

Title Smart Cites Forensics-Evaluating the Challenges of MK Smart City Forensics

Name Ebenezer Okai

This is a digitised version of a dissertation submitted to the University of Bedfordshire.

It is available to view only.

This item is subject to copyright.

University of Bedfordshire

INSTITUTE FOR RESEARCH IN APPLICABLE COMPUTING (IRAC)

MASTERS BY RESEARCH
MODE OF STUDY: FULL TIME

Ebenezer Okai
ID: 0708426

Smart Cites Forensics-Evaluating the Challenges of MK Smart City Forensics

June 2019

University of Bedfordshire

Declaration

“I, Ebenezer Okai declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University;
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- Where I have drawn on or cited the published work of others, this is always clearly attributed;
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- Where the thesis or any part of it is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- Either none of this work has been published before submission, or parts of this work have been published.

Date: 18/06/2019 Signature

Abstract

The purpose of this research identifies what challenges are associated with MK Smart Data Hub forensics. MK Smart project holds its place as one of the Smart Cities' projects in the United Kingdom and central to this project is the MK Data Hub which holds a vast amount of data from various data sources. The first phase of the project involves looking at the MK Smart project ultimately emphasising on the projects aims, objectives, achievements and some of the trial projects that have been carried on. The second phase involves in-depth research into the MK Data Hub which is the integral of the project. There will be an evaluation of data received from the different data sources such as sensors on the Data Hub. This will also examine how data is stored, types of data stored, data structure and finally evaluate these data with current digital forensic tools and techniques to see the challenges associated with MK Smart forensics. The project objectives are to perform detailed research into the MK Smart project focusing on the aims, current achievement, detailed research into the MK Data Hub which is the central infrastructure of the MK Project, analysis of different types of datasets available on the Data Hub, evaluation of the existing digital forensics tools and techniques and its limitations to Smart city forensics, evaluation of the current challenges facing Smart city forensics and evaluation of the MK Data hub and detailed research into its forensic investigation challenges. Once the objectives are met, a result will be generated to determine the proposed solution to forensically collect evidence from the MK Data Hub.

University of Bedfordshire

Acknowledgements

I am grateful to the God Almighty for helping through this project. I am forever grateful for all HIs mercies that are new every morning.

Special thanks go to my supervisors; Dr Xiaohua Feng and Dr Paul Sant for the motivation, continuous encouragement and guidance during the duration of my course.

I acknowledge my family and friends for their prayers and support through the entire duration of my research. I am also grateful to the MK Smart Development Team at Open University and to the University of Bedfordshire for the opportunity to undertake this course.

University of Bedfordshire

Table of Contents

1 Introduction 1

1.1 Background 1

1.2 The Research Problem 1

1.3 Research Questions 2

1.4 Aims 2

1.5 Objectives 2

1.6 Research Methodology 3

1.6.1 Research Method 3

1.7 Motivation for Project 3

2 Literature Review 5

2.1 Emergence of Smart Cities Technologies 5

2.2 An Overview of Digital Forensics 6

2.2.1 Computer Forensics 6

2.2.2 Network Forensics 6

2.2.3 Mobile Devices Forensics 6

2.2.4 Social and Digital Image Forensics 7

2.2.5 Digital Video/Audio forensics 7

2.2.6 Memory Forensics 7

3 Big Data and Forensics 8

3.1 Big Data 8

3.1.1 The Four Vs 8

3.2 Big Data Forensics 9

3.2.1 Challenges of Big Data Forensics 9

3.3 Big Data Forensics Vrs Traditional Forensics 10

3.3.1 Some of the Commercial Digital Forensics tools 10

3.4 The Digital forensics Process Model 13

Identification 13

Collection 13

Analysis 13

Presentation 13

3.5 Smart Cities in the World 14

3.5.1 Smart Cities Features 14
3.5.2 Benefits of Smart Cities 14
3.6 "Some of the Smartest Cities in the World 14
3.6.1 "Seattle, US 15
3.6.2 Helsinki, Finland 15
3.6.3 Barcelona, Spain 15
3.6.4 The Island City State of Singapore 15
3.6.5 Songdo, South Korea 16
3.7 The UK's current Smart Cities 16
3.7.1 Bristol is Open, Bristol City 16
3.7.2 Future City, Glasgow , ... 16
3.7.3 Smart Campus – Smart Cities, Nottingham 17
3.7.4 Future Peterborough, Peterborough 17
3.7.5 Smart Cambridge, Cambridge 17
3.7.6 Digital Birmingham, Birmingham 17
3.7.7 Smarter London Together,London 17
3.7.8 City verve, Manchester 17
3.7.9 Smart Leeds, Leeds 18
3.8 Smart Cities Challenges to Overcome 18
3.8.1 "The Challenge of Infrastructure 18
3.8.2 "The Challenge of Resources 18
3.8.3 "The challenge of Skillsets 18
3.9 "The Future of Smart Cities 18
4 Overview of MK Smart Project 19
4.1 Case Study of Milton Keynes 19
4.1.1 History of Milton Keynes 19
4.1.2 Total Population Breakdown in 2017 19
4.1.3 Milton Keynes Employment Rate 19
4.2 Milton Keynes Smart City Project 20
4.2.1 Current Status of the MK Smart Project 21
4.3 MK Data Hub 21
4.3.1 Security and Forensics Challenges to The MK Smart Project 23

4.3.2 Solutions to Some of the Challenges of the MK Data Hub....24
5 Practical Works Undertaken....26
5.1 Practical Works Introduction....26
5.2 Security Demonstration on MK Data Hub....26
5.2.1 Process for the penetration testing on MK Data hub....26
5.3 Recommendations and Critical Evaluation....32
5.3.1 Hadoop Solution....32
5.3.2 The Hadoop Architecture....33
5.3.3 The Hadoop Architecture Layers....33
5.3.4 Running Hadoop....34
5.3.5 MK Data Hub Forensic Analysis with Hadoop....34
5.3.6 Identifying MK Data Hub Evidence with Hadoop....34
5.4 Recommended Approach....35
5.4.1 Hadoop Distributed File System (HDFS)....36
5.4.2 MapReduce....38
5.4.3 Hive....38
5.5 The Forensic Process....38
Identification....38
Preservation....38
Collection....38
Analysis....38
Presentation....39
5.6 Critical Evaluation and Conclusions....39
5.7 Limitations....40
5.8 Future Works....40
6 Appendices....42
Appendix A....42
Appendix B....43
Appendix C....44
Appendix D....45
7 List of Tables....49
8 List of Figures....50

9 Publication Bibliography 51

University of Bedfordshire

1 Introduction

1.1 Background

While there seems to be various explanations of what Smart City is, sustainability remains the main objective of what it seems to promote. This has encouraged many countries to implement Smart City technologies. These countries have benefitted from using smart technologies to address some of the country's issues such as growing population which impact on social amenities of the various cities. Smart City has been defined by IBM as " the one that makes optimal use of all the interconnected information available today to better understand and control its operations and optimize the use of limited resources". (Cosgrove, 2011)

Smart Cities forensics is a new way of solving the challenges presented by cities that are implementing smart ideas such as smart parking through the use of satellite data, smart transportation system, effective eco–energy system etc. These cities are embracing new technologies that are providing solutions to the growing population and high demand for cities infrastructures. This is a confirmation of about 6 million people estimated to be living in urban in cities by 2050 as predicted by UN. (United Nations, 2014)

One of the key elements of smart city is to primarily work with data effectively. These data are collective information from various sources that are relevant to the smart city project. The data are stored normally in the cloud or physical infrastructure such as the Data Hub.

"MK Smart is joint initiative which is led by the Open University and assisted by respectable key players including The University of Bedfordshire (Milton Keynes Campus), University of Cambridge, British Telecom (BT), Playground Energy, Fronesys, Milton Keynes Council, Anglian Water, Graymatter, HR Wallingford, E. ON, Community Action MK and Satellite Application Catapult" (Okai et al., 2018)."The project is partly funded by HEFCE (the Higher Education Funding Council for England) and led by The Open University, which is developing innovative solutions to support economic growth in Milton Keynes." (MK Smart Project, 2014)

"MK Data Hub is central to the project which supports the acquisition and management of the Big Data from various data sources which are relevant to the city systems. These include data about energy and water consumption, transport data, data acquired through satellite technology, social and economic datasets, and crowdsourced data from social media or specialised apps." (MK Smart Project, 2014)

Whilst the data plays a crucial part in this project, its forensic value of the data held is also important to the investigation of this project. Data might be required to help in any forensic investigation to prove in a case of Data integrity.

1.2 The Research Problem

The project will address the problem associated with the forensic aspect of the MK Smart project. Whilst the project primarily concentrates on the Big Data usage, it is very important to also focus on the forensic side of data. There is a huge challenge for Big Data to be forensically investigated for this MK Smart Data Hub.

MK Data Hub is the integral of the MK Smart project. The Data Hub contains vast amount of data relevant to the aims of the project." The MK Data Hub is a state-of-the-art' which supports the acquisition and management of vast amount of data relevant to city systems from a variety of data sources." (MK Smart Project, 2014)

Much can be said of the importance of the MK Data Hub; however, the data integrity will be questioned unless proven to be authentic and unaltered.

Another problem the research focuses on is how the data stored on the MK Data Hub will be forensically investigated. Traditional computer forensics focuses on more familiar data sources as such data from laptops, PC and mobile devices.

Over some years, traditional forensic tools and techniques have contributed tremendously in the field of computer security, however little information is known when it comes to smart city data forensics.

Lack of tailored tools and techniques for Smart Cities data makes it very challenging for data to be forensically investigated. Such a huge challenge makes it very difficult for computer forensics investigators to carry out their jobs in order to prove the integrity of data. This is another area this project will be addressing.

1.3 Research Questions

This research will aim to investigate and provide answers to the following questions:

How would data be forensically investigated on the Mk Data hub?

How would the integrity of Data have stored on the Datahub proven?

What tools and techniques would be used to forensically investigate the Mk Data Hub?

To provide answers and solutions to the above questions, I will use my involvement in the MK Smart project to research into different areas such as the type of data stored on the data hub, security protocols used, the infrastructures used and so on. These research areas will be analysed once all the gatherings of information have been completed.

1.4 Aims

The project aims to provide answers to some of the challenges of Smart Cities forensics, ultimately using MK Smart project as a case study for the project. MK Smart project has been a huge success not just to Milton Keynes but also to the fast followers' cities that are looking to replicate what the project has produced. Despite its successes, there are some problem areas that the project has created as a result.

One of these areas is how data integrity can be proven on the Data Hub which houses huge data from different data sources.

The Project will analyse into the data structure and best possible of obtaining forensic data from the MK Data Hub.

1.5 Objectives

A list of objectives needs to be met. An extensive literature review will be carried out. Ultimately the following objectives are to be met for this project;

- Detailed research into the MK Smart project focusing on the aims, and current achievement.
- Detailed research into the MK Data Hub which is the central infrastructure of the MK Project.
- Analysis of different types of datasets available on the Data Hub.
- Evaluation of the existing digital forensics tools and techniques and its limitations to Smart city forensics.
- Evaluation of the current challenges facing Smart city forensics.
- Evaluation of the MK Data Hub and detailed research into its forensic investigation challenges.

1.6 Research Methodology

Case study methodology has been chosen for this research project. This provides the chance for one problem to have in-depth studies. “According to Robert K. Yin, who is known for his work on case study research, ‘the more a study contains specific propositions, the more it will stay within reasonable limits” (Yin, 1994)

The case study also helps to identify key issues which for instance could help introduce new solutions. Case study research has the ability for evidence collected to be investigated methodically. In summary case study is useful especially to “investigate a contemporary phenomenon within its real-life context; when the boundaries between phenomenon and context are not clearly evident”. (Yin, 1994)

1.6.1 Research Method

Most case studies are based on the quantitative method and deal with qualitative data which analyse numbers and classes. However, for this project, there will be a mixed method of quantitative and qualitative to give a better in-depth study and understanding of the research project.

Although the above methodology and method have been mentioned for this project, however, the project is not limited to the above only.

1.7 Motivation for Project

Milton Keynes has become a home to about 248,800 people and “it is one of the fastest growing cities in the United Kingdom [Census, 2011]. A recent survey suggested the City of Milton Keynes economy has grown by 1.98% beating all other cities in the UK apart from Oxford (1.99%) and Cambridge (2.19%)” (MKfm, 2017).

There have been some exciting innovations and developments that have happened in the city for the last few years. One of these innovations is the MK Smart project which I was privileged to be part of. The project aims to transform the city into a smart city and also drive a significant increase in its economic growth and serve a pioneer for cities across the world. (HR Wallingford, 2015). To effectively apply smart city solutions in the city of Milton Keynes, the city has implemented the MK Data Hub, a central infrastructure to the project and holds all the city data from various data sources.

While the Data Hub plays a significant role in the city smart project, it can also be said to have some unanswered questions. This research project aims to evaluate the challenges of digital forensics in the MK Smart Data Hub project, aiming at the central infrastructure which is the Data Hub. The motivation behind this project is based on three unanswered questions which are;

- Why do we need forensics for MK Smart City Data Hub?

- What are some of the challenges that may associate with the MK Smart Data Hub forensics?
- Are there any appropriate tools and process available for MK Smart Data Hub forensics?

2 Literature Review

2.1 Emergence of Smart Cities Technologies

According to Information Age," the idea of Smart Cities was something that was initiated through one of the Clinton foundation's events. The former US President, Bill Clinton back in 2005 challenged one of the biggest network equipment maker, Cisco to consider making more sustainable cities through their technical capabilities. A $25 million project was set up by Cisco to carry out 5-year research on how the city can become more sustainable. The project worked with cities such as Amsterdam, Seoul and San Francisco to prove the true potentials of technology. Cisco launched Smart and Connected Communities division in 2010 which was primarily to commercialise and promote the products and services that it had developed during the 5 years research with the Clinton Foundation" (Hurst, 2018). Cities such as Seattle, Helsinki, Barcelona, the Island city-state of Singapore and Songdo have followed this idea and gone on to implement their own smart city technologies.

One exciting smart city that needs to be mentioned is the City of Milton Keynes in the United Kingdom. The MK Smart project is helping innovative solutions to support rapid economic growth in the city of Milton Keynes.

Smart Cities may be present and future solutions to cities inability to keep up with the rapid increase in urban population, pressure on infrastructures, housing, energy and transport, however, there is still a long way to go to overcome the challenges.

Lack of infrastructure is another challenge for smart city projects. Many Smart Cities have benefited from smart infrastructures in the area of public health, energy, transport and education Cities need to embrace state–of-arts- technologies if they are to implement good smart infrastructures.

Most fast followers of Smart Cities project like Bristol city are shadowing project such as the MK Smart project and are getting State–of –art technologies in place to the effect of their own project. (Bristol is open, n.d.)

Ankur Aggarwal, a Smart Cities researcher, identifies one of the Smart Cities challenges to be lack of people skillsets. He elaborates further by emphasising on how different new skill sets are needed to implement these technologies as current skills but are of little or no relevance to Smart Cities projects (Aggarwal, 2016).

In response to this, many Universities and Organisations such as the University of Bedfordshire and The Open University have introduced courses that provide relevant knowledge and skills to people interested in Smart Cities projects. (University of Bedfordshire, 2016).

Lack of resources is also one of the challenges facing Smart Cities. Resources have always been a hindrance to any project especially information technology projects. Anna Lisa Boni, Secretary General of Eurocities emphasised on how difficult it is to find funds for Smart Cities projects (Hoevenaars, 2015).

There is more cost involved in turning existing cities into Smart Cities than building new cities which have full smart city compliance. Also, new smart solutions require funds for research and implementation. Lack of funds delay the smart project or even result in project failure.

One solution to these challenges might be private investors collaborating with the government on these projects. This will not only ease the pressure on the government to provide funds but also provide additional funds for smart projects.

Another challenge that has not been recognised or little has been done is Smart Cities forensics. There is a huge challenge facing the area of digital forensics when it comes to Smart Cities. Traditional digital forensic tools and methods have worked very well over the years, however, there are not well suited for Smart Cities forensics.

The project will evaluate the challenges of Smart Cities forensics using the MK Smart project as a case study for this research purposes. The Project will also look at the possible ways of providing solutions to the challenges of digital forensics in the MK Smart project.

2.2 An Overview of Digital Forensics

The 21st Century has provided more people using interconnected digital devices ranging from mobile devices to non-mobile devices. These often become a centre of investigated incidents which requires experts to extract vital information to present in the court of Law.

The presence of digital forensics has helped and is still helping to solve cases in the court of Law using data extracted which can be used as a form of evidence in cases related to digital devices. It can also be used to prove the integrity of data by analysing the data to determine if its original data has been compromised.

"Cases such as Intellectual Property theft, industrial espionage, employment disputes, fraud investigations, forgeries, bankruptcy investigations, regulatory compliance, inappropriate email and internet use in the workplace have been resolved through the use of digital forensics" (Forensic Control, 2019).

Open learn talks about digital forensics as having such a huge scientific field with many disciplines which have constantly evolved over the years (Open University, n.d.). Some of these sub-disciplines are:

2.2.1 Computer Forensics

“Forensic Control defines computer forensics as “the practice of collecting, analysing and reporting on digital data in a way that is legally admissible. It can be used in the detection and prevention of crime and in any dispute where evidence is stored digitally”. (Forensic Control, 2019). The process of collecting such information has to be carried out by a certified forensic investigator who will follow the right guidance as to the ACPO guidance.

2.2.2 Network Forensics

Network forensics, another sub-branch of digital forensics which focuses on the gathering of computer network traffic through monitoring and analysis for legal purposes. Technopedia also defines network forensics as "investigations that obtain and analyse information about a network or network events.” (Technopedia, n.d.).

2.2.3 Mobile Devices Forensics

Intaforensics defines mobile device forensics as “the digital forensic approach concerned with the systematic examination of mobile phones, tablets and satellite navigation devices and all attached media – such as SIM cards and memory storage cards.” (Intaforensics, n.d.). With the ever growing of smart devices in our current world, it is evident, more criminal activity can be found of these devices.

2.2.4 Social and Digital Image Forensics

Social media has impacted in both positive and negative ways. Social media has proven to be a key source of evidence in forensic investigations. Obtaining social media evidence has been beneficial to an investigation related to cooperate interest in employees checks, associations etc.

2.2.5 Digital Video/Audio forensics

This type of digital forensics investigation is the gathering of evidence obtained from audio, CCTV networks, cameras, VHS tapes etc. It is evident that the evidence obtained from these sources can be critical to the cases of criminal and civil cases(Cyfor, 2002).

2.2.6 Memory Forensics

This type of forensics analyses the memory dump of a computer primarily looking at the volatile data. By conducting an investigation on the memory dump can help to identify malicious behaviour that cannot be easily traced on a hard drive of a computer. (Nate and Lord, 2018).

3 Big Data and Forensics

3.1 Big Data

Big Data is described by IBM as “a term applied to data sets whose size or type is beyond the ability of traditional relational databases to capture, manage, and process the data with low-latency. It also has one or more of the following characteristics – high volume, high velocity, or high variety. Big Data comes from sensors, devices, video/audio, networks, log files, transactional applications, web, and social media - much of it generated in real time and in a very large scale.” (IBM, n.d.).

3.1.1 The Four Vs

Big Data has often been categorised by the 4Vs. These are; volume, variety, velocity and veracity. These are the main characteristics that shape the pool of information. (Infosec, n.d.) These four main properties create the scope of Big Data.

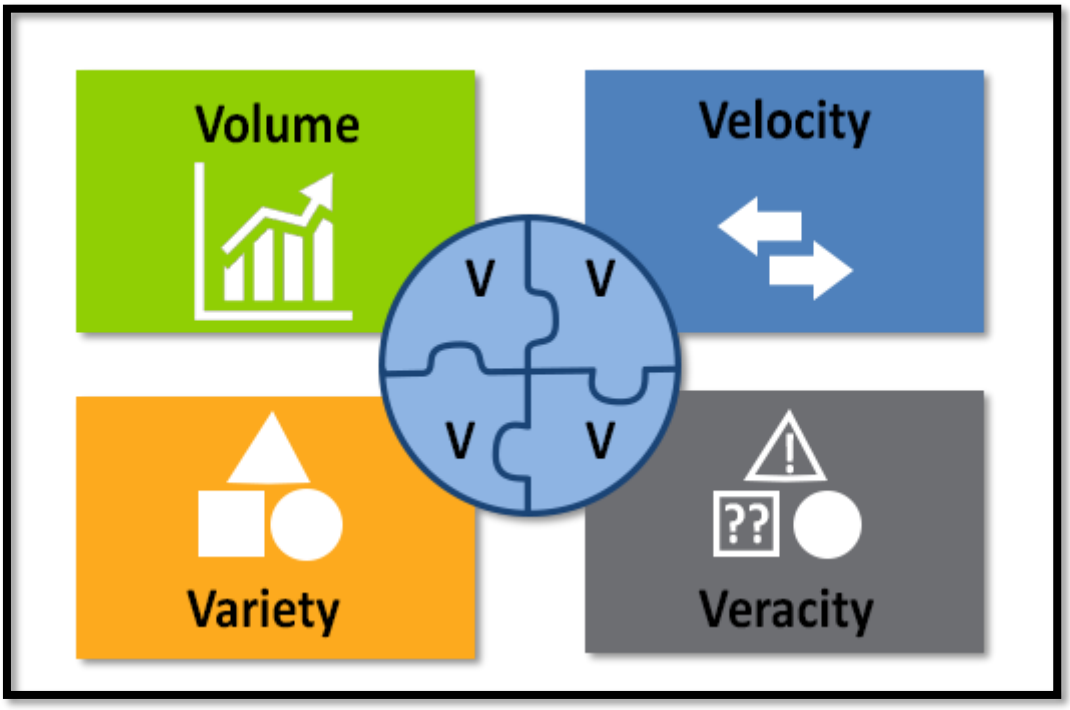

Figure 1.The Four Vs of Big Data

3.1.1.1 Volume

The collections of Big Data are so big and complex. The applications tasked with the handling overwhelm and struggle to contain the data, Storing, analysing, sharing, transferring and securing Big Data is very difficult to achieve. A typical example is the amount of data Facebook, one of the popular and most used social media platform generates within a second. “ It is also estimated that 2.5 quintillion bytes of data are created each day, and as a result, there will be 40 zettabytes of data created by 2020 – which highlights an increase of 300 times from 2005". (BigdataIdn, 2017)

What the volume of the Big Data primarily comprises the unstructured and multi-structured data making it easy to store, analyse, share and most significantly secure. The MK Data

hub hosts data from different sources which comes in either as structured or unstructured data. Volume on the Data hub is complex t making it very difficult to manage and secure.

3.1.1.2 Variety

The Variety refers to all types of data formats and data sources. Most of the data the world produce now are unstructured. Through Big Data techniques, unstructured data can be sorted structurally and examined. The variety of the MK Data Hub includes the different data sources and their different data formats. The MK Smart Data Hub has generic data processing framework that allows the creation of ad-hoc processing pipeline for any data source.

3.1.1.3 Velocity

Speed is one of the most important aspects when it comes to data processing. Velocity refers the speed used for data to be created, analysed and reused. Adapters such as the RDF, Cap and EEML provide the solution to convert the structured and unstructured data into the data hub. The MK Smart Data Hub relies on Apache ActiveMQ2 as the standard communication mechanisms.

3.1.1.4 Veracity

Veracity defines "the authenticity and credibility of the data." (The Unbelievable Machine Company, n.d.). It is essential for data coming in to be of good quality. Worthless data can cause many problems and give wrong decision making for organisations. Therefore, data will need to be authentic and credible. As part of the security requirements for the MK Smart Project, all data owners are required to register with the administrator of the Data hub. This helps to verify the owner of the dataset and also prove the authenticity and the credibility of the data.

3.2 Big Data Forensics

As previously said about the constant evolution of digital forensics, one of the areas that have emerged is Big Data forensics.” A smart city relies heavily on Big Data which plays a significant role in the process of IoT data collection.” (Aher, 2018). We cannot look at smart city forensics without looking at Big Data forensics. One of the extreme challenges facing digital forensics is Big Data.

3.2.1 Challenges of Big Data Forensics

3.2.1.1 Storing and Backup of Big Data

Storing of petabyte could be very hard with the current traditional software processors and tends to cause havoc as these systems are not built to process millions of little files or data at once. It can be mention that not all Big Data needs to be processed through the traditional method.

According to chief technology officer of EMC, “before dealing with Big Data, and protection of Big Data, knowing the type of data is crucial, for example, database data could be easily reproduced than recovered or backed up.” (Techtarget, 2013).

3.2.1.2 Forensic Applications and Implications of Big Data

Network forensic as a sub-section of Big Data is usually known as the capturing of full data packets of a security incidence in order to get the user closer to identifying the security issue and persecute the co-operate involved. Given that the functionality of the forensic

tool which is said to overlap each other in the aspect of full data capture, retention of data, acquiring data quickly and analysing packet data content. This is because Big Data consist of 4vs, it gives plenty of insight into the type of security issue(InfoSec, 2018).

3.2.1.3 Faster Indexing of Big Data

Indexing of Big Data could be very challenging as it comes in high velocity, at a very high volume in different forms, and would be very hard to scan without solid indexing.

For example

- An online game that required multiple players to interact at the same time with a scoreboard.
- A travelling app when looking at real-time travel time for departure or arrival

These are the scenarios in which requires a distributed system which gives two types of indexing which are local indexing and global indexing. (Biyikoglu, 2016)

3.2.1.4 Security in Big Data Clusters

A secure Big Data means logical and consistent support of authentication and authorisation in Big Data which is mostly done through an access control list which deals with specific access linked with users' identities. Authorisation is done after.

3.3 Big Data Forensics Vrs Traditional Forensics

Big Data is providing a way of solving the challenges of large and complex data. It has been adopted by larger organisations such as Google, Facebook, Instagram etc. As technology keeps evolving and the size of data becomes large and complex, it proves more challenging for digital forensics. The current forensics has its limitations when it comes to Big Data and this is why a new area in forensics called Big Data forensics has evolved.

Unlike the traditional forensic, Big Data forensics is a new area of digital forensics which focuses on the forensic investigation on Big Data systems. There is not clear definition of what Big Data forensics is as this is a new branch of forensics. However Big Data forensics generally refers to collection and analysis of Big Data systems (Sremack, 2015).With the complexity of data, this new area is still finding the best tools and techniques that can effectively collect and analyse Big Data. Cloud forensics provide some similarities approach for Big Data forensics as it also has features of Big Data.

Traditional Digital forensics as defined by Forensic Control is "the practice of collecting, analysing and reporting on digital data in a way that is legally admissible and also compliant to the GDPR polices. It can be used in the detection and prevention of crime and in any dispute where evidence is stored digitally" (Forensic Control, 2019). It focuses on collecting data from sources such as digital devices and machines. Traditional digital forensics has its own process, methods, tools and techniques used in forensic investigation. However, these have limitations in regard to Big Data forensics. The tools and techniques used are specifically designed for unstructured and non – complex data such as emails, documents files found on any digital device.

3.3.1 Some of the Commercial Digital Forensics tools

There are several tools used in digital forensics investigation. These tools are classified into two main categories which are commercial and non-commercial tools or the open-

source tools. Commercial tools are recognised and certified to be used in industrial and legal forensics cases. The non-commercial tools give similar but limited exposure to digital forensics investigation (AccessData, n.d.).

3.3.1.1 X Ways Forensics:

This is one of the best commercial tools available in digital forensics. X ways is a Windows-based software and it has many functionalities which play key roles in computer forensics. The software can be used in a portable mode making it very significant in live forensics. This also eradicates the requirement to install the actual software before using (X-Ways Forensics, n.d.).

“Key Features of X Ways Forensics

- Disk cloning and imaging
- Read partitioning and file system structures inside raw (.dd) image files, ISO, VHD and VMDK images
- Complete access to disks, RAIDs, and images more than 2 TB in size (more than 232 sectors) with sector sizes up to 8 KB
- Built-in interpretation of JBOD, RAID 0, RAID 5, RAID 5EE, and RAID 6 systems, Linux software RAIDs, Windows dynamic disks, and LVM2
- Automatic identification of lost/deleted partitions
- Native support for FAT12, FAT16, FAT32, exFAT, TFAT, NTFS, Ext2, Ext3, Ext4, Next3, CDFS/ISO9660/Joliet, UDF
- Superimposition of sectors, e.g. with corrected partition tables or file system data structures to parsing file systems completely despite data corruption, without altering the original disk or image
- Access to logical memory of running processes
- Various data recovery techniques, lightning fast and powerful file carving
- Well maintained file header signature database based on GREP notation
- Data interpreter, knowing 20 variable types
- Viewing and editing binary data structures using templates
- Hard disk cleansing to produce forensically sterile media
- Gathering slack space, free space, inter-partition space, and generic text from drives and images
- File and directory catalogue creation for all computer media
- Easy detection of and access to NTFS alternate data streams (ADS)
- Mass hash calculation for files (Adler32, CRC32, MD4, ed2k, MD5, SHA-1, SHA-256, RipeMD-128, RipeMD-160, Tiger-128, Tiger-16, Tiger-192, TigerTree,)
- Lightning fast powerful physical and logical search capabilities for many search terms at the same time

- Recursive view of all existing and deleted files in all subdirectories
- Automatic colouring for the structure of FILE records in NTFS
- Bookmarks/annotations
- Runs under Windows FE, the forensically sound bootable Windows environment, e.g. for triage/preview, with limitations
- Support for high DPI settings in Windows
- Ability to analyse remote computers in conjunction with F-Response." (X-Ways Forensics, n.d.).

3.3.1.2 EnCase:

Recognised globally for digital forensics, Encase provides solution of in depth and court proven investigation. This tool is powerful in processing workflows of investigations and it is not surprising that its reputation has been seen as a gold standard in digital forensics (Guidance Software, 2018).

"The lifecycle of encase ensures the importance of maintaining the integrity of evidence throughout the whole process. Encase is developed by Guidance Software now known as Open Text. It is a Windows-based computer forensic tool and can be used for Disk Imaging, reading the various file systems (NTFS, FAT, exFAT and other MAC related file systems), rebuilding the lost partitions, recovering deleted files, etc." (Guidance Software, 2018).

"Key Features of Encase

- Enhanced indexing engine

Permits investigators to perform forensic investigations process with authoritative speed , innovative index searching, language support and performance

- Easy reporting

Offers different templates that can be customised to suit examiners in creating a formal report.

- Extensibility

Provides EnScripts, which helps the examiners to be more efficient in completing investigations

- Workflow automation

Automated workflows and easy to navigate

- Updated encryption support
- Apple File System (APFS) support
- Volume shadow copy capabilities" (Guidance Software, 2018).

3.3.1.3 Forensic Toolkit (FTK):

"Built to interoperate with mobile and e-discovery solutions, FTK helps you find relevant evidence faster, dramatically increase analysis speed and reduce the backlog. It's the only

solution that utilizes a single case database, creating a clear picture of the event. Developed by AccessData, Forensic Toolkit (FTK) is another forensic tool kit that helps to find relevant helps you find relevant evidence faster, dramatically increase analysis speed and reduce backlog (FTK). One of the unique solutions is its ability to create a visible picture of events by utilising a single case database. FTK has a free version (FTK imager) which can be used for disk imaging and analysis (AccessData, n.d.).

Key Features of Forensic Tool kit

- Unmatched speed through distributed processing engines
- Unique architecture provides better stability
- Wizard-driven to ensure no data is missed
- State-of-the-art data visualization to highlight relationships and patterns
- Only solution that utilizes a single case database, reducing cost and complexity of multiple case datasets
- Faster learning with easy-to-use GUI" (AccessData, n.d.).

3.4 The Digital forensics Process Model

The digital forensic process can be classified in five stages.

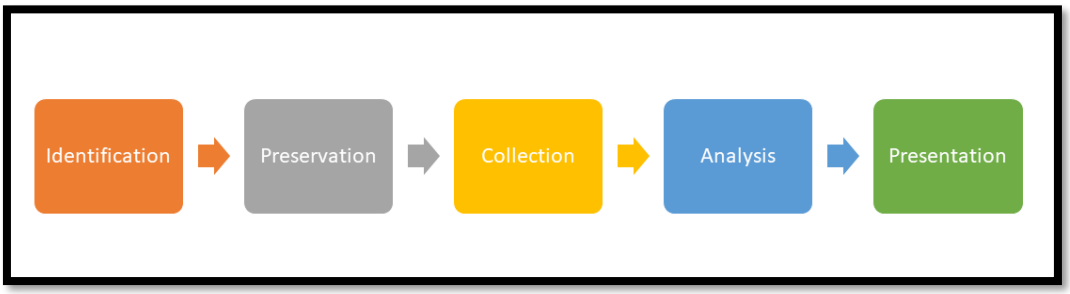

Figure 2.The Digital Forensics Process Model.

Identification – This stage identifies any vital relevance to the investigation including all digital devices and data's location.

Preservation – Vital stage of preserving evidence obtained from incident scene. It also is the stage of documenting all necessary information about the evidence obtained.

Collection – collection of digital devices that might contain potential evidence. Specialised tools are then applied to image and copy the content of the collected devices.

Analysis – In-depth analysis of evidence collected. The examination plays a key aspect of the whole process and its analysis produces the conclusion on the evidence obtained.

Presentation – Report of the whole process and final findings of the evidence obtained authentication is confirmed.

3.5 Smart Cities in the World

There are many explanations of what Smart Cities really address; however, sustainability remains the main objective of what Smart City seems to promote. Over the past years, many countries that have implemented smart technologies. These countries have used smart technologies to address issues such as growing the growing population which impact on social amenities of the city. The benefits of smart technologies can be highlighted in decision making of cities development and planning such as looking into better waste management programme, efficient renewable source of energy and so on.

3.5.1 Smart Cities Features

There are several features that distinguish Smart Cities from normal cites. These features contribute to the provision of solutions that address the problems of modern day cities. Typical features of Smart Cities include the use of technology and data information which primarily improve cities' services and infrastructure such as electricity, water, affordable homes, education, IT Connectivity and education (BW Online Bureau, 2019).

3.5.2 Benefits of Smart Cities

Smart city provides solutions to the challenges which growing cities are facing presently and in the future. As predicted by the United Nation, there will be around 6 million living in the cities by 2050. (United Nations, 2014).This puts a huge pressure on cities to meet current needs of their citizens. Among some of the benefits of Smart Cities issues are discussed below;

a) Better city planning and development

b) E-government services delivered to citizens, faster, and at a lower operating expense

c) Local economic development

d) Improved productivity and service

e) Waste management

- Waste to energy & fuel
- Wastewater to be treated
- Recycling & reduction of waste

f) Energy

- Smart meters
- Green buildings
- Renewable sources of energy" (Okai et al., 2018).

3.6 "Some of the Smartest Cities in the World

Many cities have started the implementation of smart ideas to solve some of the key issues of their cities. Below are some of the smartest cities in the world:" (Okai et al., 2018).

3.6.1 “Seattle, US

“Seattle city with its partner, University of Washington are addressing several challenges facing their urban cities. This ranges from the reduction of carbon emission, earthquake preparedness and transportation. Below are some of the features of their smart technologies.

- City of Seattle has been using analytics to reduce carbon emission reduction. This is a key to the city as about 100,000 people are present during the workdays.
- They have also implemented an adaptive transportation management system which enables traffic lights to adapt to the changes in weather and road conditions.
- RainWatch which monitors precipitation in real time and also enables alert of flood which help keep citizens and infrastructure safe.
- Another benefit is the introduction of 800 cameras for law enforcements officers
- ShotSpotter’s gunshot technology aids to deter-mine where gunshots were fired.
- New Data which can withstand natural disasters.” (Okai et al., 2018).

3.6.2 Helsinki, Finland

“The capital and also one of the busiest cities in Finland, Helsinki is piloting some smart projects which help address some of the challenges of the city. Some of its smart projects include; Car charging facilities within parking places

- An automated waste collection system which aims to reduce garbage trucks traffic by up to 90 percent.
- Smart grids and real-time energy. This will help to a 15 percent reduction in energy usage, and apps that plan the most efficient traffic routes with any type of transportation method. “ (Okai et al., 2018).

3.6.3 Barcelona, Spain

“Second biggest city in Spain, Barcelona has many smart projects happening in the city at the moment. Some of its projects include;

- Smart lighting, which aims to encourage the charging infrastructure for electric vehicles, and Wi-Fi in public transportation systems and public places.
- Smart LED streetlights where lights only activate when movement is detected.
- The sensors in the streetlights also collect environmental data collected through smart sensors. Smart garbage bins are only emptied when full and smart parking systems show vehicles to free parking spots, reducing carbon emissions and traffic congestion.
- Real time of bus times through Digital bus stops project. Buses will also have Free Wi- Fi and USB charging ports.” (Okai et al., 2018).

3.6.4 The Island City State of Singapore

“Singapore is one of the fast-growing countries in the world. Its commitment to smart technologies is helping the country grow rapidly. They are aiming to become the world’s first smart Nation. Some of the island city state’s smart technologies are as follows;

- Smart Nation Platform drives forward with trials across several sectors. This focuses on areas such as housing, transport and health.
- Smart Nation Platform collects data from sensors all over the country to analyse pedestrian and traffic flows. This also helps to conduct test for crowd evacuation simulations.
- Sensor technology and smart applications in public housing estates which sends feedback to help reduce energy and water consumptions and costs.
- The government also analyses the data produced by these sensors to help in the development of its design, planning and maintenance of these and future public housing estates." (Okai et al., 2018).

3.6.5 Songdo, South Korea

"Another country who is one of the pioneers in Smart city technologies. Among some of Songdo smart technologies include;

- Smart work centres with teleconferencing systems which helps a section of government works to work close to their homes. This technology comes fully equipped with sensor networks and Wi Fi
- Underground Waste suction systems where household dispose will run thorough underground pipes to the processing plane where it will be recycled.
- Smart devices and remote-controlled medical equipment will give check-up and medical consultation to the old and the disabled.
- OLEV (online electric vehicle technology) transportation systems which will automatically charge while the public buses move across the road (Okai et al., 2018)."

3.7 The UK's current Smart Cities

3.7.1 Bristol is Open, Bristol City

This project is a collaboration "between the university of Bristol and the Bristol City Council." (Bristol is open, n.d.)."It is funded by the national and European governments and supported by academic research and some private sectors. As recognised by UK Smart Cities index in 2017 as the UK's leading smarty city , Bristol has set path in areas such open data access, energy innovation and community engagement." (Bristol is open, n.d.)."Bristol is open is not just about smart city technology but also the management framework which supports the smart cites technologies to be successful. The project is also aiming for carbon neutrality by 2050." (Bristol is open, n.d.)."Some of its key partners include, NEC, InterDigital, Nokia etc. It also has the support from Department of Culture Media and Sports, Department for Business Innovation and Skills , Tech UK, Catapult future cities and Catapult Digital." (Bristol is open, n.d.).

3.7.2 Future City, Glasgow ,

Glasgow enjoyed its share of Smart Cities innovation after the investment of £24 million from Innovate UK in 2013. The city has enhanced its infrastructure in a large scale ever since then. Some of these projects include reduced energy costs through the implementation of intelligent streetlights apps. This app provides open plan routes and multiple systems to residents and visitors for their journeys. Glasgow has further their smart projects by implementing more projects such as street lighting pilot schemes , active travel demonstrator and social transport demonstrator specifically designed for vulnerable citizens (Computerworld UK, 2019) .

3.7.3 Smart Campus – Smart Cities, Nottingham

Another city to consider Smart Cities technologies is the City of Nottingham. It was ranked as one of the top 20 leading Smart Cities in the UK in 2017.Nottingham has deployed new smart technology solutions to tackle issues such as traffic congestion, environmental sustainability. As part of new technology solutions , the council will trial new fixed camera and sensors to reduce congestion also provide alternative methods of transport (Computerworld UK, 2019).

3.7.4 Future Peterborough, Peterborough

This project started in 2012 and it is led by Peterborough City Council and Opportunity Peterborough. The future Peterborough project include Digital Peterborough which looks at faster internet which will be 100x faster than the national average internet speed for businesses. Other projects include Dragon Patcher, pothole repair solution which uses flame to repair potholes and cracks found on the roads (Computerworld UK, 2019).

3.7.5 Smart Cambridge, Cambridge

The project is focused on the city of Cambridge's public transport. This project is managed by the greater Cambridge City Deal (Hurst, 2018).Similar to the MK Smart project, there has been the implementation of real-time bus information through the use of sensors which provide real time data. The other project looks at the traffic lights data such as number of cars crossing at a given time, rubbish bin sensors that alert once it is full etc.

3.7.6 Digital Birmingham, Birmingham

Birmingham's brilliant city system is driven by Digital Birmingham. The £10M programme directs a guide created by the Birmingham Smart City Commission, a body set up by the city council and involved delegates from businesses, the scholarly community and the open area. Open Wi-Fi is accessible through the city and a new smart technologies are created in the East Birmingham Smart City Demonstrator (Digital Birmingham, n.d.).

3.7.7 Smarter London Together, London

"This project was launched by the Mayor of London and it can be classified into five missions: more user-designed spaces, a new deal for city data, world-class connectivity, digital leadership and skills enhancement. Currently there are about 700 sets of Big Data in the city's datastore that helps to address the city's urban challenges and also solve the public services including cashless payment methods of the city's transport." (Greater London Authority, n.d.).

3.7.8 City verve, Manchester

"The purpose of this project is to explore how technology can assist in improving the people of Manchester .The project focuses on gathering and sharing information in different exciting ways to support everything form transport and healthcare to the environment and culture. The purpose of this project is based around six key themes: Live, Work, Play, Move, Learn and Organise." (CityVerve, n.d.). "The City Council together with the University of Manchester, Cisco, BT, Siemens and many interested companies have contributed to the project and have laid the foundation for the future and development for the city." (CityVerve, n.d.).

3.7.9 Smart Leeds, Leeds

The project is led by Leeds City Council and it focuses on three areas which consist of travel and transport, health and wellbeing and housing. “In Health and Wellbeing takes advantage of the city’s teaching hospital, has also taken advantage of its Assisted Living Leeds Innovation Lab (ALL INN) programme which is among the largest in Europe.” (Leeds Council, n.d.).“The Housing looks at the implementation of cost-effective and the Internet access in all homes while Travel and Transport helps to install a Clean Air Zone for vehicles in January 2020.” (Leeds Council, n.d.).

3.8 Smart Cities Challenges to Overcome

3.8.1 “The Challenge of Infrastructure

Having the right infrastructure for smart city project is one of the key challenges. A smart city preferably will provide wider infrastructure for public transport, public health, education, energy and other public utilities. To effectively implement a good approach to Smart Cities infrastructure development, Cities need to be ready for state-of-the-art technologies.” (Okai et al., 2018).

3.8.2 “The Challenge of Resources

Resource is one of the challenges facing Smart Cities. To implement new smart solutions cost significant amount of money for most cities. This results in insufficient funds to fund Smart Cities transformation projects. To effectively solve this problem, the government might need to get in additional funds from private sector who will be interested in these Smart Cities projects.” (Okai et al., 2018).

3.8.3 “The challenge of Skillsets

Having smart projects in Smart Cities need different new skillsets to be able to implement these technologies. There is a challenge to get more people with relevant skillset to undertake these projects now and in the future (Aggarwal, 2016). Essex University has provided a provision of data for researchers. They edit data documentation centrally. They offer researchers to contribute data documentation, which could be exploited on smart city data hubs analysis in the future.” (Okai et al., 2018).

3.9 “The Future of Smart Cities

Smart city provides solution to our rapid urbanization. There is a general and growing concern of the challenges cities will encounter based on the current growth rate. According to the United Nations (UN), it is expected people living in cities to be around 6 billion by 2050. It is also believed cities are consuming about 70% of the global energy usage. This has resulted in a huge strain on resources and a large degree of challenges for cities (Reeves, 2018)” (Okai et al., 2018).

4 Overview of MK Smart Project

4.1 Case Study of Milton Keynes

4.1.1 History of Milton Keynes

Milton Keynes may be a standout amongst the new towns formally designated as another town in 23 January 1967. However, the area has a rich history of human settlement back to the 2nd millennium BC (MKinspire, 2012). There was evidence of human settlement before the constructions works in the early 1960s. In 1960s the UK Government decided to build Milton Keynes which was mostly farmland and undeveloped villages. The 21st century has seen continuous development of Milton Keynes has further increased in the number of people migrating to the city as of 2019. The growth is expected to rapidly increase based on the projected population growth. Milton Keynes has become one of the UK's fastest growing cities in 2013, it was estimated to have the people living in the city to be around 255,700 (BBC, 2017). This shows an increase of about 43000 since 2001 and a growth of 20.2%. It is estimated to have population of 272612 in 2019 (Population City, 2015). Milton Keynes has roughly 25 million people that live within 2 hours or less making it a prime location for businesses.[RGS] The city of Milton Keynes celebrated its 50years in 2017 to depicts the dramatic change the city has made since then housing Minister ,Anthony Greenwood granted the permission to transform the farmland to new town (BBC, 2017). Milton Keynes has become home to some of the famous household industries in the UK and the world. Network Rail, Argos, Amazon, Volkswagen Financial Service, DHL Supply Chain, E.ON, Nissan, Unify, Deloitte, PWC, Mercedes-Benz International , Unisy, Ikea are some of the reputable companies currently in Milton Keynes (Glassdoor, n.d.).

4.1.2 Total Population Breakdown in 2017

The total population of people living in the City of Milton Keynes in the 2017 was estimated to be 545000 with the average age of 38.7 and 39.6 in population density (Plumplot, n.d.).

Milton Keynes Population History (in thousands)

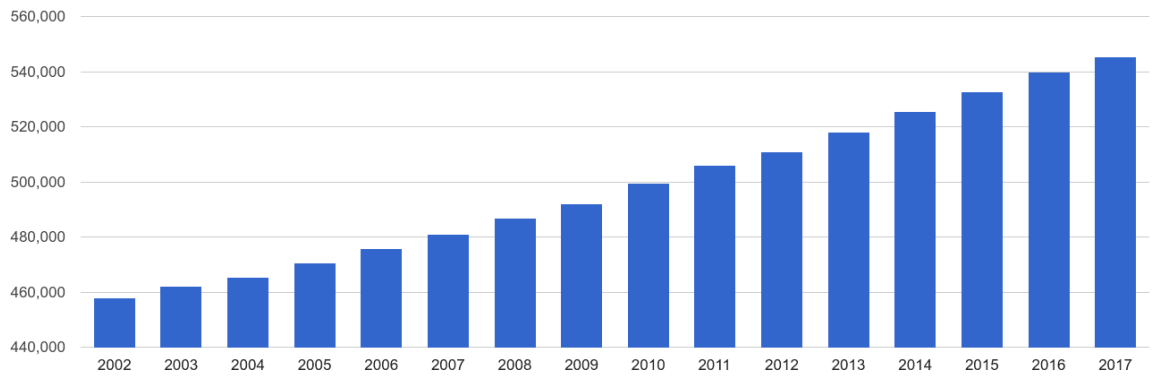

Figure 3.The Milton Keynes Population Growth in 2017 as Demonstrated by Plumplot.

4.1.3 Milton Keynes Employment Rate

A survey conducted by Nomis in 2019 shows the number of people in employment between the period Jan 2018 - Dec 2018.

All People	Number	Percentage
Economically Active	135,800	77.7
In Employment	129,100	74.0
Employees	115,000	66.3
Self Employed	13,700	7.6

Table 1.The Nomis' Breakdown of Employment Rate in Milton Keynes for the Period between Jan 2018 - Dec 2018.

With great success stories when it comes to its economy, it also has its challenges of sustaining its current growth which looks likely to exceed the targets on various areas such as carbon reduction and infrastructure capabilities. To address the ever-growing population of the Milton Keynes city, current and future challenges, the city council has adapted the smart city approach by implementing the MK Smart project (MKSmart, 2014).

4.2 Milton Keynes Smart City Project

As a joint initiative, the MK smart project is funded by HEFCE (Higher Education Funding Council for England) and led by the Open University and has support of key players such as the university of Bedfordshire, British Telecom(BT), Milton Keynes Council, University of Cambridge, community Action MK and so on. (MK Smart Project, 2014) "MK:Smart project is a £16m Higher Education Funding Council for England (HEFCE) project that emphases on the how the city of Milton Keynes can over barrier to growth through the innovation of technology that will help minimise the potential obstacles." (MKSmart, 2014).

The project was aimed to achieve solutions that will assist Milton Keynes in its growth through the ever growing of internet of Things (IoT) and become of the leaders in UK's Smart Cities.

With the help of IoT and sensors devices there is a chance to tackle the challenges of Milton Keynes city is facing. For example, sensor devices can send real-time data of buses in Milton Keynes which will help citizens to plan their journeys,

As previously discussed about the ever growing of the city's population, It is estimated to have a growth of 20% by 2026 .This raises a huge concern for the council and the city planners as there will be massive pressure on services such as water, energy, transport, education etc.

On tackling these issues, it is important to look at the three main area: energy, water and transport. As the population increases it has been predicted that there will be an increase of traffic on the city's roads by up to 60%. To tackle this, MK:Smart project will utilise the IoT and different innovations of technology to provide potential solutions.

Another area the project looks to provide solution is energy and water. The demand for these will drastically increase as a result of the current pattern of growth in population of Milton Keynes. Through smart technology which include smart sensors, water and energy management systems could benefit.

One of the significant aspects of this project is the opportunity to create partnership with businesses. To make this possible, the University of Bedfordshire, one of the key players of this project has dedicated team that are in contact with small-to-medium enterprises (SMEs) in Milton Keynes with the primary aim of helping SMEs to identify innovations that will go directly into this project and also provides solutions to current and future problems of the city.

Furthermore, the MK:Smart project is not just aiming to support the development and growth of businesses within the city, it is also aiming to improve the lives of Milton Keynes' citizens. Citizens engagements have been introduced to gather ideas from the citizens relating to Smart Cities and how their ideas could be beneficial in the city's quest to be a smart city. Works with SME such as Graymatter and Community Action MK have provided several workshops in the course of getting ideas from citizens.

Another significant area of this project is the business engagement aspect. University Campus Milton Keynes (UCMK) has partnered the University of Bedfordshire Business School to provide business engagement platform for businesses that are willing to bring technical and commercial innovation to the MK Smart project. The team is made up of Research teams from the University's Business and Management Research institute (BMRI) and Institute in Research in Applicable Computing (IRAC) and are working with SME to tackle the challenges of Smart Cities through innovations.

MK Smart is not the only important project for the City of Milton Keynes in regard to smart innovations. Milton Keynes has provided prime location for one of the leading innovation companies in the area of transportation. Transport Systems Catapult, the UKs innovation centre for intelligent mobility has introduced driverless pods (LUTZ pathfinder project) that will be carrying people from station to the main shopping centre in Milton Keynes. This project has been funded by Innovate UK, a UK innovation agency. There is also another exciting projects in Milton Keynes. Smart bins ,Low Powered Wide Area Network and smart parking have all been part of some the smart innovations in Milton Keynes (MKSmart, 2014).

4.2.1 Current Status of the MK Smart Project

The first phase of the MK smart project has been concluded in June 2007, however there is also exciting follow up initiative called the CityLabs (Knowledge Media Institute, 2019). This has been set up purposely for SMEs to build a working partnership with industry and academics leaders to develop prototypes for new products and services (Knowledge Media Institute, 2019).

4.3 MK Data Hub

MK Smart city like any other smart city relies heavily on data to function. Its state –of –art data hub stores huge amount of data from various source." Smart city provides solutions to our rapid urbanisation. There is a general and growing concern of the challenges, that cities will encounter based on the current growth pace." (Okai et al., 2018)

MK Smart is no exception to what Smart Cities are doing to help solve some of the challenges of urban cities. MK Data Hub is the central to the MK Smart project and this supports the acquisition and management of the huge data (Big Data) from various data sources which are relevant to the city systems (MK Smart Project, 2014).These include data about energy and water consumption, transport data, data acquired through satellite technology, social and economic datasets, and crowdsourced data from social media or specialised apps. (MK Smart Project, 2014). The MK Data Hub is housed at the University of Bedfordshire's Milton Keynes Campus and serve as the platform for experimentation for projects relating to the MK:Smart project.

Currently there are about 801 Datasets with several data streams available on the Data Hub. These data streams are regularly updated and keep increasing in size on the daily basis. These datasets comprise data from sources sensors, environment, water, statistics, education , businesses, etc. (MKSmart, 2014).

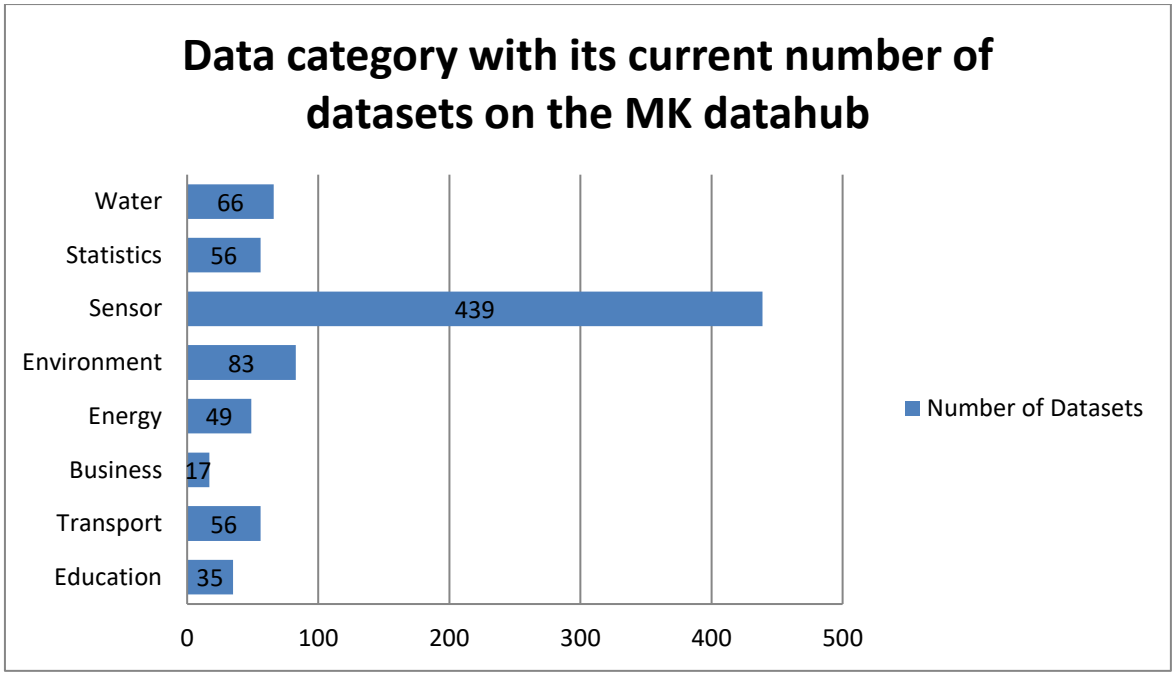

Figure 4 The Different Category with its Current Number of Datasets on the MK Data Hub

The MK Data Hub is currently open to project partners such as application developers, analytics who want to create applications using the dataset currently available on the Data Hub. Developers are presented with the opportunity to make use of the data through APIs found in Develop. There is opportunity for individuals and companies to create their own data which can be share or sell through the Data facility. Currently only registered members can access the datasets on the Data Hub. It is also worth mentioning that some datasets are private and only available to the owners.

Datasets available can be free or need to be purchased depending on the preference of the dataset owner. The Data Hub consists of several layers. Its physical infrastructure is located at the University Campus of Milton Keynes, previously known as University Centre Milton Keynes. The Data Hub uses similar configuration as private cloud to enable the use of layers and chapping of applications requirements (d'Aquin et al., 2014).

The architecture of MK Data Hub is built on three key layers. These are data import layer, the storage layer and the data delivery layer (d'Aquin et al., 2014). Design of the architecture of MK Data Hub is made in way that it does not rely on a single type storage technology. The benefit of this type of architecture design makes it easier for maintenance and also the management of storage components that are distributed to a dedicated robust servers. This is cost effective in-terms of architectural development.

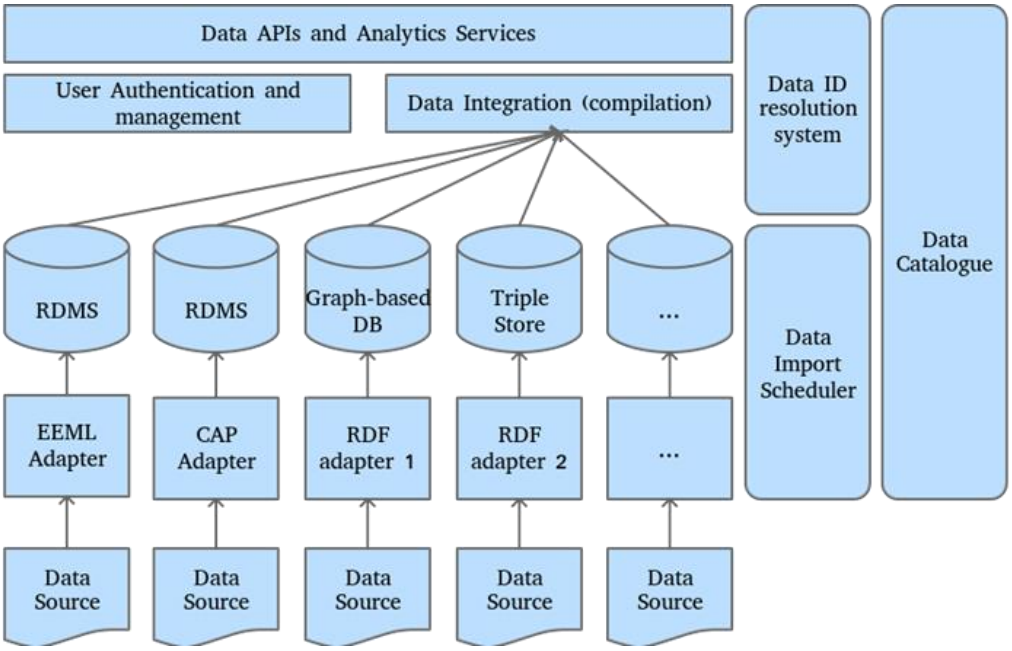

Figure 5. "Overview of the Architecture of the Data Management Layer of the MK: Data Hub" (d'Aquin et al., 2014)

One of the challenges of the Data Hub is the diversity of data source with different mode of transfer, different format , different update rate and different constraints (d'Aquin et al., 2014).

To solve this challenge, there was the need to have a generic data processing framework that will allow the creation of ad-hoc processing pipeline for any data source. MK Data Hub relies on Apache ActiveMQ2 as the standard communication mechanisms.

4.3.1 Security and Forensics Challenges to The MK Smart Project

4.3.1.1 The Security Challenges

"The MK smart project as previously explained relies heavily on data for the purpose of the project. Such huge data is not easily to manage and secure especially as data is coming from various sources" (Okai et al., 2019).

"The MK Data Hub which is the central infrastructure currently house 801 datasets. There have been challenges that needed to overcome for the project to continue to be able to host data on the Data Hub. Among some of these challenges are;" (Okai et al., 2019);

4.3.1.1.1 The Cyber-attack Challenge

"There is always a risk for cyber-attack to such a huge project. As mentioned earlier on, there are various data from various sources stored on the Data Hub. Data availability is key at all times and need to be constantly available for any project needing information on the Data Hub. The project which has involved key companies and respectable academic organisations has also encouraged various cyber-attack on the Data Hub for the past years and continue to attack on the daily basis." (Okai et al., 2019).

4.3.1.1.2 The Physical Security

“The datahub is a physical infrastructure which has a physical location as compared to other Smart city datahub which is hosted in the cloud. This exposes the Data Hub to the challenges of unauthorised access to the datacentre” (Okai et al., 2019).

4.3.1.1.3 The People Challenge

“Big project like the MK Smart project attracts and involves many people in its processes. It is hard to keep track of who is who and this can allow unauthorised to people who may be involved in the project but do not have the responsibility of looking after the Data Hub” (Okai et al., 2019)”.

4.3.1.2 The Forensics Challenges

“One of the most challenging area facing digital forensics is Big Data. Traditionally, digital forensics rely on the techniques of collecting data by the removal of the physical storage from the investigating machine and then calculating the data’s MD5/SHA-1 checksums for its integrity. There is a physical acquisition that capture all the metadata. It may be feasibly impossible to forensically acquire terabytes of data using the current tools and techniques on Big Data. This is also time consuming and involves heavy usage of resources. The current digital forensics practice is limited when it comes to Big Data. Big Data contains huge amount of data which is impossible to use the traditional forensics tools to capture evidence. Another area is the large complexity of Big Data which makes it difficult for it to be taken offline for forensics investigation. MK Smart project faces the same challenge as Big Data in this aspect. The project produces a huge amount of data from real-time applications which is stored on the Data Hub. Such huge data restricts the capabilities of the current tools and techniques used by the traditional forensics’ investigation. Challenges like this can put the integrity of data in doubt and prevent an effective forensics investigation in cases where evidences are needed in the court of Law” (Okai et al., 2019).

4.3.2 Solutions to Some of the Challenges of the MK Data Hub

“Every challenge creates opportunity for solution to evolve. Although the above challenges pose such a huge risk to the project, however, there have been some level of countermeasures to mitigate the risk. Among some of these solutions are. (Okai et al., 2019)”

4.3.2.1 Arcserve Unified Data Protection Solution

“The primary aims of Arcserve is to ensure data availability and secure data that are available on the Datahub. Arcserve Unified Data Protection is a solution to recover data over a SAN and has the capability to support environment such as heterogeneous. This solution has also safeguarded data across the physical servers of the Datahub and protected data which has been sourced from various sources across the city. The solution also automatically back up the whole data on the datahub in less than 30 minutes which saves time and resources.” (CMS Distribution, 2015; Okai et al., 2019)

4.3.2.2 Access Control Card Reader in Place & CCTV Motion Sensors

“Having access control card reader to the datacentre reduces the risk of unauthorised people gaining access to the datacentre room. There is also restriction on which group of people can have access to the datacentre room. In addition to the card reader, there is also CCTV and motion sensor that detect every movement into the data centre room.

In terms of solutions to the forensics challenges to this project, little can be said of the countermeasure as this is relatively new project and new to the digital forensics area. There is on-going research to find a better approach for Big Data forensics.” (Okai et al., 2019).”

5 Practical Works Undertaken

5.1 Practical Works Introduction

The chapter elaborates on the practical work undertaken in evaluating the Milton Keynes Smart City forensics. As previously noted, the MK Smart Data Hub which is the central infrastructure to this project will be the main project area to be examined. There would be security and forensics evaluations on the Data Hub. The evaluations will highlight the key challenge. A penetration testing and digital forensics analysis would be conducted .Today's technologies are constantly under threat of cyber-attacks. Even with the most secure network there is always a little chance of malicious attack on the network.

Whilst you can never guarantee to have a safe network at all times, it is always important to look at the aftermaths effect should an attack ever happen. This is why digital security and forensics go together. The security aspect looks to detect and prevent attacks using different techniques. The forensics on the hand investigates into the computer attacks to establish evidence.

5.2 Security Demonstration on MK Data Hub

It is important to look at the security aspects to establish how secure the MK Data Hub is about. To identify this, penetration testing also known as ethical hacking was carried on the Data Hub. It must be said this was authorised and supervised by the development team of MK Data Hub based at the Open University in Milton Keynes. The main purpose behind this testing was to find vulnerabilities could be exploited.

5.2.1 Process for the penetration testing on MK Data hub

The process for penetration testing has been described below.

5.2.1.1 Establish the IP Address for the Data hub

To be able to carry pen-testing session, it was necessary to establish the IP address of the data hub. The IP address can be identified by using the ping utility .The Ping utility tool was used to verify if a domain/server is operating and network accessible (IPaddressguide, 2005).

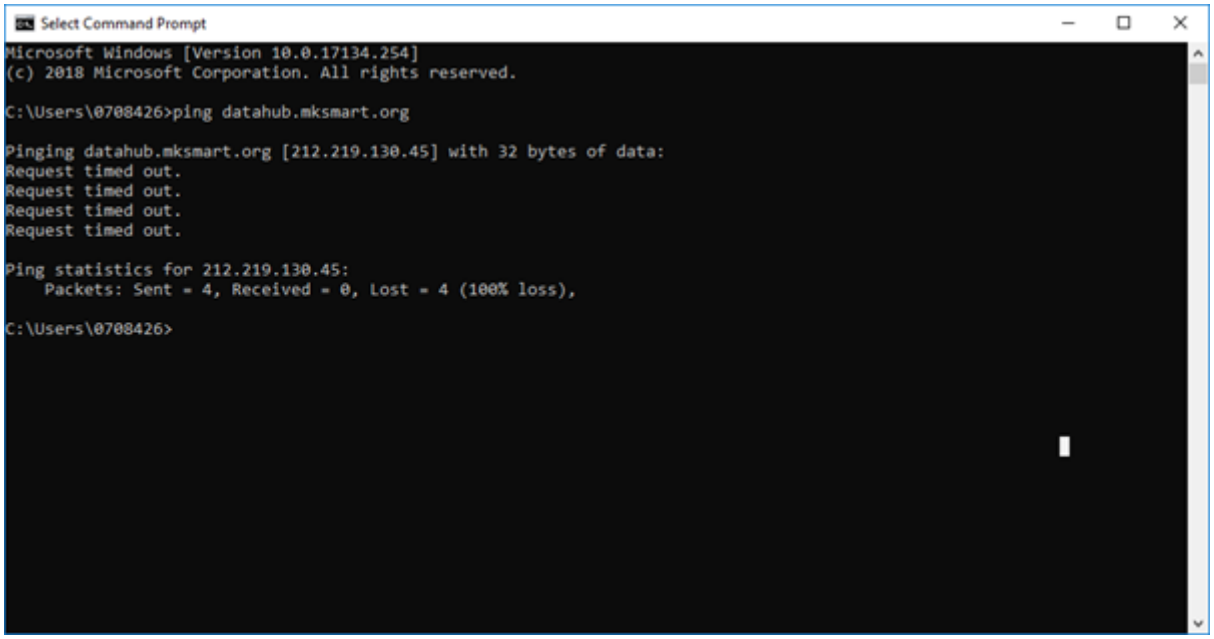

Figure 6.The Ping Details of MK Data Hub

5.2.1.2 Measures for Securing the Data Hub

Why IP is not reachable when pinged?

One of the measures of making Data Hub secure is to disable the ping function of its IP address .This prevents giving out information detailed information about the Data Hub.

5.2.1.3 Identify the Penetration tool

It was important to identify the penetration tool for this demonstration. Metasploit security tool was used for the exercise .Metasploit framework “helps security teams do more than just verify vulnerabilities, manage security assessments, and improve security awareness but it empowers and arms defenders to always stay one step (or two) ahead of the game” (Rapid7, n.d.). It was also a security framework I am competent in using for vulnerability scanning and testing. To carry out the pen-testing, Kali Linux, a Linux operating system which has pre-installation of Metasploit and GUI version (Armitage) was used.

The following are the steps followed during the penetration testing.

5.2.1.4 IP Scan

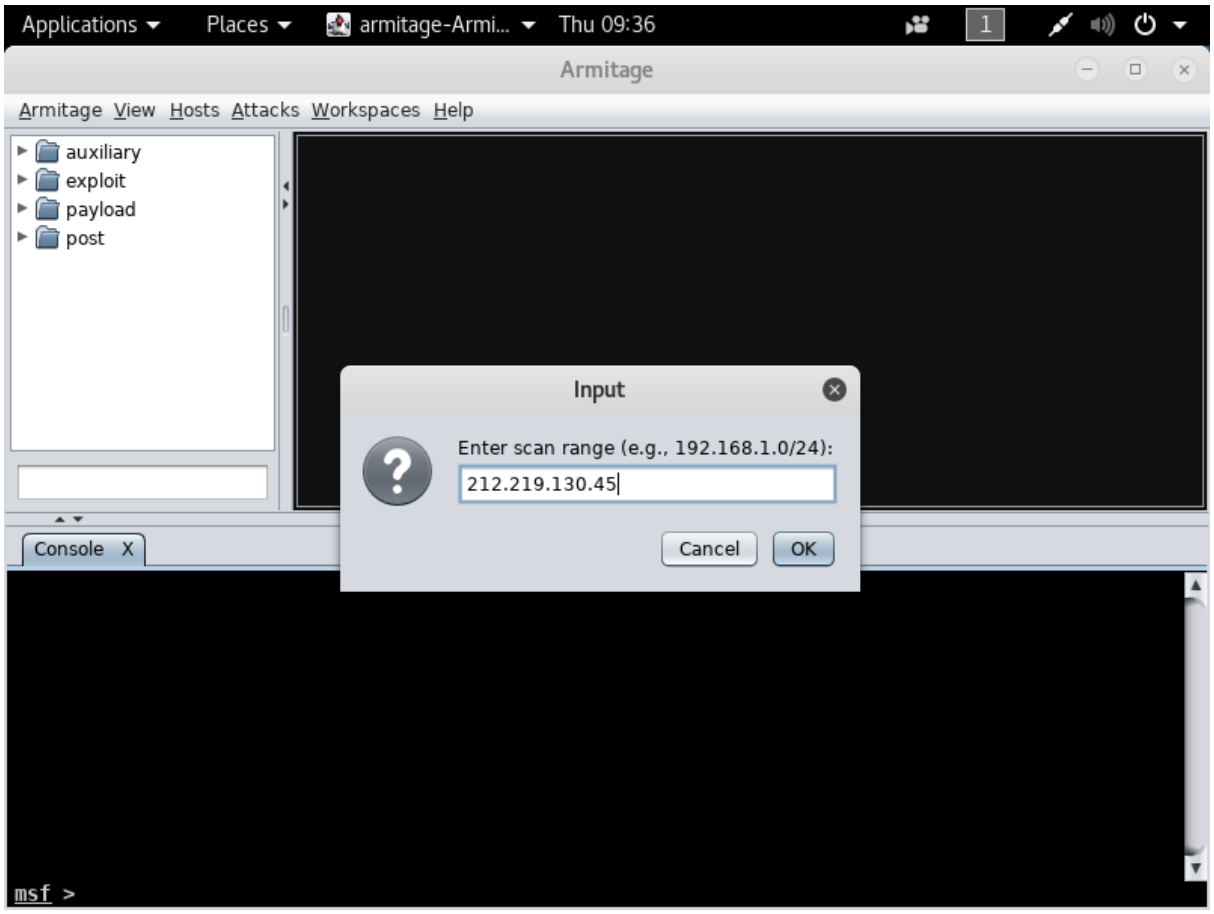

Figure 7.The Demonstration of IP Scan in Armitage

5.2.1.5 IP Scan Results

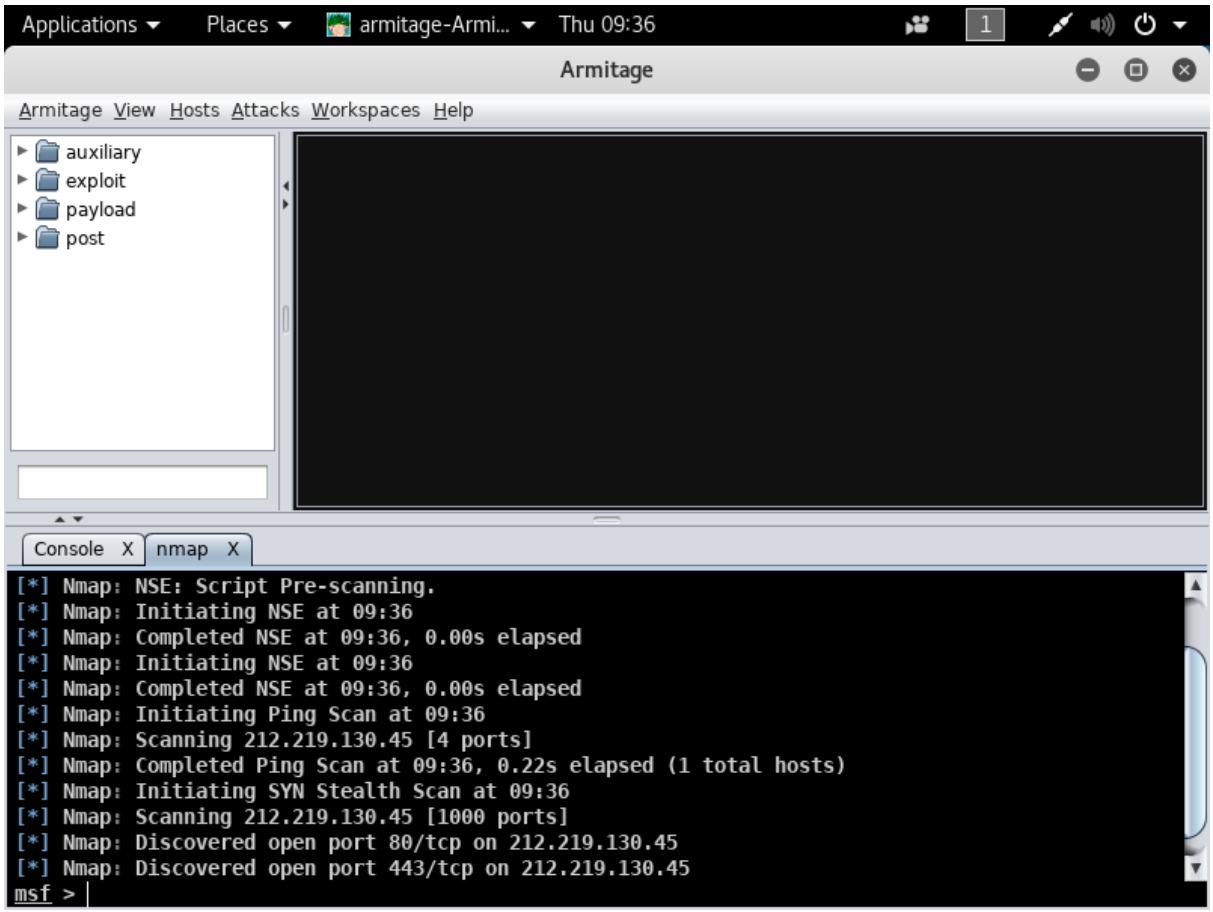

Figure 8. The Results of the IP Scan and the State of Ports on this Network.

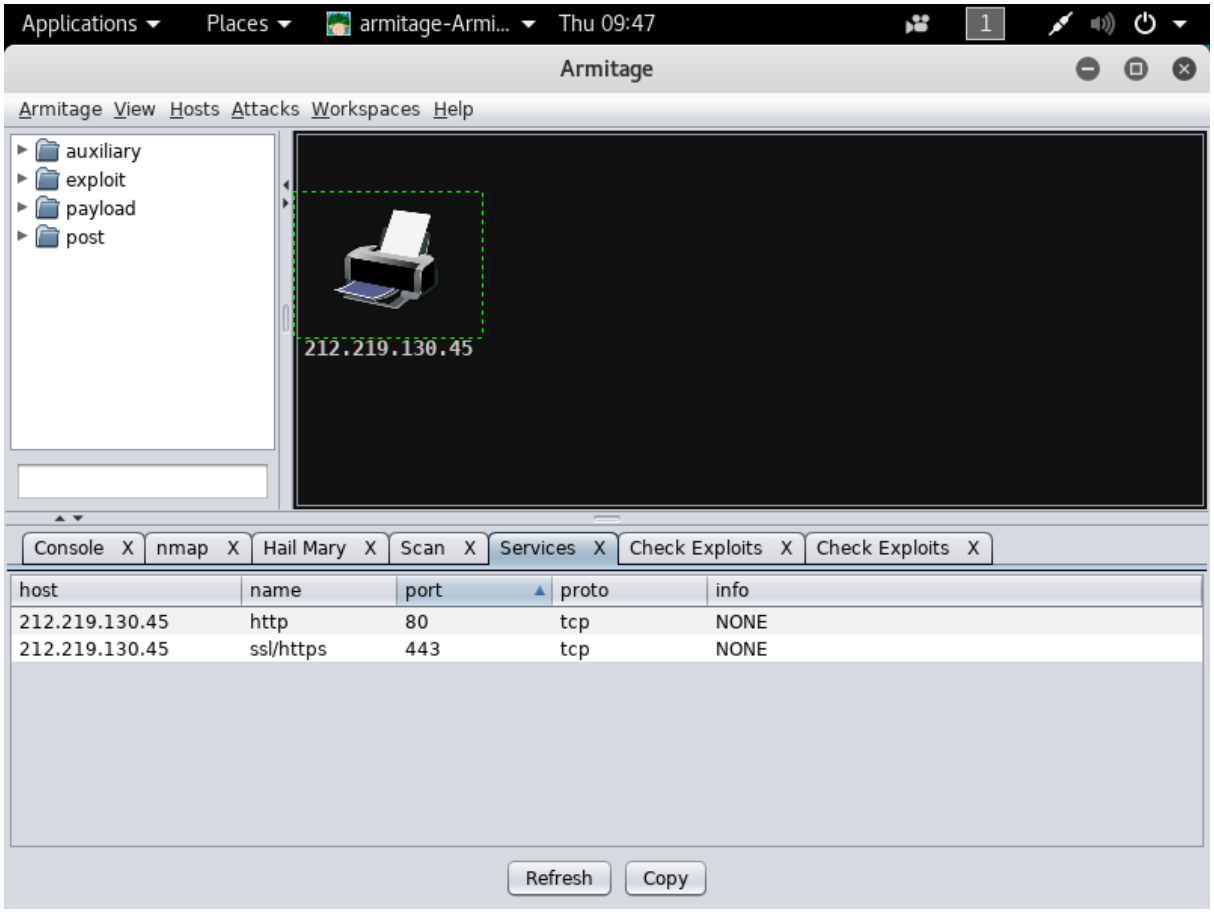

Figure 9.The Result of Ports that are Open

5.2.1.6 Vulnerability Checks

The test was to exploit task by launching an attack on the Data Hub. This was done to exploit any vulnerabilities in the system.

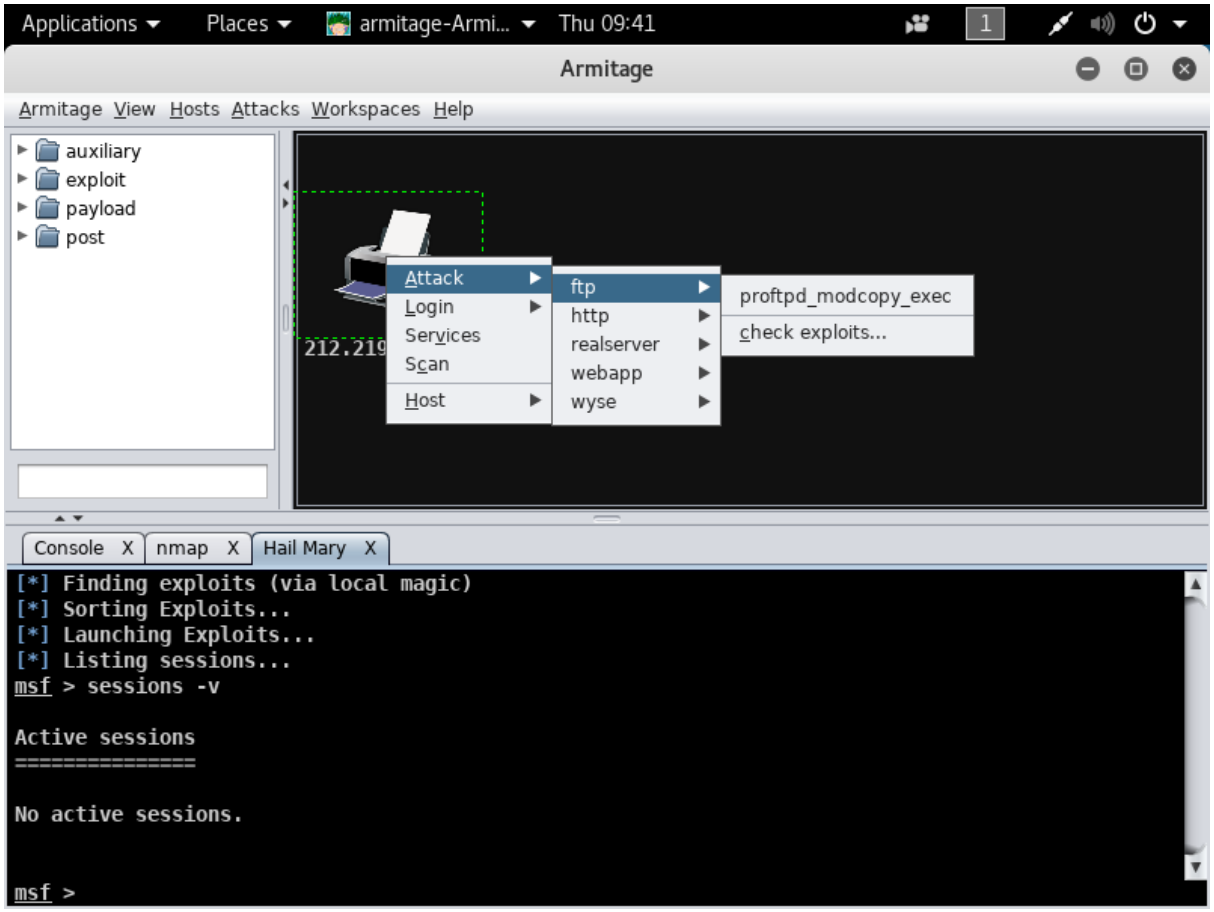

Figure 10. An Attack has been Launched to Exploit any Weakness in the System.

5.2.1.7 Final Results

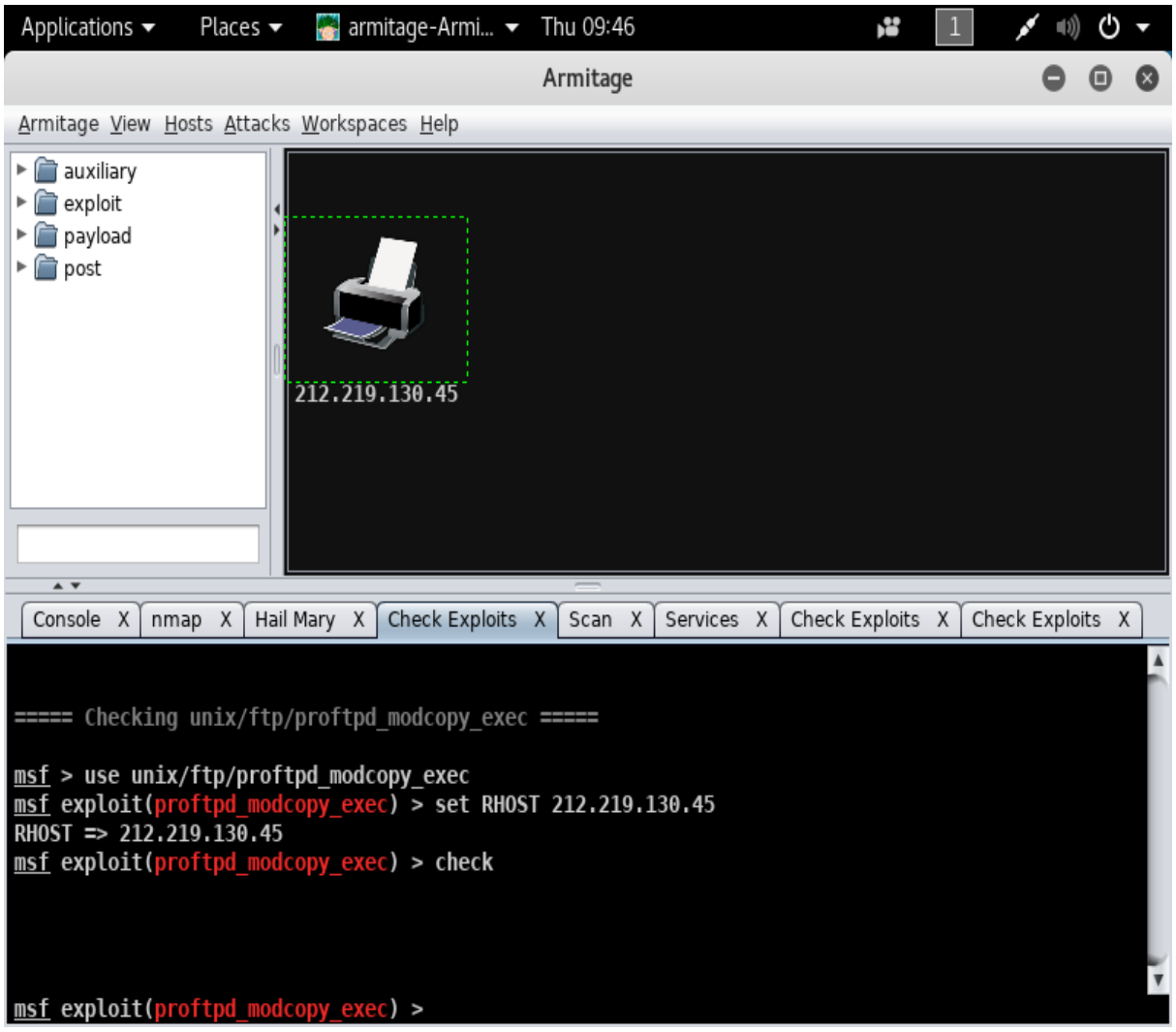

Figure 11. No Vulnerabilities to Exploit in the System

5.3 Recommendations and Critical Evaluation

One of the ways of solving Smart Cities forensics challenges is to look at Big Data forensics. Most Smart Cities rely heavily on huge amount of data (Big Data) whether stored on the local data centre or in the cloud. Big Data forensics “is the identification, collection, analysis, and presentation of the data in a Big Data system” (Sremack, 2015). It focuses on collecting data from large-scale databases and is applications.

5.3.1 Hadoop Solution

Hadoop “is an open source distributed processing framework that manages data processing and storage for Big Data applications running in clustered systems” [Rouse 2019].It is one of the most adopted platforms for Big Data and very supportive of forensics evidence on applications and data sources.

Joe Sremack explained Hadoop as “virtually synonymous with Big Data and has become the de facto standard in the industry.” (Sremack, 2015). Although there are regular new software updates for Hadoop, its architecture remains the same and it is important for forensics investigators to familiarise themselves to be able to perform data acquisition. Hadoop has its own filesystem, databases, and application layers, which can store data in different locations and different forms. The data stored can be collected as evidence in forensics investigation. With the help of the Hadoop, data evidence can be identified on multiple layers on Hadoop and can easily be extracted through Hadoop Distributed File System (HDFS), using application such as Pig. Another benefit of Hadoop is ability to collect specific data needed for evidence. This is less time consuming in the case of Big Data acquisition.

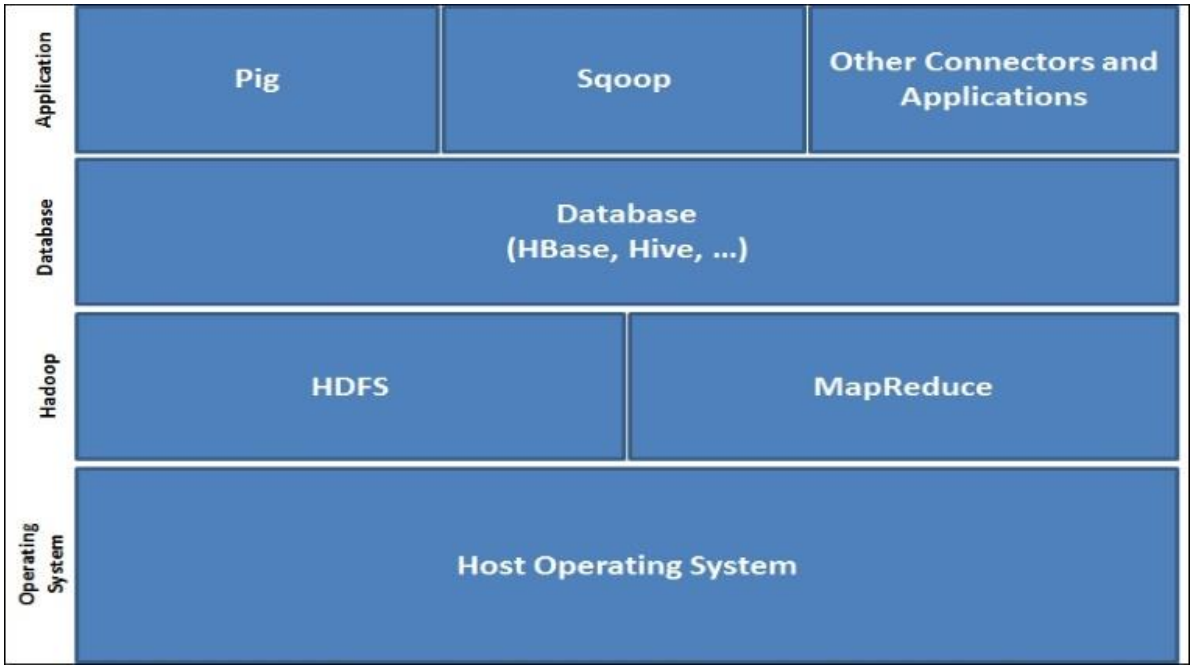

Figure 12.How Hadoop Architecture is Structured

5.3.2 The Hadoop Architecture

Hadoop has been one of the reliable systems when it comes to shared storage. It is enriched with ecosystem solution and efficient tools for Big Data .It is also a cross platform and based on java-based solutions which is built on ideas of storage and computing distribution. Hadoop has its own platform which can sit on different operating systems such as window, Linux etc. This gives Hadoop to function independently without relying on any operation system

5.3.3 The Hadoop Architecture Layers

The Operating System layer: Operating System is the first layer on the host machine. This is installed on top of the existing operating system and is compatible with all operating system.

The Hadoop layer: This layer consists of the MapReduce components and file system. It also serves as the base installation of Hadoop.

The DBMS layer: This layer houses the DBMS and other applications which are installed. Installation includes data warehousing such as Hive/HBase.

The Application layer: This is the top layer and it contains tools that are used for data management analysis. Typical tools like Pig can be installed on the application layer which can interact with the operating system and Hadoop layer directly.

5.3.4 Running Hadoop

Operating Hadoop is a straight- forward application installation that can be operated from a number of several platforms. It can operate either through a desktop machine ,through a network or as cloud -based service. It should be noted all these set up may have a vary process of collecting forensic evidence.

5.3.5 MK Data Hub Forensic Analysis with Hadoop

The primary aim of for Mk data Hub forensic is to find a data that can be used as evidence .Hadoop data can either be store on the disk or in memory. There may different data on Hadoop disk and memory that may be relevant to the forensic investigation. In the case of Mk Data hub , forensic data may fall into three categories.

Supporting information: Data on the datahub that provides in-depth information about the Hadoop configurations.

Record Evidence: Any data such as text files analysed in Hadoop will recorded.

User and application evidence

This where logs files, configuration files, analysis scripts, metadata etc , can be found. This gives information about which data has been analysed

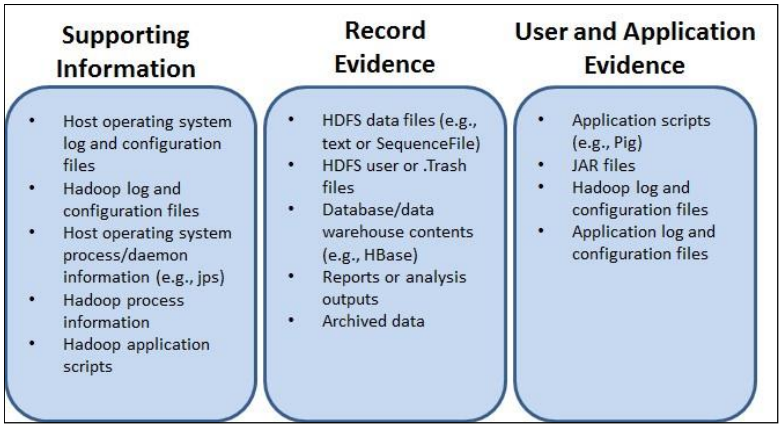

Figure 13. The Types of Forensic Evidence

5.3.6 Identifying MK Data Hub Evidence with Hadoop

Identifying evidence on such a Big Data is not only tedious but also involves many resources. This can be very costly and time consuming. Big Data contains a huge amount of data with the potential of increasing its size of data per every second.MK Data Hub is no exception to this as some of the data stored are real-time data information that keeps

changing in size. To be able to identify evidence on the data hub, it is important to look at some of the artefacts contained within Hadoop that are relevant to forensic investigation

The architecture of the MK Data Hub is based on three main layers; the data import layer, the storage layer and the data delivery layer.

The data import layer creates the channels to import the data coming from the data sources. The challenge this layer faces is the diversity of thousands of data coming from different data sources with different data formats, different constraints, different update rates etc. To solve this challenge, it was necessary to have a generic wrapper for all the data processing services and also a standard communications such as Apache ActiveMQ2

The storage layer holds the actual data for MK Smart project .The decision made in terms of the design of this layer was not to be dependent on sole storage technology but have hybrid type of data storage such as Relational Database Management System (RDMS), triple store and graphic-based .The reason behind this was the flexibility in maintenance and management of smaller storages, Its robustness compared to other and the development cost when it comes to this type of storage.

The aim of the data delivery “layer is to provide data to external applications and services in a standardised and suitable way, through both feed-centric and entity-centric Web APIs” (Mathieu d'Aquin, 2014) “The data integration in the MK Data Hub uses global identifiers (URIs) for connected objects, and then send to in format of the client’s choice (XML, RDF, CSV, JSON, etc)” (Okai et al., 2019).

5.4 Recommended Approach

MK Data hub can be related to Big Data forensics. The data Hub contains a large amount of data which can be classified as Big Data. To be able to obtain forensic evidence from the MK data hub, it is important to look at some of the existing frameworks that have helped to obtained Data evidence from other Bigdata projects. Hadoop forensics has been adopted based on its functionalities which make it easier to obtain data evidence from Big Data storages.

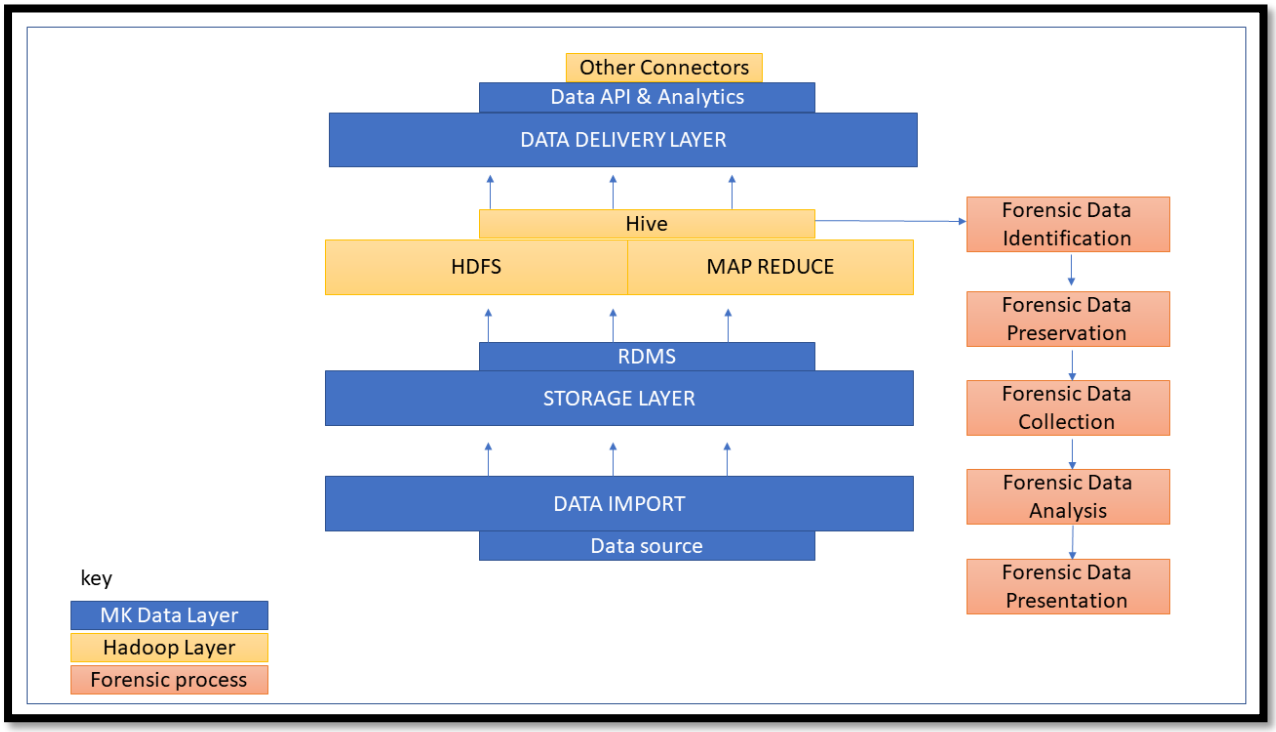

Figure 14.The proposed Hadoop Forensics Ecosystem on MK Data Hub.

As previously mentioned, there are three main layers of the MK Data hub; The Data import, the storage layer and the data delivery layer (Mathieu d'Aquin, 2014). The diagram above show the two type of layers with different colour codes. The blue layer refers to the architecture layer of the MK Data hub while the gold layer refers to the proposed Hadoop layers.

The data import takes take from different data sources and through adapters such as EEML, CAP and RDF transfer the data to the storage layer. The next layer is the delivery layer, and this is where data users can use data through different data APIs.

How Hadoop works on the MK Data hub.

There are three layers of Hadoop layer installation on the MK data hub. These are Hadoop Distributed File System (HDFS), Map reduce, and Hive (Sremack, 2015).

5.4.1 Hadoop Distributed File System (HDFS)

This is the main data storage system used by Hadoop applications. “It uses NameNode and DataNode architecture to implement a distributed file system that provides high-performance access to data across highly scalable Hadoop clusters.” (Rouse, n.d.). HDFS provides a reliable way of managing huge amount of Big Data by providing help for supporting analytics applications for Big Data (Rouse, n.d.).

HDFs together with map reduce are used to run the distributed system. Joe Sremack author of Big Data Forensics -Learning Hadoop investigation talks about the distribution

as being controlled by a master node machine which controls the slave node machine purposely for file storage ,retrieval and data analysis (Sremack, 2015).

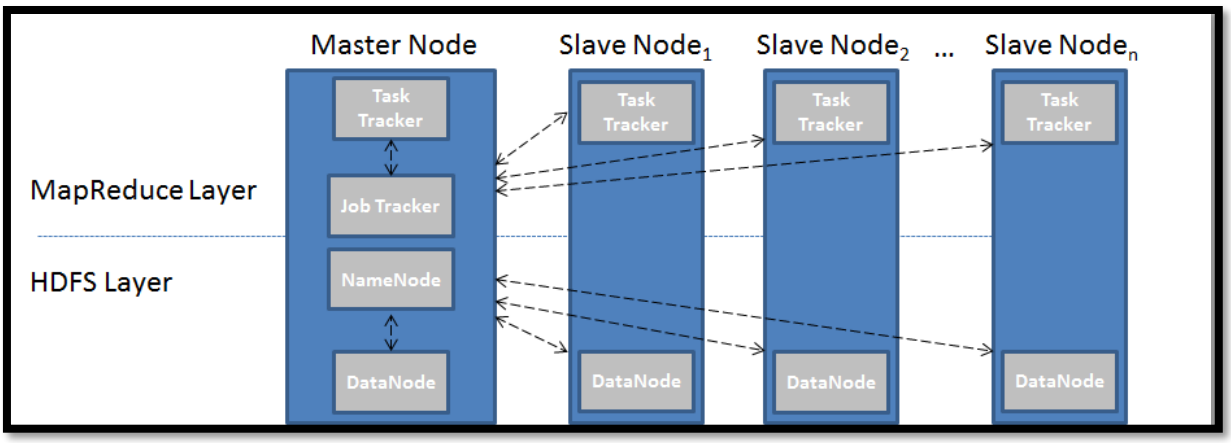

Figure 15.The Hadoop Distributed Process

Hadoop stores data for its applications through HDFS. HDFS is implemented to store data on the storage hardware in a distributed way.

Artefacts of Hadoop Distributed File System (HDFS)

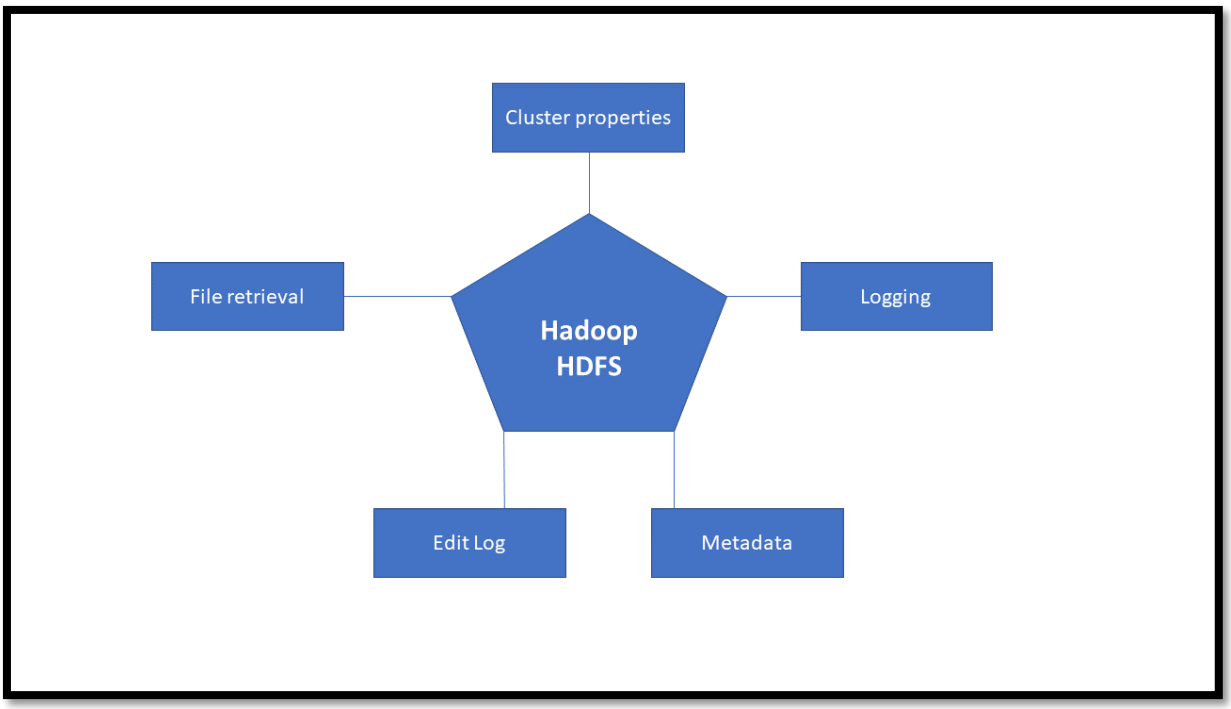

Figure 16. Artefacts of Hadoop Distributed File System (HDFS)

5.4.2 MapReduce

Another key concept of Hadoop which enables large data to be divided into map. MapReduce is additionally fault-tolerant, with every node sporadically coverage its standing to a master node. If a node doesn't respond needless to say, the master node reassigns that piece of the duty to different offered nodes within the cluster. This creates resiliency and makes it sensible for MapReduce to run on cheap goods server.

5.4.3 Hive

Hive is usually used thanks to its SQL like command language which is used to access data in Hadoop HDFS. Hive is familiar and more user friendly to SQL users using it for querying data. Its service manages the data storage and query operations, converting them jobs and execute and return a result to the query.

5.5 The Forensic Process

Identification – This stage identifies any relevance data that might be useful in the forensic investigation. In the case of MK data Hub, data can be easily identified through the use of Hadoop framework. Such Big data can easily be grouped by MapReduce and accessed using Hive to query the specific data relevant to the investigation.

Preservation –Vital stage of preserving evidence obtained from incident scene and also documenting all necessary information about the evidence obtained from the data hub. Once evidence has been identified on the data hub, it is important to preserved making sure evidence is not lost in the forensic process

Collection – At this stage evidence can be collected on the data hub using specialised tools such as FTK, Encase, X way forensics etc. The evidence collected is then imaged .This helps to preserve the original evidence and also can be re-image wherefore there is an error on the imaged evidence.

Analysis – In-depth analysis of evidence collected . The examination plays a key part in the whole process and its analysis produces the conclusion on the evidence obtained. At this stage ,the investigator may decide if the evidence from the data hub is relevant to what he/she is looking forward. Any relevant findings may be used in the court of law or in case of private investigation be presented to the client. The FTK imager shows the evidence collected from the MK Data Hub. This will help to make a copy of the data collected and also analyse the data without comprising the original data collected.

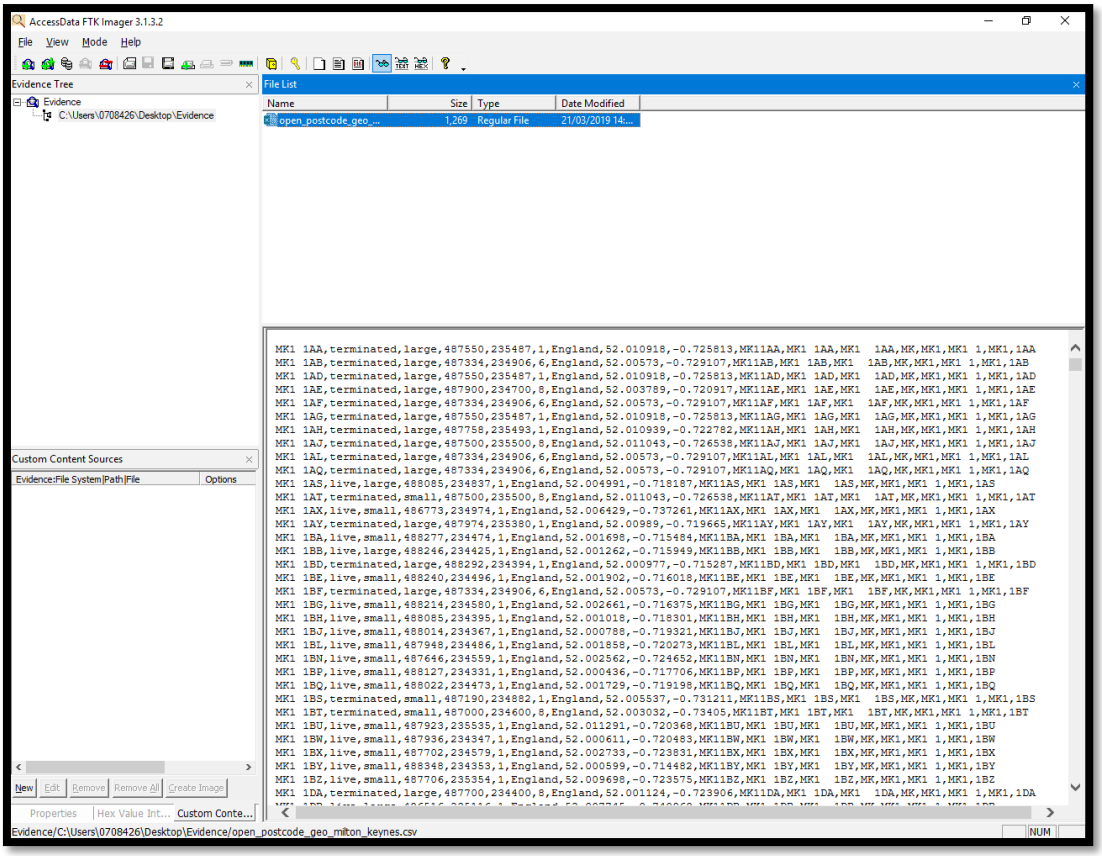

Figure 17. The Analysis Stage of Evidence Collected Using FTK Imager

Presentation – This is final report of the whole process and final findings of the evidence obtained from the Data Hub. This is a document that details the process of the forensic investigation and the outcome of the investigation.

5.6 Critical Evaluation and Conclusions

Smart Cities forensics is a new way of solving the challenges presented by cities that are implementing smart ideas such as smart parking through the use of satellite data, smart transportation system, effective eco–energy system etc. These cities are embracing the new technologies that are providing solutions to the growing population and high demand for cities infrastructures. This is no contrary to the prediction made by the UN on the number of people living in the cities to be around 6 million by 2050 (United Nations, 2014).

One of the key aspects of every smart city is the ability to work with data effectively. These data are collective information from various sources that are relevant to the smart city project. The data are stored normally in the cloud or physical infrastructure such as the data hub.

As a joint initiative ,the MK smart project is funded by HEFCE (Higher Education Funding Council for England) and led by the Open University and has support of key players such as the university of Bedfordshire, British Telecom(BT), Milton Keynes Council, University

of Cambridge, community Action MK and so on (MK Smart Project, 2014) . MK Data Hub is central to the project which supports the acquisition and management of the Big Data from various data sources which are relevant to the city systems. These include data about energy and water consumption, transport data, data acquired through satellite technology, social and economic datasets, and crowdsourced data from social media or specialised apps (MKSmart, 2014).

Whilst the data plays a crucial part of this project, its forensic value of the data held is also important to the investigation of this project. Data might be required to help in any forensic investigation to prove in a case of Data integrity.

This project has focused on the research problem associated with the forensic aspect of the MK Smart project. Whilst the project primarily relied on the use of Big Data. There are huge challenges for Big Data to be forensically investigated for this MK Smart project.

The MK Data hub which is the integral of the MK Smart project contains a vast amount of data relevant to the aims of the project. The MK Data Hub is a state-of-the-art' which supports the acquisition and management of vast amounts of data relevant to city systems from a variety of data sources (MK Smart Project, 2014).

Hadoop framework has been recommended to help tackle MK Smart Forensics. Hadoop "is an open source distributed processing framework that manages data processing and storage for Big Data applications running in clustered systems (Rouse, n.d.)." It is one of the most adopted platforms for Big Data and very supportive of forensic evidence on applications and data sources.

Although there are regular new software updates for Hadoop, its architecture remains the same and it is vital for forensic investigators to familiarise themselves to be able to perform data acquisition. Hadoop has its own filesystem, databases, and application layers, which can store data in different locations and different forms. The data stored can be collected as evidence in forensic investigation. With the help of the Hadoop, data evidence can be identified on multiple layers on Hadoop and can easily be extracted through Hadoop Distributed File System (HDFS), using an application such as Pig. Another benefit of Hadoop is the ability to collect specific data needed for evidence. This is less time-consuming in the case of Big Data acquisition. Through the Hadoop framework, MK Data hub can be mapped which makes data to be grouped through MapReduce and query by application such as hive. Through hive, forensic tools such as Encase, FTK etc can be connected to collect forensic evidence for further analysis.

5.7 Limitations

As previously discussed, there is a lack of standard when it comes to Smart Cities forensics. The current forensic tools and existing process have limitations which makes it very challenging to retrieve data evidence from Big Data. Although Hadoop has provided a mechanism to be able to help collect evidence from Big Data such as the Data Hub.

5.8 Future Works

Cloud forensics has had its share of challenges when it comes to collecting evidence that can be admissible in court. This has highlighted some of the challenges any Big Data platform will face when it comes to forensic data. Some of these challenges are identical to the ones of Smart Cities Forensics.

There is still work to be done in terms of digital forensics when it comes to the MK Smart Project. Big Data forensics continue to be a challenge to the digital forensics discipline, however, a product such as Hadoop has been very supportive of data acquisition through its layers making it easy to identify and specify data on the data hub. More tailored forensic process for Smart Cities can be beneficial for digital forensic investigation.

6 Appendices



Appendix A. The Data Hub for the MK Project



Appendix B. The Data Hub for the MK Project



Appendix C. The Data Hub for the MK Project

University of Bedfordshire

Appendix D. The Ethics form

UNIVERSITY OF BEDFORDSHIRE

Research Ethics Scrutiny (Annex to RS1 form)

SECTION A To be completed by the candidate

Registration No: 0708426

Candidate: EBENEZER OKAM

Degree of: MASTERS BY RESEARCH

Research Institute: IRAC

Research Topic: SMART CITIES FORENSICS- EVALUATING THE CHALLENGES OF MK SMART CITY FORENSICS.

External Funding: N/A

The candidate is required to summarise in the box below the ethical issues involved in the research proposal and how they will be addressed. In any proposal involving human participants the following should be provided:

- clear explanation of how informed consent will be obtained,
- how will confidentiality and anonymity be observed,
- how will the nature of the research, its purpose and the means of dissemination of the outcomes be communicated to participants,
- how personal data will be stored and secured
- if participants are being placed under any form of stress (physical or mental) identify what steps are being taken to minimise risk

If protocols are being used that have already received University Research Ethics Committee (UREC) ethical approval then please specify. Roles of any collaborating institutions should be clearly identified. Reference should be made to the appropriate professional body code of practice.

N/A

October 2014

University of Bedfordshire

UNIVERSITY OF BEDFORDSHIRE

Research Ethics Scrutiny (Postgraduate Research Students)

When completing this form please ensure that you read and comply with the following:

Researchers must demonstrate clear understanding of an engagement with the following:

1. *Integrity* - The research has been carried out in a rigorous and professional manner and due credit has been attributed to all parties involved.
2. *Plagiarism* - Proper acknowledgement has been given to the authorship of data and ideas.
3. *Conflicts of Interest* - All financial and professional conflicts of interest have been properly identified and declared.
4. *Data Handling* - The research draws upon effective record keeping, proper storage of date in line with confidentiality, statute and University policy.
5. *Ethical Procedures* - Proper consideration has been given to all ethical issues and appropriate approval sought and received from all relevant stakeholders. In addition the research should conform to professional codes of conduct where appropriate.
6. *Supervision* - Effective management and supervision of staff and student for whom the researcher(s) is/are responsible
7. *Health and Safety*- Proper training on health and safety issues has been received and completed by all involved parties. Health and safety issues have been identified and appropriate assessment and action have been undertaken.

The **Research Institutes** are responsible for ensuring that all researchers abide by the above. It is anticipated that ethical approval will be granted by each Research Institute. Each Research Institute will give guidance and approval on ethical procedures and ensure they conform to the requirements of relevant professional bodies. As such Research Institutes are required to provide the University Research Ethics Committee with details of their procedures for ensuring adherence to relevant ethical requirements. This applies to any research whether it be, or not, likely to raise ethical issues. Research proposals involving vulnerable groups; sensitive topics; groups requiring gatekeeper permission; deception or without full informed consent; use of personal/confidential information; subjects in stress, anxiety, humiliation or intrusive interventions must be referred to the University Research Ethics Committee.

Research projects involving participants in the NHS will be submitted through the NHS National Research Ethics Service (NRES). The University Research Ethics Committee will normally accept the judgement of NRES (it will never approve a proposal that has been rejected by NRES), however NRES approval will need to be verified before research can commence and the nature of the research will need to be verified.

Where work is conducted in collaboration with other institutions ethical approval by the University and the collaborating partner(s) will be required.

The **University Research Ethics Committee** is a sub-committee of the Academic Board and is chaired by a member of the Vice Chancellor's Executive Group, appointed by the Vice-Chancellor and includes members external to the University

Research Misconduct: Allegations of Research Misconduct against staff or post graduate (non-taught) research students should be made to the Director of Research Development.

October 2014

University of Bedfordshire

Applicant declaration
I understand that I cannot collect any data until the application referred to in this form has been approved by all relevant parties. I agree to carry out the research in the manner specified and comply with the statement of ethical requirements on page 1 of this form. If I make any changes to the approved method I will seek further ethical approval for any changes.

Signature of Applicant: ..... Date: 02/11/18.....

Signature of Director of Studies: ..... Date: .....

This form together with a copy of the research proposal should be submitted to the Research Institute Director for consideration by the Research Institute Ethics Committee/Panel

Note you cannot commence collection of research data until this form has been approved

SECTION B To be completed by the Research Institute Ethics Committee:

Comments:

Approved

Signature Chair of Research Institute Ethics Committee:

Date:

This form should then be filed on the student's record

If in the judgement of the committee there are significant ethical issues for which there is not agreed practice then further ethical consideration is required before approval can be given and the proposal with the committees comments should be forwarded to the secretary of the UREC for consideration.

There are significant ethical issues which require further guidance

Signature Chair of Research Institute Ethics Committee:

Date:

This form together with the recommendation and a copy of the research proposal should then be submitted to the University Research Ethics Committee

October 2014

University of Bedfordshire

Answer the following question by deleting as appropriate:

1. Does the study involve vulnerable participants or those unable to give informed consent (e.g. children, people with learning disabilities, your own students)?
   Yes **No**

   If **YES:** Have/will Researchers be DBS checked?
   Yes **No**

2. Will the study require permission of a gatekeeper for access to participants (e.g. schools, self-help groups, residential homes)?
   Yes **No**

3. Will it be necessary for participants to be involved without consent (e.g. covert observation in non-public places)?
   Yes **No**

4. Will the study involve sensitive topics (e.g. sexual activity, substance abuse)?
   Yes **No**

5. Will blood or tissue samples be taken from participants?
   Yes **No**

6. Will the research involve intrusive interventions (e.g. drugs, hypnosis, physical exercise)? **Yes** **No**

7. Will financial or other inducements be offered to participants (except reasonable expenses)? **Yes** **No**

8. Will the research investigate any aspect of illegal activity?
   Yes **No**

9. Will participants be stressed beyond what is normal for them?
   Yes **No**

10. Will the study involve participants from the NHS (e.g. patients) or participants who fall under the requirements of the Mental Capacity Act 2005?
    Yes* **No**

If you have answered yes to any of the above questions or if you consider that there are other significant ethical issues then details should be included in your summary above. If you have answered yes to Question 1 then a clear justification for the importance of the research must be provided.

*Please note if the answer to Question 10 is yes then the proposal should be submitted through **NHS research ethics approval procedures** to the appropriate **NRES**. The UREC should be informed of the outcome.

Checklist of documents which should be included:

Project proposal (with details of methodology) & source of funding	
Documentation seeking informed consent (if appropriate)	
Information sheet for participants (if appropriate)	
Questionnaire (if appropriate)	

(Tick as appropriate)

October 2014

University of Bedfordshire

7 List of Tables

Table 1. The Nomis' breakdown of employment rate in Milton Keynes for the period between Jan 2018 - Dec 2018 20

University of Bedfordshire

8 List of Figures

Figure 1.The Four Vs of Big Data 8
Figure 2.The Digital Forensics Process Model. 13
Figure 3.The Milton Keynes Population Growth in 2017 as Demonstrated by Plumplot. . 19
Figure 4 The Different Category with its Current Number of Datasets on the MK Data Hub 22
Figure 5. “Overview of the Architecture of the Data Management Layer of the MK: Data Hub” (d'Aquin et al. 2014) 23
Figure 6.The Ping Details of MK Data Hub 27
Figure 7.The Demonstration of IP Scan in Armitage 28
Figure 8. The Results of the IP Scan and the State of Ports on this Network. 29
Figure 9.The Result of Ports that are Open 30
Figure 10. An Attack has been Launched to Exploit any Weakness in the System. 31
Figure 11. No Vulnerabilities to Exploit in the System 32
Figure 12.How Hadoop Architecture is Structured 33
Figure 13. The Types of Forensic Evidence 34
Figure 14.The proposed Hadoop Forensics Ecosystem on MK Data Hub. 36
Figure 15.The Hadoop Distributed Process 37
Figure 16. Artefacts of Hadoop Distributed File System (HDFS) 37
Figure 17. The Analysis Stage of Evidence Collected Using FTK Imager 39

9 Publication Bibliography

AccessData (n.d.) *FORENSIC TOOLKIT (FTK)* [Online], USA, AccessData. Available at https://accessdata.com/products-services/forensic-toolkit-ftk (Accessed 17 May 2019).

Aggarwal, A. (2016) *Five challenges to overcome for Smart Cities to happen* [Online], India, maximumgovernance. Available at http://maximumgovernance.com/perspectives/five-challenges-to-overcome-for-smart-cities-to-happen/ (Accessed 6 June 2019).

Aher, B. (2018) *How Big Data Impacts Smart Cities* [Online], n.p., Dzone. Available at https://dzone.com/articles/how-big-data-has-the-biggest-impact-in-smart-citie (Accessed 3 April 2019).

BBC (2017) *Milton Keynes at 50: How the new town has changed over the years* [Online], London, BBC. Available at https://www.bbc.co.uk/news/av/uk-england-beds-bucks-herts-38717480/milton-keynes-at-50-how-the-new-town-has-changed-over-the-years (Accessed 11 January 2020).

BigdataIdn (2017) *Big Data:The 3Vs Explained* [Online], n.p., BigdataIdn. Available at https://bigdataldn.com/intelligence/big-data-the-3-vs-explained/ (Accessed 3 February 2019).

Biyikoglu, C. (2016) *Indexing Big Data: Global vs. Local Indexes in Distributed Databases* [Online], n.p., Dzone. Available at https://dzone.com/articles/faster-indexing-and-query-global-vs-local-indexing (Accessed 7 July 2019).

Bristol is open (n.d.) *Bristol is open* [Online], London. Available at https://www.bristolisopen.com/ (Accessed 3 March 2019).

BW Online Bureau (2019) *6 Features Of A Smart City* [Online], BW Online Bureau. Available at http://bwsmartcities.businessworld.in/article/6-Features-Of-A-Smart-City/06-02-2019-166882/ (Accessed 6 August 2019).

CityVerve (n.d.) *https://cityverve.org.uk/* [Online], Manchester, Manchester Council. Available at https://cityverve.org.uk/ (Accessed 11 January 2020).

CMS Distribution (2015) *Arcserve: MK Smart Research* [Online], London, CMS Distribution. Available at https://www.cmsdistribution.com/wp-content/uploads/2015/06/Arcserve-Open-University-MKi-GTG.pdf?utm_source=Website&utm_medium=Case%20Study&utm_campaign=Arcserve%20Study%20Case (Accessed 1 January 2019).

Computerworld UK (2019) *Future city Glasgow* [Online], London, Computerworld UK. Available at https://www.computerworlduk.com/galleries/infrastructure/best-smart-cities-in-uk-3683649/ (Accessed 12 May 2019).

Cosgrove, M. (2011) *Smart Cities series: introducing the IBM city operations and management solutions* [Online], n.p., IBM Corporation. Available at https://www.researchgate.net/profile/Leo_Bachinger/publication/311845438_Changing_Worlds_Ideologies_Utopias_and_Ambitions_in_Science_Technology/links/585d199608aebf17d38a2eb6/Changing-Worlds-Ideologies-Utopias-and-Ambitions-in-Science-Technology.pdf (Accessed 3 April 2019).

Cyfor (2002) *Digital Forensics* [Online], Manchester, Cyfor. Available at https://cyfor.co.uk/digital-forensics/audio-visual-forensics/ (Accessed 15/01/20119).

d'Aquin, M., Adamou, A., Daga, E. and Liu, S. (2014) *Dealing with diversity in a smart-city Datahub* [Online], Milton Keynes, Knowledge Media Institute. Available at https://www.researchgate.net/publication/289875110_Dealing_with_diversity_in_a_smart-city_Datahub (Accessed 10 September 2019).

Digital Birmingham (n.d.) *Smart Cities* [Online], Birmingham, Birmingham Council. Available at http://digitalbirmingham.co.uk/service/smart-cities/ (Accessed 5 May 2019).

Forensic Control (2019) *Computer Forensics: A Beginner's Guide* [Online], London, Forensic Control. Available at https://forensiccontrol.com/resources/beginners-guide-computer-forensics (Accessed 3 January 2018).

Glassdoor (n.d.) *Companies in Milton Keynes* [Online], London, Glassdoor. Available at https://www.glassdoor.co.uk/Reviews/milton-keynes-reviews-SRCH_IL.0,13_IM1128_IP14.htm (Accessed 4 March 2019).

Greater London Authority (n.d.) *Smarter London Together* [Online], London. Available at https://www.london.gov.uk/what-we-do/business-and-economy/supporting-londons-sectors/smart-london/smarter-london-together (Accessed 6 April 2019).

Guidance Software (2018) *OpenText EnCase Forensic* [Online], n.p., Guidance Software ((08/2018)10517EN). Available at EnCase Forensic (Accessed 2 May 2019).

Hoevenaars (2015) *The Biggest Challenges to Develop a Sucessful Smart City project* [Online], n.p., van Euroforum BV. Available at https://www.smart-circle.org/smartcity/blog/biggest-challenges-develop-successful-smart-city-project/ (Accessed 7 August 2020).

HR Wallingford (2015) *Smart water management for Milton Keynes* [Online], London, HR Wallingford. Available at http://www.hrwallingford.com/projects/MKSmart (Accessed 5 May 2018).

Hurst, A. (2018) *The UK's current smart city developments examined* [Online], n.p., Information Age. Available at https://www.information-age.com/the-uks-current-smart-city-developments-examined-123474013/ (Accessed 6 July 2019).

IBM (n.d.) *Big data analytics* [Online], USA, IBM. Available at https://www.ibm.com/analytics/hadoop/big-data-analytics (Accessed 12 October 2019).

InfoSec (2018) *Computer Crime Investigation using Foresic Tools and Technology: Computer Crime Investigation Using Forensic Tools and Technology* [Online], n.p., InfoSec. Available at https://resources.infosecinstitute.com/computer-crime-investigation-using-forensic-tools-and-technology/ (Accessed 6 January 2018).

Intaforensics (n.d.) *Mobile Phone Forensic* [Online], Nuneaton, Intaforensics. Available at https://www.intaforensics.com/digital-forensics/mobile-phone-and-tablet-forensics/?doing_wp_cron=1578739106.3380680084228515625000 (Accessed 4 April 2019).

IPaddressguide (2005) *IPv4 Ping* [Online], n.p., IPaddressguide. Available at https://www.ipaddressguide.com/ping (Accessed 2 February 2019).

Knowledge Media Institute (2019) *CityLabs* [Online], Milton Keynes, Knowledge Media Institute. Available at https://www.citylabs.org.uk/ (Accessed 8 January 2019).

Leeds Council (n.d.) *Smart Leeds* [Online], Leeds, Leeds Council. Available at https://datamillnorth.org/smart-leeds/ (Accessed 3 June 2019).

MK Smart Project (2014) *About MK Smart* [Online], Milton Keynes, MK Smart Project. Available at http://www.mksmart.org/about/ (Accessed 8 September 2019).

MKfm (2017) *Milton Keynes will have third-highest economic growth of all UK cities in 2018* [Online], Milton Keynes, MKfm. Available at https://www.mkfm.com/news/local-news/milton-keynes-will-have-third-highest-economic-growth-of-all-uk-cities-in-2018/ (Accessed 10 October 2019).

MKinspire (2012) *Milton Keynes: History of Milton Keynes* [Online], n.p., MKinspire. Available at http://www.mkinspire.org.uk/history-of-milton-keynes.html (Accessed 5 May 2019).

MKSmart (2014) *Data Catalogue* [Online], Milton Keynes, MKSmart. Available at https://datahub.mksmart.org/data-catalogue/ (Accessed 5 May 2018).

Nate and Lord (2018) *Memory Forensics* [Online], USA, Datainsider. Available at https://digitalguardian.com/blog/what-are-memory-forensics-definition-memory-forensics (Accessed 7 June 2019).

Okai, E., Feng, X. and Sant, P., eds. (2018) *Smart Cities Survey* [Online], London, IEEE. Available at https://ieeexplore.ieee.org/abstract/document/8623018 (Accessed 10 October 2019).

Okai, E., Feng, X. and Sant, P., eds. (2019) *Security and Forensics Challenges to The MK Smart Project* [Online], Luton, IEEE. Available at http://www.smart-world.org/2019/atc/doc/SWC2019%20-%20Conference_Program_V1.pdf (Accessed 9 August 2019).

Open University (n.d.) *Digital forensics* [Online], London, Open University. Available at http://www.open.edu/openlearn/science-maths-technology/digital-forensics/content-section-4.3 (Accessed 11 January 2020).

Plumplot (n.d.) *Milton Keynes population statistics* [Online], Milton Keynes, Plumplot. Available at http://www.plumplot.co.uk/Milton-Keynes-population.html (Accessed 2 January 2020).

Population City (2015) *Milton Keynes · Population* [Online], London, population City. Available at http://population.city/united-kingdom/milton-keynes/ (Accessed 11 January 2020).

Rapid7 (n.d.) *Metasploit* [Online], n.p., Rapid7. Available at https://www.metasploit.com/ (Accessed 2 February 2019).

Reeves, D. (2018) *How to Create a Smart City: Future-Proofed Cities That Foster Growth and Innovation* [Online], Washington, IEEE. Available at https://ieeexplore.ieee.org/document/8369448 (Accessed 6 April 2019).

Rouse, M. (n.d.) *Managing Hadoop projects: What you need to know to succeed* [Online], n.p, Techtarget. Available at https://searchdatamanagement.techtarget.com/definition/Hadoop-Distributed-File-System-HDFS (Accessed 11 January 2020).

Sremack, J. (2015) *Big Data Forensics – Learning Hadoop Investigations* [Online], Birmingham, Packt Publishing Limited. Available at https://books.google.co.uk/books?hl=en&lr=&id=lrZrCgAAQBAJ&oi=fnd&pg=PP1&dq=Joe+Sremack&ots=CnptQy1By1&sig=CXmRtYiXvfOOaB02dybySsRgSd8&redir_esc=y (Accessed 5 August 2019).

Technopedia (n.d.) *Network Forensics* [Online], n.p., Technopedia. Available at https://www.techopedia.com/definition/16122/network-forensics (Accessed 2 May 2019).

Techtarget (2013) *Guide to Big Data Analytics Tools and Best Practices* [Online], n.p, Techtarget. Available at https://searchbusinessanalytics.techtarget.com/

essentialguide/Guide-to-big-data-analytics-tools-trends-and-best-practices (Accessed 9 June 2019).

The Unbelievable Machine Company (n.d.) *What is Big Data? – A definition with five Vs* [Online], Berlin. Available at https://blog.unbelievable-machine.com/en/what-is-big-data-definition-five-vs (Accessed 6 May 2019).

United Nations (2014) *World's population increasingly urban with more than half living in urban areas* [Online], New York, United Nations. Available at https://www.un.org/en/development/desa/news/population/world-urbanization-prospects-2014.html (Accessed 8 July 2019).

University of Bedfordshire (2016) *Sensors and Smart Cities* [Online], Luton, University of Bedfordshire. Available at https://www.beds.ac.uk/howtoapply/courses/postgraduate/next-year/sensors-and-smart-cities (Accessed 11 January 2019).

X-Ways Forensics (n.d.) *X-Ways Forensics: Integrated Computer Forensics Software* [Online], n.p, X-Ways Forensics. Available at http://www.x-ways.net/forensics/index-m.html (Accessed 16 June 2019).

Yin, K. R. (1994) *Designing single- and multiple-case studies, in Educational Management through Research and Consultancy.* [Online], n.p. Available at https://books.google.co.uk/books?hl=en&lr=&id=mcUH125X6_8C&oi=fnd&pg=PA135&dq=Designing+single-+and+multiple-case+studies,+in+Educational+Management+through+Research+and+Consultancy.&ots=Xi-hIZAPH2&sig=1GqQwb1IWmsMEOCy7q7NjiWoL58&redir_esc=y (Accessed 10 October 2019).