Making Decision on Sharing Forensic Data with the Fuzzy Logic Approach*

Nancy Scheidt
School of Computing
University of Portsmouth
Portsmouth, United Kingdom
nancy.scheidt@port.ac.uk

Gulsum Akkuzu
School of Computing
University of Portsmouth
Portsmouth, United Kingdom
gulsum.akkuzu@port.ac.uk

Mo Adda
School of Computing
University of Portsmouth
Portsmouth, United Kingdom
mo.adda@port.ac.uk

***Abstract*—Internet of Things (IoT) devices advance continuously and with such development so do the issues and complexity in their forensic investigation. Worldwide investigators face challenges in retrieving evidence from the IoT environment, which includes Devices, Networks and the Cloud. Creating a report which then can be presented in the court of law is one of the most crucial steps during such forensic investigations. In this article, we propose a Hybrid Forensic IoT Server model which registers devices and stores their data for examination purposes. Furthermore, we developed models to calculate values for report accuracy, evidence accuracy and investigator expertise as well as a new way of decision-making in forensic data sharing utilising the fuzzy logic decision making system.**

Keywords—forensic investigation, IoT server, fuzzy logic, forensic data sensitivity, investigator expertise, forensic report

I. Introduction

Every day, the number of Internet of Things (IoT) devices rises rapidly per person and will continue do so and such leads to the fast development and improvement of IoT devices which also opens more doors for the variety of using these devices offer [1]. However, such developments also provide a variety of additional challenges in digital forensics and also extend the intricacy of accessing information of devices if required. Therefore, evidence collecting issues of the IoT environment, which includes devices, networks or the cloud, regarding persons of interest is something investigators face worldwide. Research on how to tackle this issue focuses mainly on solutions which define principles addressing IoT forensics, improving cloud cybersecurity and device-specific techniques due to the difference IoT devices offer when needing to be investigated [2, 3].

Additionally, [4] acknowledge challenges with large data sets and evidence collection in an IoT server environment and propose a server model which can register devices to provide further solutions to IoT forensics. Therefore, current research demonstrates that IoT Forensics is handled comparable to digital forensics concerning the investigation and evidence extraction and is not independent as of yet.

Generally, pieces of evidence include images, emails, banking details, documents, text messages found on an IoT device and a necessary report, by one or multiple investigators, needs to be created to finalise an official conclusion at the end of evidence analysis. Therefore, our solution to this problem of efficiently collecting and investigating evidence is the Hybrid Forensic IoT Server. Especially in a Smart World (SW), which we are approaching steadily, this will provide great support to investigation purposes, ease crime detection and prevention and allow a wider reach in terms of communication and teamwork due to implementing a high number of IoT servers worldwide.

Considering previous research, studies need improvement in device management, evidence storage, and its analysis and sharing as well as investigator collaboration processes to be made easier for a smoother, more organised and well developed forensic investigation procedure. These methods have not been linked together and to some degree not been developed thoroughly as we will be demonstrating. Moreover, we are not aiming to dispose of current digital forensic investigation methods and tools but rather aim to improve the investigation process in criminal cases of a larger scale and a more complex structure. To tackle and cover the issues raised and found in previous research, we aim to answer the following questions to offer a solution and which also clarify our contribution to this topic matter:

- Can we evaluate investigators' expertise in a forensic investigation?
- Can we evaluate the accuracy of evidence in a forensic investigation?
- Can we measure the accuracy of a forensic report in a forensic investigation?
- Can we use fuzzy logic in order to make a decision on sharing forensic data while the forensic data is shared among countries?

The rest of the paper is structured as follows; In section II, we discuss related work. In section III, we introduce the hybrid

forensic IoT server with its architectural structure. In Section IV, we give the mathematical modelling of this work. We then give the fuzzy logic decision making on forensic data sharing. In section VII, we conclude the paper and discuss future work.

II. Related Work

A. IoT Forensics Investigation

Several papers which are focusing on Digital and IoT forensics and how to most securely retrieve evidence have been published. For this paper, our understanding of IoT forensics is, it being an investigation of the three different levels: Cloud, Network, and Devices [5]. Such is a highly complex procedure and raises a variety of security as well as investigation issues. Research on this topic conducted by [4] as well as [6] are closest related to the aim of our paper in forms of minimising the complexity of IoT Forensics on the three-level basis. [4] research presents a Forensic Aware IoT model which supports forensic investigation in a reliable way within the IoT environment and provides a slightly different perspective to a forensic procedure by creating an environment which stores devices and its data on the cloud and enable investigators to easily retrieve evidence when physical devices are not accessible. To ease the investigation and the high amount of available investigating tools, [6] proposed a management model of forensic technologies to ease access to such. However, considering the additional research on the variety of aspects and issues surrounding IoT forensics, it is apparent that improvement is crucial for such models to work efficiently, especially in a world where Smart Cities (SC) are rapidly increasing [7].

B. Data Sharing

Due to this paper's focus being on the improvement of IoT Forensics, a crucial part of such is data sharing, be it within countries but also internationally. In this day and age, storing data in the cloud increases the ease of storage as well as sharing information with others all over the globe [8]. However, this also carries a lot of data security and privacy risks with it [9]. [8] proposed a more secure way of encryption to ensure the security of data and privacy and [10] propose a 5-step data sharing scheme to improve such further. If we take these measures a step further and consider data sharing from a police investigation point of view, additional issues need to be considered, such as the trust level of different countries with each other possibly being a hindrance when inquiring information for investigation purposes. The agreement between the UK, Australia, Canada, New Zealand and the United States regarding fingerprint data sharing would be beneficial guideline as one of the steps necessary to implement in regards to digital data worldwide [11].

III. Hybrid Forensic IoT Server

A. Architecture

As previously touched upon, our proposed Hybrid Forensic IoT server is a combination of a hierarchical and distributional system and we will refer to it as the Hybrid Server further on. Therefore, in a world where SC are implemented constantly, we suggest each country to be equipped with a main IoT server. Moreover, for a faster exchange of data, all countries will have additional servers for each of their regions (sub-server) as visualised in Figure 1.
Due to this kind of model being heavily dependent on data sharing within as well as among countries the investigation process was considered as well and was categorised into three different models.

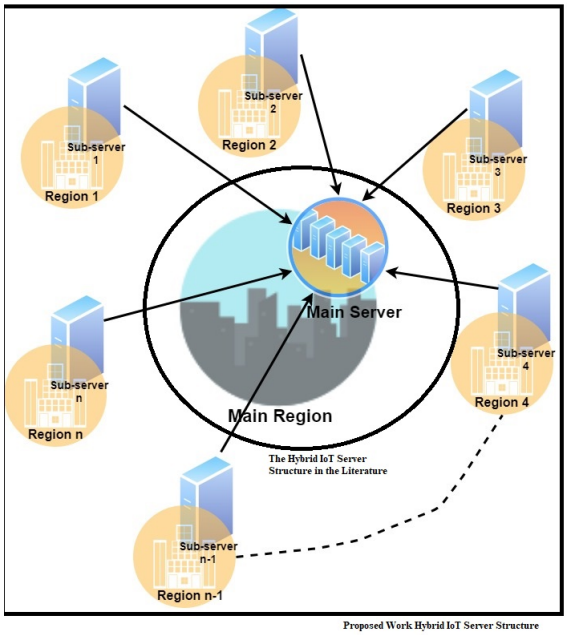

Fig. 1. Hybrid Forensic IoT Server Structure

Classical Investigation Model
Research by [12] demonstrates several classical digital forensic investigation models and shows that the majority of models present and focus on a very similar process: Collection, Examination, Analysis and Reporting of evidence which is all conducted in one laboratory [13].

Distributed Investigation Model
Building on the classical approach and considering digital forensic challenges inspired research by [14] and [15] proposing IoT digital forensic investigation models. While considering the possible issue of data sharing as well as our proposed model of the IoT server, the second model was developed.

If data sharing within or among countries is prohibited there is still the possibility to continue a police investigation. In our model of distributed investigation, this can be met with the exchange of the final analysis reports of an investigation rather than all of the evidence. This is an efficient solution to protect data from random access by individuals, as well as data contamination, and further steps against individuals only have to be taken if suspicious data was found during an examination. Figure 2 visualises a brief overview of the structure of the distributed investigation process and shows the four steps of data sharing within the Hybrid Server environment:

1) Country A requests data from Country B's Main-Server
2) Main- and Sub-Server communicate about request
3) Value of Data Sensitivity is calculated by the Server System
4) Country B decides if data is to be shared or not.

Multi-Tasking Investigation Model
Finally, if the previous steps are followed and only the investigation report is exchanged, there is the question of validating the received report. For this issue, it is proposed to utilise a Training Model beforehand to examine the investigator being employed for future investigations. This model falls under a stochastic model and will be discussed further in the following section. It can be used for different purposes but in this case, will show the employability of the individuals in regards to IoT digital forensic and will support the selection process in deciding on the most skilful examiners as demonstrated in Table I which weighs expertises of investigators on different levels.

IV. Modelling of Evaluating Investigator, Evidence, Report, and Data Sensitivity

This section introduces models of this work to evaluate investigators, evidence, and whole image/report of the process. It also answers the first three research questions of the work. As it is aforementioned, forensic investigation processes include four key factors, which are evidence, investigator, expert witness, and tools, in order to have a complete investigation process. It is very possible to have a set of those factors in an investigation process, for instance, several tools could be used, several investigators can work on the same case, and a device might include different evidence items. However, only one expert witness can take the report to the court, therefore, the expert witness can not be considered as a set. Sets of other factors can be defined as follows;
Evidence set: E=$E_1, E_2, \ldots, E_k$;
Investigator set: I=$I_1, I_2, \ldots, I_l$;
Tool set: T=$T_1, T_2, \ldots, T_n$.
Accuracy of evidence, the expertise of investigator, and the quality of tools are remarkably important factors in a forensic investigation. The process needs to be completed without a mistake or with minimum mistakes. By considering this, we have defined models, which can evaluate the expertise of investigators and accuracy of evidence. The model aims to complete the best process for minimising the mistakes and maximising the number of the files while evidence is shared among countries.

Once the accuracy of evidence and investigators' expertise values are known, the confidence value of the whole process can be calculated. In order to do this, we define a confidence function, which has two input values and produces one single output. *Confidence* is a number that ranges in [0,1].

Table I represents how investigators confidence value and accuracy value of evidence are calculated. Explanations of *0, 0.5, and 1* values are given in Table IV.

- C_{ffi} is a model that calculates the confidence value of each image in the evidence. The model is as follows;

$$C_{ffi} = \frac{\sum_{i=1}^{m}(P_i)}{m} \tag{1}$$

P_i represents the degree of expertise for each investigator or expert. *m* presents the number of experts, who are assigned to analyse evidence.

The confidence value of each file within the evidence helps us to develop the confidence model of a whole image. Model 2 is the equation that is used to calculate the confidence value of the evidence.

$$C_{fEi} = \frac{\sum_{i=1}^{n} C_{ffi}}{n} \tag{2}$$

C_{fEi} is a number in [0,1]. The evidence confidence value is calculated by dividing the summation of the confidence value of each file in the evidence with the number of files.
- C_{fIi} is a model that calculates the confidence value of the experts. The model is as follows;

$$C_{fIi} = \frac{\sum_{i=1}^{n}(d_i)}{n} \tag{3}$$

In Equation 3, d_i represents the degree of extracting the $file_i$ for the evidence and *n* represents the number of files in an evidence item. C_{fIi} is a number ranges in [0,1].
- C_{fEW} presents the confidence of the expert witness that is related to the skills of the person, who takes the forensic report to the court. In some cases, the expert witness might be the same person as the investigator. If that is the case, the value of the confidence of the expert witness is considered equal to the confidence of the investigator.

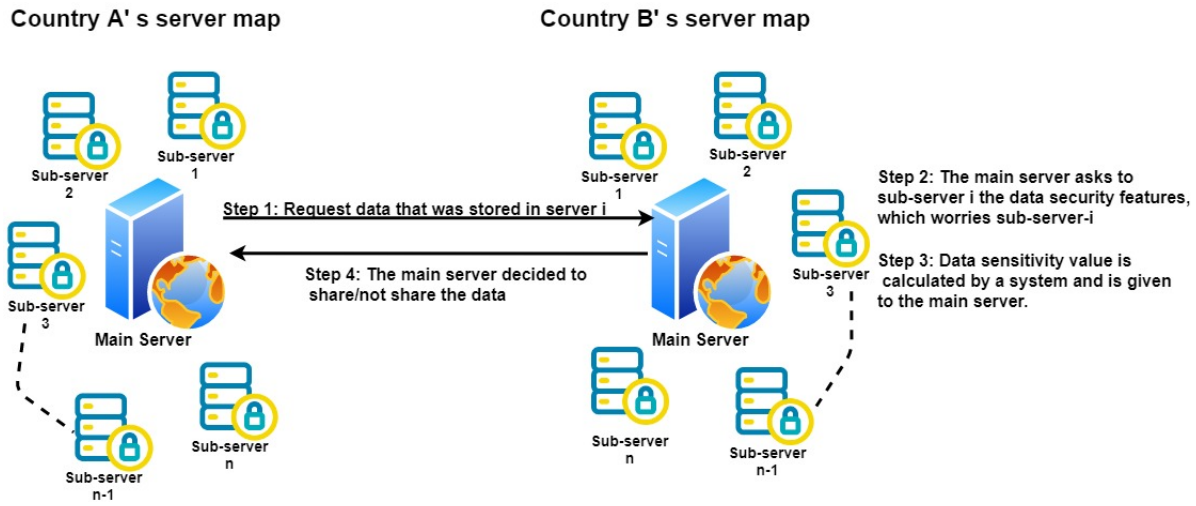

Fig. 2. Server structure with requester and owner of data

TABLE I
TABLE: WEIGHTING INVESTIGATORS' EXPERTISES FOR EMPLOYABILITY

Evidence Image E_i	$Investigator_1$	...	$Investigator_m$	Evidence Confidence Cf_{fi}
f_1	$0 \vee 0.5 \vee 1$	...	$0 \vee 0.5 \vee 1$	$Cf_{f1} = \frac{\sum_{i=1}^{m}(P_i)}{n}$
f_2	$0 \vee 0.5 \vee 1$	...	$0 \vee 0.5 \vee 1$	$Cf_{f2} = \frac{\sum_{i=1}^{m}(P_i)}{n}$
f_3	$0 \vee 0.5 \vee 1$	...	$0 \vee 0.5 \vee 1$	$Cf_{f3} = \frac{\sum_{i=1}^{m}(P_i)}{n}$
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
f_n	$0 \vee 0.5 \vee 1$	...	$0 \vee 0.5 \vee 1$	$Cf_{fn} = \frac{\sum_{i=1}^{m}(P_i)}{n}$
Investigator Confidence Cf_{Ii}	$Cf_{I1} = \frac{\sum_{1}^{n}(d_i)}{n}$	...	$Cf_{In} = \frac{\sum_{1}^{n}(d_i)}{n}$	

- The confidence value of the report depends on three values, which are the confidence value of investigators, the confidence value of the evidence, and the confidence value of the expert witness. In this work, we use *AND* conjunction for calculating the confidence value of the forensic report.

C_{fR}=f($C_{fEi} \wedge C_{fIi} \wedge C_{fEW}$)

C_{fR} is the model that presents the calculation of the correctness/confidence of the report. In order to show a detailed way of the calculation, we created Table II. There are abbreviations on the table, C_I represents the *Confidence value of the Investigator*, C_{EW} represents the *Confidence value of the Expert Witness*, C_E represents the *Confidence value of the Evidence*, and C_R represents the *Confidence value of the Report*.
Table II clearly presents that it is necessary to have all values to decide the correctness of the report. *The confidence of Expert witness (Cf_{EW})* value can be either 0 or 1 whereas *the confidence of Investigator (Cf_I)* and *the confidence of Evidence (Cf_I)* values can range in [0-1].

TABLE II
A SAMPLE PRESENTATION OF CONFIDENCE OF INVESTIGATOR, EVIDENCE, EXPERT WITNESS, AND REPORT

C_IXC_{EW}X C_E	C_R
[0-1]X0X[0-1]	**0**
[0-1]X0X[0-1]	**0**
[0-1]X1X[0-1]	**[0-1]**
[0-1]X1X[0-1]	**[0-1]**
[0-1]X0X[0-1]	**0**
[0-1]X0X[0-1]	**0**
[0-1]X1X[0-1]	**[0-1]**
[0-1]X1X[0-1]	**[0-1]**

- Sensitivity of Forensic Data: The sensitivity of a produced report is another important value to complete the process and make the decision of whether to share the report with the requester country. The data sensitivity value defines sensibility of the report, in other words, how much information might be lost if the data is shared with a server, which is not located in the same country in

which the forensic investigation is carried out. Akkuzu et al [16] developed a model that is used to calculate the data sensitivity value. We borrow their model to calculate our forensic data sensitivity value (see Equation 4).

$$S_d = \frac{\sum_{i=1}^{m}(P_i * (w_i))}{\sum_{j=1}^{n}(f_j)} \quad (4)$$

S_d represents the data sensitivity, it ranges in [0,1]. The numerator gives the summation of the data Confidentiality, Integrity, Availability, Privacy, and Possession (CIAPP) probabilities, in which P_i indicates the probability of CIAPP concerns that is selected by co-owner and w_i is the weight of the properties. The denominator indicates the total number of features (in our case five features are used see Table III).
Table III shows the data security features that are used in this work. These features can be varied depending on needs.

TABLE III
RELATED DATA SECURITY FEATURES FOR THE DATA SENSITIVITY

Subject of Protection	Discipline
Confidentiality	Information
Integrity	Information
Availability	Information
Privacy	Information
Possession	Information and Network

V. FUZZY LOGIC-BASED DECISION MAKING ON FORENSIC DATA SHARING

We use a fuzzy logic decision-making process to make a decision on the forensic data sharing in order to provide a secure forensic data sharing process. A fuzzy set is defined as (U, μ) in which U represents the universe set of elements and μ represents the membership function with the membership degrees of the elements to the set U, i.e., $x \in U \rightarrow \mu(x) \in [0, 1]$. There are various shapes of membership functions that can be chosen for a fuzzy set, such as triangle, trapezoid, and rectangle. Trapezoid functions can be viewed as a generalisation of triangular and rectangular membership functions. Therefore, we chose the trapezoidal membership function for defining whether the elements in our fuzzy sets are discrete or continuous [17].

Decision on forensic data sharing can be taken with some rules which are defined based on the confidence of the evidence and the sensitivity of the data, which are included in the forensic report. Conditions (i.e. rules) to make a decision on forensic data sharing are defined in [16], we use the same fuzzy rules since our data sharing process is considered as a co-owned data sharing process. The rules that are used to make decisions on the sharing process are as follows:

- *Rule1= If sensitivity['low'] $\wedge$ confidence['low'], decision['maybe']*
- *Rule2=If sensitivity['low'] $\wedge$ confidence['medium'], decision['maybe']*
- *Rule3=If sensitivity['low'] $\wedge$ confidence['full'], decision['yes']*
- *Rule4=If sensitivity['medium'] $\wedge$ confidence['low'], decision['maybe']*
- *Rule5=If sensitivity['medium'] $\wedge$ confidence['full'], decision['yes']*
- *Rule6=If sensitivity['medium'] $\wedge$ confidence['medium'], decision['maybe']*
- *Rule7=If sensitivity['high'] $\wedge$ confidence['low'], decision['no']*
- *Rule8=If sensitivity['high'] $\wedge$ confidence['medium'], decision['maybe']*
- *Rule9=If sensitivity['high'] $\wedge$ confidence['full'], decision['yes'].*

Through this whole process, it is very crucial to consider the possible fuzzy area during an investigation. Weighing the ability of investigators and the complexity of the evidence does not only provide high and low results and to make our method applicable to real-life cases the medium range is important to include [18].

Table IV presents the fuzzy linguistic terms with the numerical value of each term and member, which is associated with the fuzzy value, and it explains what each value means in details.

It is clearly seen that if the forensic data is not found sensitive by investigators and investigators are confident on the forensic report that they produce, then most probably the data is shared with the requester. That means that the report is related to the criminal case. On the other side, if the data is found very sensitive by investigators after their analysis and they do not have enough confidence on the report that they produce, then the data may not be shared with the requester country. The values of the fuzzy linguistic class can be specified by experts in each investigation process based on the case. In this work, high, medium, and low linguistic variables are used with expert knowledge-based decision making.
Figure 3 represents each fuzzy variables' linguistic terms and the numerical values of each linguistic variable's. *Akkuzu et al.* defined the membership range of each variable with a clustering technique (see [16]), therefore, we use the same ranges of each variable as in their work due to our conditions similarity and due to the knowledge-based fuzzy decision making is not in use anymore.

VI. DISCUSSION

The usage of IoT devices has been increasing per person day by day. They become part of people's lives due to their high

TABLE IV
FUZZY LINGUISTIC VARIABLES , ITS VALUES, AND ASSOCIATED MEMBER

Linguistic Term	Numerical Value	Member	Meaning
Yes	1	Evidence	File is not only found but also opened for analysis (All files in were opened by investigator)
		Investigator	Investigator was successful to retrieve, restore and/or access all of the evidence (out of 100 % Evidence)
Maybe	0.5	Evidence	File is found and could not be opened OR (Not all files but some files were opened by investigator)
		Investigator	Investigator was partially successful to retrieve, restore AND/OR access the evidence (out of 100 % Evidence)
No	0	Evidence	None of files is found
		Investigator	was not successful to retrieve, restore OR access any of the evidence (out of 100 % Evidence)

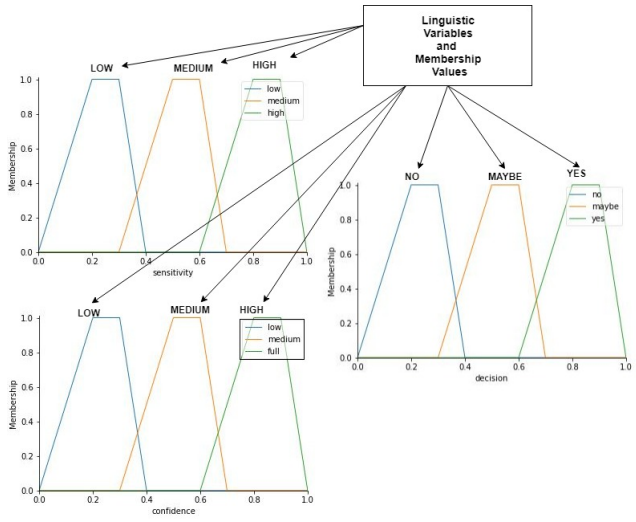

Fig. 3. Membership Values of Each Variables

dependency on their devices and provide services in daily tasks. These devices can be used for a multitude of reasons depending on the individual. Therefore, the benefits of IoT devices are irrefutable since they are the amenities for our lives. However, these devices can also be utilised in criminal activities which also require constant improvement of the forensic investigation process and tools. Therefore, we propose a model which is able to register IoT devices and stores their information, such as Owner, Subscriber, User, Device Type, Serial Number and Location, to access data if necessary and requested during an investigation. We do not eliminate current forensic investigation tools but rather propose a new method to investigate more complex criminal cases in the complicated IoT environment and steady increase of IoT devices. Moreover, investigation reports are a crucial part of a case investigation due to being the main focus of most evidence presentations in the court of law. Therefore, it is important to know how accurate the report is, what are expertise of investigators, and how correct evidence files are. Considering those needs in a forensic investigation we developed models, which give a new approach of calculating those values. We also introduce a new and efficient way of sharing forensic reports. The real-life decision expressions are not Boolean, furthermore, the sensitivity affects to decide whether to share or not to share information. In order to cover those needs, we introduce a fuzzy logic decision-making system, in which the fuzzy expressions are used and the sensitivity of the forensic data is employed to make a decision.

VII. CONCLUSION

This paper introduces a novel approach to forensic investigation processes and sharing forensic data. We propose a multilevel, distributed and hierarchical Hybrid Server Model which allows IoT devices to be registered on and which additionally stores data of these devices. During the process of forensic investigations, this Server and stored data can be used in an efficient way for analysis and report creation. A forensic investigation process includes investigators, evidence, and report, which is produced by the investigators with found evidence files. Therefore, it is important to know the accuracy of the forensic report, the expertise of the investigators who work on the report production, and the accuracy of the evidence. In this paper, we developed models for evaluating values for the accuracy of the report, the accuracy of the evidence, and the investigators' expertise. We then introduce a new way of making decisions on the forensic data sharing process. Furthermore, we use the forensic data sensitivity value and the confidence value of the forensic report as input values for the fuzzy logic decision-making system. The output of the fuzzy logic is a decision that have three values either *'yes'*,*'maybe'* or *'no'*. As it is seen the fuzzy logic provides very valuable flexibility for reasoning. In this way, we can consider the inaccuracies and uncertainties of any situation on

making decisions on forensic data sharing processes. In future work, we will be testing the Hybrid Server with mobile phones in terms of device registration and unique device identification, evidence storage and investigation analysis as well as real-time Server updates when registered devices change their information, such as updates, personal content, location.

REFERENCES

[1] B. Safaei, A. M. Monazzah, M. Bafroei, and A. Ejlali, "Reliability side-effects in internet of things application layer protocols," 12 2017.

[2] A. MacDermott, T. Baker, and Q. Shi, "Iot forensics: Challenges for the ioa era," in *2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*. IEEE, 2018, pp. 1–5.

[3] F. Servida and E. Casey, "Iot forensic challenges and opportunities for digital traces," *Digital Investigation*, vol. 28, pp. S22–S29, 2019.

[4] S. Zawoad and R. Hasan, "Faiot: Towards building a forensics aware eco system for the internet of things," in *2015 IEEE International Conference on Services Computing*. IEEE, 2015, pp. 279–284.

[5] I. Yaqoob, I. A. T. Hashem, A. Ahmed, S. A. Kazmi, and C. S. Hong, "Internet of things forensics: Recent advances, taxonomy, requirements, and open challenges," *Future Generation Computer Systems*, vol. 92, pp. 265–275, 2019.

[6] S. Jeon and S. Lee, "Digital forensics technology management platform," in *2016 International Conference on Platform Technology and Service (PlatCon)*. IEEE, 2016, pp. 1–6.

[7] S. Tanwar, S. Tyagi, and S. Kumar, "The role of internet of things and smart grid for the development of a smart city," in *Intelligent Communication and Computational Technologies*. Springer, 2018, pp. 23–33.

[8] D. Chen and H. Zhao, "Data security and privacy protection issues in cloud computing," in *2012 International Conference on Computer Science and Electronics Engineering*, vol. 1. IEEE, 2012, pp. 647–651.

[9] B. Schneier, *Data and Goliath: The hidden battles to collect your data and control your world*. WW Norton & Company, 2015.

[10] J. Shen, T. Zhou, X. Chen, J. Li, and W. Susilo, "Anonymous and traceable group data sharing in cloud computing," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 4, pp. 912–925, 2017.

[11] C. McCartney, "Forensic data exchange: ensuring integrity," *Australian Journal of Forensic Sciences*, vol. 47, no. 1, pp. 36–48, 2015.

[12] R. Montasari, "Review and assessment of the existing digital forensic investigation process models," *International Journal of Computer Applications*, vol. 147, no. 7, pp. 41–49, 2016.

[13] K. Kent, S. Chevalier, T. Grance, and H. Dang, "Guide to integrating forensic techniques into incident response," *NIST Special Publication*, vol. 10, no. 14, pp. 800–86, 2006.

[14] D. Lillis, B. Becker, T. O'Sullivan, and M. Scanlon, "Current challenges and future research areas for digital forensic investigation," *arXiv preprint arXiv:1604.03850*, 2016.

[15] S. Perumal, N. M. Norwawi, and V. Raman, "Internet of things (iot) digital forensic investigation model: Top-down forensic approach methodology," in *2015 Fifth International Conference on Digital Information Processing and Communications (ICDIPC)*. IEEE, 2015, pp. 19–23.

[16] G. Akkuzu, B. Aziz, and M. Adda, "Fuzzy logic decision based collaborative privacy management framework for online social networks," in *3rd International Workshop on FORmal methods for Security Engineering: ForSE 2019*. SciTePress, 2019.

[17] D. E. Sánchez, E. Esmi, and L. C. de Barros, "Discrete and continuous logistic p-fuzzy models," in *International Fuzzy Systems Association World Congress*. Springer, 2019, pp. 556–566.

[18] M. L. Lagunes, O. Castillo, F. Valdez, and J. Soria, "Comparison of fuzzy controller optimization with dynamic parameter adjustment based on of type-1 and type-2 fuzzy logic," in *Hybrid Intelligent Systems in Control, Pattern Recognition and Medicine*. Springer, 2020, pp. 47–56.