Framework of Confidence Values during Digital Forensic Investigation Processes

NANCY SCHEIDT and MO ADDA
University of Portsmouth
School of Computing
Portland Building, Portland Street, PO1 3AH Portsmouth
UNITED KINGDOM (UK)

Abstract: The advancement of Internet of Things (IoT) devices is continuously progressing and such development also enables a number of issues to arise which increases the complexity in the forensic investigation of the IoT. Globally, investigators are faced with challenges in ways of retrieving evidence from the different areas of the IoT environment, which includes Devices, Networks and the Cloud. One of the most crucial steps during forensic investigations is the writing up and creation of a case report which then needs to be presented in the court of law. In this paper, we propose models to estimate the confidence values of evidence, investigators and case reports to ensure case investigation accuracy and improve the evidential values of case presentation as well as evidence sharing of sensitive data worldwide.

Key Words: confidence value, forensic investigation, IoT server, fuzzy logic, forensic data sensitivity, investigator expertise, forensic report

Received: April 15, 2020. Revised: May 27, 2020. Accepted: May 30, 2020. Published: June 2, 2020.

1 Introduction

The number of Internet of Things (IoT) devices rises rapidly per person and will continue do so, such leads to the fast development and improvement of IoT devices which also opens more doors for the variety of opportunities these devices can offer in terms of usage, commercially, privately or criminally [13, 17]. Therefore, such developments can lead to a variety of challenges in digital and IoT forensics as well as increasing the complexity of accessing information of devices if forensically required. Hence, collecting evidence of the IoT environment (i.e. devices, network, the cloud) is a challenge investigators face worldwide. Research focuses on solutions expediently, however, it mainly addresses solutions to improve cybersecurity in the cloud or focuses on device-specific techniques for investigation purposes [7, 15].

Additionally, research by [18] suggest a server model to ease the investigation process due to IoT devices being registered on and information is stored on such. Managing devices and evidence of these is in need to be managed more efficiently and precisely [4]. If these steps are taken it is crucial to consider how reliable the whole investigation process has been and if sensitive data can be shared securely with other investigators which can include a number of privacy risks in this day and age [14]. Therefore, research by [3] proposes a secure encryption way to ensure data security and privacy. Furthermore, the paper by [16] suggests a data-sharing scheme which is made of 5-steps. If this research regarding data sharing is to be applied into police investigation processes, additional challenges need to be outlined and considered, such as the trust level between different countries when inquiring information for investigation purposes or the abilities of case investigators [9]. These calculations of trust levels have been implemented in research, however, focus on social media and how or if sensitive data can be shared between users [2]. However, this method is not considering the sharing of forensic data and was not applied to investigation processes. Additional research focused on the accuracy of forensic science and witness testimonies [10, 12, 5]. Other research by [5] proposes proficiency tests to assure forensic science results are accurate, however, only provide a theoretical idea by evaluating the benefits of being able to test the accuracy of forensic results. Moreover, [12]'s research focuses on the psychological factor which can influence the accuracy of evidence provided especially in terms of witness testimonies. None of the current research provides models to measure and calculate the accuracy or confidence of forensic investigation aspects, do however, stress its importance. Moreover, fuzzy logic and considering that some aspects cannot be as easily defined as by the Boolean logic 'True' or 'False' has not been linked to previous

research and are very important to consider especially in forensic investigation processes [11].

Considering previous research, studies need improvement to provide specific models to show the confidence values during an investigation such as forensic investigators, reports, evidence as well as the sensitivity of data sharing to ensure an applicability to real life cases and improve the investigation process in criminal cases. To tackle and cover the issues raised and found in previous research, we aim to answer the following questions to offer a solution which also clarify our contribution to this topic matter:

- How can the accuracy of evidence be evaluated for a forensic investigation?
- How can the accuracy of a forensic report be measured for a forensic investigation?

This paper is organized as follows. Section 2 provides the formulation of the problem of mathematical models created to calculate the confidence value of an aspect during a police investigation, such as the investigator, evidence and report while considering the results and implementing fuzzy logic rules for its solution. Section 3 demonstrates the problem solution by providing an example, implementation and results of proposed models with a short discussion of such in Section 4. Lastly, Section 5 concludes this paper's research.

2 Problem Formulation

As the IoT is expanding and crime utilising such is increasing, ways for investigation purposes need to improve and develop. It is suggested to ease the process of evidence analyses of IoT devices (i.e. smartphones, tablets, laptops). If devices' unique information, which we can refer to as the DNA of a device, are being registered on a hierarchical and distributional Hybrid Forensic Server as presented in Figure 1 the management of IoT devices can be improved [8]. The DNA of a device, as demonstrated in Table 1, provides the sub-server, which every county/city of every country will be equipped with, with the ability to store information of these devices on the server, however, such can only be accessed with a court order for investigation purposes. Part of this has been suggested in research by [18] but only on a single server level utilising it as evidence storage and investigation platform on a smaller scale. If we take this a step further and want to ensure that the process of investigation (evidence analysis, investigator expertise and case report) is of high evidently value when presented in court the confidence of such has to be considered. Therefore, in this paper, we present models to calculate the confidence values of the evidence, C_E, as shown in equation 1, confidence values of investigators, C_I, and/or expert witnesses, C_W, shown in equation 2, and finally the confidence of the outcome of a forensic written and presented report, C_O, as shown in equation 3. Where d_{ij} represents the degree level of extraction success for each evidence per investigators. There are *m* evidences per investigator, *n*, $1 \leq i \leq n$ and $1 \leq j \leq m$. This value depends on the resources used, the tools and it is also scaled by the sensitivity of sharing evidences between investigators. The sensitivity, S_{ij}, is defined in Equation 4. The level of expertise of each investigator, L_I , and each expert witness, L_W, defines the status and knowledge such as trainee, junior or senior rank. For instance, for a senior investigator of more intensive experience L_I would be set to 1.

$$C_E = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} d_{ij} S_{ij} \tag{1}$$

$$C_I = \sum_{i=1}^{n_I} \frac{L_{Ii}}{n_I}, C_W = \sum_{i=1}^{n_W} \frac{L_{Wi}}{n_W} \tag{2}$$

$$C_O = \mathrm{f} \quad (C_E \quad \wedge \quad C_I \quad \wedge \quad C_W) \tag{3}$$

Having evaluated the different aspects of an investigation, the final step is the report outcome and its value of confidence. The report (C_O) evaluates the confidence value of the evidence (C_E) and of investigators (C_I) *AND* and/or *OR* expert witnesses (C_W), considering the written and presented report. This model C_O represents the calculation process of the correctness and/or confidence of the report. An overview of possible results when calculating the confidence value of a report is shown in Table 2. To elaborate, if the confidence of the case investigator as well as the evidence ranges between *[0,1]* overall but the case expert witness is assigned a value of *0* the report's confidence value will only be *0*. Hence, the reports value is low. If the confidence of the case investigator as well as the overall evidence range in the higher threshold of *[0,1]*, and the case expert witness is assigned a value of *1* the report's confidence value will range in the higher threshold of *[0,1]* and therefore be of high evidence value. Hence, if any of the values are *0*, the overall confidence will be *0* and therefore dissolve in a court of law, as demonstrated in Table 2.

Table 1: Genes of a Device

Genes	Attributes
Owner (O)	Individual/Company who purchased a device
Subscriber (S)	Individual/Company registering device, could be owner or receiver (i.e. company phone, present)
User (U)	Individual/Company using device, could be owner or receiver (i.e. company phone, present)
Serial Number (SN)	Unique Serial Number of Devices
Location (L)	Place device has initially been registered
Device Type (DT)	Brand, Model
Connectivity (C)	Devices ability to connect to the internet

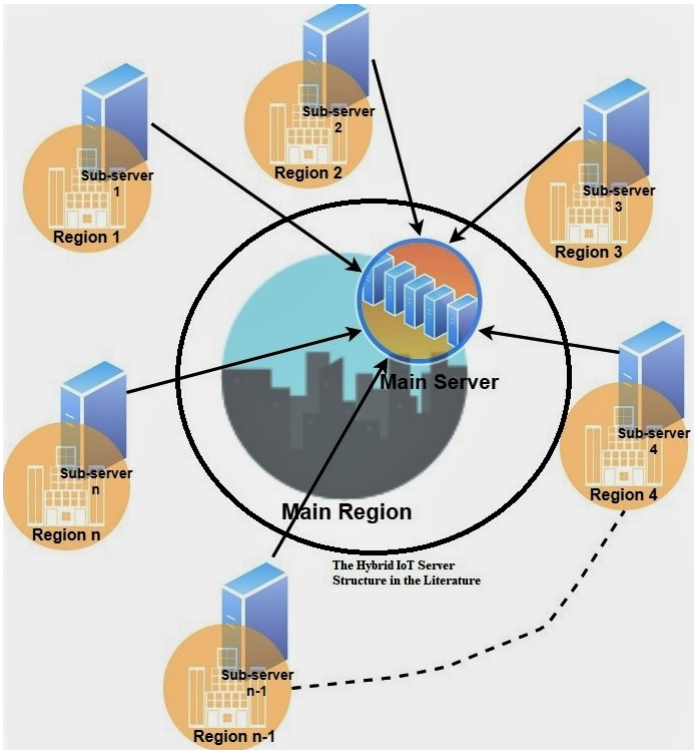

Figure 1: Hybrid Forensic IoT Server Structure

Table 2: Possible Confidence Values of Investigator, Expert Witness, Evidence and Report

$C_E \wedge C_I \wedge C_W$	C_O
[0,1] ∧ [0,1] ∧ [0,1]	**[0,1]**
[0] ∧ [0,1] ∧ [0,1]	**[0]**
[0,1] ∧ [1] ∧ [1]	**[0,1]**
[1] ∧ [1] ∧ [1]	**[1]**
[0] ∧ [0] ∧ [0]	**[0]**

Considering different aspects of an investigation and the importance in its confidence value calculations, it is also necessary to consider the sensitivity of a case. Moreover, due to the results the models provide and these results being in the range between *[0,1]*, fuzzy logic needs to be addressed. The following two models are provided in previous research by [2] and are well in implementation of this research due to sharing of co-owned data by various international servers:

- Data Sensitivity (Equation 4)
- Fuzzy Logic Rules (Table 3).

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

Results of S also range in between *[0,1]* and the summation of the numerator shows the data Confidentiality, Integrity, Availability, Privacy, and Possession (CIAPP) probabilities. P_i shows the CIAPP probabilities which are selected by the co-owner of the data (i.e. evidence) and w_i is the weight of these properties. Furthermore, the denominator presents the total number of CIAPP probabilities and features which are five in this case and these features can be varied depending on needs [2, 1]. Furthermore, making decisions on sharing forensic data/evidence considering its sensitivity as well as confidence can follow the rules demonstrated in Table 3 and is also shown in Figure 2. As our models calculate results ranging between *[0,1]* the fuzzy area of the results needs to be considered as well [6]. Therefore Table 4 provides an overview of the complexity the results can be identified as which enables these models (Equation 1-2) to be applicable to real-life investigation cases.

Moreover, if there are m number of evidences each of these will have a specific degree of extraction, d, which again relates to the probability of sharing this data, as suggested in Table 6. Hence, local investigations will always have a sensitivity value of *1*, however, investigations on an international basis will range their sensitivity value between *[0,1]*. Therefore, it is important to consider data sensitivity, S, as well as the degree of extraction, d because these two aspects can effect the outcome of a report, as further demonstrated in Table 5 as well as Section 3.

3 Problem Solution

In this section, the calculations of confidence values are applied for demonstration purposes. Assuming devices worldwide are registered on Hybrid servers such as in Figure 1, a user's, in this case a suspect's, device information will be stored on the server of the city they are using their devices in. Hence, if the suspects visits another Country B, for instance, the information will be stored on the visiting Country's B server instead, for instance, of the home country's A server. Therefore, if said suspect commits a crime, i.e. money laundering and the authorities of a Country A are able to investigate the evidence, they also need to request the evidence from the server of a Country B. This is part of a thorough investigation to portray the chain of events as well as to make sure to investigate other possible connections. Once a Country A requested the evidence from a Country B by providing a court order, the process to decide to share or not to share the forensic report and data begins in evaluating the confidence value of the evidence report and the data sensitivity as elaborated in the previous section (Equations 1 - 4).

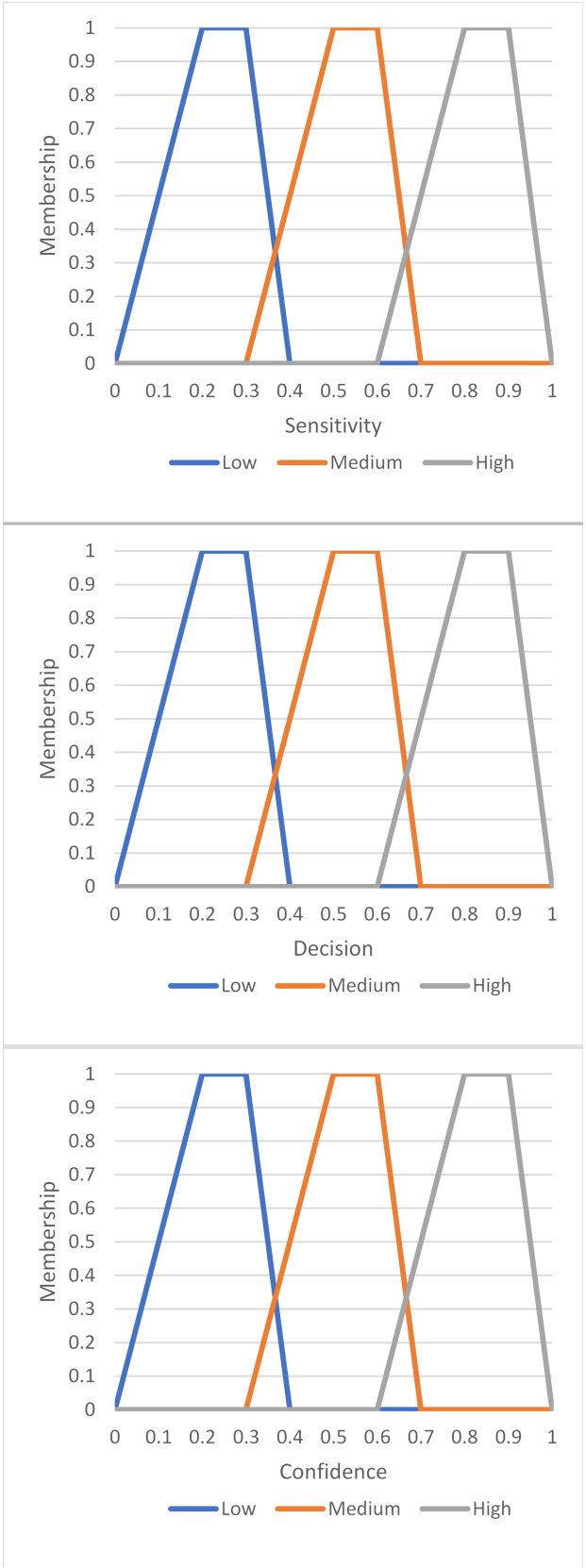

Figure 2: Membership Values of Each Variables

Table 3: Fuzzy Rules

Rule Number	**Rule**	**Decision**
Rule 1	If sensitivity['low'] $\wedge$ confidence['low']	['maybe']
Rule 2	If sensitivity['low'] $\wedge$ confidence['medium']	['maybe']
Rule 3	If sensitivity['low'] $\wedge$ confidence['full']	['yes']
Rule 4	If sensitivity['medium'] $\wedge$ confidence['low']	['maybe']
Rule 5	If sensitivity['medium'] $\wedge$ confidence['full']	['yes']
Rule 6	If sensitivity['medium'] $\wedge$ confidence['medium']	['maybe']
Rule 7	If sensitivity['high'] $\wedge$ confidence['low']	['no']
Rule 8	If sensitivity['high'] $\wedge$ confidence['medium']	['maybe']
Rule 9	If sensitivity['high'] $\wedge$ confidence['full']	['yes']

Table 4: Fuzzy Linguistic Variables , Its Values, And Associated Member

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

Due to forensic data and evidence being of sensitive nature, the numbers in Table 5 are assumed to demonstrate how calculations of the confidence models can be utilised efficiently. In this example we assume that the level of expertise of all the investigators and expert witnesses involved is high, $L_W = L_I = 1$, giving $C_W = C_I = 1$, as deduced from Equation 2. We also assume the sensitivity, S_{ij}, is equally applied to all investigators and expert witnesses, $S_{ij} = S_j$. In the example of Table 6, *m* is equal to 3 and *n* is equal to 2. In this case we ignore the additional implementation of expert witnesses. Furthermore, we assume that the evidences E_1 and E_2 belong to Investigator 1 from a Country A, whereas, the evidence, E_3, are found by Investigator 2 in a Country B. As an example, E_1 has only been able to be partially produced, E_2 not at all and E_3 fully at the sensitivity value of $S = 1$. This means the evidences are fully shared by all investigators from the countries A and B.

In Table 5 we also show differences of the effect of data sensitivity, S, where S is in the interval *[0,1]*. Which defines, as stated in Equation 4, the degree level of sharing, as demonstrated in Table 5, Column 2 of Investigator 1 and 2. Moreover, Table 5 shows the difference of considering data sensitivity, $S = 1$, compared to the original outcome, S. This calculations effect the outcome of a report and an investigation additionally and therefore need to be considered. Hence, two investigators from a Country A and B respectively have worked on this money laundering case and the information linked to a suspect. The confidence value of the investigators' ability ranges within the high threshold overall as mentioned above, see Table 2. Table 5 shows the confidence values of evidence retrieval and extraction as well as the investigators ability of this investigation regarding money

laundering and the suspect. Moreover, it demonstrates that the confidence value is effected by the sensitivity and owner ship of the evidence. Considering the sensitivity, S, of the evidence effects the outcome of the report.

After creating the case report, the investigators need to decide on the data sensitivity by selecting security features such as confidentiality, integrity, availability, privacy, and possession to decide if the evidence they analysed can be shared with Country A. As Table 6 demonstrates the data sensitivity of this case is low and therefore, the evidence can be shared with Country A to allow further investigation on their side. Moreover, according to Table 2, the third report outcome, C_O, can be applied to this case example. Going through this whole process allows a higher accuracy during an investigation process. The country who is sharing the evidence report is making sure to protect data privacy as well as case sensitivity and the countries receiving forensic reports can be ensured that the process has been done in a trustworthy and thorough manner for a valuable chain of custody.

Table 5: Representation of Calculation of Investigators Confidence and Evidence Value

Evidence E_i	Investigator1		Investigator2	
	$d_{ij} * (S = 1)$	$d_{ij} * S$	$d_{ij} * (S = 1)$	$d_{ij} * S$
E_1	0.5×1	0.5×1	0.5×1	0.5×0.5
E_2	0×1	0×1	0×1	0×1
E_3	1×1	0.5×1	1×1	1×1
C_{Ei}	0.5	0.33	0.5	0.41
$C_O = C_E$ = 0.5 (S=1), $C_O = C_E$ = 0.37 (S$\neq$1)				

Table 6: Investigators' Choices on the Data Security Features

Features	Investigator 1	Investigator 2	Probability of each feature
Confidentiality	✔	✔	$P_c = \frac{2}{2}$=1
Integrity	✖	✔	$P - i = \frac{1}{2}$=0.5
Availability	✖	✖	$P_a = \frac{0}{2}$=0
Privacy	✔	✖	$P_p = \frac{1}{2}$=0.5
Possession	✖	✖	$P_po = \frac{0}{2}$=0
			$S = \frac{2}{5} = 0.4$ (see Equation 4)

4 Discussion

Demonstrating the implementation of confidence values within a forensic investigation process showed that it is a very valuable addition to a chain of custody. Case results imply that the proposed models and their outcomes provide pivotal answers which evaluate the confidence of evidence analysis, investigators expertise and report creation. Considering results from multiple perspectives, such as incorporating data sensitivity, showed that additional aspects will effect an investigation enormously and influence the outcomes which is crucial in a court of law and present possible dissident aspects during legal processes. Therefore, the provided models not only allow to be able to categorise the investigation procedures efficiently but also demonstrate that different features will effect the results of such. Hence, it is of high importance to implement such models into forensic investigation processes to ensure a novel and improved method of evaluating case reports by considering investigators expertise, evidence extraction and data sensitivity.

5 Conclusion

This paper introduces a framework of a novel approach to forensic investigation processes and sharing forensic data. We propose models to calculate the confidence values of an investigation to ensure a highly valuable process of evidence retrieval and presentation. Areas which are of importance include investigators and/or expert witnesses, evidence, and a case report, which is produced by the investigators with found evidence files. These novel models plus the consideration of the sensitivity of evidence data enable investigators to decide if case reports and evidence can be shared with other precincts from different cities or even different countries. Furthermore, utilising the fuzzy logic decision-making system provides the efficient results in providing values within the interval *[0,1]*. Therefore, the results of our framework of proposed models enable higher accuracy and ensure applicability as demonstrated with the case example of an international investigation. In future work, the implementation of an investigators confidence value into a Hybrid System as mentioned in Section 2 will be part of our research. Implementing these confidence value models into a System environment will improve efficient IoT forensic investigation processes and allow to test its applicability further.

References:

[1] Gulsum Akkuzu, Benjamin Aziz, and Mo Adda. Advantages of having users' trust and reputation values on data sharing process in online social networks. In *The Sixth IEEE International Conference on Social Networks Analysis, Management and Security*. IEEE, 2019.

[2] Gulsum Akkuzu, Benjamin Aziz, and Mo Adda. Fuzzy logic decision based collaborative privacy management framework for online social networks. In *3rd International Workshop on FOR-*

mal methods for Security Engineering: ForSE 2019. SciTePress, 2019.

[3] Deyan Chen and Hong Zhao. Data security and privacy protection issues in cloud computing. In *2012 International Conference on Computer Science and Electronics Engineering*, volume 1, pages 647–651. IEEE, 2012.

[4] SangJun Jeon and SangJin Lee. Digital forensics technology management platform. In *2016 International Conference on Platform Technology and Service (PlatCon)*, pages 1–6. IEEE, 2016.

[5] Jonathan J Koehler. Forensics or fauxrensics: Ascertaining accuracy in the forensic sciences. *Ariz. St. LJ*, 49:1369, 2017.

[6] Marylu L Lagunes, Oscar Castillo, Fevrier Valdez, and Jose 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*, pages 47–56. Springer, 2020.

[7] Aine MacDermott, Thar Baker, and Qi Shi. Iot forensics: Challenges for the ioa era. In *2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, pages 1–5. IEEE, 2018.

[8] Ausama Majeed and Adil Al-Yasiri. Formulating a global identifier based on actor relationship for the internet of things. In *Interoperability, Safety and Security in IoT*, pages 79–91. Springer, 2016.

[9] Carole McCartney. Forensic data exchange: ensuring integrity. *Australian Journal of Forensic Sciences*, 47(1):36–48, 2015.

[10] Dawn McQuiston-Surrett and Michael J Saks. Communicating opinion evidence in the forensic identification sciences: Accuracy and impact. *Hastings LJ*, 59:1159, 2007.

[11] Hung T Nguyen, Carol L Walker, and Elbert A Walker. *A first course in fuzzy logic*. CRC press, 2018.

[12] HL Roediger, JH Wixted, and KA DeSoto. The curious complexity between confidence and accuracy in reports from memory. *Memory and law*, page 84, 2012.

[13] Bardia Safaei, Amir Mahdi Monazzah, Milad Bafroei, and Alireza Ejlali. Reliability side-effects in internet of things application layer protocols. 12 2017.

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

[15] Francesco Servida and Eoghan Casey. Iot forensic challenges and opportunities for digital traces. *Digital Investigation*, 28:S22–S29, 2019.

[16] Jian Shen, Tianqi Zhou, Xiaofeng Chen, Jin Li, and Willy Susilo. Anonymous and traceable group data sharing in cloud computing. *IEEE Transactions on Information Forensics and Security*, 13(4):912–925, 2017.

[17] Sudeep Tanwar, Sudhanshu Tyagi, and Sachin Kumar. The role of internet of things and smart grid for the development of a smart city. In *Intelligent Communication and Computational Technologies*, pages 23–33. Springer, 2018.

[18] Shams Zawoad and Ragib Hasan. Faiot: Towards building a forensics aware eco system for the internet of things. In *2015 IEEE International Conference on Services Computing*, pages 279–284. IEEE, 2015.