Digital Forensic Readiness of Information Systems: A cost-benefit variable analysis

Antonis Mouhtaropoulos
Faculty of Computing, Metropolitan College, Greece

ABSTRACT

Despite the increasing amount of research on the pre-incident side within a digital forensic investigation, little steps have been taken towards assessing the effectiveness of such a plan in terms of cost effectiveness. This research paper lays the foundations of a cost-benefit variable analysis within a digital forensic readiness context by defining a cost-benefit relationship effect model. We collect novel, primary data from organisations and institutions that implement a digital forensic readiness plan to identify cost variables of each measure and threat, and benefit variables of each measure to be taken. We conduct data analysis to portray that specific cost variables have a significant effect on specific benefit variables and present the results of the data collection process amongst organisations and institutions applying a digital forensic readiness plan. Lastly, we produce hypotheses testing results and determine the validity between each cost-benefit relationship.

Keyword: *Cost benefit analysis, security management, digital forensic readiness, digital forensic investigations, resource management.*

1. INTRODUCTION

Digital forensic readiness (DFR) is a phase within the digital forensic investigation (DFI) lifecycle that deals with pre-incident preparation, in terms of digital evidence identification, preservation, and storage. There have been many proposals, suggestions, and publications on a universally accepted digital forensic investigation; however, due to the complexity of systems, hardware, software, and legal systems, this initiative has yet to establish a common DFI. This study adopts the generic model proposed by Ciardhuáin (2004) and includes a pre-incident investigation phase, entitled digital forensic readiness (Figure 1). Despite the increasing amount of research on the pre-incident side of a digital forensic investigation, little steps have been taken towards assessing the effectiveness of such a plan in terms of cost, time, and quality (Kebande & Venter, 2019).

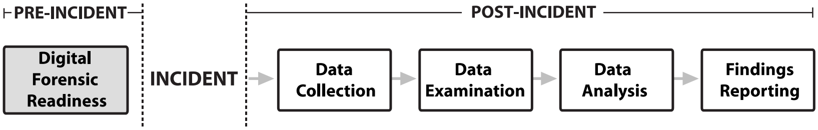


Figure 1. A generic digital forensic investigation (DFI) process

One of the few government initiatives involving the assessment of cost, time, and quality is UK's HMG Security Policy Framework (2018). According to the third principle of the framework: 'Risk management is key and should be driven from Board level. Assessments will identify potential threats, vulnerabilities, and appropriate controls to reduce the risks to people, information, and infrastructure to an acceptable level. This process will take full account of relevant statutory obligations and protections, including the Data Protection Act, Freedom of Information Act, the Official Secrets Act, Equality Act and the Serious Organised Crime and Police Act' (Cabinet Office, 2018).

The most common problem in digital forensic investigations is that the investigator can only formulate hypothesis on a component's or artifact's previous state by making indirect observations on the system. The acceptance of a hypothesis relies on the ability of the investigator to identify, preserve, extract, and interpret the data related to the crime.

According to Rowlingson (2004), discussions of the forensic process tend to ignore what happens to the object of the investigation prior to the decision to undertake an investigation. The necessary evidence either exists (and hopefully is discovered by the DFI), or it does not exist, and a suspect cannot be charged and prosecuted. This is the law enforcement view of a DFI. It begins when a crime has been committed or discovered and investigators attend a crime scene or wish to seize evidence. The quality and availability of evidence is a passive aspect of the DFI.

In a cost-oriented context however, there is the opportunity to actively collect potential evidence in the form of log files, emails, back-up disks, portable computers, network traffic records, and telephone records, amongst others. This evidence may be collected in advance of a crime or dispute and can be used to the benefit of the collecting organisation.

There have been fitful efforts to standardise digital forensics corpora in efforts to establish the ground truth (Tully et al., 2020). Similarly, the US National Institute of Standards and Technology (Chew et al., 2008), in its report published in 2008, proposed a guide to assist decision-makers in information system and program levels. Such a guide includes the definition of metrics/measures mainly for the information security domain. For the purposes of this research, the authors build on the research conducted by Chew et al. and co-opt this definition detailed below, for the digital forensics (DF) domain.

‘Information security measures are used to facilitate decision-making and improve performance and accountability through the collection, analysis, and reporting or relevant performance-related data. The purpose of measuring performance is to monitor the status of measured activities and facilitate improvement in those activities by applying corrective actions based on observed measurements’ (Singhal & Ou, 2017).

This paper highlights the lack of research output in identifying the cost aspect of a digital forensic readiness plan and presents the results of a data collection process amongst organisations and institutions applying such a plan. It presents descriptive data analysis and hypotheses testing results. It lays the foundations of a cost-benefit factor analysis by defining a cost-benefit relationship effect model.

It attempts to devise a model that will act as a decision-making tool to organisations in applying a DFR system; such a system needs to maximise the difference between the benefit derived in applying the DFR framework and the cost of implementing it. A Cost-Benefit Analysis (CBA) model will act in a decision-making context to review the effectiveness of a DFR plan. Such a relationship model will form the proposition of a digital forensic readiness planning framework that will act as a tool to aid decision-makers within an organisation. In addition, the proposed model will enable the calculation and algorithmic verification of the numerical relationship between cost and benefit variables using Bayesian networks and analysis.

The main contributions of this paper are as follows. First, a digital forensic readiness research framework is introduced. Second, the foundations of a cost-benefit analysis are laid by conducting a cost-benefit variable analysis. Third, a cost-benefit relationship effect model is presented to aid institutions in the decision-making process within a digital forensic readiness framework.

The structure of this paper takes the form of five sections, including this introductory chapter. The remainder part of the paper proceeds as follows: Section 2 introduces the reader to the concepts of digital forensic readiness, information security, digital forensics metrics, and cost benefit analysis. Section 3 presents the research methodology of this study by outlining both the hypotheses design and the data collection process. Section 4 shows the results of the descriptive analysis of the research and demonstrates the results of the analysis of variance (ANOVA) tests conducted and depicts the relationship effect between cost and benefit variables. Lastly, Section 5 concludes the paper and identifies areas for further research.

2. DIGITAL FORENSIC READINESS

Digital forensics investigations have been the central point of computer and forensic scientists during the last decade. The need for standardising the process and achieving a scientific rigor has urged researchers to discuss the digital investigation process by developing theoretical frameworks. The frameworks examined digital forensic readiness as a phase within a DFI framework as well as, individually, as a proposed proactive measure.

The inclusion of a pre-incident planning stage in a computer forensic investigation has started to grow since the Honeynet Project’s forensic challenge in 2001 (Tan, 2001). The mission of Honeynet Project, a non-profit organisation, was to raise awareness of the existing cyber threats by researching the strategies, motives, and tools of the cybercriminal community.

The project involved the digital forensic investigation and reporting of a compromised system by several forensic analysts (forensic challenge). It resulted in the forensic analysts spending over 80 hours in the investigation of a 2-hour criminal activity. The highlights of the forensic challenge were the substantial aggregate cost needed for each investigation and the disproportional time needed to examine each incident. The project’s ambition was to devise methods to automate as much of data collection and analysis in computer forensics.

Based on the Honeynet Project’s outcomes, John Tan (2001) introduced to the scientific world the notion of the forensic readiness (FR) definition and context within a digital forensic investigation. Up until then, digital forensic scientists and thinkers were only intrigued by researching post-incident forensics, i.e., all the actions taken by the investigation team after an incident has occurred. The research published identified measures, which may be incorporated into existing procedures for designing networks and deploying systems to increase digital forensic readiness. Tan (2001) defined and established two objectives (Figure 2) for a system to be forensically ready:

a) to maximise the usefulness of incident evidence data.
b) to minimise the cost of digital forensics during an incident response.

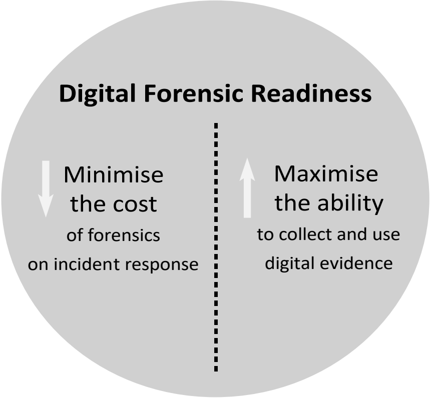

Digital Forensic Readiness
Minimise the cost
of forensics on incident response
Maximise the ability
to collect and use digital evidence

Figure 2. Digital forensic readiness aims

Additionally, Tan established five basic elements upon which a forensic readiness plan should be based:

a) how logging is done (mechanisms, time, time-stamping, permissions, reporting, retention);
b) what is logged (host/network);
c) intrusion detection systems;
d) forensic acquisition (volatile data, imaging);
e) evidence handling (chain of custody, network transport, physical transport, physical storage, examination).

In a study carried out by Carrier and Spafford (2003), the concepts and processes of physical investigations are used as a basis in proposing a process model for digital investigations. This novel study integrates both physical and digital crime scenes to devise a 5-phase investigation process which includes the first recorded study that proposes a pre-incident phase.

A systematic digital forensic readiness plan was originally presented by Rowlingson (2004) whose focal point was the study of proactive technical, procedural, legal and staff issues on pre-incident -oriented analysis. Pre-incident forensics was approached not only from a technical viewpoint, as the plan gives specific weight to procedures and processes underlining the need for organisational readiness. The study expanded Tan's theoretical output of DFR by formulating a 10-step framework. An organisation should be forensically ready to identify, preserve, collect, and extract admissible digital (and sporadically non-digital) evidence, and to maximise its exploitation. In this framework, the need for risk assessment in the organisation's critical assets along with the identification of different types of potential digital evidence is introduced.

Grobler and Louwrens (2007) discussed the overlap between information security and digital forensics, separating the proactive and reactive practice of digital forensics. They suggested that proactive digital forensics can complement information security as part of an organisation's security policy. The authors defined digital forensics readiness as the identification of all possible evidence sources and methods to gather evidence in a cost-effective and legal manner. They also suggested that the identification and collection of evidence is not enough by itself, but organisations must also implement a digital evidence record and document management system to automate document retention. Such system would provide for accessibility of retained documents, an accurate representation of the original format of the documents, as well as relevant document meta-data. The increasing need for digital evidence in organisations and the fact that very few organisations have the structures in place to enable them to conduct a cost-effective, low-impact and efficient digital investigations led researchers to propose alternative solutions (Grobler et al., 2010).

Evidence (Yeboah-Ofori and Brown, 2020) suggests that there is a general tendency for digital evidence to be identified and collected at a later part of the cycle; this subsequently leads to lost or damaged evidence. To determine the significance of incident response, the guide proposed a DFR plan and clearly defined the process of selecting and securing digital evidence at an early stage of an investigation. This study complemented the previous research by providing a more concise way of selecting and preserving digital evidence. A major part of this process is risk analysis but targeted to the identification and importance of digital evidence for an organisation. More specifically, the guide suggests an eight-process forensic readiness plan.

3. METRICS

Existing literature on DFR focuses on first line incident response, training requirements, tools enhancement and digital evidence management (Stoyanova et al., 2020). It is quite true that setting up the above will increase the forensic readiness of an organisation, yet in a profit-oriented market the most important variable would be minimising any costs.

Even though the cost of a DFI has been a continuous topic of academic discussions (Caviglione et al., 2017), investment decisions on a DFI have been limited up until now. Research has only been thoroughly conducted on metrics on investments on information security. Gordon et al. (2016) introduced a dynamic cost-benefit analysis model, the Gordon-Loeb model to aid practitioners and academics in deciding the level of investment in cybersecurity related activities. Throughout their highly influential work, the authors underly the importance of an organisation's ability to identify their own information sets. This process results in the assessment of an organisation's investment by evaluating and calculating individual cost and benefit variables.

Cavusoglu et al. (2004) presented an alternative cost benefit variable model based on IT technologies to respond to the ever-increasing need of quantifying the investment decision process. Bohme (2010) compared several scientific approaches and reviewed the impact of different investment methods and approaches on security budgetary requirements underlines that the key to a (security) investment methodology is the cost/benefit ratio.

Digital Forensic Readiness' basic objective is to maximise an organisation's ability to collect and use (admissible in-court) digital evidence. As a result, digital evidence should be available before an incident occurs. However, since each organisation is profit-oriented, the research on proactive forensics implementation should be directed towards cost-effectiveness. This is measured by identifying relevant costs (information security investments) and benefits (containment, recovery, recourse to litigation) and deciding on whether each measure is cost feasible.

Gordon & Loeb (2002) developed a mathematical framework (Gordon-Loeb model) to explore the optimal level of cyber security investments by considering the potential loss from a cyber security breach, the probability of a breach, as well as the reduction of the probability of the breach caused by investments in cyber security.

Similar to the Gordon-Loeb model, the optimal level of the application of a digital forensic readiness plan and activities should consider total costs and expected benefits. Figure 3 depicts the optimal level of digital forensic readiness activities that should consider risk assessment, business continuity planning and staff training effects.

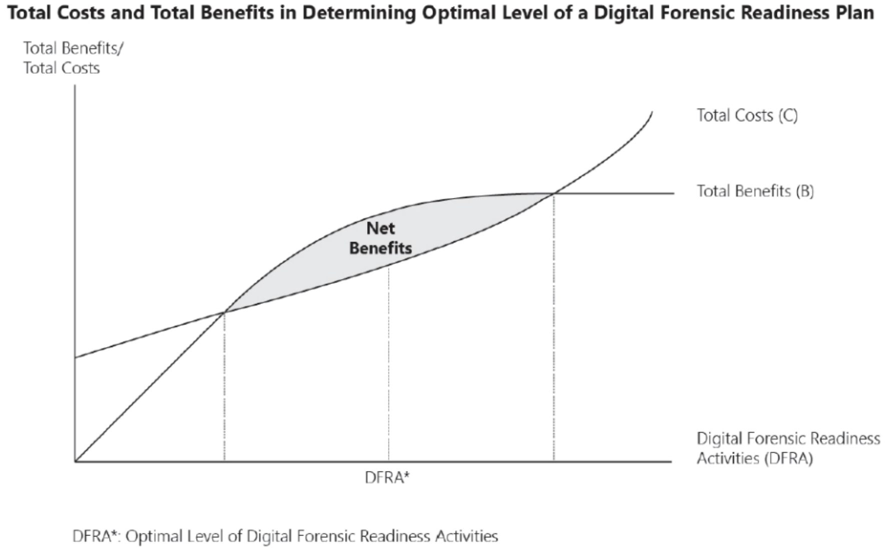


Figure 3. Assessment of a digital forensic readiness plan
(adapted from the Gordon and Loeb GLEIS model – Gordon et. al, 2016)

The identified cost and benefit variables need to be evaluated; a mixed methods approach was chosen by collecting both qualitative and quantitative data to evaluate cost and benefits of the application of a forensic readiness framework and assess the potential investment. The results, through a set of ANOVA tests, form the basis of a cost-benefit relation model. The model enables the calculation and (algorithmic) verification of the numerical relationship between cost and benefit variables. Such a relationship model forms the proposition of a digital forensic readiness planning framework that acts as a tool to aid decision makers within an organisation. The implementation of such a plan (Rowlingson, 2004) would incur the following benefits (each benefit variable is given a code name):

B_1	*Define business scenarios that require digital evidence*
	Evidence collection capability of the organisation in terms of risk and threats and vulnerabilities.
B_2	*Identify sources and types of potential evidence*
	Sources, equipment, application software, monitoring software etc.
B_3	*Determine the evidence collection requirement*
	Selection of sources and types of B_2 based on the risks, threats, and vulnerabilities of B_1
B_4	*Establish a capability to securely gather (legally admissible) evidence*
	Establishment of processes that ensures business continuity and legally admissible evidence.
B_5	*Establish a policy for storage and handling of evidence*
	Long term policies to ensure evidence integrity.
B_6	*Ensure monitoring and auditing mechanisms are in place*
	Event correlation in collaboration with intrusion detection mechanisms and honeypots.
B_7	*Specify triggering mechanisms that will lead to a digital forensic investigation*
	Decision criteria should be put in place to escalate any event to formal digital forensic investigation
B_8	*Staff training*
	Training to all staff members to raise awareness on each staff member's role pre, during and post incident.
B_9	*Evidence-based case presentation*
	Production of a policy that defines the development of an evidence-based case.
B_{10}	*Ensure legal review*
	The review should consider the evidence available and should also propose follow-up actions.
B_{11}	*Resolve a commercial dispute*
	Provision of any type of evidence to resolve any dispute.
B_{12}	*Support employee sanctions*
	Provision of digital evidence to support employee sanctions.

On the other hand, activities where costs would be incurred are (each cost variable is given a code name):

C_1	*Monitoring /tools and staff time*
C_2	*Software*
C_3	*Hardware*
C_4	*Staff training*
C_5	*Systematic gathering of potential evidence / to classify, index, prepare digital evidence*

C_6	*Secure storage of potential evidence*
C_7	*Cost of digital investigation*
C_8	*Legal advice*
C_9	*Policies updates*

4. METHODOLOGY

4.1 Hypothesis design

In quantitative research questions, hypotheses are predictions the researcher makes about the expected outcomes of relationships among variables (Creswell & Creswell, 2017). As described on the previous section, this study examines hypotheses between the relationships of cost (C) and benefit (B) variables. Each cost and benefit variable is assigned a code number (C_{1-9} and B_{1-12}) based on previous research output (Rowlingson, 2004). This section considers two sets of hypotheses: the first set examines nine (9) hypotheses (H_A), each hypothesis examining the effect that each cost variable has on the total benefit (TB) variable. The second set considers one hundred and eight (108) hypotheses (H_{CB}), each hypothesis examining the effect of each cost variable to each benefit variable. For all hypotheses, the H_0 null hypothesis is outlined to examine whether there is a significant correlation between the two variables. The significance of the correlation will be determined by comparing the probability value (p-value) and the level of significance assuming the null hypothesis (H_0) is true. The significance level has been set at $p = 0.05$. The hypotheses considered are listed in the following section.

4.2 Hypotheses list

Hypothesis H_{A1}

H_0 - Null Hypothesis. Monitoring (C_1) and total benefit (TB) are not related; the monitoring (C_1) cost variable has no effect on total benefit (TB).

H_1 - The monitoring (C_1) cost variable has an effect on total benefit (TB).

Hypotheses $H_{Ax\ (x=2,3,4,5,6,7,8,9)}$ have been designed in the same way as above.

Hypothesis H_{C1B1}

H_0 - Null Hypothesis. Monitoring (C_1) and business scenarios ($B_{\ 1}$) are not related; the monitoring (C_1) cost variable has no effect on business scenarios (B_1).

H_9 - The monitoring (C_1) cost variable has an effect on business scenarios (B_1).

Hypotheses $H_{C1Bx\ (x=2,3,4,5,6,7,8,9,10,11,12)}$ have been designed in the same way as above.

Hypothesis H_{C2B1}

H_0 - Null Hypothesis. Software (C_2) and business scenarios (B_1) are not related; the software (C_2) cost variable has no effect on business scenarios (B_1).
H_9 - The software (C_2) cost variable has an effect on business scenarios (B_1).
Hypotheses $H_{C2Bx\ (x=2,3,4,5,6,7,8,9,10,11,12)}$ have been designed in the same way as above.

Hypothesis H_{C3B1}

H_0 - Null Hypothesis. Hardware (C_3) and business scenarios (B_1) are not related; the hardware (C_3) cost variable has no effect on business scenarios (B_1).
H_9 - The hardware (C_3) cost variable has an effect on business scenarios (B_1).
Hypotheses $H_{C3Bx\ (x=2,3,4,5,6,7,8,9,10,11,12)}$ have been designed in the same way as above.

Hypothesis H_{C4B1}

H_0 - Null Hypothesis. Staff training (C_4) and business scenarios (B_1) are not related; the staff training (C_4) cost variable has no effect on business scenarios (B_1).
H_9 - The staff training (C_4) cost variable has an effect on business scenarios (B_1).
Hypotheses $H_{C4Bx\ (x=2,3,4,5,6,7,8,9,10,11,12)}$ have been designed in the same way as above.

Hypothesis H_{C5B1}

H_0 - Null Hypothesis. Log files (C_5) and business scenarios (B_1) are not related; the log files (C_5) cost variable has no effect on business scenarios (B_1).
H_9 - The log files (C_5) cost variable has an effect on business scenarios (B_1).
Hypotheses $H_{C5Bx\ (x=2,3,4,5,6,7,8,9,10,11,12)}$ have been designed in the same way as above.

Hypothesis H_{C6B1}

H_0 - Null Hypothesis. Secure storage (C_6) and business scenarios (B_1) are not related; the secure storage (C_6) cost variable has no effect on business scenarios (B_1).
H_9 - The secure storage (C_6) cost variable has an effect on business scenarios (B_1).
Hypotheses $H_{C6Bx\ (x=2,3,4,5,6,7,8,9,10,11,12)}$ have been designed in the same way as above.

Hypothesis H_{C7B1}

H_0 - Null Hypothesis. Cost of digital investigation (C_7) and business scenarios (B_1) are not related; the cost of digital investigation (C_7) cost variable has no effect on business scenarios (B_1).

H_9 - The cost of digital investigation (C_7) cost variable has an effect on business scenarios (B_1).

Hypotheses $H_{C7Bx\ (x=2,3,4,5,6,7,8,9,10,11,12)}$ have been designed in the same way as above.

Hypothesis H_{C8B1}

H_0 - Null Hypothesis. Legal advice (C_8) and business scenarios (B_1) are not related; the legal advice (C_8) cost variable has no effect on business scenarios (B_1).

H_9 - The legal advice (C_8) cost variable has an effect on business scenarios (B_1).

Hypotheses $H_{C8Bx\ (x=2,3,4,5,6,7,8,9,10,11,12)}$ have been designed in the same way as above.

Hypothesis H_{C9B1}

H_0 - Null Hypothesis. The policies updates (C_9) cost variable and business scenarios (B_1) are not related; the policies updates (C_9) cost variable has no effect on business scenarios (B_1).

H_9 - The policies updates (C_9) cost variable has an effect on business scenarios (B_1).

Hypotheses $H_{C9Bx\ (x=2,3,4,5,6,7,8,9,10,11,12)}$ have been designed in the same way as above.

4.3 Questionnaire design

The questionnaires developed were self-administered (SAD), explicitly designed to be completed without the researchers' intervention. The questionnaire was disseminated to the selected organisations sample electronically (e-mail), where it was asked by respondents to return the completed questionnaires within 8 weeks of their receipt. The current research is based on assessing potential costs and benefits of organisations of four types (Healthcare, Financial, Services, Information Technology). The selected organisations were large enterprises (>250 employees)The response rate of the questionnaires was 68% (115/170) over a period of two months.

To test the hypotheses described in Section 4.2, two types of questionnaires were designed and distributed in two separate periods of time. This study is limited in analysing and discussing the data and results of the first questionnaire distributed during the first period. The questionnaire was designed to be self-administered for two reasons:

(a) for the organisations to be able to respond to the questions truthfully, without any time constraints.
(b) self-administered questionnaires are unbiased in terms of social desirability responding regarding financial behaviour and organisational strategy (Kelly, 2015). The methodology and design of this research paper was based on valid and reliable scientific research tools widely used in academic research (Creswell & Creswell, 2017).

The questionnaires consisted entirely of closed-ended questions and scales. A scale score is a range of values representing all possible answers within a continuum. It is designed in such a way for all the variables within the specific research area to be aligned with the above values. Also, a scale should meet the three following criteria (Kent, 2015):

(a) all possible answers should be included.
(b) all answers should be mutually exclusive.
(c) all answers should only refer to one dimension.

This research utilises a five-level Likert type scales, which represent five linear responses within a continuum. Respondents were asked to respond on the level of importance of the topic in question, ranging from ‘not important’ to ‘very important’ and from ‘low’ to ‘high’. Each response was given a specific number to assist in the quantification of data and in further inferential statistical analysis:

(a) 1-not important; 2-slightly important; 3-moderately important, 4-important; 5-very important.
(b) 1-low; 2-below average; 3-average; 4-above average; 5-high.

The questionnaires consisted of three main sections:

(a) main demographics.
(b) cost-benefit metrics.
(c) digital forensic readiness technical details.

The questions developed were based on Rowlingson's (2004) research, who identified costs and benefit variables that incur when an organisation applies a digital forensic readiness plan.

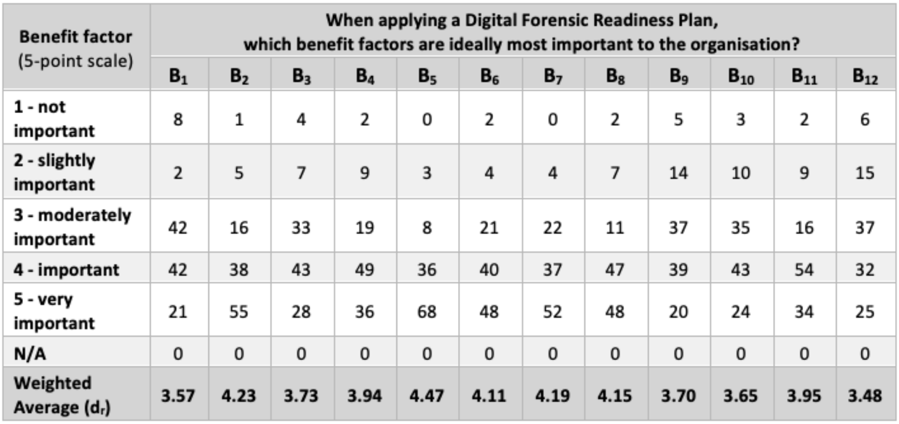
5. DATA ANALYSIS

5.1 Cost benefit variable analysis - Benefit variables

The respondents were asked to rate, on a scale of 1 to 5, the level of benefit on two different states: the optimal level of benefit organisations are expected to obtain when applying a digital forensic readiness plan (Table 1), and, the current level of benefit obtained when applying the same plan (Table 2).

The assessment of the benefit variables was conducted according to the list of 12 benefit variables (Bx, where x=1, 2, …,12) provided in Section 2.2, where the organisations rated each benefit variable on a Likert scale (1 to 5). In Table 1 (optimal level of benefit variables), 1 represents 'not important', and 5 represents 'very important', while in Table 2 (actual level of benefit variables), 1 represents 'low' and 5 represents 'high'.

Table 1. Respondents' data on benefit variables collected: optimal assessment

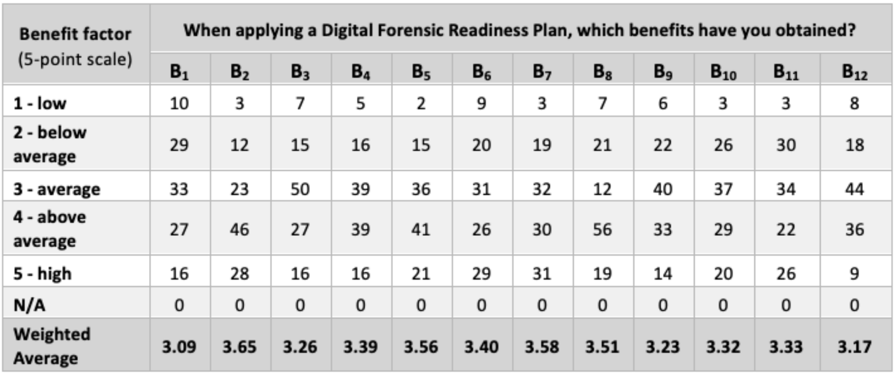


Benefit factor (5-point scale)	**When applying a Digital Forensic Readiness Plan, which benefit factors are ideally most important to the organisation?**											
	B_1	**B_2**	**B_3**	**B_4**	**B_5**	**B_6**	**B_7**	**B_8**	**B_9**	**B_{10}**	**B_{11}**	**B_{12}**
1 - not important	8	1	4	2	0	2	0	2	5	3	2	6
2 - slightly important	2	5	7	9	3	4	4	7	14	10	9	15
3 - moderately important	42	16	33	19	8	21	22	11	37	35	16	37
4 - important	42	38	43	49	36	40	37	47	39	43	54	32
5 - very important	21	55	28	36	68	48	52	48	20	24	34	25
N/A	0	0	0	0	0	0	0	0	0	0	0	0
Weighted Average (d_r)	**3.57**	**4.23**	**3.73**	**3.94**	**4.47**	**4.11**	**4.19**	**4.15**	**3.70**	**3.65**	**3.95**	**3.48**

The above table (Table 1) shows the responses on the ideally most important benefit variables to the organisation. B_2 (sources and types of digital evidence), B_5 (policy for storage and handling of evidence), B_6 (monitoring and auditing), B_7 (triggering mechanisms), B_8 (staff training) have gotten higher rankings. The organisations questioned believe that the above benefit factors are more important than others when applying a digital forensic readiness plan.

The weighted average variable calculates the proportional average comparison of the results reveal that organisations responded that B_2, B_5, B_6, B_7, B_8 were the ideally highest rated benefit variables within the application of a digital forensic readiness plan, while B_2, B_5, B_7 (as shown in Table 2) were the variables that are beneficial to organisations when applying a DFR plan. This data portrays that organisations do consider some variables more beneficial than others within the digital forensic readiness domain. B_2 (sources and types of digital evidence), B_5 (policy for storage and handling of evidence), B_6 (monitoring and auditing), B_7 (triggering mechanisms), B_8 (staff training) have gotten higher rankings.

Table 2. Respondents' data on benefit variables collected: actual assessment



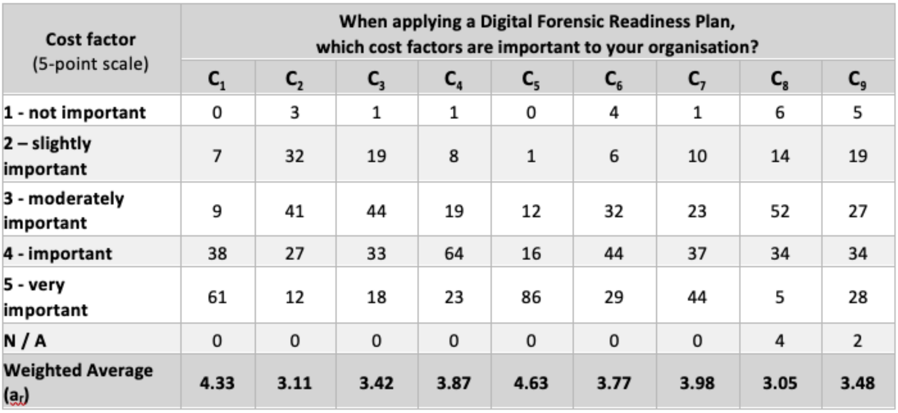
Benefit factor (5-point scale)	When applying a Digital Forensic Readiness Plan, which benefits have you obtained?											
	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}	B_{11}	B_{12}
1 - low	10	3	7	5	2	9	3	7	6	3	3	8
2 - below average	29	12	15	16	15	20	19	21	22	26	30	18
3 - average	33	23	50	39	36	31	32	12	40	37	34	44
4 - above average	27	46	27	39	41	26	30	56	33	29	22	36
5 - high	16	28	16	16	21	29	31	19	14	20	26	9
N/A	0	0	0	0	0	0	0	0	0	0	0	0
Weighted Average	**3.09**	**3.65**	**3.26**	**3.39**	**3.56**	**3.40**	**3.58**	**3.51**	**3.23**	**3.32**	**3.33**	**3.17**

5.2 Cost benefit variable analysis - Cost variables

The respondents were asked to rate, on a scale of 1 to 5, the level of cost on two different states: the level of costs involved in organisations when applying a digital forensic readiness plan (Table 3), and the current level of cost spent when applying the same plan (Table 4).

The assessment of the cost variables was conducted according to the list of 9 cost variables (C_x, where x=1, 2, …,9) provided in Section 3, where the organisations rated each cost variable on a Likert scale (1 to 5). In Table 3 (importance of cost variables), 1 represents 'not important', and 5 represents 'very important', while in Table 4 (actual level of cost variables), 1 represents 'low' and 5 represents 'high'.

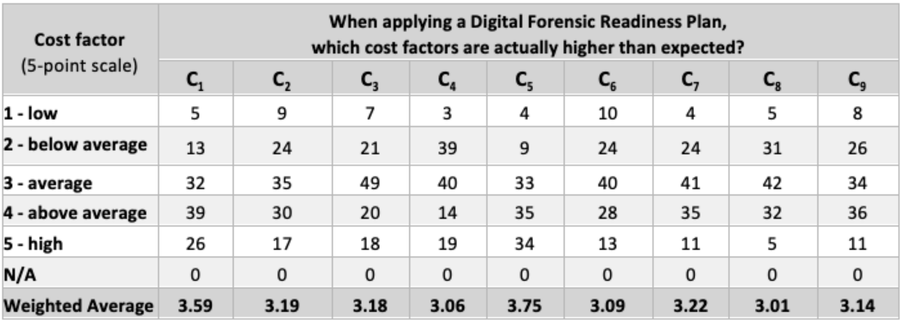
Table 3. Respondents' data on the importance of cost variables collected: optimal assessment



Cost factor (5-point scale)	**When applying a Digital Forensic Readiness Plan, which cost factors are important to your organisation?**								
	C_1	**C_2**	**C_3**	**C_4**	**C_5**	**C_6**	**C_7**	**C_8**	**C_9**
1 - not important	0	3	1	1	0	4	1	6	5
2 – slightly important	7	32	19	8	1	6	10	14	19
3 - moderately important	9	41	44	19	12	32	23	52	27
4 - important	38	27	33	64	16	44	37	34	34
5 - very important	61	12	18	23	86	29	44	5	28
N / A	0	0	0	0	0	0	0	4	2
Weighted Average (a_r)	**4.33**	**3.11**	**3.42**	**3.87**	**4.63**	**3.77**	**3.98**	**3.05**	**3.48**

Organisations responded that C_1 (monitoring) and C_5 (systematic gathering of potential evidence) were the highest rated cost variables within the application of a digital forensic readiness plan. Thus, an investment on the above cost items would be more likely to take place when applying a DFR plan. Similarly, Table 4 portrays that C_1 (monitoring) and C_5 (systematic gathering of potential evidence) were the variables that are costing more than expected to the organisations when applying a DFR plan.

Table 4. Respondents' data on cost variables collected: actual assessment



Cost factor (5-point scale)	**When applying a Digital Forensic Readiness Plan, which cost factors are actually higher than expected?**								
	C_1	**C_2**	**C_3**	**C_4**	**C_5**	**C_6**	**C_7**	**C_8**	**C_9**
1 - low	5	9	7	3	4	10	4	5	8
2 - below average	13	24	21	39	9	24	24	31	26
3 - average	32	35	49	40	33	40	41	42	34
4 - above average	39	30	20	14	35	28	35	32	36
5 - high	26	17	18	19	34	13	11	5	11
N/A	0	0	0	0	0	0	0	0	0
Weighted Average	**3.59**	**3.19**	**3.18**	**3.06**	**3.75**	**3.09**	**3.22**	**3.01**	**3.14**

5.3 Total cost and total benefit

In this section we introduce the concepts of total cost (TC) and total benefit (TB). The two concepts are defined here to stress the importance and different weighting of each benefit and cost variable. Both TC and TB take into account the weighted average and thus do not treat each variable evenly. The actual cost and benefit of each organisation is calculated to signify the total costs and benefits through the application of a digital forensic readiness plan.

The Total Cost for each organisation is defined by:

$$\underline{\mathbf{C}} = (\sum_{\mathbf{r=1}}^{\mathbf{p}} \mathbf{a_r} \bullet \mathrm{C_r})/\mathrm{p}, (\mathrm{p} \leq 9)$$

where p is the number of cost factors, a is the value of each cost factor, C is the weighted average of each cost factor; r can have a value of 1, 2, 3, 4, or 5.

Table 1. Total benefit clustering

TOTAL BENEFIT			
Range	**Frequency**	**%**	**Average TB**
7-12	13	11.3	10.21
12,01-15	22	19.1	14.07
15,01-17	47	40.9	16.05
17,01+	33	28.7	17.47

The Total Benefit for each organisation is defined by:

$$\underline{\mathbf{B}} = (\sum_{\mathbf{r=1}}^{\mathbf{p}} \mathbf{d_r} \bullet \mathrm{B_r})/\mathrm{p}, (\mathrm{p} \leq 12)$$

where p is the number of benefit factors, a is the value of each benefit factor, B is the weighted average of each benefit factor; r can have a value of 1, 2, 3, 4, or 5.

Table 2. Total cost clustering

TOTAL COST			
Range	**Frequency**	**%**	**Average TC**
7-12	14	12.2	10.12
12,01-14	22	19.1	13.37
14,01-16	60	52.2	15.06
16,01+	19	16.5	16.61

Following the calculation of the total benefit and total cost for each organisation, the results were classified into four arrays and presented in Table 1 The Total Benefit for each organisation is defined by:

$$\underline{\mathbf{B}} = (\sum_{\mathbf{r=1}}^{\mathbf{p}} \mathbf{d_r} \bullet \mathrm{B_r})/\mathrm{p}, (\mathrm{p} \leq 12)$$

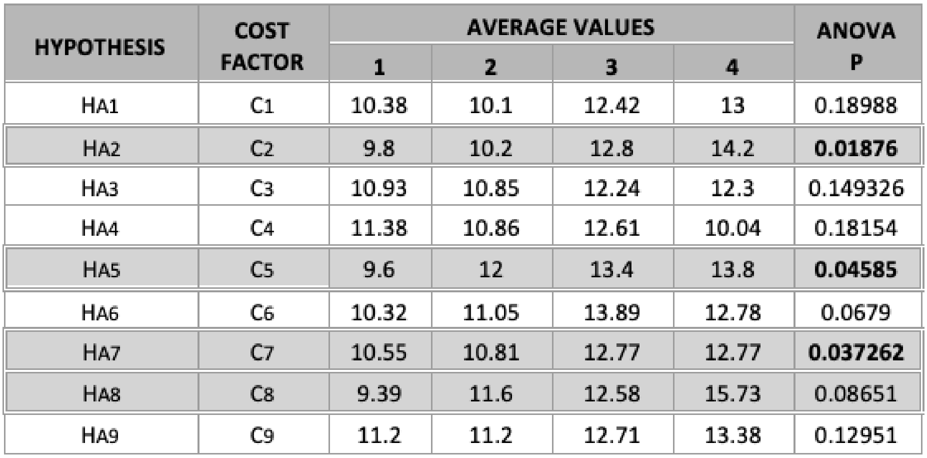
where p is the number of benefit factors, a is the value of each benefit factor, B is the weighted average of each benefit factor; r can have a value of 1, 2, 3, 4, or 5.

Table 2.

5.4 Cost benefit variable analysis

Hypotheses H_{AX} (where x=1, 2, …, 9), detailed in Section 3.2, test the effect each cost variable has on the TB variable. To test the hypotheses, several ANOVA tests were conducted to determine the significance of each effect based on testing whether $p < 0.05$. Cost variables were the dependent variables, and the total benefit was the independent variable. Table 7 depicts the results of the analysis among the four arrays, where C_2, C_5, C_7, and C_8 have got a p-value less than 0.05 ($p<0.05$). This means that hypotheses H_{A2}, H_{A5}, H_{A7}, H_{A8} are accepted, while H_{A1}, H_{A3}, H_{A4}, H_{A6}, and H_{A9} are rejected ($p>0.05$). This result supports the statement that increasing investment in cost variables C_2, C_5, C_7, and C_8 will significantly increase the Total Benefit variable compared to cost variables C_1, C_3, C_4, C_6 and C_9. Thus, an increase in the investment of variables C_2, C_5, C_7, and C_8 will lead to a substantial increase in the Total Benefit variable within an organisation when applying a DFR plan.

Table 7. ANOVA testing between TB and CX variables (x=1, ... , 9)

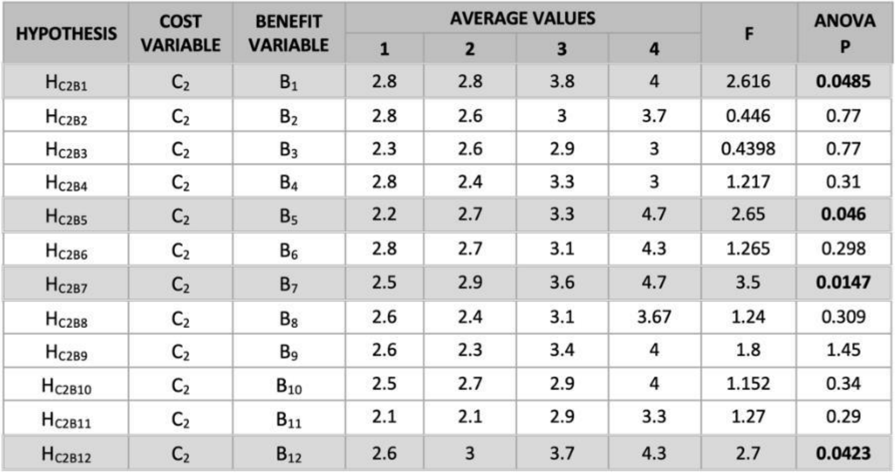


HYPOTHESIS	COST FACTOR	AVERAGE VALUES				ANOVA P
		1	2	3	4	
HA1	C1	10.38	10.1	12.42	13	0.18988
HA2	C2	9.8	10.2	12.8	14.2	**0.01876**
HA3	C3	10.93	10.85	12.24	12.3	0.149326
HA4	C4	11.38	10.86	12.61	10.04	0.18154
HA5	C5	9.6	12	13.4	13.8	**0.04585**
HA6	C6	10.32	11.05	13.89	12.78	0.0679
HA7	C7	10.55	10.81	12.77	12.77	**0.037262**
HA8	C8	9.39	11.6	12.58	15.73	0.08651
HA9	C9	11.2	11.2	12.71	13.38	0.12951

The above analysis examined the effect of each variable on the Total Benefit variable; to determine the effect each cost variable has on each benefit variable 108 hypotheses (H_{CB}), detailed in Section 4.2 were tested. For this reason, 108 ANOVA tests were conducted to evaluate the relationship between each one of the nine cost variables (dependent variable) and twelve benefit (independent variable) variables. The significance level has also been set at p = 0.05.

Table 8 displays the ANOVA results in testing hypotheses H_{C2} B_{X}, examining the effect of the C_2 cost variable to each benefit variable B_X (x=1,2, …, 12). The assessment results show that hypotheses H_{C2B1}, H_{C2B5}, H_{C2B7}, and H_{C2B12} are accepted (p<0.05), while H_{C2B2}, H_{C2B3}, H_{C2B4}, H_{C2B6}, H_{C2B8}, H_{C2B9}, H_{C2B10}, H_{C2B11} are rejected because p>0.05. These results show that increasing investment in C_2 will significantly increase the benefit B_1, B5, B7, B12 obtained by an organisation in a digital forensic readiness context.

Table 8. ANOVA testing between C_2 and B_X variables (x=1, …, 12)



HYPOTHESIS	COST VARIABLE	BENEFIT VARIABLE	AVERAGE VALUES				F	ANOVA P
			1	2	3	4		
H_{C2B1}	C_2	B_1	2.8	2.8	3.8	4	2.616	**0.0485**
H_{C2B2}	C_2	B_2	2.8	2.6	3	3.7	0.446	0.77
H_{C2B3}	C_2	B_3	2.3	2.6	2.9	3	0.4398	0.77
H_{C2B4}	C_2	B_4	2.8	2.4	3.3	3	1.217	0.31
H_{C2B5}	C_2	B_5	2.2	2.7	3.3	4.7	2.65	**0.046**
H_{C2B6}	C_2	B_6	2.8	2.7	3.1	4.3	1.265	0.298
H_{C2B7}	C_2	B_7	2.5	2.9	3.6	4.7	3.5	**0.0147**
H_{C2B8}	C_2	B_8	2.6	2.4	3.1	3.67	1.24	0.309
H_{C2B9}	C_2	B_9	2.6	2.3	3.4	4	1.8	1.45
H_{C2B10}	C_2	B_{10}	2.5	2.7	2.9	4	1.152	0.34
H_{C2B11}	C_2	B_{11}	2.1	2.1	2.9	3.3	1.27	0.29
H_{C2B12}	C_2	B_{12}	2.6	3	3.7	4.3	2.7	**0.0423**

This section reports the detailed results of hypothesis H_{C2BX} (12 ANOVA tests), which shows the positive effect that C_2 has on B_1, B_5, B_7, B_{12}. Similarly, all H_{CB} hypotheses test the effect relationship between C_X (x=1,3, …, 9) and B_X (x=1,2, …, 12). The complete relationship effect diagram depicting all 108 ANOVA tests is presented in Figure 4. Each arrow represents a significant effect that a cost variable has to the benefit variable it points. It is worth stating that each cost variable has got an effect, direct or indirect, on each variable to some degree, however certain cost variables have a stronger effect on benefit variables and will increase them accordingly.

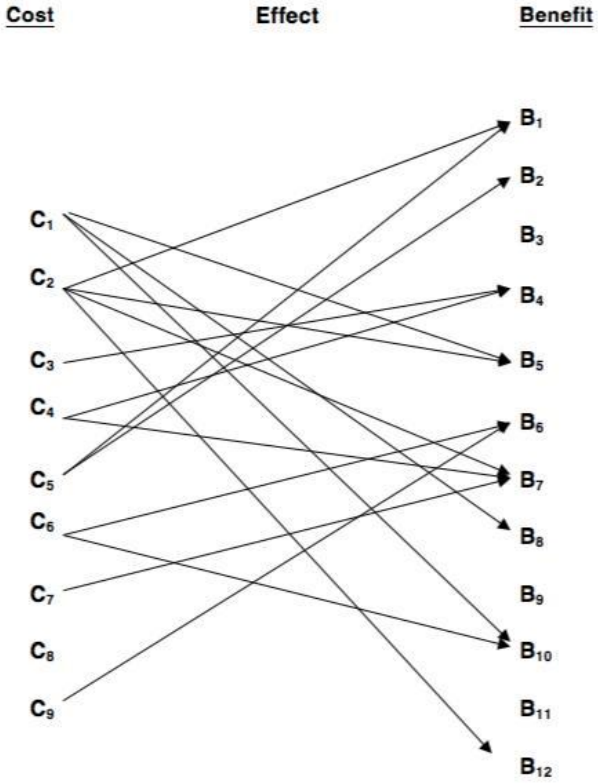


Figure 4. Digital forensic readiness planning: a cost-benefit relation model

The cost-benefit relation model can act as a decision-making tool when organisations would want to decide on the level of investment to be made on all cost variables. For instance, in a digital forensic readiness context, should an organisation want to improve the benefit variable B_7, it would increase the investments made to C_2, C_4, and C_7. Likewise, increasing expenditures on C_1, the benefits obtained for B_5, B_8, B_{10} would also be improved.

6 FUTURE RESEARCH & CONCLUSIONS

Discussions of the digital forensic process tend to ignore what happens to the object of the investigation prior to the decision to undertake an investigation. The necessary evidence either exists (and hopefully is discovered by a digital forensic investigation), or it does not exist, and a suspect cannot be charged and prosecuted (Kebande et al., 2020). This is the law enforcement view of a digital forensic investigation. It begins when a crime has been committed or discovered and investigators attend a crime scene or wish to seize evidence. The quality and availability of evidence is a passive aspect of an investigation.

In a cost-oriented context however, there is the opportunity to actively collect potential evidence in the form of log files, emails, back-up disks, portable computers, network traffic records, and telephone records, amongst others. These pieces of evidence may be collected in advance of a crime or dispute and can be used to the benefit of the collecting organisation.

In this paper we have expressed the need for increasing awareness on the cost-side of digital forensic investigations. Our aim was to conduct a cost benefit analysis and propose a model that will act as a decision-making tool to organisations in applying a digital forensic readiness system.

The identification and evaluation of benefit and cost variables within the application of a digital forensic readiness plan has been one of the key contributions of this article. The terms total cost and total benefit were also introduced together with their overall frequency and distribution table.

The results of the data collection process amongst organisations and institutions when applying a digital forensic readiness plan were firstly introduced, while descriptive data analysis and hypotheses testing results were presented. Our main aim was to determine the validity between each cost-benefit relationship. For this reason, we have proposed a cost-benefit variable relationship model based on the digital forensic readiness concept.

Our future work aims to further develop the cost benefit relation model by designing a Bayesian Network to depict cost-benefit relationship effects and estimate conditional probability distributions. The model will be verified by performing inference and applying a Junction-Tree algorithm.

Overall, this work is part of my ongoing research in presenting a holistic cost-benefit model to aid institutions in the decision-making process within a digital forensic readiness framework.

7 REFERENCES

Böhme, R. (2010). Security metrics and security investment models. In I. Echizen, N. Kunihiro, R. Sasaki (Eds.), *Advances in information and computer security*, *IWSEC 2010. Lecture Notes in Computer Science Volume 6434*. Springer.

Cabinet Office (2018). *HMG security policy framework: version 1.1.*

Carrier, B., & Spafford, E. H. (2003). Getting physical with the digital investigation process. *International Journal of Digital Evidence*, *2*(2), 1-20.

Caviglione, L., Wendzel, S., & Mazurczyk, W. (2017). The future of digital forensics: challenges and the road ahead. *IEEE Security & Privacy 15*(6), 12-17.

Cavusoglu, H., Mishra, B., & Raghunathan, S. (2004). A model for evaluating IT security investments. *Communications of the ACM*, *47*(7), 87-92.

Chew, E., Swanson, M., Stine, K., Bartol, N., Brown, A., & Robinson, W. (2008). *Performance measurement guide for information security*. NIST.

Ciardhuáin, S. Ó. (2004). An extended model of cybercrime investigations. *International Journal of Digital Evidence, 3*(1), 1-22.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Gordon, L. A., & Loeb, M. P. (2002). The economics of information security investment. *ACM Transactions on Information and System Security, 5*(4), 438-457. http://dx.doi.org/10.1145/581271.581274

Gordon, L. A., Loeb, M. P., & Zhou, L. (2016). Investing in cybersecurity: insights from the Gordon-Loeb model. *Journal of Information Security*, *7*(2), 49-59.

Grobler, C. P., & Louwrens, C. P. (2007). Digital forensic readiness as a component of information security best practice. In H. Venter, M. Eloff, L. Labuschagne, R. von Solms (Eds.), *IFIP International Federation for Information Processing Volume 232* (pp. 13-24). Springer.

Grobler, C. P., Louwrens, C. P., & von Solms, S. H. (2010). A framework to guide the implementation of proactive digital forensics in organisations. In *2010 International conference on availability, reliability and security* (pp. 677-682). IEEE.

Kebande, V. R., Phathutshedzo, P. M., Ikuesan, R. A., Venter, H. S., & Choo, K.-K. R. (2020). Holistic digital forensic readiness framework for IoT-enabled organizations. *Forensic Science International: Reports 2*, 100-117.

Kebande, V. R., Venter, H. S. (2019). A comparative analysis of digital forensic readiness models using CFRaaS as a baseline. *WIREs Forensic Science*, *1*(6).

Kelly, N. L. (2015). *Examining social desirability bias in measures of financial behaviour* (Publication No 464) [Master's thesis, Illinois State University].

Kent, R. A. (2015). *Analysing quantitative data: variable-based and case-based approaches to non-experimental datasets*. Sage publications.

Rowlingson, R. (2004). A ten-step process for forensic readiness. *International Journal of Digital Evidence*, *2*(3), 1-28.

Singhal, A., & Ou, X. (2017). Security risk analysis of enterprise networks using probabilistic attack graphs. In L. Wang, S. Jajodia, A. Singhal (Eds.), *Network security metrics* (pp. 53-73). Springer.

Stoyanova, M., Nikoloudakis, Y., Panagiotakis, S., Pallis E., & Markakis, E. K. (2020). Survey on the Internet of Things (IoT) forensics: challenges, approaches, and open issues. *IEEE Communications Surveys & Tutorials*, *22*(2), 1191-1221.

Tan, J. (2001). Forensic readiness. *Cambridge, MA:@ Stake*, 1-23.

Tully, G., Cohen, N., Compton, D., Davies, G., Isbell, R., & Watson, T. (2020). Quality standards for digital forensics: Learning from experience in England and Wales. *Forensic Science International: Digital Investigation*, *32*.

Yeboah-Ofori, A., & Brown, A. D. (2020). Digital forensics investigation jurisprudence: issues of admissibility of digital evidence. *Journal of Forensic, Legal & Investigative Sciences*, *6*(1), 1-8.

BIOGRAPHICAL NOTES

Antonis Mouhtaropoulos is a computer security and digital education researcher and is currently serving as the Dean of Digital Learning and EdTech at Metropolitan College, Greece. He was previously employed at De Montfort University, Leicester UK. Antonis has got more than fourteen years of experience in teaching and researching in the higher education computer science domain. His research interests lie in the overlap of information security, digital forensics, and economics, as well as on current trends in the digital learning ecosystem. He is also examining the application of proactive computer-system forensics across organisations in the fields of event reconstruction, event sequencing and data carving. He has been involved, as a general chair, in the organisation of four international conferences, and of several workshops. In addition, he has delivered keynote speeches in Japan, Malaysia, and Greece, has served as a reviewer for peer-reviewed scientific journals and as a member of the program committees for several international conferences. He is a Fellow of the Royal Society of Arts and, also, a Member of the IEEE, ACM, and the British Computer Society.

REFERENCE

Reference to this paper should be made as follows: Mouhtaropoulos, A. (2021). Digital forensic readiness of information systems: A cost-benefit variable analysis. *International Journal on Cyber Situational Awareness*, Vol. 6, No. 1, pp23-45.