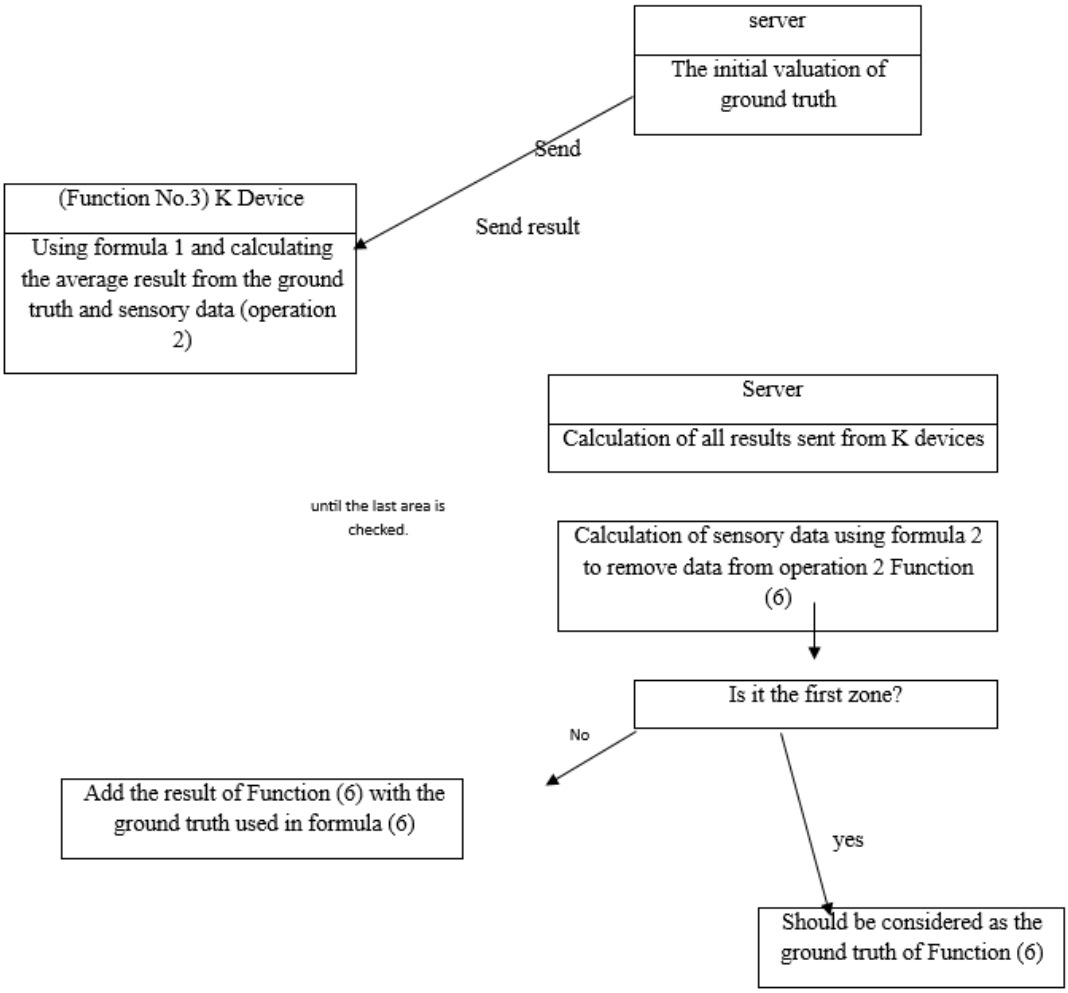


Figure 1: General BTD framework

The initial run in the server is done by the data-collecting group. Using this framework, the random value will not affect the ground truth at all. Yet, it is suggested to use a random value that creates the contextual concept. The initial value is used in a way that no individual sensory data is sent to the server. Algorithms Nos.1 and 2 are reliable for processing the calculation user's zone in a specific zone.

The estimated ground truth the agent is going to use is sent to User K. These devices represent a zone and are processed together. Therefore, no individual sensory information is used by the server. Number K represents the zone size. the value of K can be chosen by the collection group. A quantitative value for K guarantees that the estimated ground truth is repeatedly updated. The method of K devices for a specific zone completely depends on the collection group for decision. There might be a logic by which we can divide the population into n zones with K size, based on the location, device specifications, or other cases that can be easily accessed by the server. The unknown degree required varies from text to text. Any user device K calculates their sensory its own sensory computability and the ground truth established by the agent. This collection is

calculated in a way that if there are only two sections, they are evaluated by 50%, and the agent is evaluated by the remaining 50%. The formula is as follows:

Data sensor Xk = (1)* 0.5 + x * 0.5 cc (1)

Any of the user devices K sent the data to the agent when the individual calculations are completed. The agent collects the results. The simulated BTD framework used for analysis easily finds the results without weight validity. The zone's processing user hides the data at the individual level, however, the bigger image (one of the K devices) is still clear. For example, the location and change of location of a group of people in New York can be investigated. If someone tries to learn the location and habits of someone else, the data in the group cannot be detected.

Estimation Method:

The initial valuation of the ground truth is done randomly (in a content-based zone). If it is not investigated correctly, the issue of data validity is presented. The BTD framework addresses this problem using a method that collects the ground truth from the K users in a zone of sensory data. As a reminder, it should be said that the calculation of the sensory data user divides the device user into two equal parts with the ground truth presented by the agent. Thus, the collected sensory data from the K user devices in a zone, the sensory data of the semi-user device, and half of the ground truth are estimated. The following equation estimates the ground truth and the agent aggregates it with the sensory data.

$$sz = xz * 0.5 + [xz - 2(x - xz)] * 0.5 \quad (2)$$

Where sz is the collective sensory data of zone z, xz is the results collected from the user's devices (i.e., before extraction), and x is the current estimated ground truth. This estimation is not only used for extraction of the random value, but also it adjusts the estimated ground truth to the initial state. Also, the following equation is used to update the ground truth:

$$x = sz * k/c + x * (c-k)/c \quad (3)$$

k is the number of user's devices in a zone and c is the number of devices of a user that have participated in the population measurement system including the k user's device in the processing zone. The estimation method used in the equation allows for preservation of the privacy without losing data integrity through the random initial valuation and hiding the sensory data of a user's device. The results obtained from this method are theoretically a precise copy of the sum of all the sensory data of c devices.

Probable Changes for Security:

The BTD framework does not create a specific encryption method. Yet, it is suggested to use an encryption system based on the BTD to correctly preserve the privacy of all parties involved. A potential threat can exist if a third party separates the data sent from the agent to the device user and from the device user to the agent. If an invader who has eavesdropped the data gets to know the nature of the BTD, it can use the estimation method to obtain the sensory data.

The BTD framework is a truth detection method that can calculate the reliability of the user's devices and determine the truth by the device to check whether the data are changed or not. Each device has an equal rule (decision-making) in the integration of the ground truth. The weight reliability theoretically protects the data integration. Common weighting methods can be found in [11], [16], and [17]. For example, [18] provides a reliable weighting method by the use of which any information about the device weight should be sent to the user device, while the BTD needs to be calculated by such method and be merely used on an agent. In the BTD, we calculate

the value of the sensor state to check whether a person is changed during transmission or not. The main idea is that if the device's transmitted value is close to the estimated ground truth, the value of the device state can be presented at a high level. Usually, the values of the sensor state are calculated as follows:

$$S_k = \log\left(\frac{\sum_{k'=1}^{K}\sum_{m=1}^{M} d(x_m^{k'}, x_m^*)}{\sum_{m=1}^{M} d(x_m^k, x_m^*)}\right) \quad (2)$$

d (.) is the distance function that estimates the difference between the values of sensors' observations and ground truth. xm * [4]. d relies on specific sensitive applied scenarios (events or schedules). The proposed framework is considered to be employed with applied population measurement programs which are the continuous sensory data. We use the following normalized approximate distance function:

$$d(x_m^k, x_m^*) = \frac{(x_m^k - x_m^*)^2}{std_m} \quad (3)$$

When using these equations together, we can put the weight of a user's device on the standard deviation and the normalized squared distance function. Also, it should be noted that the device's reliability may vary from context to text [19]. To ensure the best possible data integration, it is suggested to calculate the weight reliability capability by the agent for each user tool in any context (i.e., if the agent finds two or several ground truths by various sensors).

Findings:

The proposed BTD framework was evaluated to preserve the balance between data integrity and privacy.

Algorithm 3: Enhanced Method (User calculation)

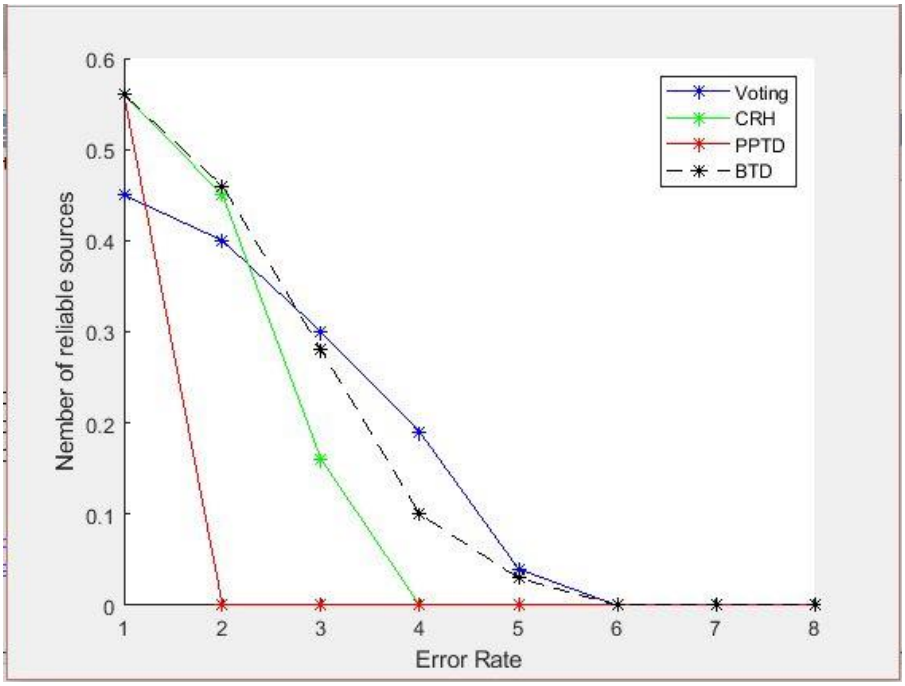
Input: Estimated ground truth: *x*, Variance:
Output: Result

```
1 if v is not a sufficient value do
2         Replace v with desirable value;
3 end
4 weight = random between 50-v and 50+5;
5 Result = (sensory data * weight/100) + (x * 100-
          weight/100)
6 return Result;
```

Simulation Configurations:

We have run the BTD framework in a CloudSim format (network simulation farmwork) that supports dual radios for each node [12]. Each of the simulations is run for 500 virtual minutes and is iterated 12 times. In addition, since we focused on the mobile phones and their sensors (i.e., the mobile population), each node of the simulated sensor is programmed to randomly move 4 meters per minute. The target zone is 50×50 meters. The results of the testing of the real

data were obtained from the sensory systems [11]. For testing and measurement, the simulation environment that runs the BTD framework uses three classes: 1) Simulation class, 2) agent, and 3) device. The agent class defines an agent thing that owns the required methods to imitate the behaviors of an agent using the BTD introduced in the present study. The device class similarly imitates the behaviors of a user. The simulation class provides the agent with devices required for the calculation of the ground truth. The simulation program is generally equipped with population sensor simulation using the BTD framework, with/without the enhanced method. We provide a set of simulation samples with a simulation set:
500 devices, zone size: variable, zone count: 500 devices/ zone size, variance" +/-0.5%. The Simulation No.2 objective: Calculation of the effects of zone size when using the enhanced method on the data precision. Zone sizes 2-50, used in simulation. The device storage is constant.



As shown in Figure 3, the simulation results indicate that the increase in zone size leads to an increase in difference (reduction in precision), while it preserves a fixed device account. The value of effects is small (about +/-0.01% precision in zone size of 30). However, some specific contexts may require higher precision or quantitative storage of devices.

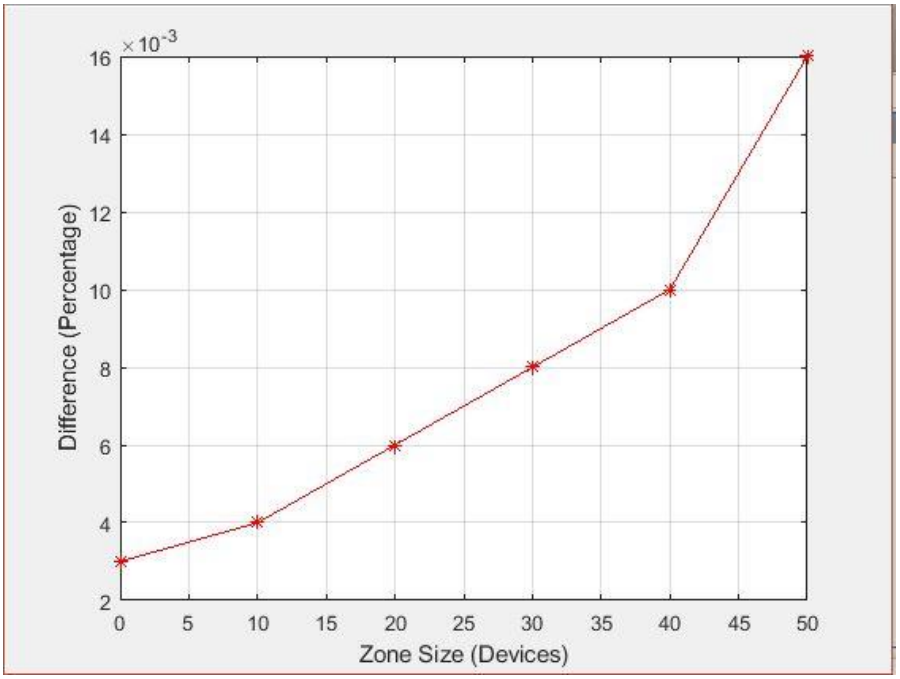


Figure 3: Zone size compared to fixed device storage difference

The truth discovery value is executed using the Threshold Security Paillier toolbox [11. We also consider the voting framework for the comparison [8]. Voting to eliminate conflict is used to make decisions based on the collected data, which is used to perform majority voting, so that the information with the highest number of occurrences, mean, or median is considered the correct answer. In voting, it is assumed that all sensors are equally reliable, and therefore votes from different sensors are evaluated uniformly. We employed a network-based voting algorithm from [13]. The tests for PPTD and BTD, which were previously discussed, are not the same. The results of the two cannot be compared directly. Yet, the conclusion can be made from the results of PPTD tests and BTD simulation as well as the analysis of the CRH [15] and a population-sensitive framework used in PPTD.

Figure 2 shows the results of a comparison between the BTD and other functions. It shows a minimum error level of 0.70-0.71. Regardless of the parameter of rounding L, when the mean absolute error (measured by the mean absolute gap between the estimated results and the ground truth) in PPTD is mixed with CRH (blue line), the error is 0.70-0.71.

Figure 3 indicates that as long as there is at least one reliable source, the BTD (black line) has an error level of 0. It shows that PPTD creates an approximate error of 0.71 without enhancing privacy preservation. The simulations indicate that BTD has an error level of 0.05 and a variance of +/-5%. “at worst”. Yet, the use of lower-weight variances can bring an error level of much smaller than 0.001. As mentioned before, we cannot directly compare the results of BTD simulation and PPTD tests. The errors produced by the PPTD still provide evidence for improvement of the precision of the data provided by the BTD. We should consider that BTD also estimates steps to guarantee data integrity through the protection of the users with the intention of changing the ground truth.

RESULTS

In the present study, a Balanced Truth Definition (BTD) framework was proposed to maintain the balance between the initial data and integrity in the IoT. The enhanced method shows this problem by making the individual's sensory data unclear. This method allows the users to use a random weight when combining their sensory data with the estimated ground truth provided by the agent. Investigating the simulation process, we faced two cases: Mean Error Rate and Difference Zone. The first one is indicative of the error level obtained for the data exchange, whose values have been considered per change in the number of reliable sources. However, the second one, which is the source difference, is considered per change in the size of the zone undergoing the changes.

Evaluating the results obtained from the simulation, we found that as long as there is at least one reliable source, the BTD (black line) has an error level of 0. It shows that PPTD creates an approximate error of 0.71 without enhancing privacy preservation. The simulated BTD framework depicts the results simply and clearly, and without weight reliability, for ease of analysis. The zone processing user hides the data at the individual level, however, the bigger image (one of the K devices) still remains clear. The running simulation of the BTD indicates that precise data can be obtained per mille while preserving the device user's privacy.

Regarding the ever-increasing developments in IoT advanced technologies and the increase in the data exchange between smart devices, the accuracy of receiving data will increase to the extent that we can reduce the number of errors created during data exchange between devices.

CONCLUSION

This study revealed that the natural compounds Digoxin and Warifteine among the selected plant compounds have better binding free energies with the 6Y2F protein of SARS-CoV-2. Although the molecular binding results of Ganoderic acid C2, Ursolic Acid, Lupeol, Kuwanon B, Emodin-8-glucoside, Adonitoxin, Kuwanon E, and Isohemiphloin are lower than the first two compounds, the analysis of RMSD parameters, interactions, number of hydrogen bonds, and RO5 criteria and their non-toxic properties showed better performance. These compounds have a better potential as antiviral plant chemicals and to solve respiratory, inflammatory, infectious, and coagulation problems, which may prevent the proliferation of the virus or help to treat this disease. These 9 inhibitors are appropriate candidates as drugs for inhibiting the activity of the primary enzyme of the SARS-CoV-2 coronavirus for clinical and laboratory studies. However, the conducted studies are theoretical. Experimental work is required to ensure the accuracy of the data, and the results of this research alone cannot claim that the introduced compounds can inhibit the COVID-19 protease.

REFERENCES

1. . Zhou Book:internet of things in the cloud ,a middleware prespective Honbo zhou,2013
2. Fadele Ayotude Alab,mazliza Othman , Ibrahim Abaker Targio Hashem ,Faiz Alotaibi "internet of things security :A survey.2017
3. IJCN-265:Towards Internet of things survey and Future vision , O.said , M.masud international Journal of computer networks. 2013
4. S.Li, L.Da Xu , and S.Zhao , "the internet of things: ," information systems frontiers , pp.1-17, 2014
5. Nishtha Madaan , Mohd Abdul Ahad, Sunil M.Sastry : Data integration in Iot ecosystem : information linkage as a privacy threat. 2017
6. Rolf H.weber " internet of things :privacy issues revisited ".2015
7. Ovidiu Vermasan & Peter Friess, internet of things –from research and innovation to market deployment ,June 2014

8. Md Zakirul Alam Bhuiyan, Tian Wang, Thaier Hayajneh, and Gary M. Weiss, Department of Computer and Information Sciences, Fordham University, New York, 10458 USA College of Computer Science, and Technology, Huaqiao University, Xiamen, Fujian 361021, China,2017:" Maintaining the Balance between Privacy and Data Integrity in Internet of Things"
9. Rifkin, Jeremy, the zero marginal cost society: the internet of things, the collaborative commons ,and the eclipse of capitalism Macmillan ,2014
10. Ryan Roontz ,the internet of things "smart" products demand a smart strategy ,2015
11. Miao, C., Jiang, W., Su, L., Li, Y., Guo, S., Qin, Z., Xiao, H., Gao, J., and Ren. K. 2015. Cloud-Enabled Privacy- Preserving Truth Discovery in Crowd Sensing Systems. In Proc. of ACM SenSys, 2015.
12. Helgason, O., and Kouyoumdjieva, O. 2011. Enabling multiple controllable radios in omnet++ nodes. In Proc. of ICST, 2011, pp. 398–401.
13. Pai, H. T., and Han, Y. S. 2008. Power-Efficient Direct- Voting Assurance for Data Fusion in Wireless Sensor Networks. IEEE Transactions on Computers, vol. 57, no. 2, pp. 261–273, 2008.
14. Bhuiyan, M. Z. A., Wang, G., and Choo, K. R. 2016. Secured Data Collection for a Cloud-Enabled Structural Health Monitoring System. In Proc. of IEEE HPCC 2016.
15. Wang, S., Wang, D., Su, L., Kaplan, L., and Abdelzaher, T. F. 2014. Towards cyber-physical systems in social spaces: The data reliability challenge. In Proc. of IEEE RTSS 2014.
16. Bhuiyan, M., and Wu, J. 2016. Trustworthy and Protected Data Collection for Event Detection Using Networked Sensing Systems. IEEE Sarnoff, 2016.
17. Li, Q., Li, Y., Gao, J., Su, L., Zhao, B., Demirbas, M., Fan, W., and Han. J. 2014. A confidence-aware approach for truth discovery on long-tail data. Proceedings of the VLDB Endowment, 8(4):425–436, 2014.
18. Bhuiyan, M. Z. A., and Wu, J. 2016. Event Detection through Differential Pattern Mining in Internet of Things. In Proc. of IEEE MASS 2016.
19. Ma, F., Li, Y., Li, Q., Qiu, M., Gao, J., Zhi, S., Su, L., Zhao, B., Ji, H. and Han, J. 2015. Faitcrowd: Fine grained truth discovery for crowdsourced data aggregation. In Proc. of ACM SIGKDD, 2015.