THE USE OF BIG DATA IN THE FIELD OF DIGITAL FORENSICS INVESTIGATIONS
(COMPARATIVE STUDY BETWEEN DIGITAL FORENSICS IN UK AND NIGERIA)

Dr Uma Mohan & Saminu Salisu

Abstract: The growing reliance on digital devices has increased data created and stored on daily basis around the world exponentially, with more and more data analyst keeping logs about what data is stored and how the data is used, it is becoming a big challenge for digital investigators to process large amounts of data easily.

This paper would take a closer look into the challenges of using big data in the filed of digital forensics investigation, the study further suggests alternative ways to process large amounts of data effectively with the use of open source techniques and tools.

Keywords: Big Data, Digital Forensics, Digital Investigation, Digital Evidence, Digital Forensics Nigeria, Digital Forensics UK

Track: Big Data In the field of digital forensics

I. Introduction to Research Field

Digital Forensics investigation is a fairly new science of study and majority of the investigation involves the sourcing of digital information stored on mobile devices, computers, game console and other various media storage mediums for the purpose of relevance to civil and criminal investigation.

The common approach in every digital investigation involves the conventional process which includes Identification, Acquisition, Preservation, Examination and Presentation of findings to the chief investigator, court of law and other stakeholders by expert investigators where decision are made on the outcome of an investigation hence, making it a highly delicate process on the outcome of many modern cases, software’s such as Forensics Tool Kit (FTK) and ENCASE are used to carry out this process.

The above conventional process limits the scope of investigation and examination of evidence by to mainly the examiners and investigators, and these with no surprise raises challenges with the proliferative nature of big data spreading across different sectors and professions, this is where e-Discovery as an aspect of digital forensics is applicable to bridge these gaps.

e-Discovery approach enables rapid examination of information from different sources, sectors and different devices with the added advantage of involving stakeholders from other sectors to collaborate with law enforcement agencies during the investigation process.

Available online @ www.ijntse.com
Big Data
The growing reliance on digital devices has increased data stored around the world exponentially, with more and more data analyst keeping logs about what data is stored and how the data is used, its commonly know as “Data about data” technically called metadata.

Big data is not only about volume of data nevertheless also about Velocity and Variety commonly abbreviated as the 3-Vs (Volume, Velocity and Volume), this ever-growing technique relies on collecting, storing large volumes of data with related metadata for filtering and cleansing. Making sense of big data has been a significant factor in various sectors for example, the U.S government’s homeland Security collect store and analyse terabytes of data every day for intelligence on treats to national security, and also in the UK where big data is being used to solve crimes both domestically and internationally with evidence to the on going missing 3 yrs. old Madeleine McCann who disappeared when on holiday with her parents in 2007, the latest breakthrough in the investigation by Scotland Yard recently has been the use of big data which included phone records and is believed to hold the key to solving the case.

Advancement in the field of digital forensics has made the extremely hard task of analysing large amount of data considerably easy and in a cost effective with limited time scale with the help of techniques such as E-Discovery, Levenshtein Algorithm.

Research Aim
The project will evaluate the current state of digital forensics investigation in the UK with emphasis on law enforcement agency. It will analyse the information’s collected from various stages of the research and suggest cost effective techniques and model for carrying out effective digital investigation with Big Data.

II. Literature Review

Digital Forensics
Digital Forensics is a novel fields hence the limited academic publications on the study area. The following section would be analysing and discussing the literature reviews on digital forensics.

(Ruibin, Yun and Gaertner 2005)’s paper on is an improvement on the work carried out by Beebe and Carrier, the author suggested several new concepts that emphasises on the reuse of knowledge from by investigator form previous cases the concept reads ‘seek knowledge’ and ‘knowledge reuse’. The author stresses that seek knowledge as the investigative clue that defines analysis of data, while ‘knowledge reuse’ it he use of know knowledge across multiple cases.

Case relevance as describedby (Ruibin, Yun and Gaertner 2005) is a concept flow that is used to measure the strength of the 6 common investigative questions the “who, what, where, when, why, and how” in a digital investigation. The diagram below shows the level and direction of case significance used to show the various case-relevance in computer forensics investigations.

Figure 5: (Ruibin G., 2005)

Discussion
With consideration on the time of publishing of the (Pollitt, 1995) paper, digital forensic has come a long way with more steps now included the process of evidence extraction. The paper can be excused for not including Evidence Preservation and Reporting stages of digital investigation, that are essential parts of any successful modern
forensics investigation. However, (Kent, Chevalier, Grance and Dang 2006) showed a strong understanding of digital forensics investigation stages, a proper arrangement of the processes was not present, the author also eliminated identification and preservation stages as major steps in carrying out any digital forensics investigation.

**Benefits of Digital Forensics**

(Weiss 2009) suggested emphasises on the power of digital footprint in the modern day and age, digital trails are already used by investigator to know the history of a suspect or what a suspect carried out before, what he is doing in the present and nevertheless the interesting part would be to be able to predict future crimes by the past history and present state of a suspect which raises the question “can digital trail be used to predict actions of an individual before they happen in the future?” (Weiss, 2009).

The author suggests that digital forensics investigation could be used to help counter crimes that are yet to happen rather than wait for them to happen before investigating. Though the idea sounds like a highly effective way of preventing harmful events from happening privacy law and civil activist fight the use of required amount of private information necessary to effectively monitor and predict events.

**Discussion**

However, (Weiss, 2009) further elaborates on the amount of digital traces left behind after the use of devices in the modern society, with every action we make or anything we touch now leaves digital trail that can be traced back to us which is why privacy is non-existence. (Weiss, 2009). Online advertising agencies such as Google remarketing are now using ‘behavioural targeting’ model to anticipate a user’s future purchase based on their purchase history.

**Big Data**

Big Data as defined by (Chaudhary & Shah, 2013) is said to be a large volume of data sourced from various mediums such as Social Media, Health Care, Transport, ecommerce, mobile phones, satellite communications, GPS systems, media sharing through handheld devices and other modern day means of communication.

However, (Baoan, 2014) also defined Big Data as large volume of data with emphasis on big data research being carried out around the world, the author also emphasises on the key problems of big data which are Data Storage, Data Inquiries, Real-Time Data Processing, Large Data Analysis and Effective Management of data.

**Examples of Big Data**

Big data examples come in structured and unstructured formats, structured data format are those data that are stored in tuples and have relationship with other data while unstructured is a stream of data that are not stored in a tuple or table and may not have any form of relationship with one another.

**Key Elements of Big Data**

The three key elements of big data includes the 5 V’s synonymous with Big Data namely; Volume, Velocity, Variety, Viscosity and Virility.

**Benefits of Big Data**

Despite the above challenges posed by big data, it is still has tremendous amount of advantages as listed below:

a) Benefits in analysing almost all media types under one investigation that offers as much resource as needed

b) Taking advantages of data such as social network feeds and live feeds created through GPS and other devices to help proactive and reactive investigations
c) Access to large set of data that can be analysed based on the profile of a given suspect of information being searched for by the investigators.

Despite being a considerably new term in the field of technology, Big Data has played some key role in the modern day advances in the IT field, one such key role is Decision making support for organisation by enabling them to harness relevant data and use it to make better decisions.

Other recent benefits include the mining large set of data to look for key information that can be used in investigation and help refute or support a claim or put together a missing piece, this has seen rapid increase in the field of digital investigation.

Technical Aspects of Big Data
Metadata can be said to be set of information recorded automatically by the system in relation to registered values, memory, timers, network events that are considerably useful to the course of digital investigation. This information can be used to recreate all the events that happened on a system within various spaces of times. This information include title, Authors, Size, Date modified, tags, categories, content status, content types, date created, date last accessed, user information etc.

Big Data Platforms
Some of the platforms big data can be sourced from includes nevertheless not limited to Sensors, Computer Networks, Data Storage, Cluster Computing Systems

Challenges with Big Data

A. Storing and backups of Petabytes of information

The large set of data created as a result of big data concept of analysing large set of data always creates the issues of data storage, trillions of data sets are created on daily and to analyse this set of data, the investigation team has to had them stored in some form or capacity.

B. Faster indexing of huge amounts of data

Due to the size of data that needs to analysed, faster device and and methods are needed to be able to analyse data within a given time frame. Faster Indexing of data has always been a challenge for investigators when it comes to big data compared to small se to of data.

C. Methods for presenting large amounts of data to the court / visualisation methods

Presenting findings from research has always been a debatable aspect of digital forensics with emphasis on technical knowledge of the judge and other courtroom members or stakeholder. With the emergence of big data, wider varieties of data are often needed to be presented after investigation or during various stages of the investigation. The large size of data always makes it a difficult task for investigators to be able to present the whole data without cutting it down to smaller pieces and potentially omitting key information

Discussion

In comparison with (Chaudhary & Shah, 2013) definition of big data which focuses on the 5 V's Volume, Velocity, Variety, Viscosity and Virility, (Baoan, 2014) further explains some of the key benefits of big data which investigators are likely to come in contact with in real life investigation, some of does points are outlined below
Section Summary

The above chapter has analysed literature on the research study from various authors with emphasis on research areas such as the digital forensics where literatures are listed in chronological order followed by some key components digital forensics which are big data and e-discovery with emphasis on the research objective and then the state of digital investigation in the UK, US and Nigeria with comparison and key factors to look out for creating the bases for recommendation in chapter 5 of the study.

III. Research Methodology

Considering the uniqueness of the research objective which includes comparative studies of digital investigation between and developed country and a developing one, and the nature of the objectives to be addressed, large part of the research would adopt a Qualitative research methodology approach due to the social and physiological nature of the research objectives.

Inductive research

Inductive research generally works the other way round from observation to a theory; it is informally referred to as the “bottom to top” approach.

In inductive approach, research is expanded from a specific reasoning to a more generalised and theories.

Inductive Research is best suited to the nature of the research study being carried out due to substantial amount of observations required from the study group to answer some of the research objectives such as analysis of key elements of digital forensics and benefits of Big Data to digital forensics in general (M.K, 2006).

IV. Data Collection Method

Online Questionnaire Survey

The adopted data collection method of questionnaire used for the study was presented to participants in the form of web based and email questionnaires due to the remote locations of participants and cost of printing.

Interviews

Open-ended Interview question where used to conduct the data collection for the study due to the nature of the research target group and participant who will require evaluation and response from the participants his or her words. These sets of participants involve Digital forensics investigators, Examiners and experts.

VI. Data Analysis

Quantitative Data Analysis

Result from close ended questions presented to Digital Investigation firms with the below result presented based on question

Respondents Institution and Position

The research result demonstrates the variety of participant from various commercial Investigation firms from around the world with each of them giving their expert view on the state of digital investigation from there answers with the majority of the poll showing a \(\%70\) Digital Investigators and \(\%20\) being Examiners.
Awareness of Big Data
The result from the research shows the percentage of participant’s awareness to Big Data with the majority of the poll showing a %50 Strongly Agreeing and further %25 Agreeing to know about Big Data with the rest of the %25 left opting for Disagree.

Qualitative Data Analysis
Coding method involves the transcription of data collected from interviews through audio or video recording to short written scripts that can be used to make analysis as seen in Chapter 4 interview analysis. Data collected using qualitative method such as the interviews with the TargetGroupA are analysed in coding by segmenting the response into meaningful variables and assigning those variables into categories know as a “code”.

The below responses with be tallied to the study aims as outlined earlier and how the various responses would be used to answer the research questions raised by the study.

Codes
1. Key elements of digital forensics investigation
   After the interview with the digital forensics experts from Nigeria and UK, the researcher was able to gain an insight into the current state of digital forensic in a developed country and a developing country. The key elements of digital forensic derived from the interview are as below:
   a) Both experts from the UK and Nigeria agree that every aspect of digital forensics is as significant as the other
   b) Every country has a different definition of digital investigation
   c) Every country has different laws relating to digital investigation and crime

2. Benefits of Big Data in Digital Investigations
   There is more awareness of big data in UK and US than in West Africa. The concern of large set of data being created from different devices on daily bases and how this data can be analysed in short period of time which relates to the key objective of the study is not taken into practice in west Africa as oppose to UK and US.

3. Nature of Digital investigation in the UK and Nigeria
   The response from the interview with the experts that practised and studied in the UK and UAE Adnan Aftzal indicates that the state of digital investigation is in in higher level compared to the level of investigation Nigeria. Some of the key points to further prove these points are below:
   a) The use of ACPO guideline as a standard has further solidified the level of digital investigation in the UK
   b) Majority of the investigators from the quantitative research believe there is enough digital resources create din the UK to successfully win the first against terrorism

The interview with the expert from Nigeria further highlights area which can improved, some of them are as below:
   a) The standard of digital investigators in west Africa
   b) Acceptance of digital evidence in court as a standard form of evidence
   c) The government adoption and support for agencies such CFIN would improve the level of investigation in west Africa
   d) Introducing acts similar to the evidence ACT 2011 and guidelines similar to ACPO for the West African population

Available online @ www.ijntse.com
4. Challenges in carrying out investigation

Some of the challenges to digital investigation in both UK and Nigeria with reference to big data are analysed below with result mainly from the interview from experts and also general view commercial investigator around the world. These points would be further analysed as part of the recommendations.

a) Cost of investigation
b) Communication between examiners and other key stakeholders
c) Real-time data analysis

Section Summary

The above section has reported the result from the findings with both qualitative and quantitative analysis technique used to explain the key findings from respondent profile, to interview analysis, questionnaire analysis and key “codes” analysed form the interviews.

V. Discussions of Main Findings

Discussion and findings from the context of the literature review

Findings in agreement with literature review: In the context of the literature review with reference to (Chaudhary & Shah, 2013) a key aspects of digital forensic investigation with emphasis on e-discovery would drastically reduce the cost and time spent on investigations by clearly outlining all the steps needed to ensure every aspect of investigation is left untouched with emphasis on all key stakeholders which is not currently the case with conventional investigation method.

Big Data as a resource information for any digital investigation would allow investigators opportunity to analyse large set of data during an investigation with emphasis on the usefulness of the information to the case at hand.

Also, (Baoan, 2014) further states that some of the major challenges of using big data is data storage which perfectly agrees with some findings from the research carried out in the study.

Findings Not in agreement with literature review: Despite all the advantages posed by the use of big data and outlined in the literature review by (Chaudhary & Shah, 2013), some of the challenges not outlined by (Chaudhary & Shah, 2013) where clearly mentioned by (Baoan, 2014) which can be seen below;

a) Data Storage: Storing large sets of data collected to be analyzed can always be a challenge especially when the data sizes are in petabytes.

b) Slow Indexing: The use of indexing technique of files to allow easier searching and filtering of large amount of data is always slow due to the size of data created by when dealing with Big Data.

c) Presenting results and finding: Due to the large sets of data processed during investigation, result presentation tends to be a challenge for investigators especially when large sets of results are involved.

d) Privacy: Processing big data during investigation often results in sourcing data from different mediums such as GPS locations, Social network profiles and databases, the issue of privacy is paramount to the data being sourced.

Discussion and findings from the context of the research objectives

Findings in agreement with research objectives: In the context of research objective which includes the analysis of key digital forensics techniques, compares guidelines governing digital investigation in the UK, US and Nigeria, examining the use big data in digital forensics and suggesting cost effective ways of carrying out digital investigation; the below points would detail where the research objectives are carefully achieved;
Key Benefits of Big Data to Digital Forensics investigation: The research objective agrees that some of the key benefits of big data that was mentioned as part of the results and findings of the study includes but not limited to: Collecting information from digital communication devices and GPS tracking facilities such as mobile phones and handheld devices could enable investigators to gain access to wide variety of resources on a single target.

Cost effective ways of carrying out digital investigation: The research objective also agrees that use of key elements of digital investigation such as the e-Discovery reference model (EDRM, 2015) can reduce the amount of cost and time spent on an investigation by including sages such as the information governance stage to enable easy information recovery during investigation.

Findings Not in agreement with research objectives: In the context of the research objective the following points do not agree with the findings from data captured and analysed in relation to the set out objective of the study as below; According to the research objective, consideration of Big data in any digital investigation being carried out has huge potential to emerging challenges such as the large amount of data created on daily basis and the challenges to storing and processing such volumes of data in a given time frame.

VII. Recommendation and Benefits
The bellow recommendation would be categorised based on policies and tools, the view of the researcher are on how to improve digital investigation in West Africa based on the research and knowledge gained from the research study on digital investigation in the UK and US

Key knowledge gained from Digital Investigation in UK

1) Use of strict Guidelines to achieve maximum result
2) The adoption of e-discovery to ease cost of investigation
3) Loosening the law on privacy to allow investigators more access to digital resources
4) Adoption of Big Data to aid search for clues during investigation (e.g. the missing case of Madeline McCann in Portugal)

Key knowledge gained from Digital Investigation in US

1) License for digital investigators before they can practice
2) Outsourcing to the private sector
3) Recruiting, mentorship and promotion
4) Proactive investigation as oppose to reactive investigations.

Recommended Policies

A. Use of Big Data as a key source of resource
Big Data can benefits the current state of digital investigation in Nigeria by taking advantage of key features such as resource information for use with criminal and civil cases in Nigeria

Criminal Case: Search For boko haram Islamist
Big can prove very useful the current fight against extremist called (Boko haram) in West Africa with the wide spread od digital communication devices and GPS tracking facilities on social media files regularly posted by these extremist.
The use of such technique could lead to subduing one West Africa’s biggest extremist concern in recent times.
Civil Cases: Petroleum Subsidy Probe Corruption Case
Big data in the filed of digital investigation can also be used to further eradicate corruption in civil cases such as the on-going Nigerian Senate house probe of the past government on the Fuel Subsidy and missing funds belonging to the federal government.

The use of proper information governance in the ERDM method of e-discovery would ensure further investigations into such cases are smooth and would always result in high quality result as oppose to reliance on the conventional methods of investigation.

B. Adopting “Information Governance” as a compulsory procedure for any governmental setup
All government setups adopting compulsory information governance as outlined in the first step of ERDM (Electronic Discovery Reference Model) model would ensure future extraction of digital information from this organisation is easy and cost effective, with reference to the steps included in ERDM method of e-discovery.

C. Adoption of License for Practitioners
The use of similar licensing for forensic examiners would ensue quality investigations and presented in court rooms and increase the level of the society trust in the Digital forensics experts thereby making it extremely difficult for unqualified professional to practice as Digital Forensics professionals

D. Use of Open source investigation tools to save cost
Patronising and teaching student and other practitioners the use of open source investigation tools such as Hadoop with Katta, and HFSS along side widely popular but expensive commercial tools like FTK and Encase would further make the use of Big Data in digital investigation cheap and accessible.

Recommended Tools
The tools that would be analysed below, as part of the recommendation for easy searching of data when using Big Data in digital investigation. Recommended tools would are cost effective to implement and open source. Some of processes would include processing large set of Data in the processing stage of digital investigation with emphasis on compatibility with Hadoop.

Katta
Katta is a data storage tool that can be used to store large set of data for real time access; it enables easy indexing, scalability and failure tolerant data storage.
It can be implemented with Hadoop map file to reduce some of the challenges posed by the use of Big Data in digital investigation such as Data Storage problems, slow indexing problems, real time access to data.

HFSS: Forensics Indexed Search System
HFSS is a digital forensics index search system developed by the ETRI Electronic and Telecommunication Research Institute of Korea. The indexing benefits of HFSS include the web-based searching with the help of Hadoop.

Conclusion
In light of the research question asked by the research study with the aim of lessons that can be learned by the Nigerian law enforcement services from the use of Big Data in the field of digital forensics from the UK is presented above, various factors have been analysed by the research study with Introduction section presenting Digital Forensics research field followed by the benefits of the research, research question, research objectives and target the users.
The research also analyses the literature on the study subject in section 2 with emphasis on the key elements of the research study big data and a key element of digital investigation e-discovery. The section 3 of the research paper presents the study research methodology with emphasis on the different types of the research approach, data collection methods and data analysis methods. Section 4 of the research paper presents the result from the research methodology with reference to the key objectives set out in the beginnings of the research with both qualitative and quantitative data analysed and presented.

The last part of the research paper recommends the best possible solution form the lessons learned in the implementation of Big Data in the field of digital forensics with key emphasis on how to overcome the challenges posed by the use of big data by recommending policies and tools to tackle these challenges resp

REFERENCES


Redmayne, M. (2001). summarises research which suggests that as expert evidence becomes more complicated, jurors shift their focus and rely on peripheral indicia of reliability such as the expert’s qualifications or demeanour. 110.


