EXPLORATION OF CLUSTERING OVERLAPS IN A RANSOMWARE NETWORK BASED ON LINK STRUCTURES AND CONTENT RELEVANCE

BERNARD CHUKWUEMEKA OGAZI-ONYEMAECHI
PhD, MEng, MSc, BSc

Doctor of Philosophy (PhD)

SCHOOL OF COMPUTING, SCIENCE AND ENGINEERING
UNIVERSITY OF SALFORD

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy

MARCH 2019

Contents

List of figures

List of Appendices

Acknowledgments

The task of accomplishing a four-year Ph.D. research work had been an uphill one. This was against the background that the path trailed in this great academic adventure was a dark tunnel with the expectation of a great light at the end. It was said, “Behind every cloud, there is a silver lining”. In realization of the golden and seasoned saying “where there is a will, there is always a way” I willed it within me that I must overcome all challenges in the pursuit of this academic goal. Today, it is a reality. The requirements for the degree of Doctor of Philosophy have been accomplished.

However, the will and consequent accomplishment would have been a mirage if a capable supervisor were not available. May all the glory be to God that Dr. Rob Aspin took over the project at the time he did. Therefore, I acknowledge the great opportunity he has given to me to accomplish this task under his short period of guidance. It has been a great experience to work with him. I acknowledge also the groundwork made by Dr. Ali Dehghantanha to this project.

I express my heartfelt thanks and love to my wife, Mrs. Chinyere G. Ogazi, and my children, Divine E. I. Onyemaechi, Stephanie E. C. Onyemaechi, Chelsea U. C. Onyemaechi and Malvyn C. C. Onyemaechi for their love, and for always been a source of inspiration during this period of studentship. My brothers and sisters are well appreciated for been available for me.

To the evergreen memory of my parents, Late Longinus Sunday Onyemaechi and Late Elizabeth Alunma Onyemaechi, I say peace of GOD.

In all situations, GOD is SUPREME. I thank HIM for HIS WILL in my life. To HIM be all the GLORY, HONOR and ADORATION, AMEN.

Dedication

To the Glory and Majesty of GOD, I dedicate this Thesis to my children, Divine Ebubechukwu Ihechidere Onyemaechi, Stephanie Ezichidindu Chinelo Onyemaechi, Chelsea Uchechi Chidiogo Onyemaechi, and Malvyn Chichetaram Chukwuemeka Onyemaechi. May this be one of the sources of inspiration for them to achieve full accomplishments in their chosen careers.

"There is nothing constant in the universe. All ebb and flow, and every shape that's born *bears in its womb the seeds of change." Ovid*

YOU CAN BE WHAT YOU CHOOSE TO BE.

Declaration

I, Dr. Bernard C. Ogazi-Onyemaechi, do declare that the contents of this Thesis are my own independent work. In addition, I certify that I have duly acknowledged the literary or content use and quotes of the works of others from other sources, where appropriate.

Signed: Date: 31 March 2019

Abstract

The advancement in technology makes it easy and effective to transmit and spread ransomware to different devices. The result is increased ransomware threats to different sectors of the World Economy. The reason for the spread of these threats is that Ransomware developers are trying to increase their revenue by infecting victims of specific (industry) sectors of the world economy with targeted ransomware to breach their security and steal valuable data. To counter these threats, industries employ different security measures to prevent ransomware-related losses, yet losses continue to occur because of the ever-changing dynamics of ransomware. Consequently, industries are continuously, searching for effective measures to control ransomware attacks. Forensic and security experts and law enforcement operatives also have limitations in the control of such security threats.

The study of cluster analysis in none ransomware domains (e.g. complex social networks links and contents) has proved invaluable for the detection of cluster hubs, authorities, and communities in complex social networks. This has helped in targeted marketing activities and in the detection, identification, and prediction of terrorist and criminal hubs (gangs) within a network. That ransomware distribution and spread have similar complex network configuration to the social network, application of cluster analysis on the ransomware becomes a possible area of interest in the fight against ransomware threats. Compared to the social network, ransomware structure comprises of a set of links and contents, whose nodes (vertices) represent ransomware families, IP Addresses, URLs, Host, Registrar, ASNs, Countries, status or other entities embedded in the distribution. Real-time Active Cluster Overlap Profiling and Tracking (ACOPT) of ransomware network overlapping cluster trends presents an opportunity to prevent a successful attack. The study reveals there is active threat when the network events activity peaks at 53.53% with a prior gradual increase from 10.19% through 27.38%. The threat happens when the number of overlapping clusters reaches the highest maximum threshold preceded by the lowest minimum threshold. At the onset of threat, the clustering elements and the percentage values between the active cluster node and terminal cluster node are equal (29.11%); and the difference between them and the highest percentage cluster node (41.77%) is -12.11% and 12.11%. In addition, the onset is characterized when the percentages of the cluster intensity of the active cluster node and terminal cluster node reaches respective values of 31.03% and 24.89% and the difference between them and the cluster node with the highest value (44.08%) becomes equal to -13.05% and 14.19%. The active threat therefore, occurs when the active cluster node and terminal cluster node records respective 27.38% and 39.29% in the number of clustering objects, while other cluster nodes record equal values of 11.11%. The active overlapping cluster, therefore, is identified to be the cluster that has the most regular, consistent and closely distributed number of clustering objects, measures of centrality and intensity values in all the cumulative periods of the time series of the ransomware network.

Therefore, the present investigation by exploring temporal events and overlapping cluster formation in a Ransomware network identified an active cluster-overlap, which could be removed to timely dislodge potential Ransomware threat. The active cluster-overlap was tracked through cluster profiling in a time-series and periodic network clustering analysis of Ransomware Network to establish pattern consistency. The consistency tracking and validation were achieved using the key performance parameters of the network, cluster intensities, and the frequencies of clustering objects. The removal of the active Cluster Overlap (node 1) was proved effective in dislodging the ransomware network and controlling threat before it attacks. Hence, the study proposes a real-time Exploratory Machine-learning Cluster Overlap System **(EMCOS)** for links and contents cluster analysis in a complex ransom-ware network as a tool to control threat.

CHAPTER 1

GENERAL INTRODUCTION

1.0 Introduction

Industries and organizations take many security measures to prevent ransomware threats. Despite the huge capital investments to protect products and properties, organizations still face the risk of ransomware-related thefts and breaches. Ransomware threats do not seem to have terminal solutions as the cyber-space experience development in infrastructure, technology, and cyber-intelligence services rapidly expand and evolve. Considering the cost of ransomware threat, this study seeks an exploratory machine-learning platform to help industries to make quick decisions to counter and control the threat. Therefore, the self-healing approach (cycle) to ransomware control aims to identify the active cluster overlap, which is the originating source (node) of ransomware traffic, that can be removed in a ransomware network to prevent an attack. The desired system (Figure 1) imports network data and conducts network analysis to identify the key performance parameters that influence the formation and development of the network clusters.

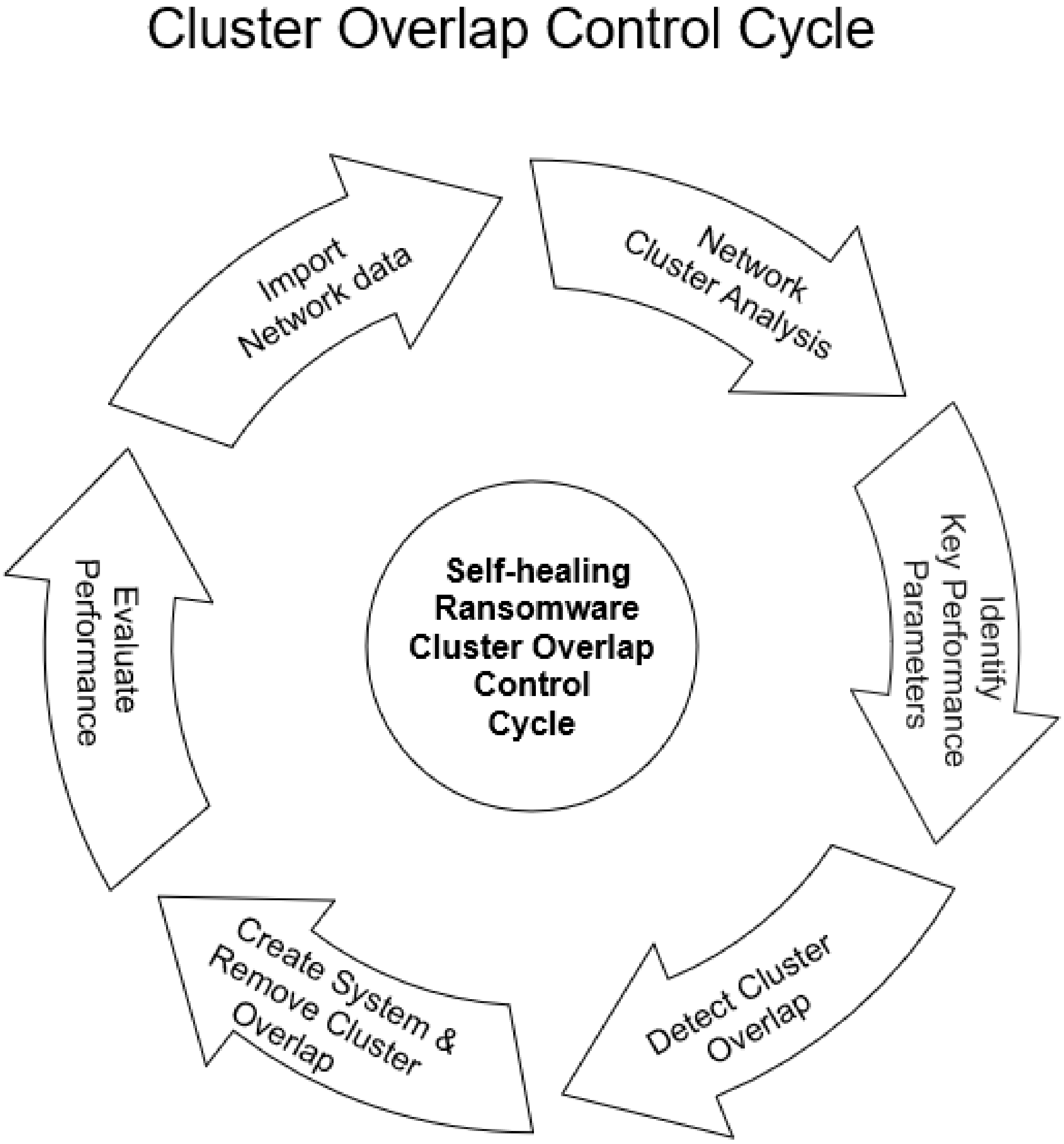

Cluster Overlap Control Cycle
Import Network data
Network Cluster Analysis
Identify Key Performance Parameters
Detect Cluster Overlap
Create System & Remove Cluster Overlap
Evaluate Performance
Self-healing Ransomware Cluster Overlap Control Cycle

Figure 1: Self-healing Ransomware Cluster Overlap Control Cycle, Showing the cycle from data import to performance evaluation

The above understanding helps to detect cluster overlaps and identify the active node to remove to dislodge the network and control threat.

To appreciate the urgent need for this study, an understanding of the menace of modern malware, namely ransom-ware is imperative. The emergence of ransomware dates to 2005. Ransomware has so much grown in popularity that 2015 recorded an estimated average of 407,000 attempted ransom-ware attacks [1,2]. The consequent estimated value of 325 million (USD) was extorted from victims [1,3,4]. The year 2017 recorded an annual average cost of 11.7 million (USD) [5], while the predicted cost for the end of 2019 is 11.5 billion (USD) [6].

Ransom-ware consists of two words namely **ransom**, and **ware** deriving from **software**.

- **Ransom**, as the name implies, is money paid to release a person or thing that is forcefully held [7]; conversely,
- **Ware** simply stands for software or malware application used to execute restriction for collecting ransom.

Specifically, ransom-ware is a type of malware developed to infect computing systems with the criminal intent to encrypt the files in such a system; consequently, denying legitimate access to them by retaining the decryption key until the victim pays the required ransom amount [1,8,9]. Ransom-ware has its root and origin from malware (malicious software), and in most cases are the variants of malware. Basic among the ransom-ware includes crypto and locker ransom-ware that have emerged as one of the most troublesome categories of malware in modern history. While crypto encrypts files in a victim's device, locker locks the victims completely out of their devices [10]. Ransom-ware targets a variety of economic, industrial, social, financial and banking sectors of the World economy. Other sectors targeted include educational, public, and hospital, etc. [11]. In recent years, the most often preferred mode of ransom payment includes Bit-coin, the cryptographic digital currency based on Block-chain distributed ledger technology, which offers a secure, anonymous, and untraceable method of making and receiving payments [1,12–16]. The ransom was estimated to cost about one Bit-coin, approximately 450 (USD), per infected machine at the time of record [1,4,10,14,17,18]. Technological development in media devices and network connectivity (Figure 2), enhance the spread and effectiveness of ransom-ware attacks.

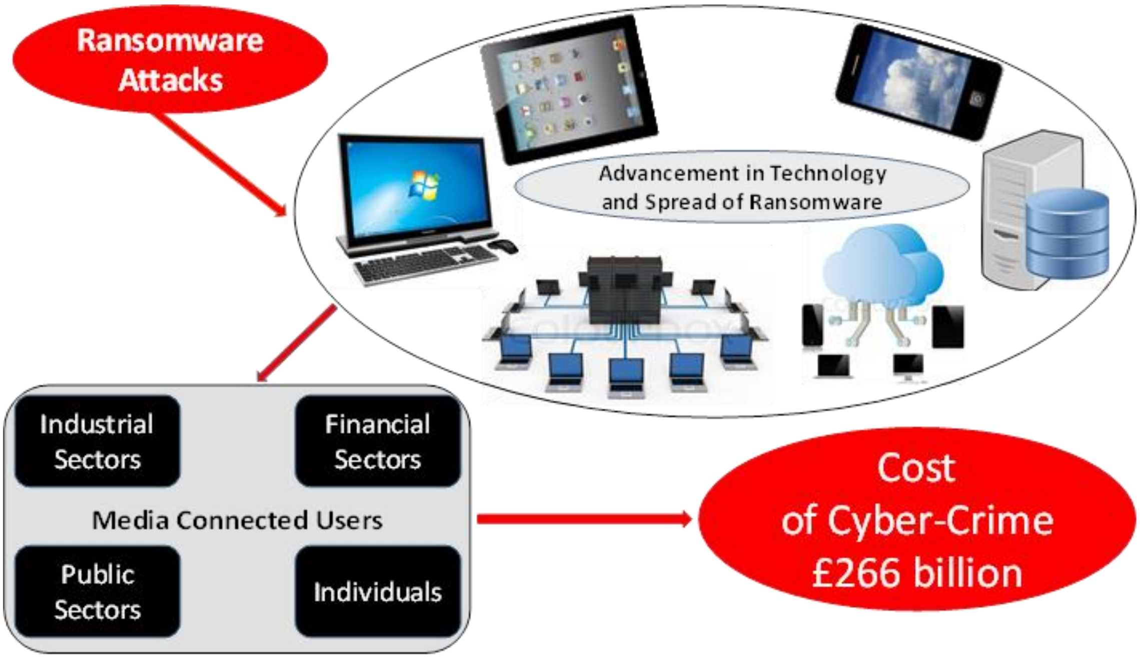


Figure 2: The effect of advances in modern technology on the Spread and cost of Ransom-ware

Analysis has been conducted on ransom-ware (malware) distribution, aiming to provide an insight into the activities of malware, origin, and distribution [19–22]. Studies have focused on analysing types of malware such as smartphone malware, worms, viruses and other propagating malware (e.g. spyware, keystroke loggers, information theft malware; botnet attacks, detection/tracking and defence; and rootkit and virtualization techniques) [19].

The primary target of this research was the creation and distribution of malware. Some malware trend analyses were conducted using APK Auditor that uses static analysis to characterize and classify android applications to target Android platform [23]. A study of malware (virus) propagation research proposed a delayed computer virus propagation model and studied the dynamic behaviours of malware [24]. One of the threats to network security is malware propagation, and topological scanning seen as a type of malware that spreads based on topology information [25]. The focus was on modelling the spread of topological malware to understand the potential damages they cause, and to develop countermeasures for protecting the network infrastructure. Assessing the suitability of internet provider's countermeasures against malware, a study [26] used an agent-based model, called ASIM, to investigate the impact of policy interventions at the Autonomous System level of the Internet [27].

A further study [28,29] developed an automatic malware detection model by training an SVM classifier based on behavioural signatures to overcome the problem of classification accuracy regarding unspecified malware detection when using signature-based analyses [29]. In 2008, a study presented an approach that enabled economic modelling of information security risk management in contemporary businesses and organizations [20]. The work focused on the prevention of heavy losses that may happen due to cyber-attacks and other information system failures in an organization. The prevention of such losses, the study observed, is associated with continuous investment in different security management measures and purchase of data protection systems [30–32], including:

- Identification of business assets
- Identification of threats
- Assessment of damages that result from a successful attack
- Identification of security vulnerabilities that could be exploited in the system
- Assessment of security risks
- Measures to minimize the risk and implementation of appropriate controls
- Evaluating the performance of the implemented controls

The study [30–32] introduced methods to identify the assets, the threats, the vulnerabilities of the ICT systems and proposed a model that enables the selection of the optimal investment of the necessary security technology based on the quantification of the values of the protected systems. This study also provided the model for external insurance-use based on quantified risk analyses [20]. Cyberark Labs classified some set of ransom-ware by their behaviour to determine the strategies that could be more effective in mitigating damages caused by ransom-ware attacks. The mitigating strategies discussed, include application whitelisting, application blacklisting, application grey-listing, least privilege, and backup and recovery [1]. This study did not specify the use of any classification or machine learning technique; however, they simulated the experiment in a controlled environment. Equally, this research did not provide any feature-selection classification model to compare the trend of the static features between benign-ware and ransom-ware and to identify the features that have more impact on the trend of the ransom-ware. Furthermore, this work did not provide a comparison of the trend of static features among different ransom-ware families, notwithstanding the huge and growing cost of successful attacks by different families of ransom-ware on the respective victims. These studies also did not model the degree to which the static features of ransom-ware contrast with benign-ware that makes them evade different security measures that result in successful attacks. The consequence of this gap is that security industries are caught unprepared to prevent malware attacks. These industries are not aware of the trend of the impacts of the attacks of each malware features because there are no known framework (models) on malware and ransom-ware trend and classification prediction. The lack of such frameworks, therefore, make it difficult to have effective control mechanisms to prevent malware attacks resulting in high economic consequences to different economic sectors affected.

From the above background, this research focuses on Clustering in Complex ransom-ware families Network Based on Content Relevance and Link Structures. This study employs the Clustering machine-learning algorithm to build cluster maps and analyse different cluster profiles. The study uses Machine-learning algorithms to "learn" information directly from the ransomware data without relying on a predetermined equation as a model. The algorithms adaptively improve their performance as the number of samples available for learning increases [33]. Two techniques are identifiable in machine learning as depicted in Figure 3. These include unsupervised and supervised machine learning techniques.

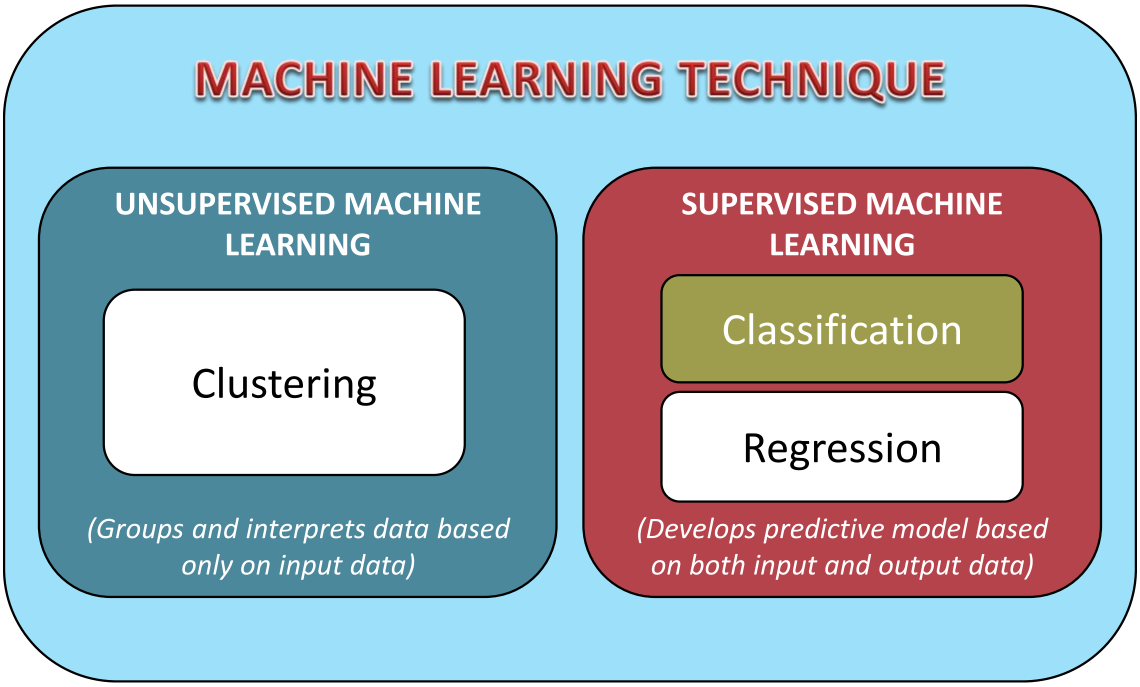

MACHINE LEARNING TECHNIQUE
UNSUPERVISED MACHINE LEARNING
Clustering
(Groups and interprets data based only on input data)
SUPERVISED MACHINE LEARNING
Classification
Regression
(Develops predictive model based on both input and output data)

Figure 3: Supervised and Unsupervised Machine Learning Approaches

Unsupervised machine learning finds hidden patterns or intrinsic structures in the input data. Subsequently, it groups and interprets the data based only on input data. This means that unsupervised learning finds hidden patterns or intrinsic structures in a dataset and uses it to draw inferences from data sets consisting of input data without labelled responses. Clustering is the most common unsupervised learning technique. It is used for exploratory data analysis to find hidden patterns or groupings in a data set. Applications for clustering include gene sequence analysis, market research, and object recognition [33]. Conversely, supervised machine learning technique trains a model on known input and output data so that it can predict future outputs [33]. Supervised learning typically uses two techniques, classification, and regression, to develop predictive models. Classification techniques predict categorical responses; while, regression techniques predict continuous responses [33]. Considering the dataset (ransom-ware) used in this research, which has input data only, the focus of this research is on the use of unsupervised machine learning technique for clustering and supervised machine learning for classification. The unsupervised machine learning clustering algorithm maps the overlapping clusters that are required for analysis in this study. The decision to use an unsupervised machine learning technique for cluster exploration is demonstrated in Figure 4 adapted from Scikit-Learn [34].

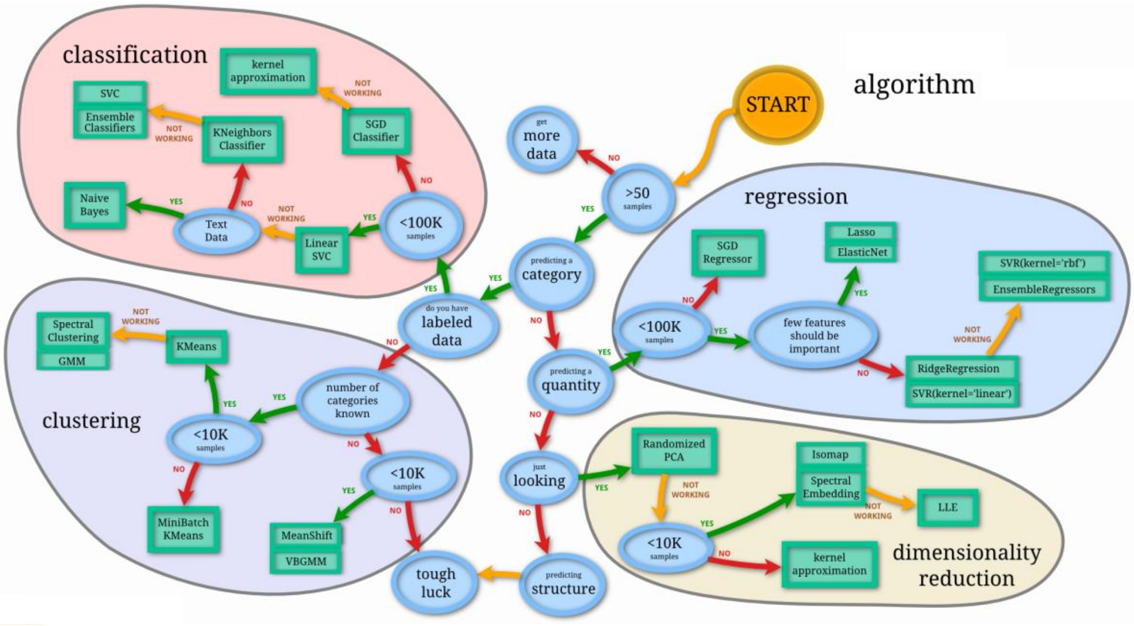


Figure 4: Process of Selecting Machine Learning Technique (Scikit-Learn)

Previous research has applied machine-learning techniques to clustering in complex social networks based on content relevance and link structures. Fuzzy Clustering Algorithm for Complex Networks was proposed in some of the studies [35]. Unlike the hard clustering algorithm (Exclusive Clustering) that allows objects to belong to one cluster, the Fuzzy clustering, Soft Clustering (Overlapping Clustering), allows objects to belong to more than one cluster with different degrees of membership [35,36]. The graphical representation of clusters is important to reveal very important topological information about sub-clusters or sub-graphs in a given network [37]. The graphical representation is needed to develop understanding of the complex nature of the data system. The main target for ransom-ware cluster analysis, like social networks, is to understand the topology, clustering formation and the dynamics of the network by visualizing the community structure of ransomware families, https://ransomwaretracker.abuse.ch/feeds/csv/ (Figure 5).

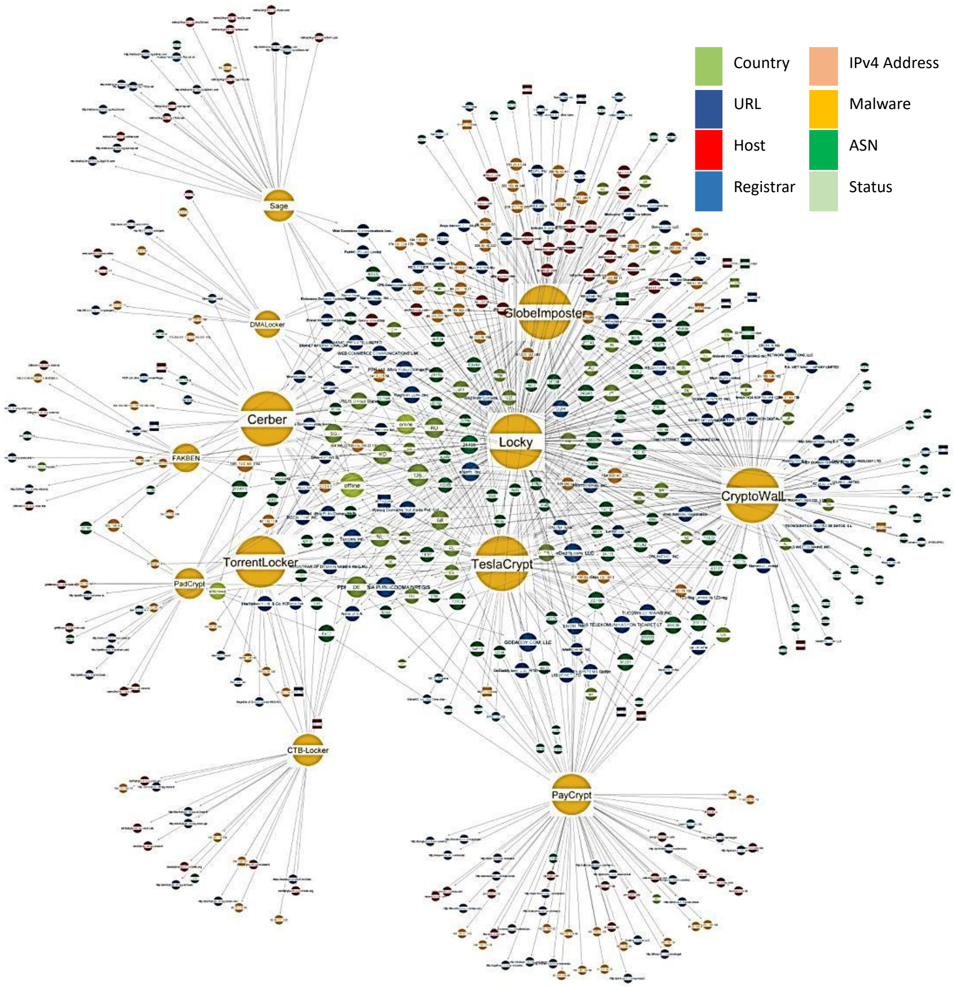


Figure 5: Ransomware Communities Network:
Showing cluster nodes of different ransomware families in the wild

Cluster visualization helps to visually identify hubs, authorities or overlaps and outliers in the network [37–39]. Therefore, it would be helpful to adopt a good method or algorithm to detect the modules (nodes) of a network connection with high sensitivity, accuracy, and reliability [40]. Considering the menace of malware, especially ransom-ware, research need to investigate the complex nature of the ransom-ware network in order to explore content relevance and link structures in the detection and termination of threats [1]. An understanding of the links and content relevance of the Network is vital to develop effective counter-measures to potential attacks as poor understanding would result to ineffectiveness in the control of threats. Therefore, the exploration of temporal events and formation of clusters and cluster overlaps gives an insight on the dynamics of ransomware threats. To achieve this objective, this investigation collects real-time ransom-ware data and passes this to machine learning algorithms for analysis. Clustering algorithm identifies overlaps within the clusters for the determination of links association and content relevance. The identified active overlap cluster node becomes the preferred node to remove to dislodge a potential ransomware threat.

1.1 The Motivation for this Investigation

Events Networks have existed from the very start of internet invention. Perhaps the prominence of Networks was not predominant then as it has been today, potentially because internet services were limited due to limited access to enabling technology. As technologies developed, internet activities gradually evolved into different visible communities and networks. The development of Social Networks in early 1990 does not negate the existence of malware and ransomware. Just like the Social Networks, it took several years for the malware and ransom-ware network communities to become prominent in the modern internet platform as a hub to distribute dangerous applications with the intent to steal private and valuable data from individuals and organizations. The leap in adoption of the internet, resulting from modern technology, has led to new paradigms in Social Network Communities. Following these paradigm shifts in Social Network Communities, efforts are made to understand and analyse the various dynamics and topologies of the communities and sub-communities. The structures of these social Network Communities, referred to as clusters and sub-clusters, and their links and contents, describe the activities of the networks. Links and contents explain both the social, economic, religious, political, and criminal cleavages of associating memberships. Many types of research have been conducted in the social network communities especially with the leap in the development of internet technology [41–45].

Unlike social networks, research in the field of complex ransom-ware networks is still developing. In Social Networks, people of different background have the freedom of association. This means they both have the freedom to join a social network community platform or to leave it. Such freedom does not apply to ransom-ware networks. Instead, the threats they pose are imposed on various categories of users. The implication is that each user will either spend resources to force them out from the ransom-ware attacking network or risk losing valuable data and money. However, with an increasing amount of ransom-ware network traffic (attacks), there is an urgent demand on effective and efficient approaches to handling large and dynamic ransom-ware connected networks. Therefore, an effort is made to seek machine-learning approaches that would learn patterns, behavioural characteristics or community associations to identify threats and take immediate actions against them. This research seeks to propose the real-time Exploratory and Machine-learning Cluster Overlapping System **(EMCOS)** to links and contents cluster analysis of the complex ransom-ware network. Thus, this investigation seeks to focus on finding the best ransom-ware network cluster overlaps and outliers to detect and terminate, in real-time, the threats of ransom-ware. To this end, clustering is one method to discover hidden values (knowledge) in a complex network, hence as a preliminary study of the ransom-ware network, partitioning algorithm such as k-means and fuzzy c-means, which are known for their dependence on distance measures are used to recognize clusters in the ransom-ware dataset [46]. Subsequently, random (decision) forest algorithm is used to build independent k-d trees of the forest to determine the similarity degree and overlap between sub-clusters, and clustering unordered sub-clusters thereby discarding singles elements [47].

1.2 Problem Statement

Cybercrime against different industrial sectors of the World economy has been identified to result in huge loss of revenue – estimated at an average of £266 billion per year. Ransom-ware contributes a large percentage of such losses. It is among the various malware designed purposely to gain access, steal valuable information or cause damage to the host device without the knowledge of the victim, subsequently demanding ransom from the host victims. Advancement in technologies makes it easier and more effective to transmit malware and infect devices more rapidly and consequently, to commit cybercrime. Transactions are done, mostly, online via digital devices thereby providing a ready conveyor through unprotected machines. Different sectors of the World economy make huge investments to secure their assets, yet they incur losses through malware attacks – especially those sectors with high-risk attack value. The worldwide attack of WannaCry ransom-ware of May 2017 corroborates the destructive nature of ransom-ware. In the United Kingdom, WannaCry attacked the systems of National Health Service (NHS) [48]. There is a widespread speculation that WannaCry was among the huge leak of NSA hacking tools [49]. The severity of the attack prompted Microsoft to release an emergency security update to counter WannaCry Ransom-ware [50]. The general level of impact of ransom-ware on different industries could depend on the lack of adequate security measures and lack of good knowledge of the trend and static and network properties of ransom-ware.

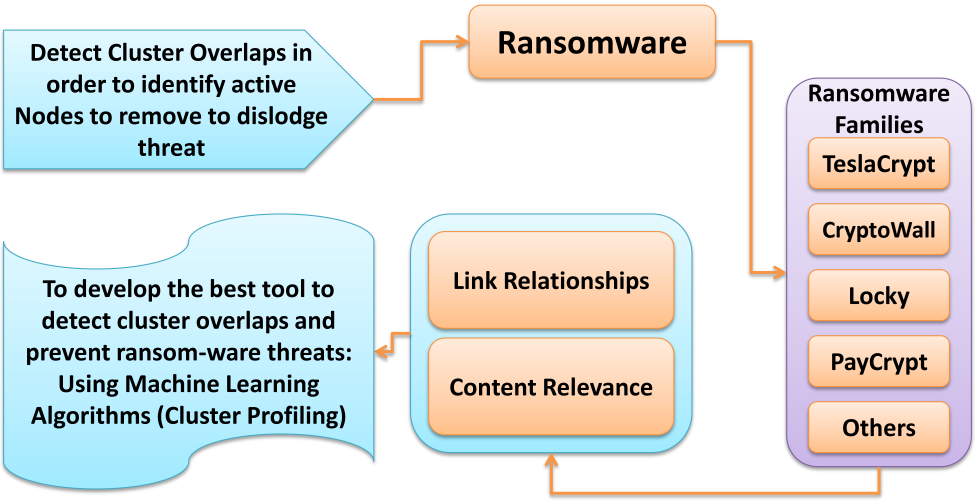
Previous researchers have conducted several analyses on malware properties, aiming to provide insight into the behaviour of different malware. The primary target of most research, among others, was to analyse the malware properties using the various features to create detection models. To develop an automatic malware detection model by training some classifiers based on the clusters, links, and contents of the ransom-ware network. This helps to overcome the problem of classification accuracy regarding unspecified malware detection when using signature-based analyses. However, these did not provide any analysis of the cluster, links, and contents among different families of ransom-ware in the network. Consequently, these previous works, among others, lack the frameworks to differentiate between different families of ransom-ware. They equally lack the capability to detect the best active cluster nodes to remove to dislodge a potential ransomware threat. In other words, they lack a tool for real-time detection and termination of potential ransom-ware threats.

The major problem statement of this investigation is to evaluate the possibility or otherwise of developing a system for real-time detection of overlaps and outliers in a dynamic network cluster. Such a system will help to detect the best network cluster nodes to remove, to dislodge a real-time potential ransom-ware threat. The development of such system will help forensic and security experts in the timely determination and control of ransom-ware threats. The results will help to inform the conclusions drawn from this investigation and its consequent recommendations. This research will choose to apply K-means, and Random Forest based clustering algorithm, which builds patterns of intrusion (cluster objects and outliers) over training data, thereafter matching the intrusion against network activities [43], to determine the degree of association of a node with the cluster that it belongs, to better identify clusters in complex networks. Consequently, this will help to identify overlapping clusters.

1.3 Research Aim and Objectives

The aim is to develop an understanding and knowledge of the formation of ransomware threat for a future development of a system to automate the detection and removal of active overlapping cluster node to control ransomware threat at the time of emergence, thereby efficiently and effectively reducing the impact of developing threats within a continuous and proactive monitoring approach. This will be achieved through the application of data science techniques to analyse evolving network cluster overlaps and pattern recognition to identify the threat profiles.

The objectives undertaken to achieve the aim are progressively structured and cumulatively address the aim (Figure 6) through progressively building understanding of the network clustering process and approaches to analysing and identifying the emergence of threats.



Detect Cluster Overlaps in order to identify active Nodes to remove to dislodge threat
Ransomware
Ransomware Families
TeslaCrypt
CryptoWall
Locky
PayCrypt
Others
Link Relationships
Content Relevance
To develop the best tool to detect cluster overlaps and prevent ransom-ware threats: Using Machine Learning Algorithms (Cluster Profiling)

Figure 6: Schematic Representation of Research Aims to detect cluster overlap using link relationships and content relevance in ransomware (families) network

1) To explore and identify the topology (structure) and Cluster (Nodes) of Ransomware Community (Family) Networks to develop an understanding of the patterns exhibited by developing ransomware attacks (threats).
2) To detect Ransomware Cluster Network overlaps (hubs) and outliers.
3) To explore the use of machine learning approaches to identify, detect and analyse the key parameters affecting cluster evolution in Ransomware Networks.
4) To compare different network key performance parameters to identify the active cluster overlap for the real-time detection and termination of Ransomware (Network) threats.
5) To identify the most impactful parameters (links and contents) to consider when analysing a Ransomware Network for the detection of Ransomware threat.
6) To make recommendations on the key characteristics of developing threats for future approaches to control ransomware threats.

The objectives are expected to help in gaining an understanding into the topology of the Network clusters (nodes), links (vertices) and contents of different consisting ransom-ware families with network topology, meaning the arrangement of the various elements such as links (vertices), nodes, and contents of the communication network. This step helps to gain an insight into the basic concepts behind Ransom-ware Networks and the formation and developments of clusters and cluster overlaps with the intention to identify key performance parameters, clustering objects intensities and active cluster overlaps. This is to detect the central communication hubs termed overlaps connecting the various sub-communities (clusters). Without the overlaps or association, the Ransom-ware community network has no connection and is bound to disintegrate, isolated or go into extinction. The detection of these overlaps would give insight on how forensic and security experts can use them to investigate and prevent Ransom-ware attacks.

The steps defined by these objectives seek to investigate the applicability of machine learning algorithms to analyse ransom-ware Networks based on the overlaps, outliers, content relevance and link structures for developing a platform suitable for the real-time detection and prevention of ransom-ware threats.

In these research steps, the focus is on the development of a dynamic machine-learning detection system that can analyse Ransom-ware Network for real-time detection and prevention of threat. The system explores the topological configuration of different families of ransomware network to map out the subgraphs (clusters). The determination of the key performance parameters of the clusters of the network becomes needful to detect the active cluster overlap node that originates and distributes the ransomware traffic to other cluster overlap nodes. The detection and removal of this active cluster overlap node helps to dislodge the ransomware threat and control attack. The tool is expected to be self-healing and runs in cycle.

1.4 Objectives of the Investigation

The achievement of the aim of this investigation depends largely on pursuing a concise and objective plan of action. The plan of action outlines and explains the procedures to execute to achieve the desired results to satisfy the set aims. The schematic of these procedures is presented in Figure 7 with further explanations in sections 1.4.1 – 1.4.3.

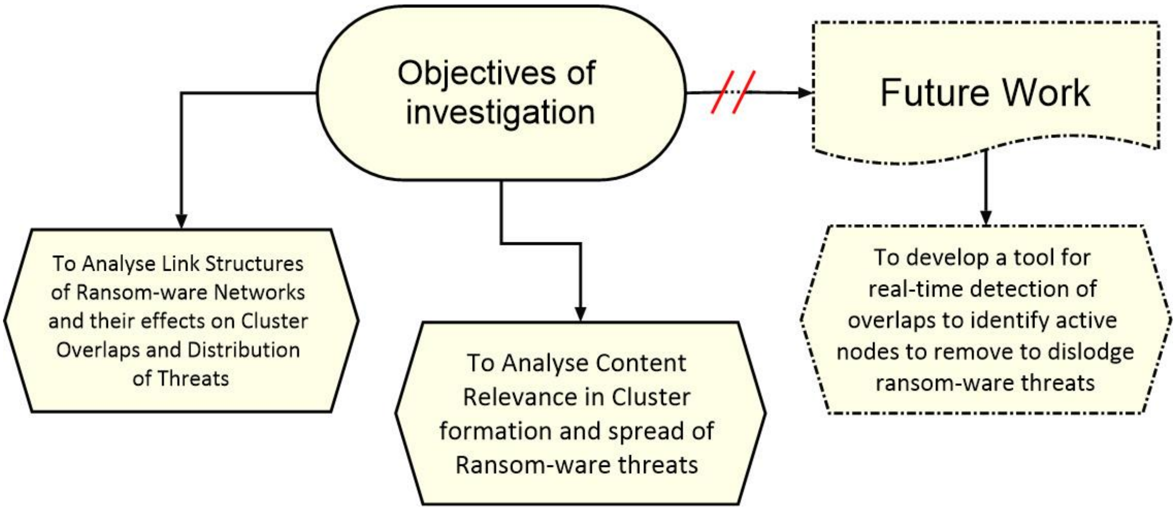

Objectives of investigation
Future Work
To Analyse Link Structures of Ransom-ware Networks and their effects on Cluster Overlaps and Distribution of Threats
To Analyse Content Relevance in Cluster formation and spread of Ransom-ware threats
To develop a tool for real-time detection of overlaps to identify active nodes to remove to dislodge ransom-ware threats

Figure 7: Schematic Representation of Research Objectives, namely to map network clusters (overlaps) using link distribution and content relevance to identify active cluster node, and develop a tool to remove the active cluster node

1. **To analyse link structures of ransomware networks and their effects on cluster overlaps and distribution of threats**

This objective seeks to analyse the link structures of ransomware Networks. The focus is to understand the communication mechanism between the link structures of different families of ransomware and the distribution of ransom-ware threats through the internet ecosystem. In addition, an understanding is sought on how these communication relationships affect the formation of different clusters and cluster overlaps. The understanding will help (to develop a system) to track ransom-ware threats and remove them before their actual attack.

2. **To analyse content relevance in cluster formation and spread of ransomware threat**

The knowledge of the degree of importance and relevance of the contribution of network contents in the topology and clustering of network community helps to understand the contents to give critical attention in developing a counter system to ransomware threats. In other words, there could be contents that are more consistent in the network clustering formation of overlaps. These contents could be identified by the degree of variable importance associated with them. Variable selection processes in machine learning algorithm can equally identify such contents.

3. **To develop an understanding for a real-time detection of overlaps to identify active nodes to remove to dislodge ransom-ware threats**

The end objective to achieve the set aim of this investigation is to develop a knowledge and understanding that could help in real-time to make quick decision to dislodge and undermine the ransomware threats even before a potential attack. The system will utilize the different information or clustering patterns and overlap behaviours learned from the exploration of temporal events and formation of cluster overlaps in a given ransomware network.

Summarily, the pursuit of these objectives is to use the understanding of the topological dynamics of the network to identify the key performance parameters in the formation and development of ransomware network and cluster overlaps, and their role in the distribution and spread of ransomware threats across the network devices of different sectors of the World Economy. Subsequently, the understanding is used to create a tool for management decision to prevent and control threats before they attack.

1.5 Research Contributions

Every investigation contributes to produce new knowledge or to improve upon an existing one. Therefore, the completion of this research is expected to produce the contributions depicted in Figure 8.

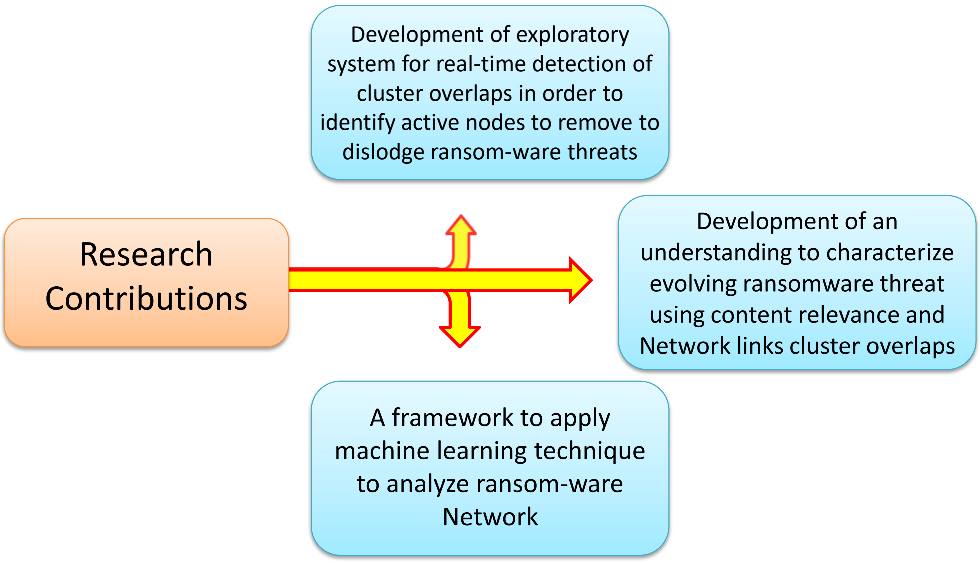

Development of exploratory system for real-time detection of cluster overlaps in order to identify active nodes to remove to dislodge ransom-ware threats
Research Contributions
Development of an understanding to characterize evolving ransomware threat using content relevance and Network links cluster overlaps
A framework to apply machine learning technique to analyze ransom-ware Network

Figure 8: Schematic Representation of Research Contributions showing the contributions of: a dynamic machine learning detection tool for the detection of threat; a tool suitable for implementation in social and other networks; and a platform that provides reference point for comparisons of future studies

The contributions of this current research investigation are:

a) The development of an understanding of machine-learning detection approach that can analyse Ransom-ware Network for real-time detection and removal of active overlaps nodes for the prevention of ransom-ware threats.

b) The successful application of the platform to ransom-ware makes it a contribution and suitable tool for use in social networks and networks in other sectors of the economy for the detection of cluster overlaps, criminal hubs, economic/operational links, and content clusters

c) The study creates recommendations on the impactful network key parameters to consider in the future development of tools for the control and prevention ransomware threats.

1.6 Research Limitations and Scope

This research currently applies to unidirectional network link where the structure allows a given node to receive an incoming connection (communication) but cannot connect (communicate) back to the same originating node. However, it can connect (communicate) to different nodes in the network. This means there is no inward and outward connection between two nodes. The data used in this research is limited to the feeds imported from https://ransomwaretracker.abuse.ch/feeds/csv/. Data from other sources have not been applied yet, but there is hope that other sources of data would be applied in the future.

However, the effectiveness of the practical application of any system that is developed based on this principle is beyond the scope of the present study because of social, ethical, and need for moral network, etc.; and is left for future studies to characterize. In addition, a more detailed understanding of how the ransomware events join a given cluster overlap node or leave it to form a new one would be the focus of a future study.

1.7 Structure of Thesis

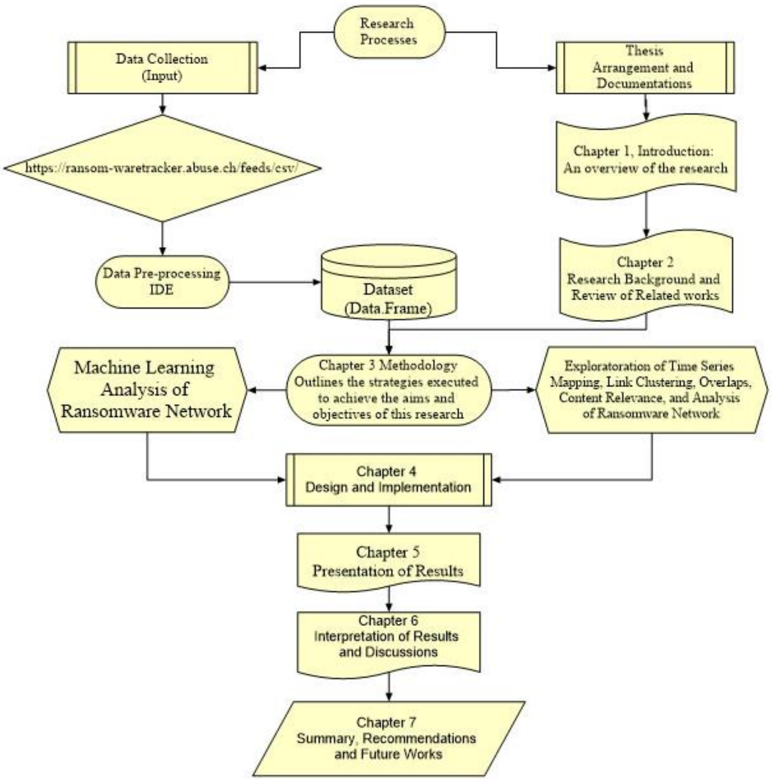

Research Processes
Data Collection (Input)
Thesis Arrangement and Documentations
https://ransom-waretracker.abuse.ch/feeds/csv/
Chapter 1, Introduction: An overview of the research
Data Pre-processing IDE
Dataset (Data.Frame)
Chapter 2 Research Background and Review of Related works
Machine Learning Analysis of Ransomware Network
Chapter 3 Methodology Outlines the strategies executed to achieve the aims and objectives of this research
Exploratoration of Time Series Mapping, Link Clustering, Overlaps, Content Relevance, and Analysis of Ransomware Network
Chapter 4 Design and Implementation
Chapter 5 Presentation of Results
Chapter 6 Interpretation of Results and Discussions
Chapter 7 Summary, Recommendations and Future Works

Figure 9: General Project Plan and Structure of Thesis: showing data input, pre-processing/dataset, and project arrangement and documentation merging into methodology; then machine learning of dataset and exploration of machine-learned network parameters merging into design and implementation and continuing unto summary, recommendations and future work

General Project Plan and Structure of Thesis are shown in Figure 9. The investigation introduces the research in Chapter 1 to describe the focus of the study. It gives a broad overview of Social Networks and Ransomware Networks, as well as the motivation of the study, statement of the problem, research aim and objectives, research contributions, and limitations. It highlights the prevalent level of Ransomware threat in different sectors of the global economy. Further to this, is differentiation between Ransom-ware Network and social Network vis-à-vis the links, contents, and topology of the two platforms. This highlights the differences and evaluates the prospects of using Ransom-ware Network features to control internet security threats.

Chapter 2 undertakes a historical review into the historical context of ransomware and gives an in-depth evaluation of existing works (literature review) on the emergence of ransomware threat and control; and compares existing tools, which allow users to (join and exit a given Social Network) detect and terminate ransom-ware attacks. In this chapter, these features will be defined to give an understanding of the parameters used by various tools in their Network Cluster graphs and analysis.

The methodology applied in the execution of the research aim and objectives are presented in Chapter 3, while the design and implementation of the project are presented in Chapter 4. Chapter 5 presents the results of the investigation. Chapter 6 deals with the analysis, interpretation, and discussions of the results. Chapter 7 is dedicated to summary, conclusions and recommendations and a preview of a possible future work (see Figure 9).

CHAPTER 2

REVIEW OF RELATED LITERATURE

2.0 Background of the Study

The May 2017 attack of WannaCry ransomware became the latest worldwide cyber-crime incident reported to have affected many key infrastructures and services. It spread spontaneously but was short-lived as a killer key was purportedly triggered accidentally by an IT professional [50]; fear of it still remains [51]. Figure 10 shows how different sectors of the economy are infected by ransomware for a short period, January 2015 – April 2016 [4]. The figure shows that the most infected sector is the services sector, while the least infected are the Mining, Agriculture, Forestry, and Fishing sectors [4].

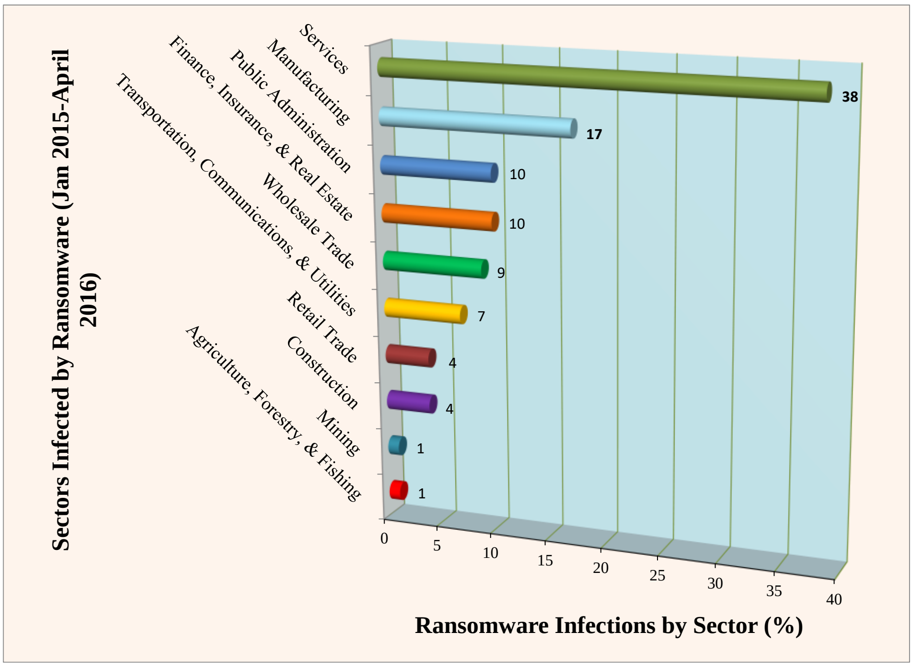


Figure 10: Infection of Ransomware by different Sectors of the Economy showing the services sector at the highest risk and infection level while the mining, agriculture, forestry, and fishing are least infected sectors

The returns to cyber criminals are great, and besides prosecution, the risks are low. It is estimated that the likely annual cost to the global economy from cybercrime is more than $400 billion. A conservative estimate would be $375 billion in losses, while the maximum could be as much as $575 billion [52–54]. Even the smallest of these figures is more than the national income of most countries. Unfortunately, governments and companies underestimate how much risk they face from cybercrime and how quickly this risk can grow [52,55–57]. This literature review will attempt to cover the classification techniques of malware and especially different ransomware families using their static and dynamic features in comparison with the features of “good” ware. In addition, there will be a review on the Performance of Android Forensics Data Recovery Tools *(Published Book Chapter, 2016)* [58]*;* to gain an understanding of works already done on the subject of classification techniques as a tool for determining the trend of ransomware attacks.

Generally, the growing menace of cyber-crime is facilitated through particular media. Therefore, smart mobile devices, particularly smartphones, are seen to be increasingly popular in today's Internet-connected society and present the potentials for the spread of cybercrime [59–61]. In 2010, shipments of smartphones grew by 74% to 295 million units [61]. Unsurprisingly, sales of smartphones have been increasing since [62,63], and it has been estimated that 1.5 billion smartphones will be sold by 2017 with 1 billion mobile subscribers by 2022 [64–72]. Such devices are generally used to make phone calls, send SMS messages, web browsing, locate places of interests, map navigation, image and video capture, entertainment (e.g., gaming and lifestyle), business and economic transactions (e.g., internet banking), take notes, create and view documents, etc. [63,73–75]. Due to their widespread adoption in corporate businesses, these devices are a rich source of information (e.g., corporate data and intellectual property) [76–79]. The potential to target such devices for criminal activities (e.g., malware such as banking Trojans) or to be used as an attack launch pad (e.g. used to gain unauthorized access to corporate data) [80–87], makes it important to ensure there is the capability to conduct a thorough investigation of such devices [75,79,88–93].

While there are a small number of forensic tools that can be used in the forensic investigation of smart mobile devices [94], the extent to which data can be recovered varies, particularly given the wide range of mobile devices and the constant evolution of mobile operating systems and hardware [95,96]. For example, recovering data from the internal memory of a smartphone remains a challenge. Further to these challenges is the requirement to create forensically sound and effective tools and procedures [77,92,94]. Therefore, it is essential that the forensic community keeps pace with forensic solutions for smart mobile devices [60,61,97,98]. This is the focus of this chapter. Specifically, the effectiveness of five popular mobile forensics tools is studied, namely: Phone Image Carver, AccessData FTK (Forensic Tools Kit), Foremost, Recover My Files, and DiskDigger, in recovering evidential data from a factory-restored Samsung Galaxy Note 3 running Android Jelly Bean version 4.3. The need to conduct a literature survey of Regression Techniques for (Malicious) software prediction is more urgent now to give security managers an insight into the future trend of malware attacks. The emergence of malicious software (malware) as an attack and valuable-information-stealing tool has added new dimensions to the subject of digital cybercrime. The menace of malware was not given a thought when Charles Babbage invented the first computing machine in 1832, the consequent arrival of modern computers in the 1950s [99] and subsequently the development of other media devices. The software has been developed since the 1960 but the emergence and spread of malware through the internet was never imagined when the internet came to life in 1969 [100,101]. However, computer and media device users became aware of malware through the widespread infections caused by Melissa and LoveLetter in 1999 and 2000 respectively [102]. From the sudden attack of Melissa and LoveLetter, different types of other malware have taken turns to cause economic havoc in the World of internet computing, media communication, and business transactions. Over the years, different sectors of the economy have experienced different types of malware attacks, which make this survey imperative.

The urgency to develop a malware prediction and control model comes from the increase in targeted malware attacks, especially with the advancement of technology. Most industrial sectors of the world economy are more at risk for these targeted attacks than other sectors because of the value of the resources they hold in digital form. Hence, these high-risk industrial sectors suffer huge economic losses. In targeted attacks, there is evidence that the attacker has specifically selected the recipients of the attack. It might be that the attacker suspects that the intended victim holds or has access to high-value information that the attacker wishes to compromise, or the compromised systems can be used to launch further attacks against other high-value systems or individuals that could be of economic value [103]. Therefore, modelling the trend and economic impact control of malicious software (malware) attacks in the industries demands immediate attention. Malware developers are exploiting the opportunity of the advancement in technology to spread malware attacks on their victims. Malware is found to be responsible for attacks affecting computer users. These are very dangerous as they take advantage of security vulnerabilities, errors in applied programming interfaces (API), memory corruption-based vulnerabilities to execute on the host device [104]. The growing incidence of malware attacks has become a great challenge to different sectors of the world economy. The sectors are affected by malware threats at different risk levels ranging from high to low risk levels depending on the value of the resource they hold. The immediate effect of Malware threats is the breach of Confidentiality, Integrity, and Availability of private information. Such breaches consequently translate to real economic costs within the various sectors that are subjects of attack. Malware threats thrive most when there are transmitting medium. The advancement in internet technology provides this medium and brings with it an increasing rise in digital crimes. An estimated annual cost of cybercrime on the global Economy is £266 billion ($445 billion) [105–107]. Stealing of individual's personal information and cyber espionage affected more than 800 million people in 2013 [52,108]. In estimating the consequent global financial cost of cybercrime, these losses could result in the loss of an estimated 150,000 jobs in Europe [52]. In this context, the race to protect valuable and sensitive information, overcome the digital crimes and cyber criminals and possibly prosecute them pose very great research challenges. Information security and media forensics as an emerging research area not only seek to provide security for valuable and sensitive data but also applies computer investigation and analysis techniques to detect these crimes and gather digital evidence that could be used for prosecution in courts. However, the use of the internet and other information technologies had grown rapidly all over the world in the 21st century. Directly correlated to this growth is the increased momentum of criminal activities involving digital crimes or e-crimes worldwide [109]. These digital crimes impose new challenges on the confidentiality, integrity, and availability of personal information, hence increasing the task of prevention, detection, investigation, and prosecution of corresponding breaches.

2.1 Malware Evolution and Timeline

The evolution of malware reviewed in the context of the background study of this project is captured in a Ransomware Timeline (Figure 11). The Ransomware Timeline covered the malware evolutionary period from 1970 – 2008; the evolutionary timeline for the period 1980 – 2017 is presented in Appendix 1.

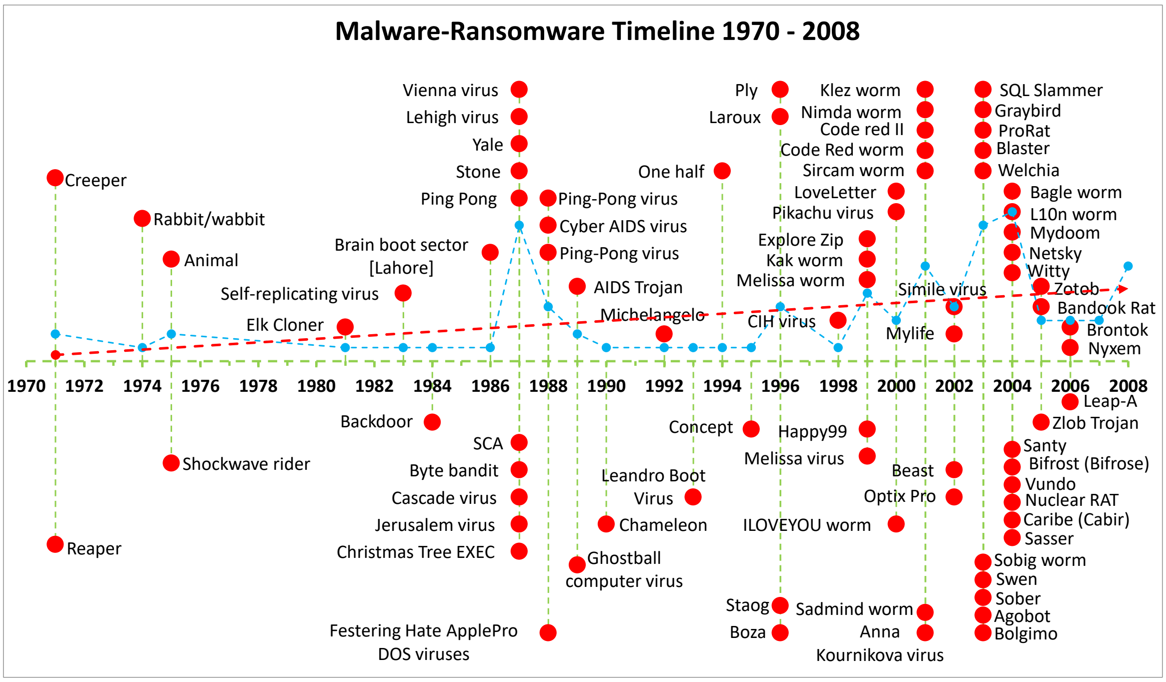


Figure 11: Ransomware Timeline showing the names of malware against their dates of release, the dotted blue curve shows the annual frequency of malware occurrence, while the dotted red line is the trend showing an increasing number for the period 1970 - 2008

The subject of malware attack was not known until computer and media device users became aware of malware through the widespread infections caused by Melissa virus and LoveLetter worms in 1999 and 2000 respectively [102]. The sudden attack of Melissa in Aberdeen Township, New Jersey, and LoveLetter in the Philippines resulted in different types of other malware taking turns to cause economic havoc in the World of internet computing, media communication and transactions [90,110,111]. Prior to the Melissa and LoveLetter attacks, the “Theory of self-reproducing automata” by John von Neumann [112] was tested in 1971 using the self-replicating Creeper system malware written by Bob Thomas at BBN Technologies [112,113]. Creeper infected DEC PDP-10 computers running the TENEX operating system. Creeper gained access via the ARPANET and copied itself to the remote system with the message "I'm the creeper, catch me if you can!" [114]. Creeper was an experimental program, and the Reaper malware program was later (1971) created to delete Creeper [115,116]. The rabbit (or wabbit) virus (created in 1974) was known to spread by creating multiple copies of itself [90,115]. The next malware recorded in 1975 included the ANIMAL, the first Trojan, and the shockwave rider. While the program PERVADE acted as a catalyst to create a copy of itself and ANIMAL in every directory in the host system [90,117–119], Shockwave Rider spread through a network of computers [120–127]. Elk Cloner was a boot sector virus that appeared in 1981 and Elk Cloner caused the first large-scale outbreak of computer virus in history [90,128]. Shortly after Elk Cloner was the creation of the self-replicating computer virus in 1983. The virus spread to other programs by the activity of modifying them to include an evolved copy of itself [90,129]. The evolution of malware appears to happen in quick succession with the backdoor introduced in 1984. The backdoor attacks the host system by modifying the C-compiler to embed into the login command, thus, the backdoor insertion code is activated when the C-compiler is eventually used to compile itself. This happens even if neither the backdoor nor the backdoor insertion code was present in the source code [113,130–135]. Later, was the introduction of Brain boot sector (virus) [aka Lahore, Pakistani, Pakistani Brain, and Pakistani flu] in 1986. It was created in Lahore Pakistan and became the first IBM PC compatible virus that was responsible for the first IBM PC compatible virus epidemic [90,136,137]. In 1987, the Vienna virus and Lehigh virus were introduced. The Vienna virus attack was on the IBM platform [138], while the Lehigh virus (boot-sector virus) did not spread because it was discovered and stopped at Lehigh University [138]. Also in the same 1987, the viruses, Yale, Stone, Ping Pong, SCA and Byte Bandit, (boot-sector virus) were introduced [139]. The cascade was the first encrypting file virus and the Jerusalem virus that attacked and destroyed all executable files on infected machines [139]**.** In the same year, the Christmas Tree EXEC, a computer worm disguised as a benign holiday greeting was introduced. It spread rapidly via email and clogged up networks worldwide [140]**.** In 1988, the Ping-pong virus, an MS-DOS boot sector virus, was introduced [141]. In addition, the Cyber AIDS and Festering Hate ApplePro DOS viruses were introduced in the same year. These viruses spread from underground pirate BBS systems and started infecting mainstream networks [141]. Furthermore, the Morris worm was introduced, and this was an Infected DEC VAX and Sun machines running BSD UNIX__that are connected to the Internet. It became the first worm to spread extensively "in the wild", and one of the first well-known programs exploiting buffer-overrun vulnerabilities [141–146].

In 1989, Ghostball Computer virus was introduced. It was the first multipartite virus designed to infect both the executable and COM-files and boot sectors on MS-DOS systems. It captured specific information entered or saved by the user, with the corresponding threat to privacy and caused the loss of information stored on the computer on either specific files or data in general. It affected the productivity of the computer and the network to which it was connected or other remote sites. It decreased the security level of the computer but did not automatically spread itself [147,148]. In addition, AIDS Trojan (ransomware) was introduced; this virus was mailed to subscribers of PC Business World magazine and a WHO AIDS conference mailing list. This DOS Trojan lay dormant for 90 boot cycles, and then encrypts all filenames on the system, displaying a notice asking for $189 to be sent to a post office box in Panama in order to receive a decryption program [138]. In 1990, the polymorphic virus (chameleon family) was launched. It remained in the wild for almost 20 years and reappeared afterward; during the 1990s it tended to be the most common virus in the wild with 20 to more than 50 percent of reported infections [134,137,149,150].

In 1992 was the introduction of Michelangelo virus; the virus was designed to infect DOS system but did not engage the operating system or make any OS calls. However, it operated at the BIOS level. It was rumoured the virus would wipe out information from millions of computers. However, the effect of the virus was seen to be minimal [151]. 1993, marked the launch of Leandro Boot Virus (Leandro & Kelly). This boot virus emerged in pirate MS-DOS operating system software. The virus infected the boot when it was turned on and would infect any disc inserted into the PC. It spread quickly due to the popularity of BBS and shareware distribution [113]. In 1994 and 1995, One Half and Concept were introduced. The former was DOD-based polymorphic computer virus while the later was the first Macro virus that attacked the French version of Microsoft Word document [98]. However, in 1996 the Ply, a DOS 16-bit based complicated polymorphic virus appeared with built-in permutation engine. Furthermore, Boza and Laroux were introduced. Boza was the first virus designed specifically for Windows 95 files to arrive while Laroux was the first wild-spread macro virus to infect Excel spreadsheets to appear [98]. Staog, the first Linux virus that attacked Linux machine was introduced [114,115].

In 1998, CIH virus was formed; it was the first known virus able to erase flash ROM BIOS content [116]. Whereas Happy99 worm was introduced in 1999, it invisibly attached itself to emails and displays fireworks to hide the changes being made, ultimately wishing the user a happy New Year. It modified system files related to Outlook Express and Internet Explorer (IE) on Windows 95 and Windows 98 [98]. In the same year also, Melissa worm Targeting Microsoft Word and Outlook-based systems, and creating considerable network traffic was introduced [117,152,153]. Similarly, Explore Zip worm and Kak worm were introduced; the former destroyed Microsoft Office documents while the latter was a JavaScript computer worm that spread by exploiting a bug in Outlook Express [118]. In

2000, ILOVEYOU worm (Love Letter, or VBS, or Love Bug worm) was introduced. It was a computer worm believed to be created by a Filipino computer science student. Written in VBScript, Love Letter infected millions of Windows computers worldwide within a few hours of its release. Using social engineering techniques, it was considered as one of the most damaging worms ever [34,119]. In the same way, Pikachu virus was introduced. This virus was believed to be the first computer virus targeting children. It contained the character "Pikachu" from the Pokémon series and distributed in the form of an e-mail titled "Pikachu Pokemon" with the message: "Pikachu is your friend." The attachment to the email has "an image of a pensive Pikachu", along with a message stating, "Between millions of people around the world I found you. Don’t forget to remember this day every time MY FRIEND." Along with the image, there was a program, written in Visual Basic 6, called “pikachupokemon.exe” that modified the AUTOEXEC.BAT file and added a command for removing the contents of directories C:\Windows and C:\Windows\System at computer's restart. However, a message would pop up during startup, asking the user if they would like to delete the contents. The affected operating systems were Windows 95, Windows 98 and Windows Me [120]. In 2001, the Anna Kournikova virus was created. This virus hit e-mail servers hard by sending an e-mail to contacts in the Microsoft Outlook address book [121]. It was also the year that Sadmind worm and Sircam worm were created. The former spread by exploiting holes in both Sun Solaris and Microsoft IIS [154–156], whereas the later spread through Microsoft systems via e-mail and unprotected network shares [134,137,157].

Code Red worm and Code red II (Code Red worm) started also in 2001 [158,159]. The Code Red worm was known to attack the Index Server ISAPI Extension in Microsoft Internet Information Services [126,127]. On the other hand, code Red II began aggressively to spread onto Microsoft systems, primarily in China [127–129]. In the same year, Nimda worm spread through a variety of means including vulnerabilities in Microsoft Windows and backdoors left by Code Red II and Sad-mind worm [128,130] while Klez worm exploited a vulnerability in Microsoft Internet Explorer and Microsoft Outlook and Outlook Express [137]. In 2002 Simile virus and Beast were introduced; Simile virus was a metamorphic computer virus written in assembly [137] while Beast was a Windows-based backdoor Trojan horse, more commonly known as a RAT (Remote Administration Tool). It was capable of infecting almost all versions of Windows [137]. Mylife and Optix Pro were equally introduced in 2002; Mylife was a Windows-based backdoor Trojan horse, more commonly known as a RAT (Remote Administration Tool). It was capable of infecting almost all versions of Windows [160,161]. Optix Pro was a configurable remote access tool or Trojan, similar to Sub Seven or BO2K [162,163].

In 2003, SQL Slammer worm was introduced; it was known as Sapphire worm, Helkern and other names, attacked vulnerabilities in Microsoft SQL Server and MSDE becomes the fastest spreading worm of all time (measured by doubling time at the peak rate of growth), crashing the Internet within 15 minutes of release [164]. Graybird was a Trojan horse also known as Backdoor. Graybird was also introduced in the same year [165,166]. ProRat and Blaster worm was introduced in the same year. However, ProRat was a Turkish-made Microsoft Windows-based backdoor Trojan horse, more commonly known as a RAT (Remote Administration Tool) [167], while Blast worm (Lovesan worm), rapidly spread by exploiting a vulnerability in system services present on Windows computers [137].
Similarly, Welchia (Nachi) worm was introduced in 2003 [168]; the worm tried to remove the Blaster worm [169] and patch Windows. Sobig worm (technically the Sobig.F worm) was also introduced in the same year as the Welchia. The worm spread rapidly through Microsoft systems via mail and network shares [170]. Others include; Swen, a computer

worm written in C++ [171] and Sober worm [172], which was first seen on Microsoft systems and maintained its presence until 2005 with many new variants. The simultaneous attacks on network weak points by the Blaster and Sobig worms caused massive damage. The Agobot and Bolgimo were also introduced in the same year [137]. The Agobot was a computer worm that can spread itself by exploiting vulnerabilities on Microsoft Windows. Some of the vulnerabilities are MS03-026 and MS05-039. Conversely, Bolgimo was a computer worm that spread itself by exploiting a buffer overflow vulnerability at Microsoft Windows DCOM RPC Interface [137].

In 2004 was the introduction of the Bagle worm and L10n worm (Lion). Bagle worm a mass-mailing worm affecting all versions of Microsoft Windows. There were two variants of Bagle worm, Bagle.A and Bagle.B. The Bagle worm was discovered on February 17, 2004 [173]. Moreover, L10n worm a Linux worm that spread by exploiting a buffer overflow in the BIND DNS server. It was based on an earlier worm known as the Ramen worm (commonly, albeit incorrectly referred to as the Ramen Virus) which was written to target systems running versions 6.2 and 7.0 of the Red Hat Linux distribution [174]. In addition, Mydoom and Netsky worm were discovered. MyDoom worm currently holds the record for the fastest-spreading mass mailer worm [175], while Netsky worm, was a worm that spread by email and by copying itself to folders on the local hard drive as well as on mapped network drives if available. Many variants of the Netsky worm appeared [176]. In the same year, the Witty and Sasser was discovered. The Witty worm was a record-breaking worm in many regards. It exploited holes in several Internet Security Systems (ISS) products. It was the first internet worm to carry a destructive payload and it spread rapidly using a pre-populated list of ground-zero hosts [176]. Sasser merges by exploiting a vulnerability in the Microsoft Windows LSASS service and causes problems in networks, removing MyDoom and Bagle variants, even interrupting business [177]. In addition, Caribe (Cabir), a computer worm designed to infect mobile phones that run Symbian OS was introduced. Caribe was the first computer worm that can infect mobile phones. It spread itself through Bluetooth [178]. Equally, Nuclear RAT (Nuclear Remote Administration Tool) was also discovered. This backdoor Trojan infected Windows NT family systems (Windows 2000, Windows XP, and Windows 2003) [179]. Still, in 2004, Vundo (Vundo Trojan, Virtumonde or Virtumondo, MS Juan) and Bifrost (Bifrose) were introduced. This was a Trojan known to cause popups and advertising for rogue antispyware programs, and sporadically other misbehaviour including performance degradation and denial of service with some websites including Google and Facebook [180]. Bifrost (Bifrose) was a backdoor Trojan that can infect Windows 95 through Vista. Bifrost uses the typical server, server builder, and client backdoor program configuration to allow a remote attack [179]. Finally, in 2004 Santy (webworm) was introduced. This was the first known "webworm" to be launched. It exploited the vulnerability in phpBB and used Google to find new targets. It infected around 40000 sites before Google filtered the search query used by the worm, preventing it from spreading [181].

Furthermore, in 2005 Zotob [182] the copy protection rootkit was launched; this was a rootkit deliberately and surreptitiously included on music CDs sold by Sony BMG and exposed. The rootkit creates vulnerabilities on affected computers, making them susceptible to infection by worms and viruses [183]. Also Zlob Trojan; This was a Trojan horse program masquerading as a required video codec in the form of the Microsoft Windows ActiveX component and was first detected in late 2005 [167]. Subsequently, Bandook or Bandook Rat (Bandook Remote Administration Tool) was also detected in 2005. It was a backdoor Trojan horse, which infected the Windows family. It used a server creator, a client, and a server to take control over the remote computer. It used process hijacking/kernel patching to bypass the firewall, and let the server component hijack processes and gain rights for accessing the Internet [167].

In 2006, Nyxem worm was detected. Nyxem spread by mass mailing. Its payload, which activates on the third of every month, starting on February 3, attempts to disable security-related and file sharing software, and destroy files of certain types, such as Microsoft Office files [173]. However, in the same year, OSX/Leap-A or OSX/Oompa-A and Brontok variant N were detected. The former was the first-ever malware for Mac OS X, a low-threat Trojan-horse, while the later was a mass-email worm and the origin for the worm was from Indonesia.

2.2 Economic Impact of Malware Attacks

The emergence of malicious software (malware) as an attack and valuable-information-stealing tool added new dimensions to the subject of digital cybercrime. Over the years, different sectors of the economy have experienced different types of malware attacks. Some sectors are subject to more attacks than others are in comparison. In 2013, the Internet Security Threat Trends identified the risk level of malware attacks in some sectors as follows: Mining sector (risk ratio: 1 in 2.7), Public Administration (Government) (risk ratio: 1 in 3.1), and Manufacturing (risk ratio: 1 in 3.2) are considered high risk. Considered to be Medium risk sectors include Wholesale (risk ratio: of 1 in 3.4), Transportation, Communications, Electric, Gas & Sanitary Services (risk ratio: 1 in 3.9), Finance, Insurance & Real Estate (risk ratio: 1 in 4.8). Low risk sectors include: Services – Non-traditional (risk ratio: 1 in 6.6), Construction (risk ratio: 1 in 11.3), and Agriculture, Forestry & Fishing (risk ratio: 1 in 12.0) [184]. The risk ratios for these different economic sectors are presented in Figure 12 and Figure 13. The values of the risk ratios suggest that Mining, Public Administration, and Manufacturing, in that order, are more at the risk of malware attacks.

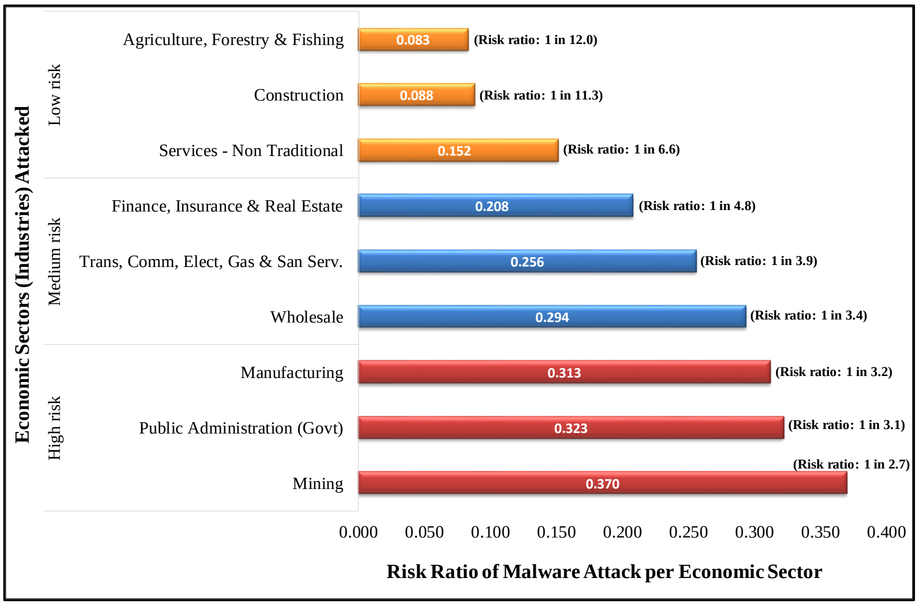


Figure 12: Risk Ratio of Malware Attack per (Industry) Economic Sector (Semantec 2014)

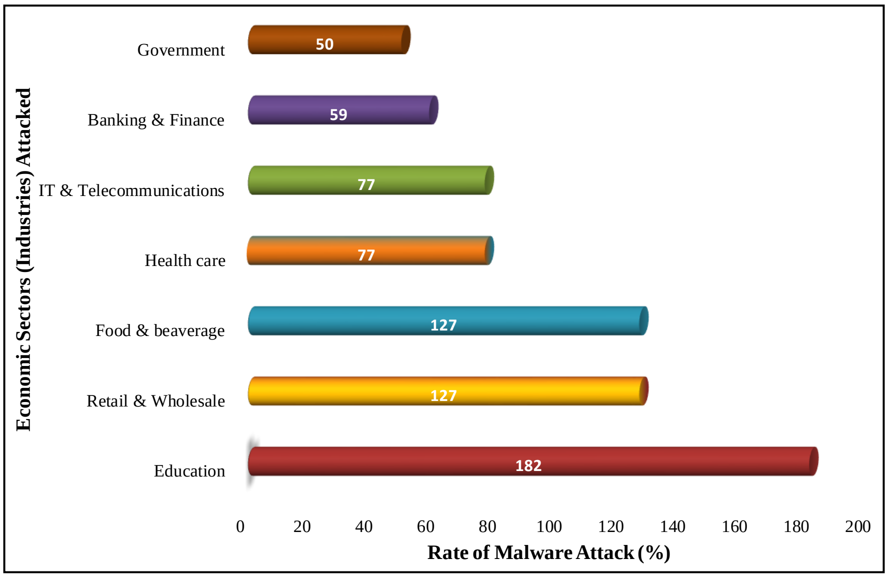
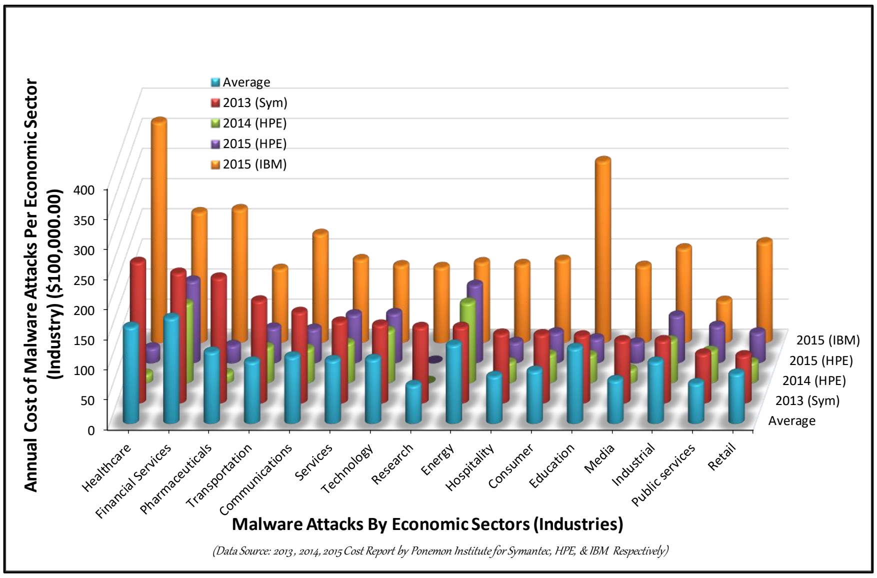


Figure 13: Rate of Malware Attack (%) per Economic Sector (Industry) (ContentKeeper 2013)

To investigate the sectoral economic impacts, research was conducted on the Global cost of cybercrime. In 2013, a study was conducted on the Global Analysis of Cost of Data Breach [55]. In 2014 and 2015, similar studies were conducted on the Global Cost of Cyber Crime [185], (PonemonInstitute 2015). Further study was conducted in 2015 on the Global Analysis of Cost of Data Breach [187]. These studies revealed variations and how each industry is impacted by malware attacks as shown in Figure 14 and Figure 15. While Figure

14 shows the average and yearly costs of attacks to industries, Figure 15 shows the contributions of the different types of Cybercrime to these costs.



(Data Source: 2013, 2014, 2015 Cost Report by Ponemon Institute for Symantec, HPE, & IBM Respectively)

Figure 14: Annual economic cost of malware attacks on Sectors of the World economy: showing the cost for the various sectors for 2013-2015 and their collective average cost for the same period

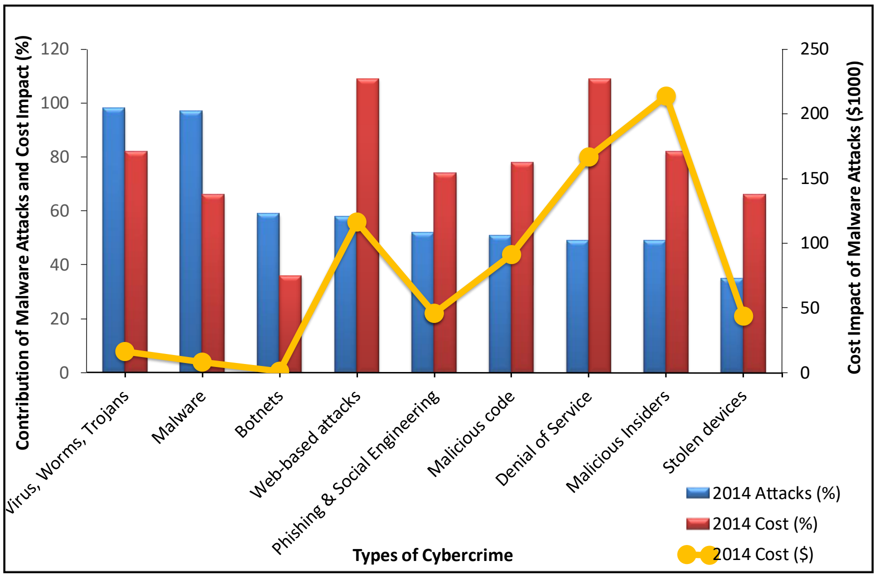


Figure 15: Cybercrime attacks & cost (%) and cost of attacks ($1000.00): x-axis is the different malwares, primary y-axis is the annual %-cost contribution of each malware, secondary y-axis is annual cost impact of each malware in dollars

2.3 How Do Industries Control Ransom-ware threats

The urgency to prevent ransomware attacks has resulted in many types of research on signature-based and behaviour-based methods. Signature-based methods rely mainly on the identification of known malware. It is easy for algorithms to scan and determine the digital signature of objects because, in computing, every object has attributes that are used to create a unique signature. Objects identified as malicious are added to repositories making it easy to use for future matching. However, the appearance of new versions of malicious codes that evade the recognition of signature-based technologies has rendered the signature-based techniques ineffective for the control of malware. In other words, signature-based techniques can only track known malware. To track and detect those malicious codes that evade signature-based technologies, a behavioural analysis approach needs to be adopted. In behaviour-based approach, objects are evaluated based on their intended actions before they execute the behaviour. The behavior-based approach, though evolving, still has limitations. This is the ability of a malware to obfuscate and avoid detection by attempting to curtail malicious activities [188,189]. In the effort to overcome malware threat, andronomy was introduced as a framework to detect malware in Android mobile devices. Andronomy, as a Host-based malware detection system, monitors various features and events on the host device and applies Machine Learning to classify collected data as benign or malicious [190].

2.4 Clustering Overlap and Machine Learning Approach to the Control of Ransomware

Clustering overlap and machine learning techniques have been applied, severally, to analyse different datasets including social network datasets and others. Traditional clustering algorithms such as k-means have the limitation of producing disjoint and exhaustive clustering. This means that data points are assigned to one cluster ignoring there could be overlaps and outliers. Consequently, the Fuzzy k-means algorithm was proposed for overlapping clustering [191,192]. However, the Non-Exhaustive Overlapping K-means (NEO-K-means) was introduced to tackle the problem of overlapping clustering and outliers in a unified way, which was suggested to prove effective to detect overlapping community in a dataset [191]. Attempting to overcome time series clustering problems associated with distance measure for efficient clustering of the dataset with input-output time series, an extension of Martin cepstral distance was proposed. The extension allows the efficient clustering of these time series and applies it to the simulated dataset [193].

Different supervised and unsupervised machine learning approaches were used to investigate overlapping clusters in ransomware network dataset. The **HP Neural network** creates multilayer neural networks that pass information from one layer to the next to map an input to a specific category or predicted value. The HP Neural node enables this mapping to take place in a distributed computing environment, which enables you to build neural networks on massive data sets in a relatively short amount of time [194]. A neural network consists of units (neurons) and connections between those units. There are three types of units namely the Input Units, Hidden Units and Output Units. The Input Units obtain the values of input variables and standardize those values; they can be connected to hidden units or to output units. Hidden Units perform internal computations and provide the nonlinearity that makes neural networks powerful; and they can be connected to other hidden units or to output units. Output Units compute predicted values and compare those values with the values of the target variables; and they cannot be connected to other units [194]. However, the HP Neural network failed to create a distinct visualization of overlapping clusters that is required in the current investigation. **Hierarchical** clustering method (Ward’s clustering) and a non-hierarchical clustering method (k-Means clustering) were used to visualize and analyze overlapping clusters in a ransomware network dataset to identify the active cluster overlap in an online network environment [195]. The hierarchical and non-hierarchical clustering showed cluster overlaps, but the visualization was very fuzzy and did not show distinct classification (groupings) required to identify the active cluster overlap. In addition, non-hierarchical k-Means is applicable only if the mean is defined for categorical data, and it is sensitive to outliers and requires the specification of the number of clusters (k) by the user.

Neural Networks are statistical learning models which are modeled based on the information processing procedure found in the brain [196]. Neural Networks have evolved from modeling simple problems to a wide variety of complex problems and this has been fueled by the availability of computation ability and novel algorithms. Neural networks, just like the brain can solve complex problems such as image recognition, speech processing, and natural language processing [196]. The artificial equivalents of biological neurons and synapses are the nodes and weights respectively [196]. Several different types of Neural Networks exist based on their application, including a simple three-layer feed forward backward propagation network as it is a popular multilayer perceptron used to model nonlinear data for prediction and classification tasks [196–198]**.** While Neural Networks is good for classification, it lacked the clustering and visualization advantage of presenting the overlapping clusters in a network dataset that the present investigation seeks to explore to profile and identify the active cluster overlap in a ransomware network. This research investigated the use of **Random Forests** to identify overlapping clusters in a ransomware network. Random Forests constructs many classification trees and thus the name 'forest'. For each pathway, the input data for Random Forest would be a ransomware network variables expression matrix of the ransomware network variables belonging to the pathway by the number of subjects in the dataset. Every tree in a Random Forests is built using a deterministic algorithm and the trees are different from the ordinary tree algorithms (e.g. CART) owing to two factors. First, at each node, a best split is chosen from a random subset of the predictors rather than all of them. Second, every tree is built using a bootstrap sample of the original observations. A subject is put down a tree for classification using the input vector of ransomware network variable expression for ransomware network variables within a pathway. The tree gives a classification and decides which class this subject belongs to. In the end, the forests choose the class that gives the majority votes for each subject. The out-of-bag (OOB) data, approximately one-third of the observations, are then used to estimate the prediction accuracy. Small classification error based on ransomware network variable in a given pathway would indicate the pathway as potentially interesting [199,200]. However, **Random Forests** did not produce an identifiable visual mapping of distinct overlapping clusters in the network community. The use of machine learning algorithms to extract valuable information from the wild exposes the algorithms to the threat of data poisoning; a situation of a coordinate attack in which a fraction of the training data is controlled by the attacker and manipulated to subvert the learning process [201]. As a counter measure to data poisoning, a back-gradient optimization algorithm was proposed to compute radiant of interest through automatic differentiation and the reversal of the learning procedure to drastically reduce the attack complexity [201]. The **Software-Defined Networking (SDN)** based detection approach utilizes the characteristics of ransomware communication, namely the HTTP messages' sequences and their respective content sizes [202,203]. However, the preventive and reactive security measures can only partially mitigate the damage caused by modern ransomware attacks [204]. Pure-detection approaches (e.g., based on analysis sandboxes or pipelines) are not sufficient nowadays, because often we do not have the luxury of being able to isolate a sample to analyze, and when this happens it is already too late for many users [204]. A forward-looking solution is to equip modern operating systems with practical self-healing capabilities against this serious threat. Therefore, ShieldFS, an add-on driver that makes the Windows native filesystem immune to ransomware attacks was proposed. For each running process, ShieldFS dynamically toggles a protection layer that acts as a copy-on-write mechanism, according to the outcome of its detection component. Internally, ShieldFS monitors the low-level filesystem activity to update a set of adaptive models that profile the system activity over time [204]. Unfortunately, **ShieldFS** is not cluster based and lacks the capability to profile overlapping clusters. The proactive approach, known as **Cyber Defense** focuses on finding out methods to prevent threat incidents from occurring rather than analyzing threat incidents after they had occurred. They use the predictive analysis method to find what would be the actions that would be used by an attacker to compromise the system [205].

RansomWall, a layered defense system for protection against Cryptographic Ransomware follows a Hybrid approach of combined Static and Dynamic analysis to generate a novel compact set of features that characterizes the Ransomware behavior. The presence of a Strong Trap Layer helps in early detection. It uses Machine Learning for unearthing zero-day intrusions. When initial layers of RansomWall tag a process for suspicious Ransomware behavior, files modified by the process are backed up for preserving user data until it is classified as Ransomware or Benign [206]. Present day malware shows stealthy and dynamic capability and avails administrative rights to control the victim computers [207]. Malware writers depend on evasion techniques like code obfuscation, packing, compression, encryption or polymorphism to avoid detection by Anti-Virus (AV) scanners as AV primarily use syntactic signature to detect a known malware. To overcome these evasion techniques, an approach based on semantic aspect of PE executable that analyses API Call-grams to detect unknown malicious code was proposed [207]. Another study proposed a platform that derives the common execution behavior of a family of malware instances. A clustering graph is constructed for each instance that represents kernel objects and their attributes, based on system call traces. The method combines these graphs to develop a super-graph for the family. The super-graph contains a subgraph, called the Hot-Path, which is observed during the execution of all the malware instances. The proposed method is scalable, identifies previously-unseen malware instances, shows high malware detection rates, and false positive rates close to 0% [208,209]. In a comparative study of machine learning techniques, the use of **linear classifiers**, **ensembles, decision trees** and various **hybrid techniques** to detect malware was investigated. The study proves the ensemble algorithm provides the best malware detection rate, but it also has the highest false positive rate. To use this algorithm in practice, it requires a combination of a method for filtering false positive, such as file white-listing [210]. The signature-based malware detection is not effective during zero-day attacks. Until the signature is created for new malware, distributed to the systems and added to the anti-malware database, the systems can be exploited by that malware. But Machine learning methods, Association rule, Support Vector Machine (SVM), Decision tree, Random forest, Logistic Regression and Naive Bayes, that learns from the header data of PE32 files can be used to create more effective antimalware software, which can detect previously unknown malware and zero-day attack [211,212]. In a clustering approach to malware detection, a scalable system for network-level behavioural clustering of HTTP-based malware aims to efficiently group newly collected malware samples into malware family clusters [213]. The objective is to obtain malware clusters that can aid the automatic generation of high quality network signatures, which can in turn be used to detect botnet command-and-control (C&C) and other malware-generated communications at the network perimeter [214].

The **Anubis** is another platform applied for a dynamic malware analysis. It is designed to execute binaries submitted in a controlled environment. The system analyses the execution of the binaries by monitoring the invocation of important Windows API calls and system services and records the network traffic to track its data flow, thereby examining the influence of code polymorphism on evolution [215]. Nowadays, malware writers use polymorphic, metamorphic and obfuscation techniques to evade detection from commercial anti-virus and anti-spyware that use signature-based techniques [216]. To overcome this problem of evading detection, a machine learning framework to automatically analyse malware behaviour was proposed. The framework, **incremental approach for behaviour-based analysis**, identifies new classes of malware with similar behaviour (**clusters**) and assigns the unknown malware to the discovered class (**classification**) [217]. Relying on the analysis of instruction frequency and function-based instruction sequences, an **Automatic Malware Categorization System (AMCS)**, the **principled cluster ensemble** framework for combining individual clustering solutions based on the consensus partition was developed. The framework automatically groups malware samples into families that share some common characteristics using a cluster ensemble by aggregating the clustering solutions generated by different base clustering algorithms [218]. In another study, **BotMiner**, **Clustering Analysis of Network Traffic for Protocol** - and Structure-Independent Botnet Detection was developed to overcome shortcomings of the “botnet command and control (C&C) protocols and structures” in preventing cyber-threats such as spam, distributed-denial-of-service (DDoS), identity theft, and phishing. The BotMiner platform is C&C independent and on evaluation, the BotMiner prototype was claimed to have very low false positive rate [219]. Data mining was introduced to detect malware using three different static features for malware classification: Portable Executable (PE), strings and byte sequences. In the PE approach, the features (like list of DLLs used by the binary, the list of DLL function calls, and number of different system calls used within each DLL) are extracted from DLL information inside PE files. Strings are extracted from the executables based on the text strings that are encoded in program files. The byte sequence approach uses sequences of n bytes extracted from an executable file. They used a data set consisted of 4266 files including 3265 malicious and 1001 benign programs [205,220]. A rule induction algorithm called Ripper [221] was applied to find patterns in the DLL data. A learning algorithm Naive Bayes was used to find patterns in the string data and n-grams of byte sequences were used as input data for the Multinomial Naive Bayes algorithm. The Naive Bayes algorithm, taking strings as input data, gives the highest classification accuracy of 97.11%. The authors claimed that the rate of detection of malwares using data mining method is twice as compared to signature based method [205,212,220].

2.5 Summary of Literature Review

The background of the study reviewed a wide range of issues relating to the menace of ransomware and its spread through technological advancement. Among the issues reviewed was the evolution of different kinds of malware from 1971 to 2008. Also reviewed was the propagation pattern and termination of the identified malware. The economic impact of the attack of the identified malware to different sectors of the World economy was highlighted. Consequently, several measures were adopted to control malware attacks and minimize losses. Despite these measures, malware attacks continue to be on the increase. The malware detection techniques in use, including machine learning approaches are signature-based, runtime-based or hybrid. These focus on malware properties and processes under inspection. The existing malware detection algorithms rely on the properties and behaviours of malware (signature, rule-set, and processes under inspection or runtime) to detect malware. Because of the reliance on post-attack history of new malware, most detection can take place only when such properties, signatures and runtime behaviours are analysed and registered in the malware database. In this case the malware had attacked already. In these techniques malware still evade detection through obfuscation (dead-code insertion, register reassignment, subroutine reordering, code transportation, and code integration), fragmentation and session splicing, application specific violations, protocol violation, inserting traffic at IDS, denial of service, and code reuse attacks [222,223].

Among these measures to control malware threat, there was no **network content and link structure** (**network variable and content)** based approach to control malware. There was no approach dedicated to Exploratory Machine Learning using overlapping clusters of network content and link structure to control ransomware threat. **The active cluster overlap** approach in this research focuses on finding out methods to prevent threat incidents from occurring rather than analyzing threat incidents after they had occurred using overlapping clusters and identifying thresholds. Therefore, this study proposes a real-time Exploratory Machine-learning Cluster Overlap System **(EMCOS)** for links and contents cluster analysis in a complex ransom-ware network to build an understanding and knowledge for future development of a tool to control malware threat.

CHAPTER 3

BACKGROUND OF E-SECURITY THREATS AND ATTACKS AND RESEARCH METHODOLOGY

3.0 Background of E-security Threats and Attacks

Cybersecurity has become a global issue depicting a complex social and technical challenges for different sectors of the World economy. The spread of security threats derives from technological advancements and the inter-connectivity of the cyberspace. It is important to classify e-security threats and attacks to understand their purpose, different behaviours and mode of propagation. Common types of malware and cybersecurity attacks include the following: Phishing, SQL Injection (SQLi), Cross-Site Scripting (XSS). Man-in-the-Middle (MITM), Malware, Denial-of-Service (DoS), Spear Phishing, Whaling Phishing, and Brute-Force and Dictionary attacks [224,225].

The category of malware (attacks) refers to various forms of harmful software, such as viruses and ransomware and they can be classed as follows:

a) **Viruses** pose considerable problems for cyber-connected devices. They are malicious applications that activate and replicate (reproduce) themselves in a user's computer device without permission. In this process they modify and infect other legitimate software applications. They infect other programs by modifying them to include an evolved version of it. Replication is seen as an important characteristic of a computer virus [138,226,227].

b) **Worms** are independent programs that can to spread copies of itself or of parts of itself to other computers, commonly across network connections. These copies are themselves fully functional independent programs, which are capable either of spreading further or of communicating with the parent worm to report back results of some computation [90,227,228].

c) **Trojan Horses** are self-contained programs. Unlike the computer viruses, trojan horses do not need to attach itself to other programs to attack a computer device. Trojan horses present themselves as benign software and deceive computer users to believe they are downloading benign application. Trojan horses may have functions of use to the user [132,217,229–231]**.**

d) **Adware** is purposely designed to display advertisements on a computer. Adware do not have the capability to replicate itself; it can be seen as a subclass of spyware and will unlikely lead to dramatic results [226].

e) **Spyware** is a malicious application used for the purposes of espionage. It tracks the user's search history to send personalised advertisements and sells them to third parties [211,232,233].

f) **Rootkits** enable the attacker to have unauthorized access to data with higher permissions than is allowed. For example, rootkits can be used to give an unauthorized user administrative access. Rootkits always hide their existence and quite often are unnoticeable on the system, making the detection and therefore removal incredibly hard [226,234].

g) **Backdoors** as a malware provide additional secret "entrance" to a device system for attackers. It does not cause any harm by itself but provides access to attackers, therefore, they are not used independently but precede other malware attacks [229,232].

h) **Keyloggers** are malware class that aims to harvest sensitive data by recording any typed in information. It logs all keys pressed on a given device by the user. Through this process, it stores all user input data such as passwords, bank details and other private and sensitive information. Keyloggers strive to hide their presence using rootkit-like techniques to evade detection by antivirus and other system protections [235].

i) **Ransomware** is designed to encrypt all data on a device, and then asks the victim for a ransom to be paid in exchange for a decryption key. Ransomware hinders working of a computer and restricts user access to your computer or your files and displays a message that demands payment for the restriction to be removed [236,237]. Three types of identifiable ransomware exist, namely the Scareware, locker, and Crypto-ransomware [237]. The Scareware poses the least security threat. It only posts a message on your screen to say the your computer is locked but does not actually encrypt any file or data in your computer [237–239]. The locker locks up the system and demands a ransom. It denies the user access to certain programs or to the whole computer till ransom is paid. The severity of this ransomware is medium [237,240]. The crypto-ransomware has a severe security impact. The crypto-ransomware encrypts the user data and denies access to them until the victim pays ransom [237,240–243]. The most popular variants of crypto-ransomware include: Crypto-wall, CTB-Locker, Torrent-Locker, Tesla-Crypt, and Cryp-Vault [237,241,244,245].

3.1 Methodology

The achievement of the aims and objectives in section 1.3 follows some strategies laid out in this section. The strategies outline a plan of actions to (ensure that the investigation is robust, reliable and repeatable, and subsequently) achieve the aim of this study. In other words, research methods (methodology) describes the specific techniques applied in a specific study [246]. The strategies adopted are mainly exploratory and machine learning to achieve the aims of this study [246]. The research techniques adopted aims to define the strategies adopted, the algorithms employed and the concepts and frameworks to achieve a robust, reliable and reproducible (repeatable) investigation. These include a review of related literature, implementation of research concept (theory), modeling and evaluation [247,248]. Having identified the aim of the study, the first step is to define the criteria to achieve them. In this case, the initial step is to decide on the source and nature of data. Then decide whether the data fits a supervised or unsupervised machine-learning algorithm as shown in Figure 3 and Figure 4. The algorithm in Figure 4 suggests the suitable use of Clustering, which is an unsupervised machine learning technique. Clustering models (e.g. K-means) groups and interprets data based only on input data [249]. The data for this investigation, therefore, fits into the clustering model. The schematic of the methodology is shown in the Strategies for the Achievement of Research Aims (Figure 16).

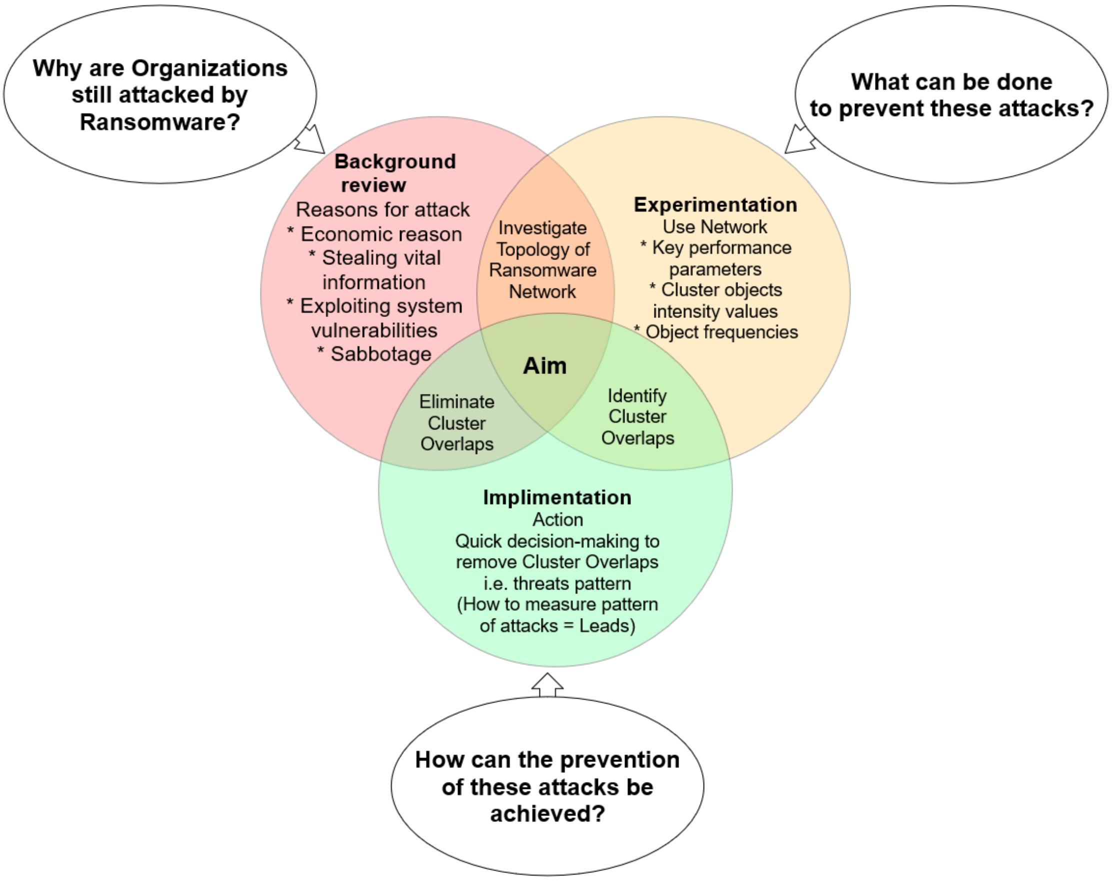

Why are Organizations still attacked by Ransomware?
What can be done to prevent these attacks?
Background review
Reasons for attack
* Economic reason
* Stealing vital information
* Exploiting system vulnerabilities
* Sabbotage
Investigate Topology of Ransomware Network
Experimentation
Use Network
* Key performance parameters
* Cluster objects intensity values
* Object frequencies
Aim
Eliminate Cluster Overlaps
Identify Cluster Overlaps
Implimentation
Action
Quick decision-making to remove Cluster Overlaps i.e. threats pattern
(How to measure pattern of attacks = Leads)
How can the prevention of these attacks be achieved?

Figure 16: Methodology: Strategies for the Achievement of Research Aims

The execution of these research strategies satisfies the research questions of “Why, What and How” (Figure 16), which is further expanded in the schematic representation of strategies to achieve the research aim (Figure 17) presented in the implementation Chapter 4.

i. The “Why” question of the research investigates the reasons behind continued attacks of ransomware on different sectors of the world economy, the motivations and factors that enable the spread of these threats. This is against the background that organizations make huge investments in security against these threats (Figure 2). The success in removing active cluster overlaps helps security practitioners and organizations to take quick decisions and mitigating actions against ransomware threats even before they attack. Consequently, access to user’s valuable data and information is controlled.
ii. The “What” question of the research looks at what approaches the research should take to create solutions to counter the threats of ransomware, in other words, to help in quick decision-making and preventing ransomware attacks. The main concept is the use of exploratory machine learning approaches to investigate ransomware network systems to identify and understand key performance parameters in the link structures and content relevance of the network that lead to the formation of clusters and cluster overlap to build knowledge and understanding of the dynamic system operation profile. This helps to identify the active cluster overlap node to remove to dislodge ransomware network threats even before they attack. This procedure includes graphical visualization of ransom-ware network clusters, investigation of nodes intensity values and other key performance parameters in the formation of cluster overlaps (Figure 16). The exploratory system, when implemented, will continually monitor the events’ activities (or traffic) in

the ransomware network to identify and remove every evolving and consistent active overlapping cluster node.

iii. The "How" question of the research defines the approaches undertaken to execute, measure, validate the results, and ensure the aim of the investigation is satisfied. This seeks to profile and track the active overlapping cluster nodes over cumulative half-yearly segments of the entire time series to establish and understand temporal morphology consistency (Figure 16). Inconsistent cluster overlaps would be indications of change in event activity within the network and suggest the entry or exit of an entity (ransomware connection) within the network community, thereby denoting the initiation or formation of an attack and prompting a need for action.

The relationship, thereby, established to detect, validate and remove the active overlapping cluster (AOCLust) follows the adaptation of the Davies-Bouldin index [250]:

$$AOCLust = \frac{1}{n}\sum_{i=1}^{n} \max_{j \neq i}\left(\frac{\sigma_i + \sigma_j}{d\left(s_i, s_j\right)}\right) \qquad Eqn(1)$$

$where\ n$ is the number of overlapping clusters,

s_x is the total intensity (strength) of the overlapping cluster $x,$

σ_x is the average intensity (or total intensity) of all elements in the overlapping cluster x to intensity $s_x,\ d(s_i, s_j)$ is the difference (distance) between intensity s_i and s_j.

Since algorithms that produce clusters with low intra-cluster intensities (high intra-cluster similarity) and high inter-cluster intensities (low inter-cluster similarity) will have a low Davies–Bouldin index, the clustering algorithm that produces a collection of clusters with the smallest Davies–Bouldin index is considered the best algorithm for active overlapping cluster (AOClust) based on this criterion. Hence, this research adopts Davies-Boulin index to quantify and identify the active overlap cluster node.

CHAPTER 4

DESIGN AND IMPLEMENTATION

4.0 Introduction

The Chapter discusses the design and implementation of the Exploratory Machine-learning Cluster Overlap System **(EMCOS).** The EMCOS is designed to identify the active cluster overlap node for timely removal to dislodge ransomware-network-event activities and prevent the attack. First, this section specifies the requirements of this project schematically represented in the process map (Figure 17). The requirements include:

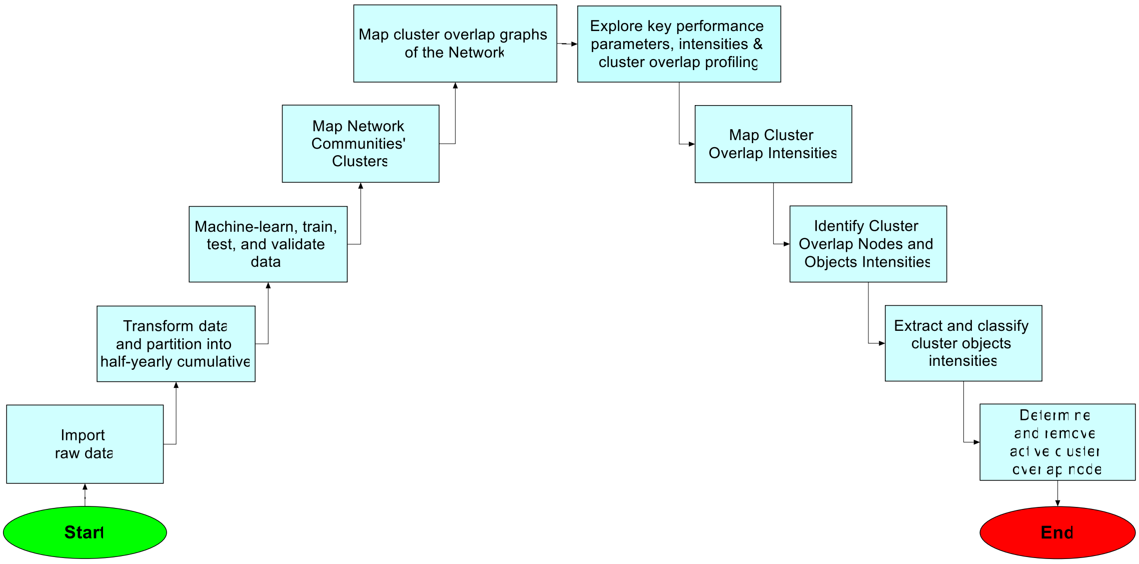

Start
Import raw data
Transform data and partition into half-yearly cumulative
Machine-learn, train, test, and validate data
Map Network Communities' Clusters
Map cluster overlap graphs of the Network
Explore key performance parameters, intensities & cluster overlap profiling
Map Cluster Overlap Intensities
Identify Cluster Overlap Nodes and Objects Intensities
Extract and classify cluster objects intensities
Determine and remove active cluster overlap node
End

Figure 17: Process Map of Exploratory Machine-Learning Cluster Overlap System (EMCOS)

a) Acquiring the raw data and pre-processing it to the required format suitable for other processes.

b) Mapping the Network communities' clusters to understand the topology (structure) of the Network

c) Mapping different sub-graphs (sub-clusters) of the Network events and the cluster overlap object intensities

d) Identifying cluster overlap nodes and the cluster objects intensities

e) Extracting and classifying the cluster objects intensities into different clusters

f) Determine and remove active overlap cluster (AOCLust) node

Further to outlining the requirements, this section outlines the source of data collection and the pre-processing methods. The chapter will highlight the reasons relating to some technical decisions taken in the investigation. The section also discusses the procedures for the Network clustering, profiling, and tracking of the active cluster overlap nodes to establish consistency and validate the system. Key performance parameters considered in the profiling and tracking of active cluster overlap nodes are outlined in this implementation chapter. These parameters define the various network cluster graphs that are presented in the general investigation, and how they are used to establish understanding of pattern consistency in the active cluster overlap node for subsequent validation of the Active Cluster Overlap Profiling and Tracking System **(ACOPTS)** for the prevention of threat.

All the methods or steps applied in this project were developed using R and SAS Enterprise Miner 14.3 environment, specifically the exploratory link analysis node in SAS. The steps include the partitioning of the network data into cumulative half-yearly sub-sets, applying cluster, and time series algorithms to map the ransomware network clusters to track the cluster formation and development, and to establish topological patterns at different stages of the network time series. These identify the overlapping clusters and the extraction of the intensity values of all the elements (objects) in the various cluster nodes. Through exploratory operations, the study classifies and tracks the cluster overlaps and their intensities over different cumulative datasets to establish the consistency of the active cluster overlap node. Subsequently, the intensities and other performance parameters (such as Out-degree Centrality of Node, Weighted Eigenvector Centrality, Clustering Coefficient Centrality, Weighted Closeness Centrality, Weighted Betweenness Centrality, and Weighted Influence Centrality, et cetera) of the cluster overlap nodes are extracted (Figure 18). Through the exploration of cluster overlap nodes (vertices) intensity values and other key performance parameters, the study will profile and track the cluster formation and development to identify active cluster overlap that can be removed to control the threat.

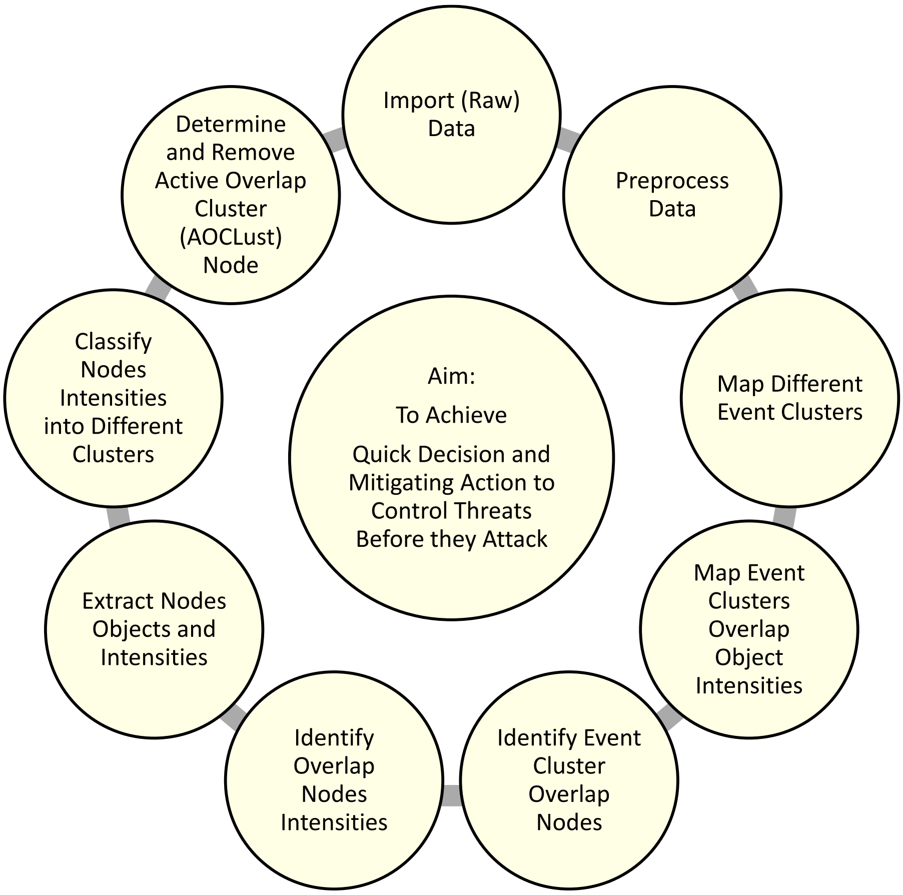

Import (Raw) Data
Preprocess Data
Map Different Event Clusters
Map Event Clusters Overlap Object Intensities
Identify Event Cluster Overlap Nodes
Identify Overlap Nodes Intensities
Extract Nodes Objects and Intensities
Classify Nodes Intensities into Different Clusters
Determine and Remove Active Overlap Cluster (AOCLust) Node
Aim:
To Achieve
Quick Decision and Mitigating Action to Control Threats Before they Attack

Figure 18: Schematic Representation of Strategies to Achieve Research Aim from the import of raw data to the removal of the active overlap cluster node

4.1 Ransom-ware Data Collection and Processing

Data collection is based on data feeds from the website, https://ransom-waretracker.abuse.ch/feeds/csv/. The website gives a constantly refreshed list of online and offline ransom-ware families. This represents a dynamic data environment. The feeds CSV file is imported directly into RStudio (R) where the data was transformed and converted into appropriate data frame and format by running required algorithms. RStudio (R) was chosen for pre-machine learning processes and analysis of k-means clustering.

4.2 Supervised Machine Learning Approach

Preliminary processes with k-means clustering in R did not show the desired clarity and distinction of clustering and clustering overlaps (Figure 19 and Figure 20).

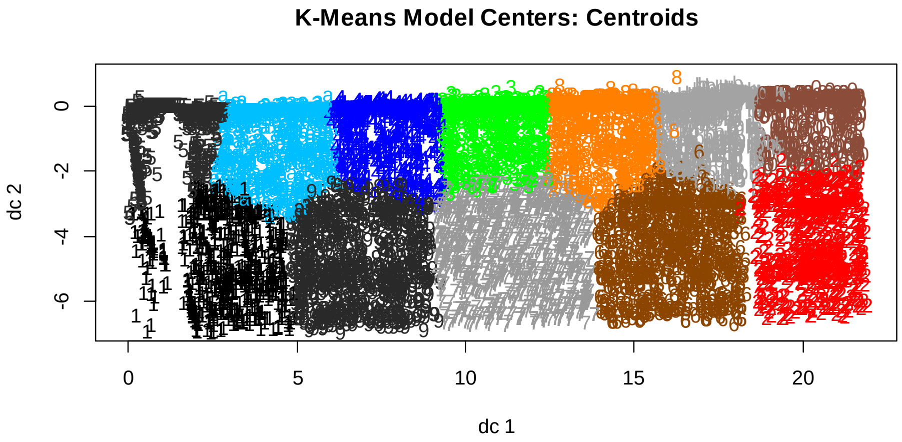


Figure 19: Detection of overlaps by use of kmeans centers – Centroids (k-means in R): where K=12 clusters represent the various ransomware families overlapping one another without clarity

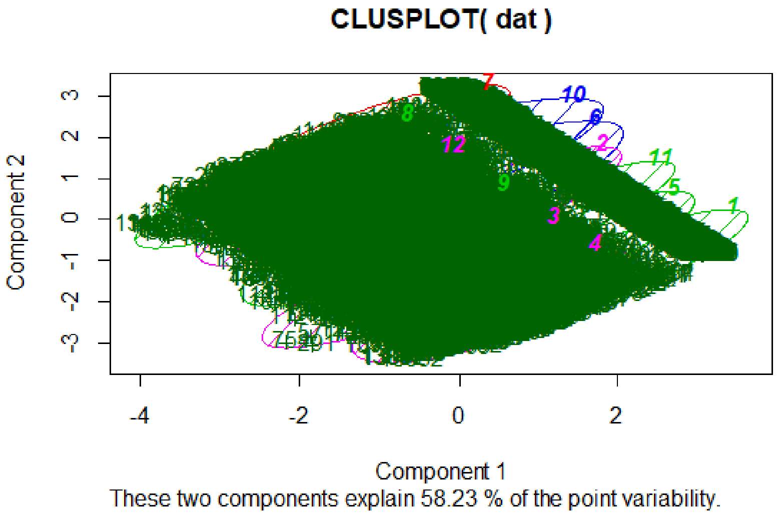


Figure 20: K-Means Components Cluster Plots (topology) showing community network of different families of ransom-ware (k-means in R): where K=12 clusters represent the various ransomware families overlapping one another without clarity

HP Neural node in the SAS Enterprise Miner Workstation 14.3 was used to analyse the ransomware network data to understand the clustering pattern of the algorithm using the ransomware network data.

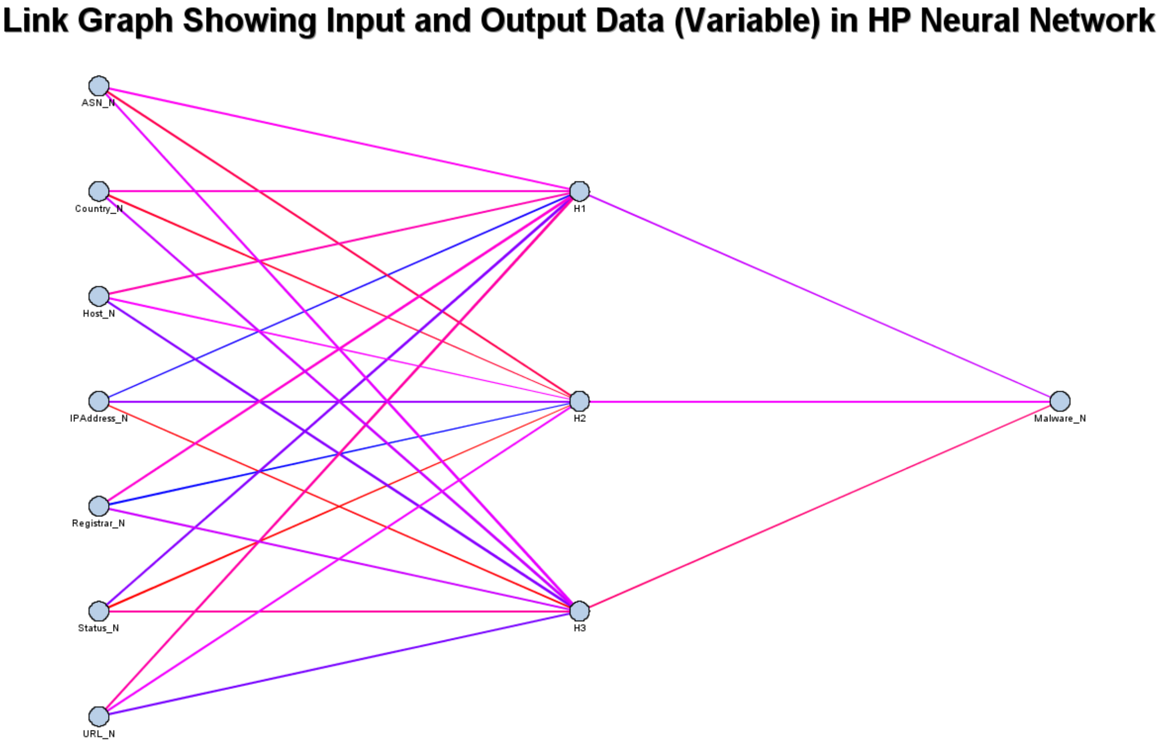

Link Graph Showing Input and Output Data (Variable) in HP Neural Network
ASN_N
Country_N
Host_N
IPAddress_N
Registrar_N
Status_N
URL_N
H1
H2
H3
Malware_N

Figure 21: Link Graph Showing Input and Output Data (Variable) in HP Neural

The link graph (Figure 21) shows the input and output data (variable) in the HP Neural Network. The link graph indicates the algorithm received seven input variables, ASN, Country, Host, IPAddress, Registrar, Status and URL for process. The computation showed three hidden neurons and three hidden layers in the processing of the input data. The process shows the output data to be Malware. The output value suggests that HP Neural is more suitable for prediction analysis. In addition, the HP Neural Network algorithm did not produce a distinctive cluster overlap that is required for profiling and analysing the effect of overlaps in the timely control of ransomware threat. The **HP Forest** node was also used to analyse the ransomware network data. The leaf plot (Figure 22) and Iteration plot (Figure

23) show there was no visible clustering in HP Forest. Therefore, the HP Forest algorithm is not suitable for further analysis of overlapping cluster. Likewise, the **HP Principal Component, HP Tree** and **Neural Network** did not produce any explorable overlapping clusters for further profiling and analysis to determine active cluster overlap to remove for a timely control of ransomware threat.

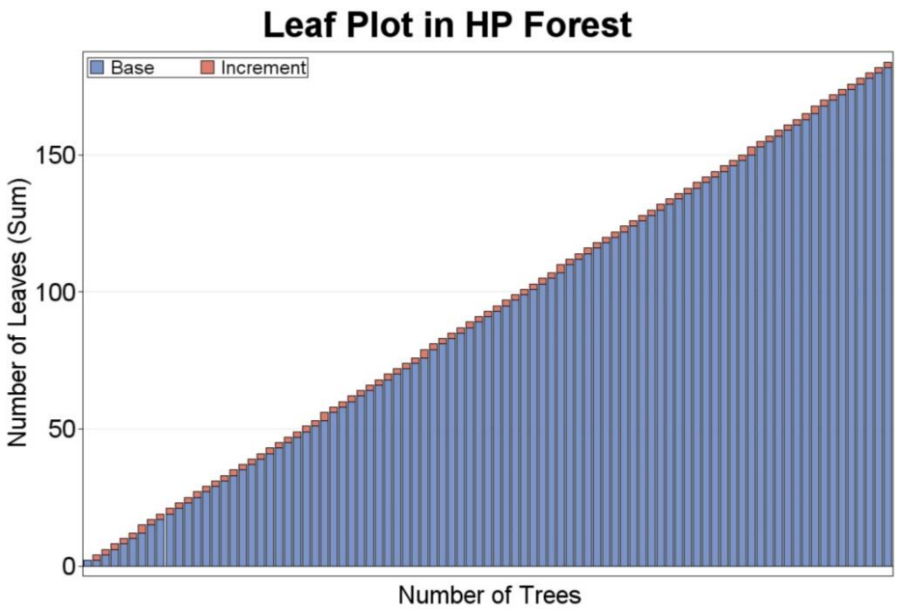


Figure 22: Leaf Plot Showing Base and Incremental Number of Leaves in HP Forest

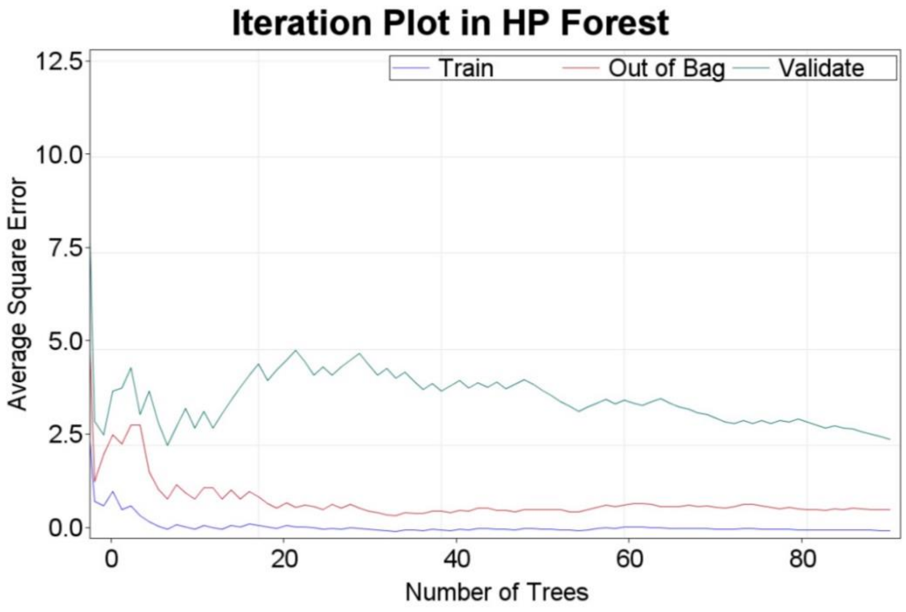


Figure 23: Iteration Plot Showing the Average Square Error of Ransomware Network Data in HP Forest

To achieve the desired clarity and distinction in the cluster and cluster overlap plots for analysis, this investigation continued with **Unsupervised Machine Learning approach (Clustering)** in the graphing and analysis with SAS Enterprise Miner 14.3, Base SAS 9.4 and SAS Enterprise Guide 7.1.

While exploratory machine learning investigations are done in SAS, the graphs and results run in other tools like R, Python, Matlab, Maltego, and Gephi would be shown to bring clarity where they are required to present more visual understanding and comparisons. To achieve the overall aims and objectives of this investigation, the imported dataset, **RansomwareFeed** (Source: Ransom-waretracker.abuse.ch), is converted into data.frame for use in the different machine learning algorithms. The dataset is converted from text data to a numerical dataset to fit the required format of most machine-learning algorithms. Subsequently, the dataset is standardized to avoid overfitting. The resulting dataset is randomly partitioned, where required, into 70% train dataset to train and model the data to establish network characteristics and patterns, and 30% test dataset to test the accuracy and validity of the patterns/model using machine learning. Further processes carried out to achieve the aims and objectives of this research are stated in sections 4.3 – 4.5 below as previously depicted in Figure 18 above.

4.2 Cluster Graphing to show different Communities (sub-cluster) of Ransomware Network

Pursuant to steps, 1.3.1 and 1.3.2 the general Network cluster representation will help to understand the topology (structure) of Ransom-ware Community (Family) Networks (Figure 24). The Network Communities cluster presents the general structure of the network, showing the nodes, vertices and the direction of links. However, it fell short of defining the sub-communities or sub-clusters and the overlapping clusters by only presenting the cluster of objects around the input variables.

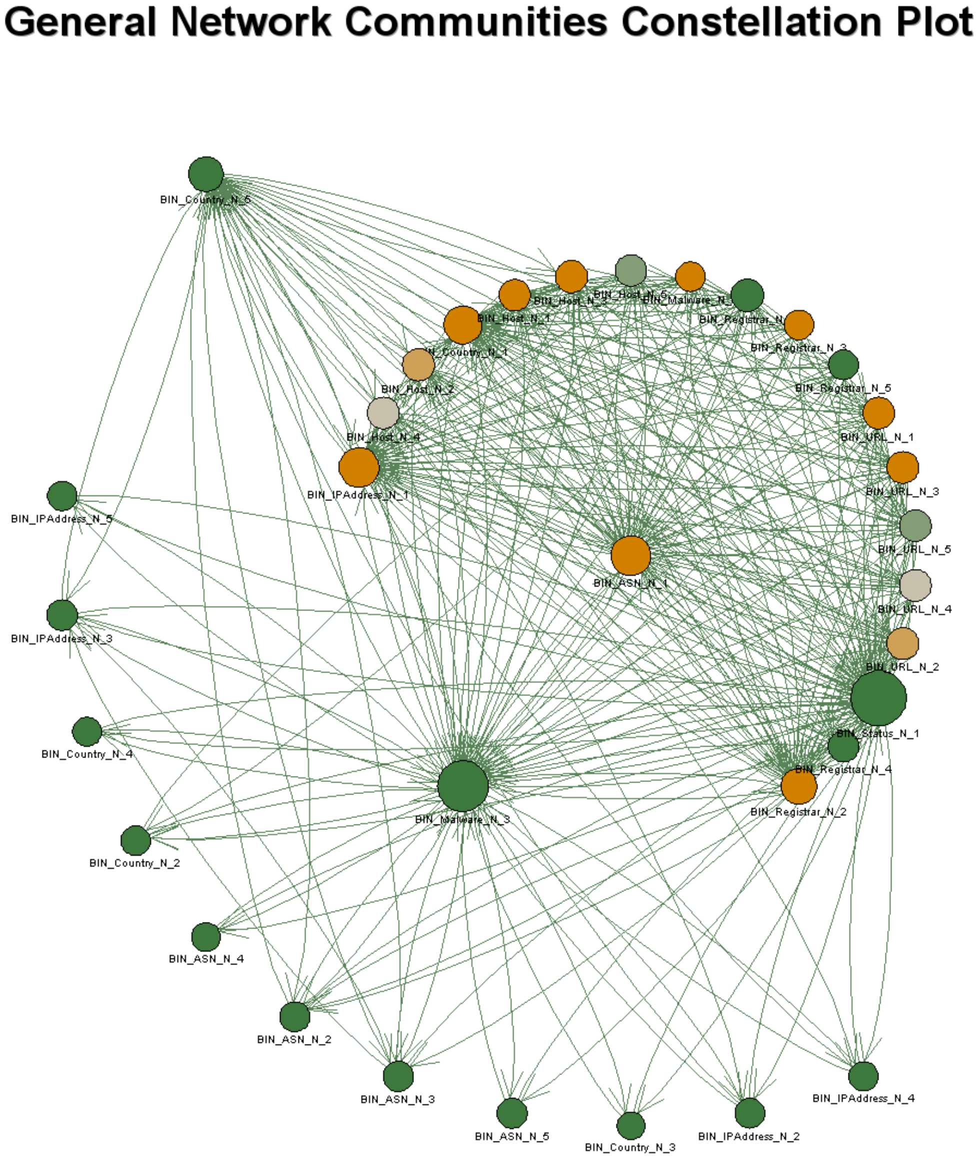


Figure 24: General Network Communities Constellation Plot 2015Jan-2018Jun Showing different cluster node and clustering objects in the network

To make up for the shortcoming above, specific graphical cluster representations will be presented. Such cluster graphs are designed to show the relationships between the nodes (clusters), vertices (links) and contents of the Network. These identify the key performance parameters in the formation and development of clusters and cluster overlaps. Such parameters as the Cluster Nodes Intensity, Out-degree Centrality of Node, Weighted

Eigenvector Centrality, Clustering Coefficient Centrality, Weighted Closeness Centrality, Weighted Betweenness Centrality, and Weighted Influence Centrality, et cetera and other Unweighted Centrality Measures of the Network will be determined. The Ransom-ware Network Communities graph attempts to show the topology and different ransomware network elements (objects) clusters (nodes) as depicted in Figure 24. The network associations and content relevance will be analysed using the above listed key parameters (components).

4.3 Detection of Ransomware Cluster Network Overlap (Hubs) and Outliers using Machine Learning Algorithms

The objective of using exploratory and machine learning of links and cluster analysis is to detect the overlaps (hubs) of different nodes of Ransom-ware Network sub-cluster and outlier nodes to satisfy step 1.3.3. Cluster overlaps, (e.g. k-means centres), also known as the centroids, in the representative-cluster overlap plot in Figure 25 show the cluster overlap node 1 as the active cluster overlap node of the ransomware networks (variable clusters). Figure 25 shows that cluster overlap node 1 originated all the outward traffic links of the ransomware threat and did not receive any inward traffic link. Thus, it is the primary originating and distribution link to nodes 2, 3, 4, and 5. This means that all elements of the ransomware threat patterns coalesce and overlap at node 1 before distribution (transmission) to other nodes. This is further illustrated in the representative cluster-constellation plot in Figure 26 that shows the link connections (relationships) of different nodes, hence satisfying step 1.3.3. The connections, in Figure 26, show that cluster node 1 (active cluster node) originates all primary connections to other cluster nodes and does not receive any incoming link. Furthermore, the representative cluster-constellation plot in Figure 26 shows the link connections (relationships) of different nodes, hence satisfying step 1.3.3. The connections, in Figure 26, show that cluster node 1 (active cluster node) originates all primary connections to other cluster nodes and does not receive any incoming link.

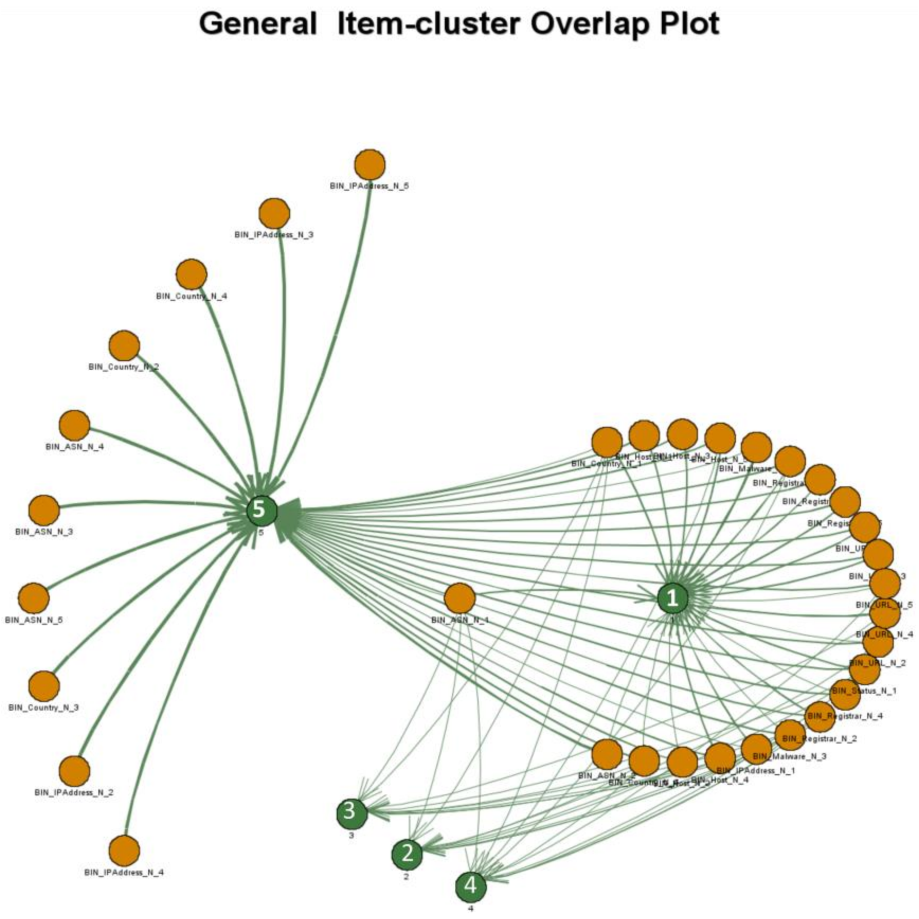


Figure 25: Representative General Cluster Overlap Plot (Jun2015):
Showing cluster overlap nodes 1, 2, 3, 4, and 5

Node 1 is the source originator of the developing ransomware threat. It transmits the threat through other nodes, while node 5 is the major node through which the threat is distributed to network-connected user devices. Consequently, Figure 27 shows the resultant effect of disconnecting or deleting the active cluster Link node 1. This shows the network events traffic was effectively dislodged resulting in the timely control of threat.

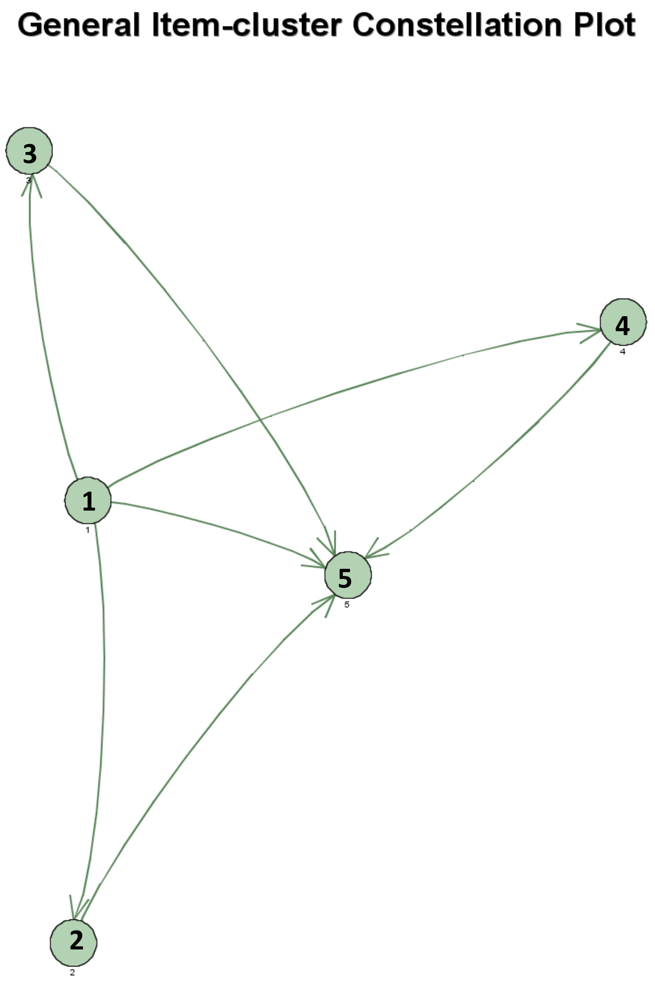


Figure 26: Representative General Cluster Constellation Plot: Showing link connections among cluster nodes 1, 2, 3, 4, and 5 where cluster node 1 generates all outward links, receives no inward links whereas cluster node 5 receives all inward links, and does not give outward link

The same result occurs in Figure 28 when the active cluster overlap node 1 is disconnected or deleted. The result suggests the dislodgement of the ransomware network and prevention of threats before it attacks its target industry device. Consequently, Figure 27 shows the resultant effect of the disconnection or deletion of the active cluster Link node 1. This shows the network events traffic was effectively dislodged resulting in the control of threat. The same result occurs in Figure 28 with the disconnection or deletion of the active cluster overlap node 1. The result suggests the dislodgement of the ransomware network and prevention of threats before it attacks its target industry device.

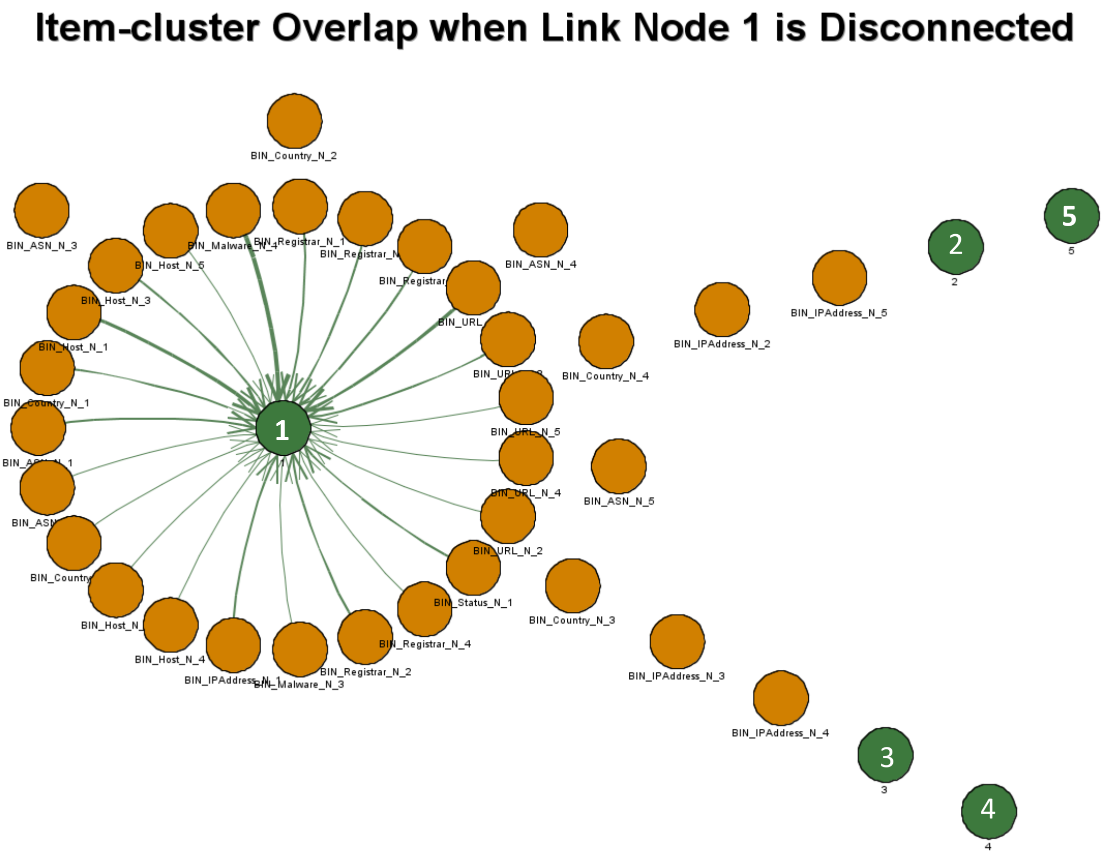


Figure 27: Representative Cluster Overlap showing the dislodgement of the network when Link Node 1 is disconnected

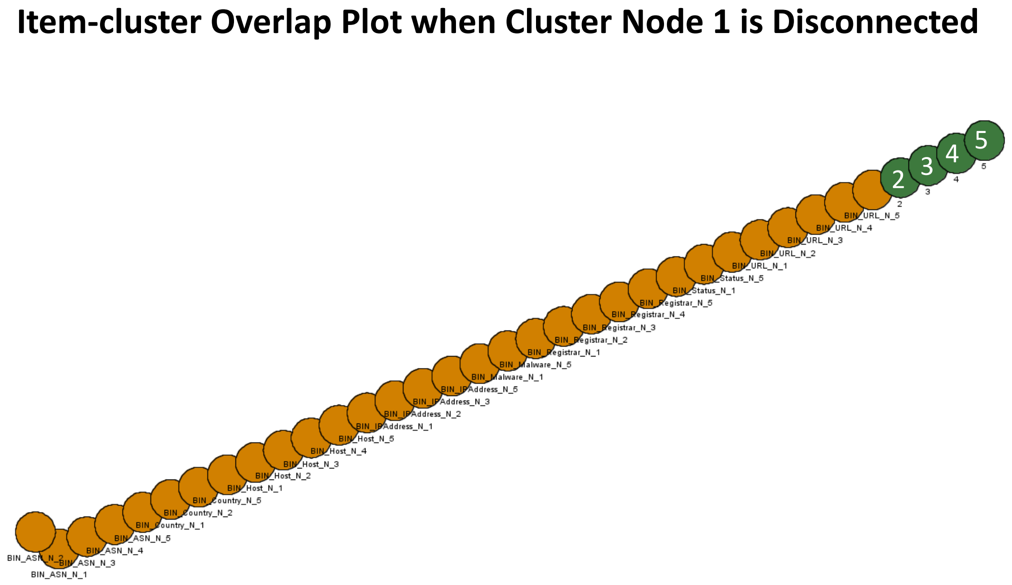


Figure 28: Representative Overlap Cluster showing the dislodgement of the network when cluster node 1 is disconnected

4.4 Implications of Removing the Active Cluster Overlap Node in Validation Analysis of the Ransomware Threat

The procedure above demonstrates that removal of the active cluster overlap node in the ransomware network traffic, timely and effectively dislodges the network traffic and prevents the ransomware threat from attacking the network-connected devices. Figure 25 and Figure 26 clearly depicts the origination and destination of ransomware network traffic. Whilst it is evident that cluster overlap node 1 originates the entire traffic in the network, it is also true that all traffic terminates at cluster overlap node 5 (terminal node). Through the terminal node 5, the ransomware threat is transmitted into connected user devices. Whilst there are inward and outward traffic (links) in cluster overlap nodes 2, 3, and 4, these nodes are mere traffic by-pass and do not transmit into user devices. Figure 26 clearly shows that cluster overlap node 1 had four outward links, each to cluster overlap nodes 2, 3, 4 and 5. However, it does not receive any inward links. Cluster nodes 2, 3, and 4 respectively received one inward links from cluster node 1 and gave out the same one outward links, respectively to cluster node 5. In the next sections, this investigation would take steps to prove and validate cluster node 1 as the originating and active cluster node; while cluster node 5 or any other terminal node would be proved to be the ransomware threat dispersion node.

4.5 Summary

In section 4.4, the study suggests that cluster overlap node 1 is the active cluster overlap node, therefore, its removal effectively and timely dislodges the ransomware network threat. Conversely, cluster node 5 or any other terminal node is the terminal cluster node that disperses ransomware to connected user devices through the IPAddresses. While there could be the temptation to consider removing other cluster overlap nodes, if cluster node 2 is removed, the ransomware network traffic would still reach the terminal node (node 5) through the by-pass of nodes 3 and 4. The same situation occurs if cluster node 3 or node 4, respectively, is removed, suggesting that the network traffic is not dislodged. Therefore, cluster overlap node 1 is the active and appropriate node to remove to timely stop the flow of ransomware traffic to the terminal node and consequently preventing dispersion of threat into the network connected user devices and systems.

CHAPTER 5

ANALYSIS AND GRAPHICAL PRESENTATION OF TIME SERIES RESULTS

5.0 Introduction

This research has taken several approaches to investigate the topology, link relationship and content relevance of a complex ransomware network [251]. These approaches aim to explore the dynamics of temporal events and cluster formation in a ransomware network. The aim is to gain an understanding of the propagation of clusters and cluster overlaps in a complex ransomware network. The exploratory approach attempts to profile and track active cluster overlap nodes to determine their consistency as the node to remove to dislodge ransomware threats. In other words, this helps to detect the active cluster node to remove to prevent threats before the attack. The results are presented in stages following a logical order of investigation. In this chapter, the investigation presents the results of time series, link clustering, and content relevance visualization and analysis to prepare the ground and provide the necessary data to build discussions in the succeeding chapters.

5.1 Time Series of Ransomware Data

A network time series has been defined as a multivariate represented graphically to describe the connection between the variables (or nodes) [252]. In this study, the time series of the target data was mapped to reveal the pattern and distribution of temporal events in a ransomware network over the available time January 2015 – June 2018 [253–255]. Prior to mapping and further time series exploration, the network data was pre-processed with the

TS Data Preparation node within SAS. Using the graph explore node of time series in SAS enterprise miner, the "Multiple Time Series for all Variables Events Activities in the Ransomware Network" (Figure 29) was mapped to display the distribution pattern for the time series for all variables in the network.

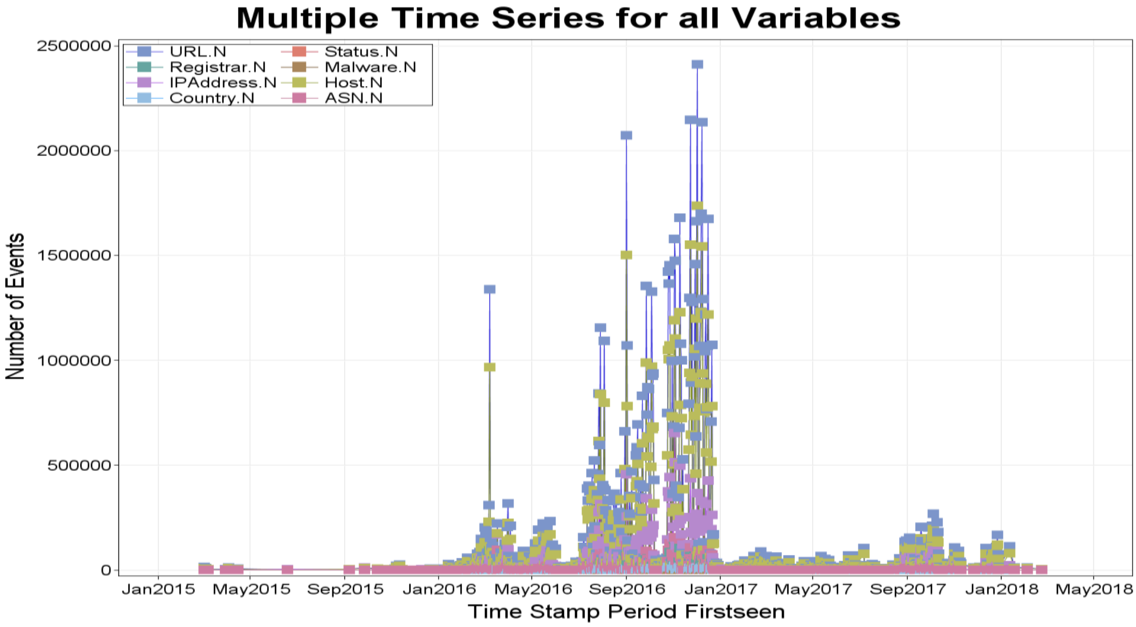


Figure 29: Multiple time series for all variable events activities in Ransomware Network

The time series graph (Figure 29) shows the distribution of the activity level of events within the different quarter, half-year, and annual period [256,257]. Prior to mapping the time series distribution, the ransomware network data was prepared using the SAS time-series data preparation node. The Time Series Events Activities Jan2015 - May2018 (Figure 29, Appendix 3) and the percentage values of the events' variables (Figure 30, Appendix 4) suggest a stationary low activity of events ranging from 0.00 – 0.50% from the start of Jan2015 to Jan2016. There was an increase in the events' activities across all variables to 8.67 – 12.91% in the period ending May2016. The period ending Sep2016 recorded a range of 24.96 -2 9.17% increase in the events' activities. However, there was a large increase in the events activities for the period ending Jan2017 (Figure 30, Appendix 4). At this period, the events activities increased to 51.00 – 56.77% among all the events variables. After this increase, the events activity level decreased to the range of 0.42 – 2.50% for the period ending May2017. The periods Sep2017 and Jan2018 recorded an increase to 3.05 – 4.94% in the events' activities and decreased to 0.01 – 0.02% for the last period ending May2018. The percentage values (Figure 30) show there was onset of active events activities from May2016 that climaxed in the period ending Jan2017. The reasons for the overall distribution pattern of the time series in relation to the threat will be discussed in later chapters.

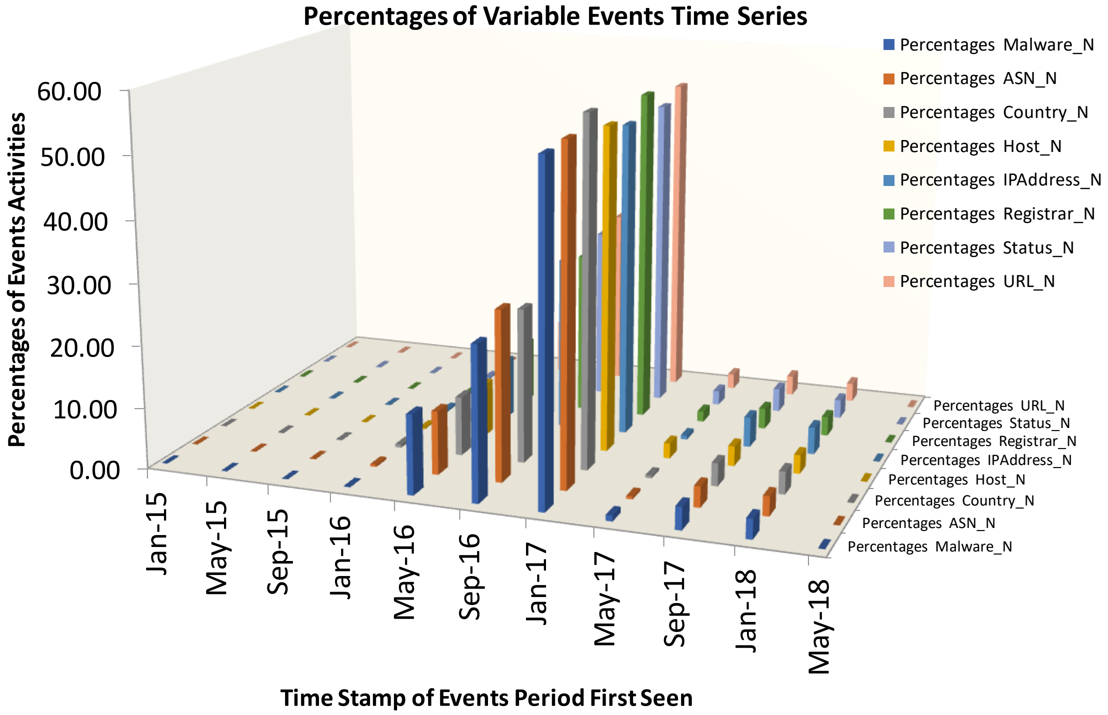


Figure 30: Percentages of Variable Events Time Series

5.1.1 Decomposition of Time Series

To investigate the components that influence the dynamics of the network, there is a need to decompose the time series plot. Decomposition was achieved using the time series decomposition node of SAS. The decomposition follows a general mathematical approach shown in Equation 2 [258,259]:

$$O_t = f(TC_t, S_t, I_t) \qquad Eqn(2)$$

where

O_t is the time series value (actual data) at time *t*;

TC_t is a deterministic trend-cycle or general movement component;

S_t is a deterministic seasonal component;

I_t is the irregular (remainder or residual) stationary component.

The decomposition resulted in many components of the series. Figure 31 shows the original network distribution time-series.

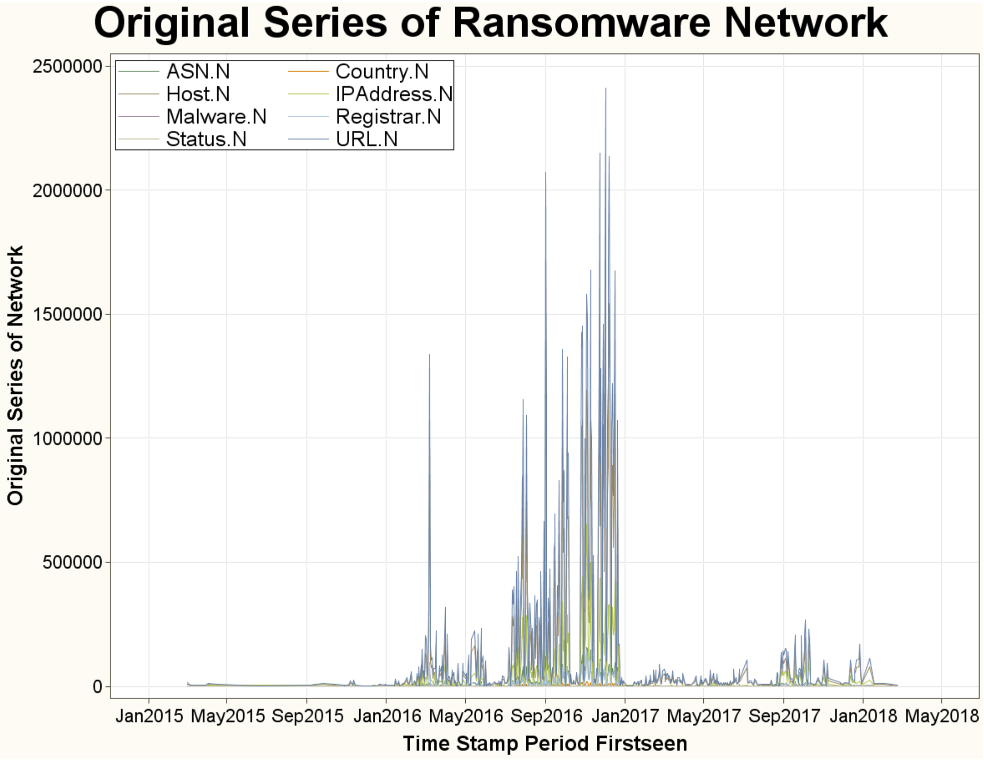


Figure 31: Original Time Series of Ransomware Network

The Event trend-cycle component (TCC) of the series is shown in **Figure 32**. This measures the long-term pattern of the time series after filtering out the high and low-frequency events of the series [260,261]. Also called the short-term trends, the trend-cycle Component (TCC) is useful for short-term analysis of time series [262]. TCC is measured by the centered moving average of the original data series (O_t). Another component is the seasonal-irregular component (SIC) of time series (**Figure 33**). The SIC is computed using Equation 3:

$$SI_t = S_t I_t = O_t / TC_t \qquad Eqn(3)$$

where the components are as defined in Equation 2.

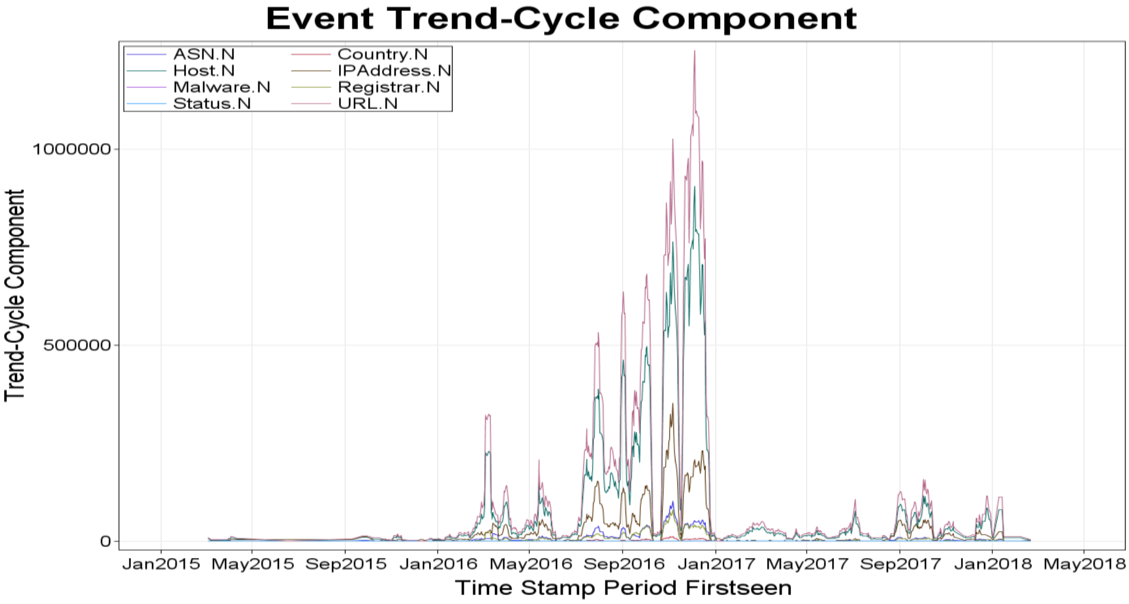


Figure 32: Event Trend-Cycle Component

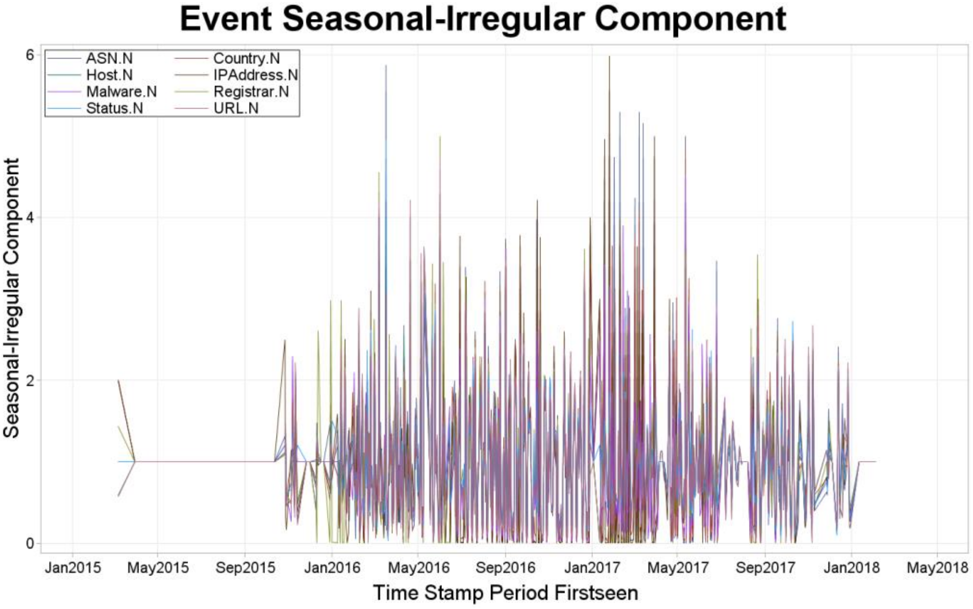


Figure 33: Events Seasonal-Irregular Component

Further components of the time series include the Event Seasonal Component (SC), Event Trend-Cycle-Seasonal Component (TCSC) and Events Irregular Component (IC) shown in **Figure 34**, **Figure 35** and **Figure 36** respectively. **Figure 34** is the seasonal component of the time series of the network. This seasonality component shows when there are regular fluctuations in the time series during corresponding times over a given number of years. The seasonal component (SC) is computed from the seasonal averages of SI_t. Conversely, **Figure 35** shows the Event trend-cycle-seasonal component (TCSC) of the time series of the ransomware network. The TCSC is a cyclic seasonal pattern in the data that explains fluctuations that are not of a fixed period but are influenced by seasons [263].

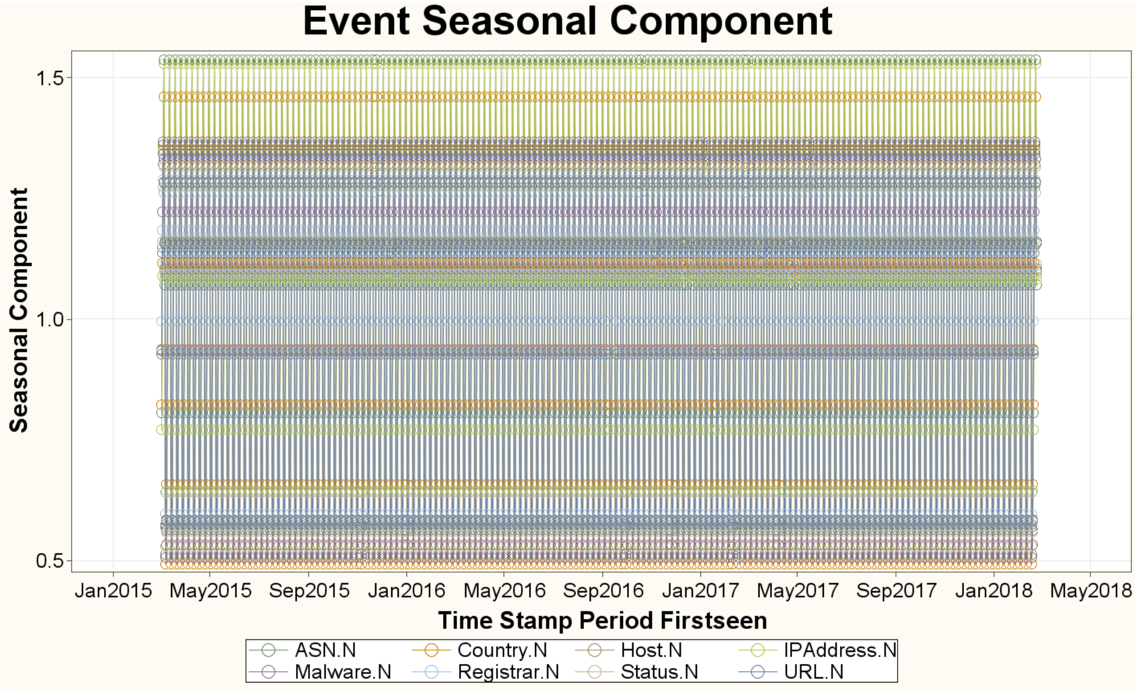


Figure 34: Event Seasonal Component of Network Time Series

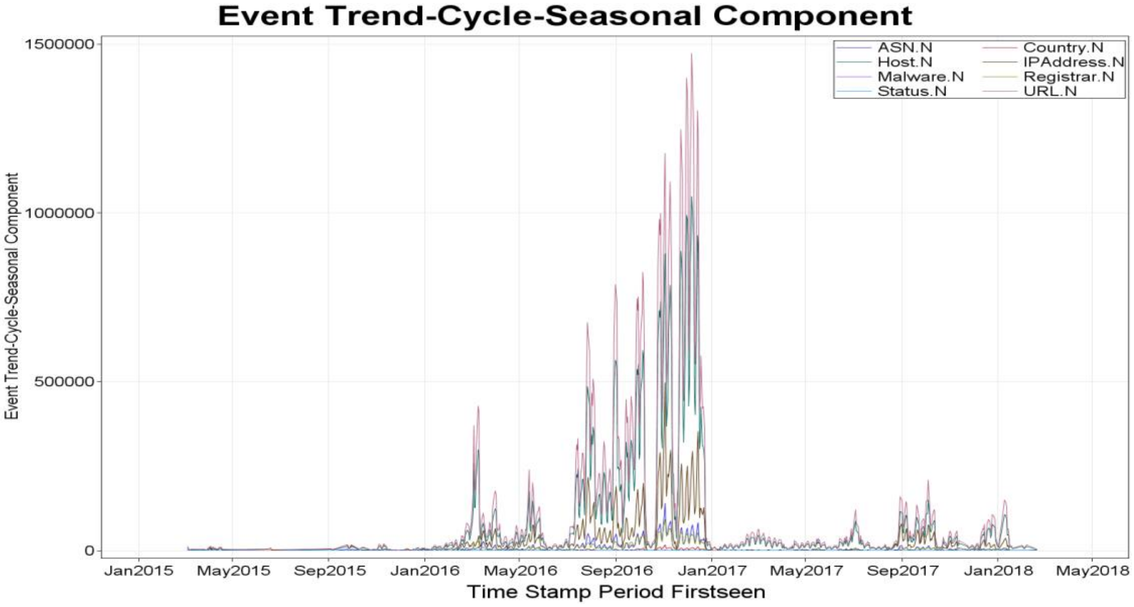


Figure 35: Events Trend-Cycle-Seasonal Component of Network Time Series

The Events Irregular component (Figure 36) is a measure of the unpredictable behaviour of the ransomware network or events in a time series [264]. The unpredictable nature of this component makes it a random variable. Modeling, therefore, requires this component to be the only random component while ensuring that other components are consistent. The Event Trend Component (ETC) of the time series is shown in Figure 37. The ETC explains the overall and persistent long-term structure, behaviour, pattern or movement of a time series. This means the time series assumes a fixed pattern. The ETC could assume a positive or negative value depending on the direction of change, either increasing or decreasing growth. The trend component becomes stationary if there is no change in growth [265,266].

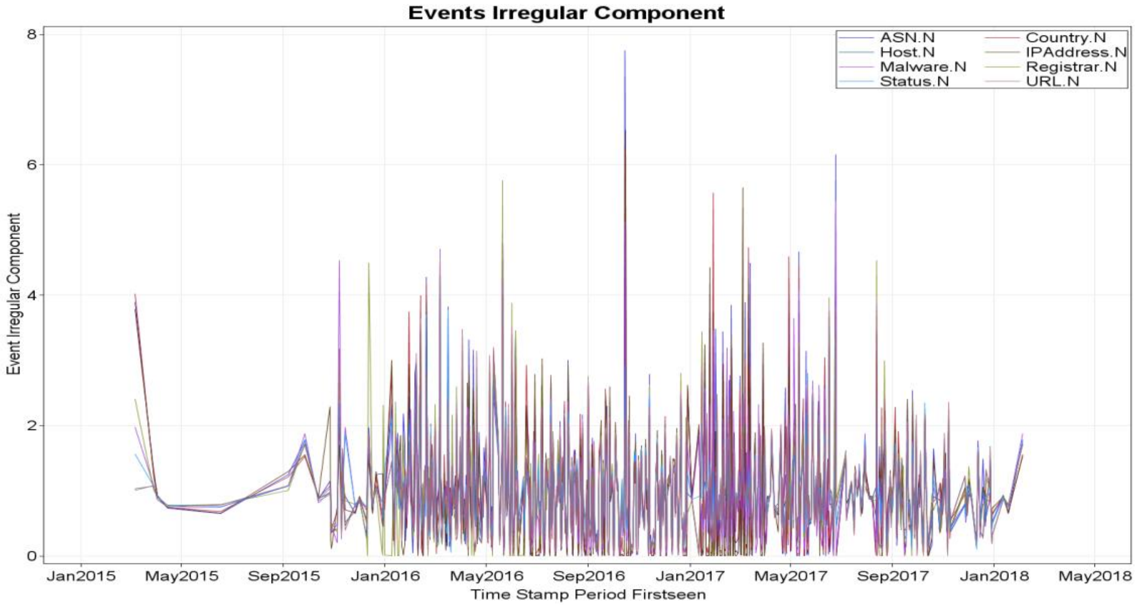


Figure 36: Events Irregular Component of Network Time Series

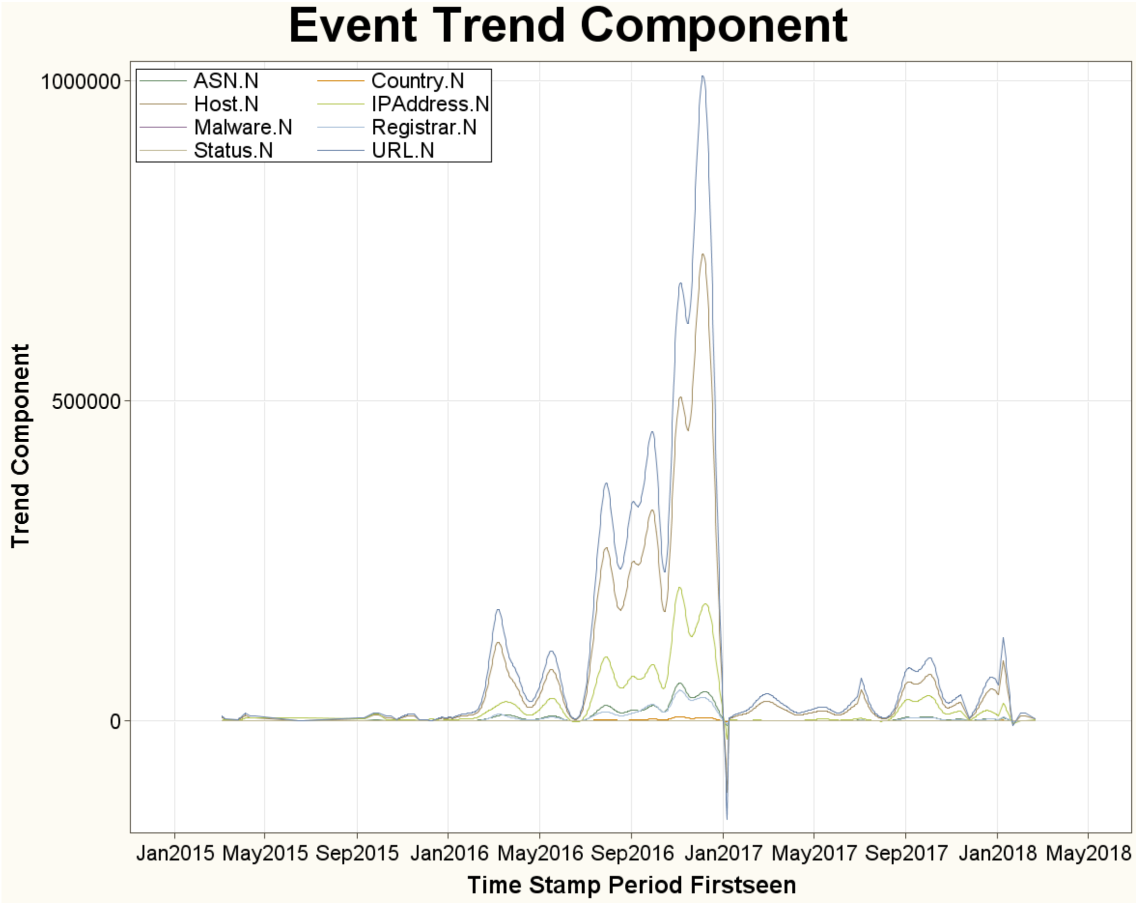


Figure 37: Event Trend Component of Network Time Series

Figure 38 shows the trend-cycle seasonal component of the time series of ransomware feeds dataset. The seasonality component shows when there is a regular fluctuation in the time series during corresponding times over a given number of periods (years). The Irregular component in Figure 39 is a measure of the unpredictable behaviour of any network or events in a time series [264]. The unpredictable nature of the irregular component makes it a random variable. Modeling, therefore, requires the irregular component to be the only random component while ensuring that other components are consistent.

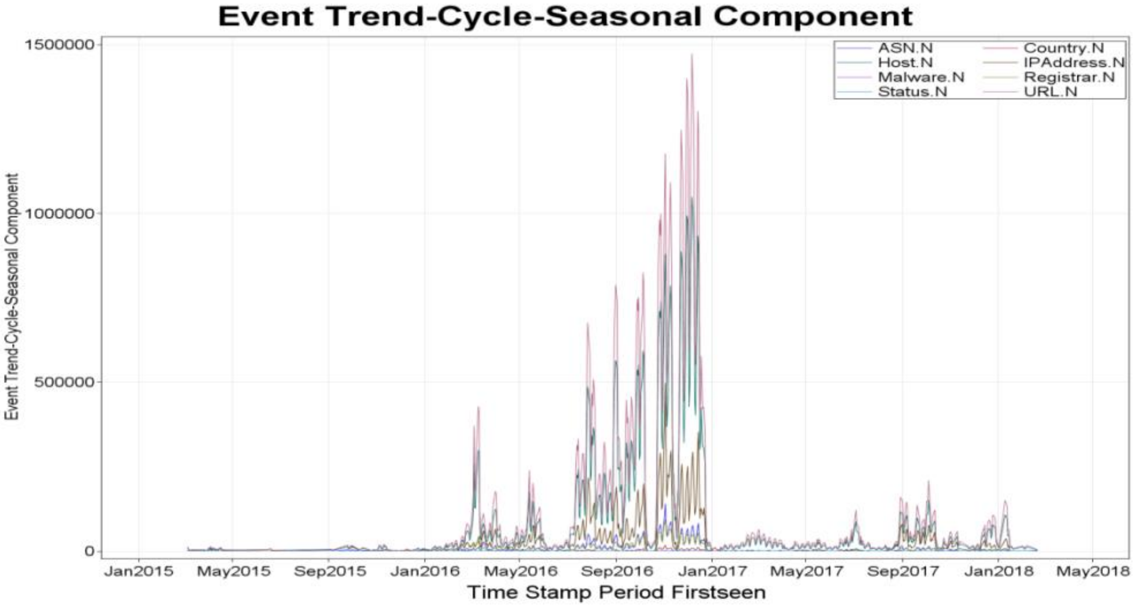


Figure 38: Event Trend Cycle Seasonal Component

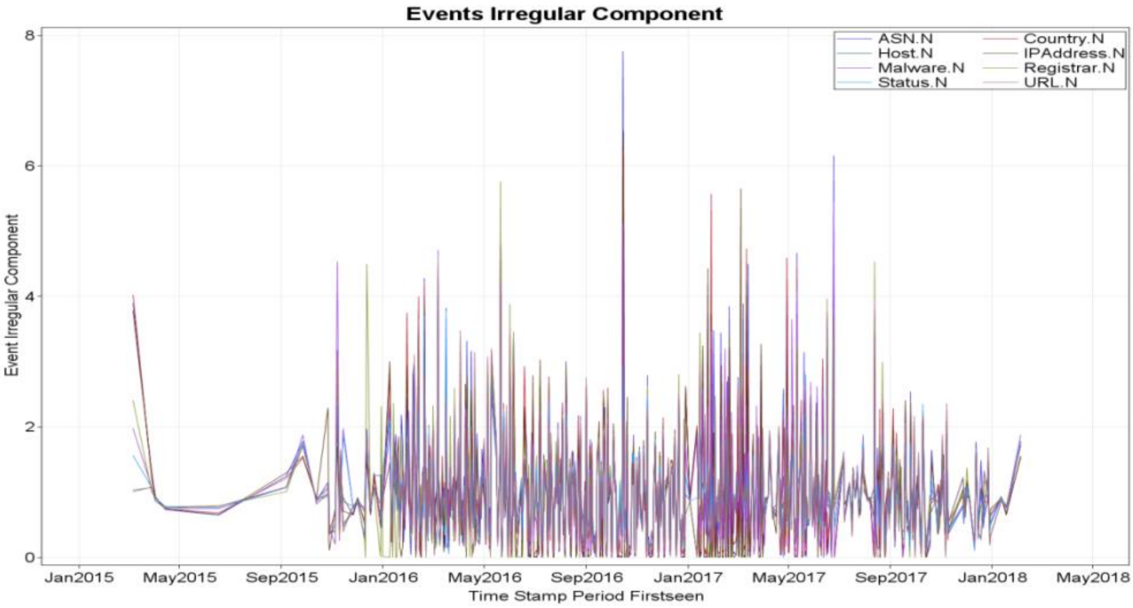


Figure 39: Events Irregular Component

Other components to consider are the seasonally adjusted series, and percent change seasonally adjusted series (PCSA) in Figure 40 and Figure 41. These components are important because seasonally adjusted series in Figure 40 helps to estimate and remove movement in a time-series caused by regular seasonal variations in events activity [267–269], while the PCSA in Figure 41 plots the percentage changes in the seasonally adjusted component of the event time series [259].

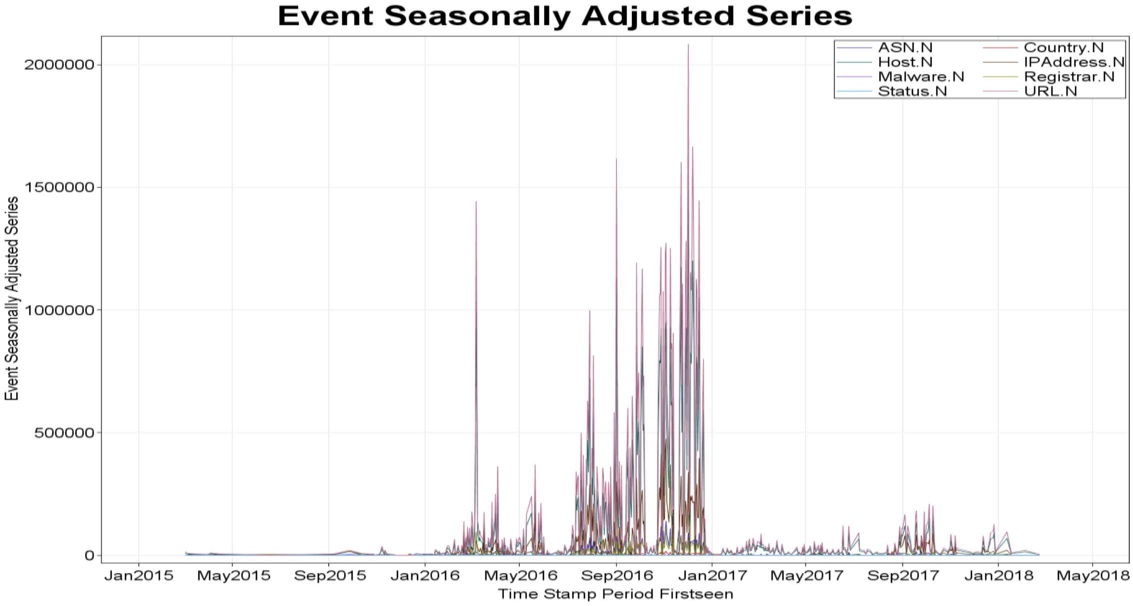


Figure 40: Event Seasonally Adjusted Series

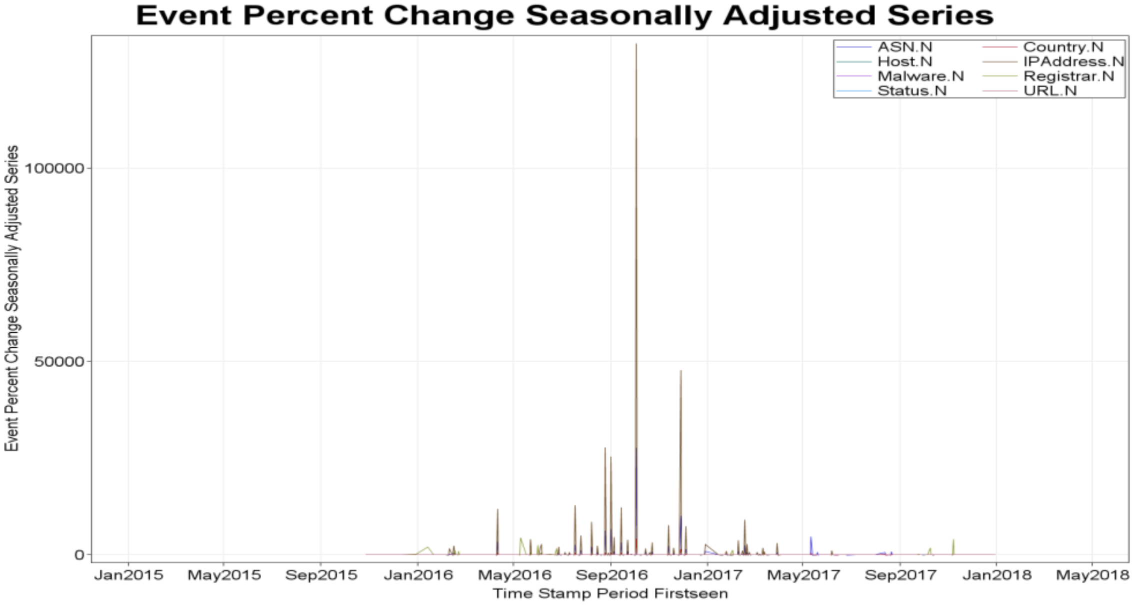


Figure 41: Event Percent Change Seasonally Adjusted Time Series

The Cycle Component (CC) of the time series is shown in Figure 42. Cycle Components are fluctuations or patterns that do not have fixed periods; they are not associated with the effects of calendar periods. CC is quite different from seasonal components that are associated with fixed effects relating to calendar periods [270]. Whilst seasonal components measure the effects of the same (corresponding) calendar periods over an event, the Cycle Component (Cyclic pattern) shows the fluctuations (rises and falls) that are not of a fixed nature, which occur in an event (data) for at least two years and over.

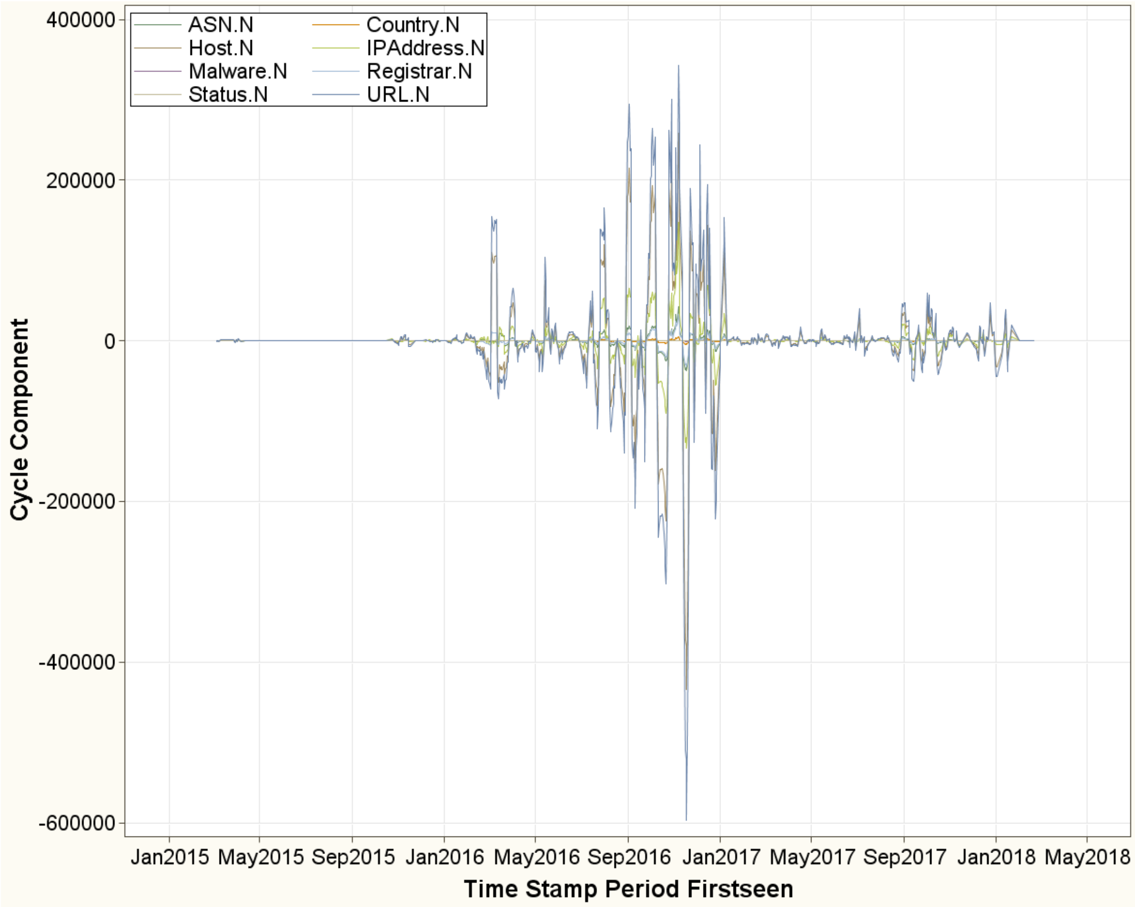


Figure 42: Event (Trend) Cycle Component (ECC)

5.1.2 Time Series Similarity Measures

The time-series similarity node of SAS enterprise miner 14.3 was used to generate the Malware versus Input Series in Figure 43, and Distance Measure Map in Figure 44.

Similarity procedure (SAS) computes similarity measures associated with time-stamped data, time series and other sequentially ordered numeric data. PROC SIMILARITY computes similarity measures for time-stamped transactional data (transactions) with respect to time by accumulating the data into a time series format, and it computes similarity measures for sequentially ordered numeric data (sequences) by respecting the ordering of the data [271]. The Similarity Measures evaluate and quantify similarities between objects or variables in a (ransomware) dataset by assigning real-values to them.

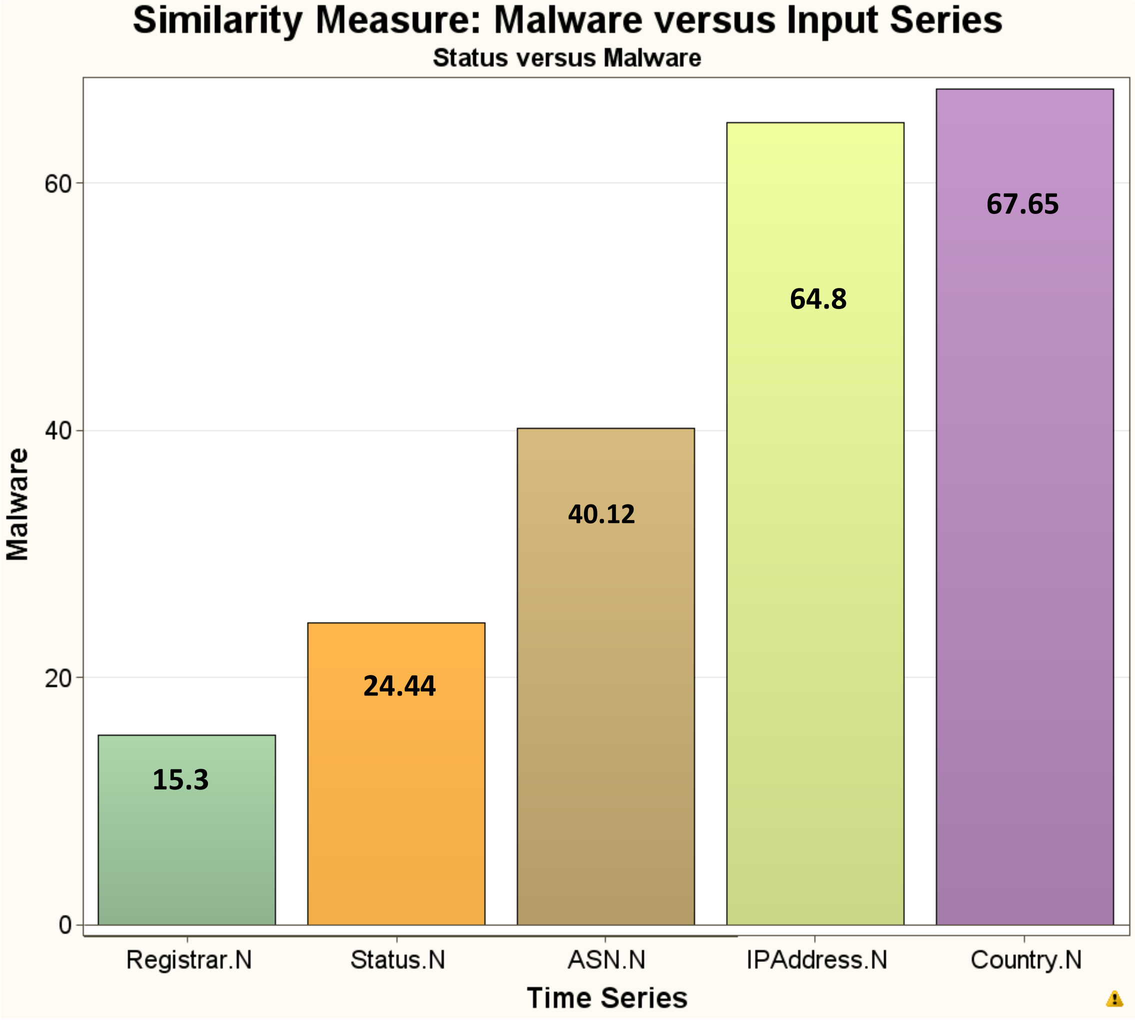


Figure 43: Similarity Measure: Malware versus Input Series

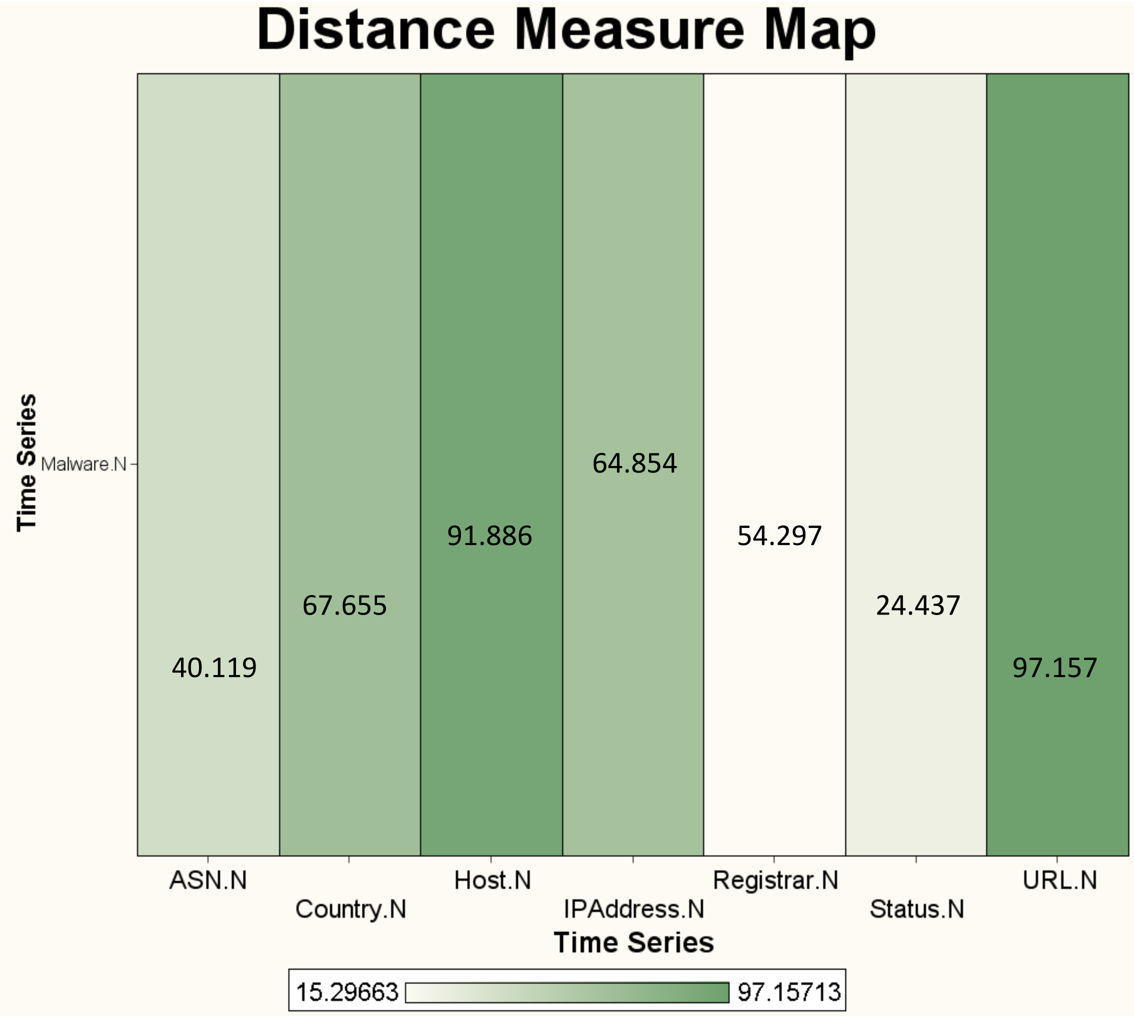


Figure 44: Distance Measure Map

Figure 43 is a similarity measure assigning real-values to the ransomware variables or contents and quantifying them for the determination of their content relevance, contribution and closeness with other objects in a given network. It compares each of the input variables against the target variable in the time series. The Similarity Measure: Malware versus Input Series (Figure 43) demonstrates the threat of malware in the cyberspace is similar and more likely to cluster, in the following order: Country, IPAddress, ASN, Status, and Registrar, respectively. The higher the similarity value of an object, the more likely it is clustered with the malware threat (object). Similarly, Figure 44 presents the distance map of the objects in the overlapping clusters. The values of the distance measure map also show how objects in the ransomware network (dataset) are likely to group together in a cluster. The distance map indicates that URL, Host, Country and IPAddress are more likely to cluster together in that order. While the map indicates that ASN and Status appear to cluster together, the Registrar appears to cluster separately.

5.1.3 Implications of the Components of the Time Series on the Development of Ransomware Threats

Time series analysis helps to identify trends, cycles and seasonal variations to aid in the forecasting (predicting) of future events using a collection of data input at specific intervals over a period. Time series analysis gives an understanding of the underlying forces leading to a given trend of events in the time series data point. It is useful in forecasting and monitoring the data points by fitting appropriate models to it. The percentages of variable (ransomware) events time series (Figure 30) clearly shows the evolving ransomware threat was stationary at almost 0.00% from the period, Jan2015 – Jun2016. The development (onset) of the threat increased to 8.67 – 12.91% across all the events variables in May2016. These values doubled to 24.96 – 29.17% in Sep2016 suggesting a more intense activity. The threat activities doubled further to 51.00 – 56.77% in Jan2017, indicating the period of active attack. The activity level dropped and became apparently stationary at 0.01 – 4.44% for the period May2017 – May2018. The original time series of the ransomware network data (Figure 31) reflects the percentage pattern description of the time series (Figure 30). The description is also true for the Trend-cycle components of the time series (Figure 32). Because the ransomware network data is cumulative, the Trend-cycle does not appear to be affected by season as shown by the Events Seasonal-Irregular Component (Figure 33), which appears to be stationary. This interpretation is corroborated by the stationary nature (regularity) of the Event Seasonal Component of the Network time series (Figure 34). This suggests that ransomware activities (threats) in the network are not influenced by seasons or periods but are purely ransomware traffic. The Events Trend-Cycle-Seasonal Component (Figure 35) mirrors exactly the Event Trend-Cycle Component (Figure 32) to suggest there are no seasonality in the ransomware network traffic activities. This means the only influence on the traffic is the onset of ransomware threat and the actual attack. The Events Seasonal-Irregular Component also shows stationarity in the amplitudes of the waves (Figure 33) to mean there was no extraneous factors, other than ransomware threat, in the network trend. The distribution pattern is seen in other components of the Time Series including Events Irregular Component of the Time Series (Figure 36), Event Trend Component of Network Time Series (Figure 37), Event Trend Cycle Seasonal Component (Figure 38), Event Seasonally Adjusted Series (Figure 40), Event Percent Change Seasonally Adjusted Time Series (Figure 41) and Event (Trend) Cycle Component (ECC) (Figure 42). The decomposed Time Series Components show the only influence on the ransomware network Time Series is the onset and real attack events of ransomware threat, which occurred between Sep2016 and Jan2017. The Similarity Measure (Figure 43) and Distance Measure Map (Figure 44) show the degree closeness and betweenness of the variables and how likely they can cluster together. The measure of the distance measure map shows that URL has a highest degree of 97.16. This is followed by the Host (91.89), Country (67.66), IPAddress (64.85), Registrar (54.30), ASN (40.12) and Status (24.44) respectively. The understanding that ransomware is the only influence in the Time Series will help in the profiling and validation of the overlapping clusters in the network time series in Chapter 6.

5.1.4 Summary

The results of the decomposed components of the Time Series has demonstrated that ransomware threat was the only influence in the time series of the ransomware network traffic. The knowledge of this trend will be valuable in Chapter 6 to profile, proof and validate the overlapping clusters in the network.

CHAPTER 6

DISCUSSION OF RESULTS AND SUMMARY

6.0 Introduction

The main concept of this investigation is to establish the dynamics and pattern consistency in the active cluster overlap node and to show the effectiveness of its removal to dislodge impending ransomware threats. Focusing on the aim of this investigation, the achievement hinges on the relationship between the General Network clusters, and the traceable and consistent formation and development of the active cluster overlap node from one cumulative period to another period in the time series of the network. These periods are taken six-monthly cumulative from the first time-stamp in the first year to the last time stamp in the last year of the ransomware dataset. Different graphical plots of the ransomware network clusters, and other derived plots to establish the aims of this research will be presented and discussed. The discussions will prove the effectiveness of the principles of exploratory and machine learning approach to the identification of cluster overlaps in the control of ransomware threat, and the validation (evaluate) of its performance. The discussions are based on the results of the experiments and analysis, which were conducted. It is important to state, however, that the effectiveness of any system built from this principle in the wild is beyond the scope of this investigation. This will be left for future studies and implementation to characterize.

6.1 Visualizing and Exploring the Topology of Ransomware Network

The visualization of data has become indispensable in the business world. It also finds application in social networking and reveals hidden information. Visualization of the ransomware network data helps to reveal, not only the obvious structures but also the hidden patterns of the network. Simply said, it is highly informative, and gives a clear understanding of the interaction of events within the network. In the Introduction of Chapter one, this study presented the visualization (graph) of the ransomware communities' network (Figure 5). For ease of reference, the ransomware communities' network graph is represented (Figure 45).

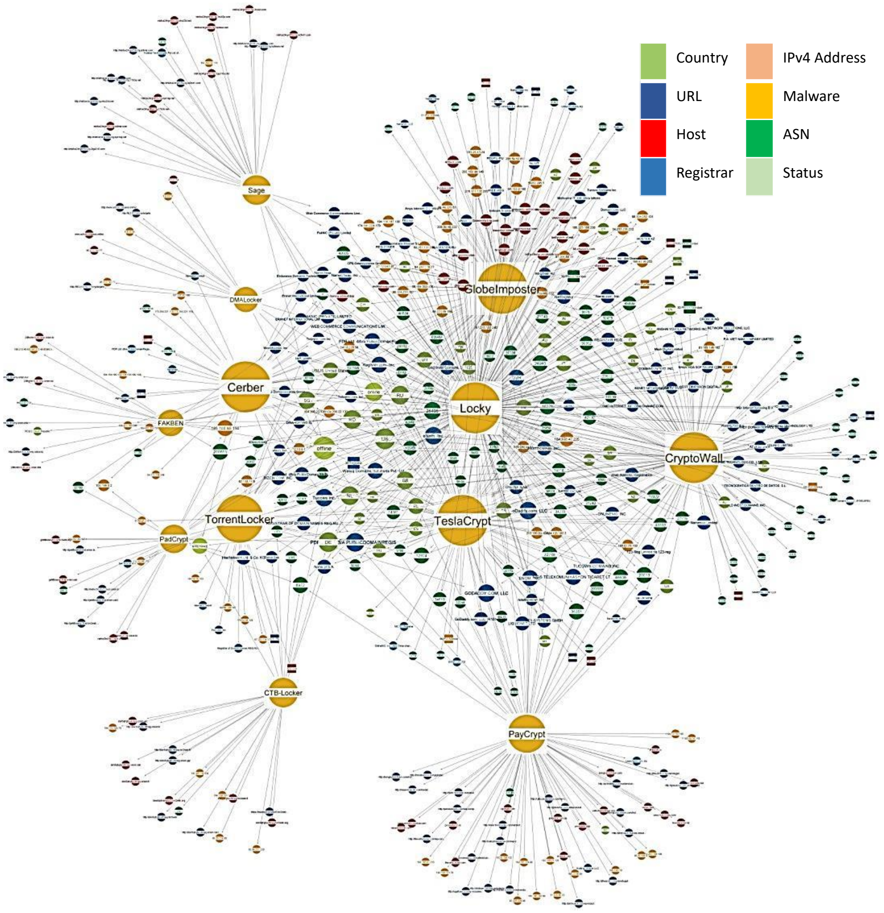
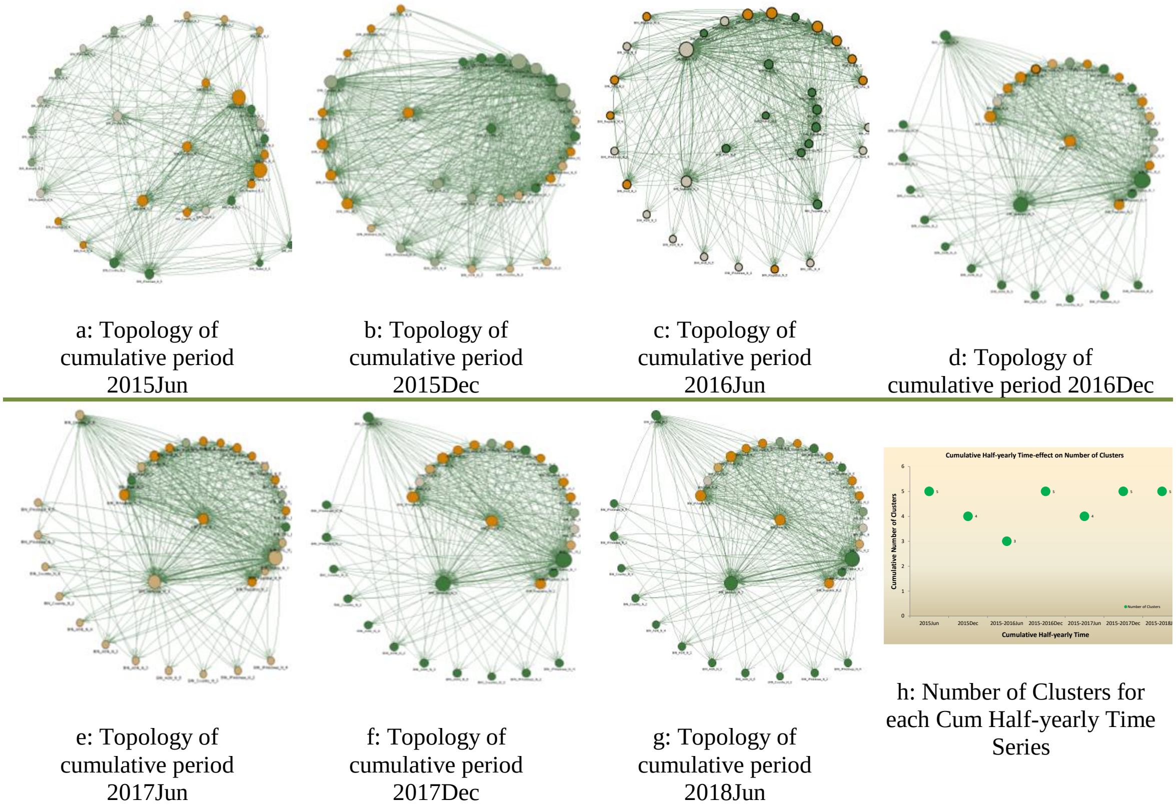


Figure 45: Ransomware Communities Network:
Showing cluster nodes of different ransomware families in the wild

The figure shows the topology of the ransomware network in the cyberspace (wild) prior to pre-processing as captured by the Maltego visualization tool. It is evident from the graph that there are clusters of twelve different ransomware families showing distinct partitions with various connecting links and distribution sources. The ransomware is in different families namely: Sage, GlobeImposter, DMALocker, Cerber, Fakben, and Locky. Other ransomware families include Cryptowall, TeslaCrypt, TorrentLocker, PayCrypt, CTB-

Locker, and PadCrypt. The contents and connecting links of the network come from the various entities such as the ransomware family, originating country of the ransomware, the Registrar, the Host, the URL (Universal Resource Locator), the status of the ransomware (online/offline), and the ASN (autonomous system number) of the ransomware.

Further processing and clustering show different structure (topology) of the Network. The cumulative half-yearly events' time series are plotted to reveal the topological development and changes between network clusters on different periods along the entire time series (Figure 29). For ease of reference and clear visual understanding, the structural formation and development of the clusters of the various cumulative half-yearly periods are shown in the "Progressive topology of cluster formation and development in the various cumulative time series of the Network (a-h)" (Figure 46a–h).


Cumulative Half-yearly Time-effect on Number of Clusters
Cumulative Number of Clusters
0
1
2
3
4
5
6
5
4
3
5
4
5
5
Number of Clusters
2015Jun
2015Dec
2015-2016Jun
2015-2016Dec
2015-2017Jun
2015-2017Dec
2015-2018J
Cumulative Half-yearly Time

a: Topology of cumulative period 2015Jun

b: Topology of cumulative period 2015Dec

c: Topology of cumulative period 2016Jun

d: Topology of cumulative period 2016Dec

e: Topology of cumulative period 2017Jun

f: Topology of cumulative period 2017Dec

g: Topology of cumulative period 2018Jun

h: Number of Clusters for each Cum Half-yearly Time Series

Figure 46: Progressive topology of cluster formation and development in the various cumulative time series of the Network (a-h)

The progressive topology of cluster formation and development in the various time series of the network (Figure 46a–h) visualizes the structural changes in the network clusters that occurred during the various cumulative periods of the ransomware network time series. Without an explanation at this point, the sequence of these graphs suggests that the number of clusters continues to decrease until it reaches a minimum number (Figure 46c, h) suggesting an onset of a threat. Following immediately is a spike in the number of clusters to a maximum suggesting a full-blown active ransomware attack (Figure 29, Figure 30 and Figure 46d, h).

The cluster graph of the first half-yearly period of the time stamp 2015Jun is displayed in Figure 47a. Contrary to the Ransomware Communities Network: Showing cluster nodes of different ransomware families in the wild (Figure 45), the node-links connections and density distributions (Figure 47a) reveal that the formation of the clusters is based on the items and contents of the network, including ransomware. The clustering was not just around the ransomware families as was evident in the community's network in the open cyberspace (wild). The elements of the network are converted to bin items prior to clustering. However, the number of clusters cannot be determined from this plot until the presentation of the plots, which show:

- The number and size of the clusters for the cumulative half-yearly period 2015Jun (Figure 47b)

It is clear from the "Network cluster graph showing (a) the node-links connections and density distribution, (b) Number and size of clusters for cumulative half-yearly period 2015Jun" (Figure 46b), that this period has five clusters, apparently because of the weight values of the clusters as shown in the key performance parameters for the period ending 2015Jun (Figure 48). To achieve a good representation plot, the gap between the highest and lowest actual average values of the key performance parameters for all cumulative periods are normalized using the following relations:

$$KPP = 6 + \log(x) \qquad Eqn(3)$$

$where\ KPP\ is\ the\ derived\ key\ performance\ parameter$

$x\ is\ the\ average\ value\ of\ the\ clustering\ elements, 6\ is\ a\ constant$

It is important to observe that the network is a directed network. The connections suggest an even distribution among the nodes with higher concentration (distribution density) on a few nodes. This is evident in the time series for the first half-year cumulative period Jan2015-Jun2015 (Figure 29). Summarily, the distribution density is reflected in the key performance parameters of the network and their corresponding percentage values (Figure 48, Figure 49 and Appendix 5) paying special attention to the following average values: cluster weight, out-degree centrality, weighted eigenvector centrality, weighted closeness centrality, weighted betweenness centrality, and clustering coefficient centrality. A visual assessment of the cluster graphs (Figure 47a, b) cannot point to the dominant cluster node until further analysis is conducted.

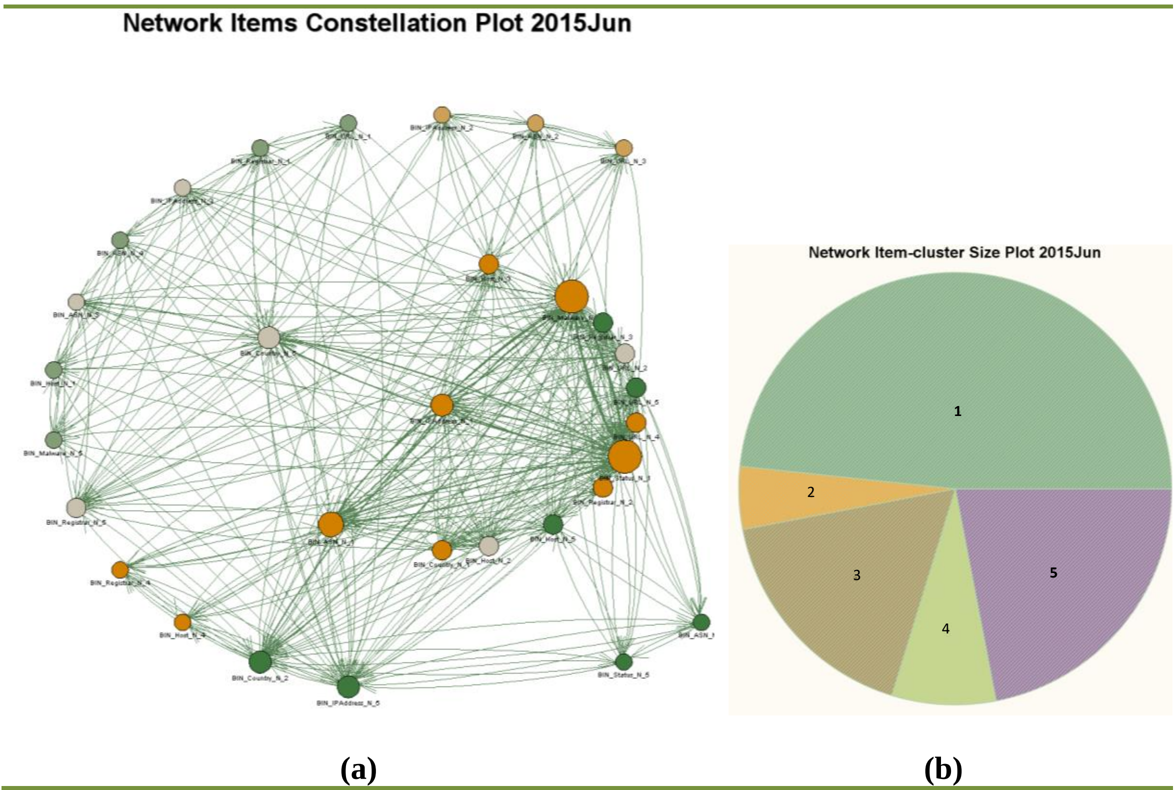


Figure 47: Network cluster graph showing (a) the nodes-links connections and density distribution, (b) Number and size of clusters for cumulative half-yearly period 2015Jun

The interest of this study is not the dominant cluster in the general network but the dominant (active) overlapping cluster, which would be discussed later in this chapter. At this point, it is difficult to understand the pattern (onset) of ransomware threat without the analysis of the behaviours of cluster formation and development of subsequent cumulative time series. However, the percentage values of the key performance parameters (Figure 49, Appendix 5) show that cluster node 1 recorded the highest percentage (16.89 – 84.96%) distributions in all the measures of centrality than those recorded by cluster nodes 5, 3, 2 and 4, respectively, in that order. The percentage values show, on average, a 34.33% distribution for all variable activities in cluster node 1 with the highest percentage distribution of 84.96% and 54.91% on weighted betweenness centrality and unweighted betweenness centrality respectively.

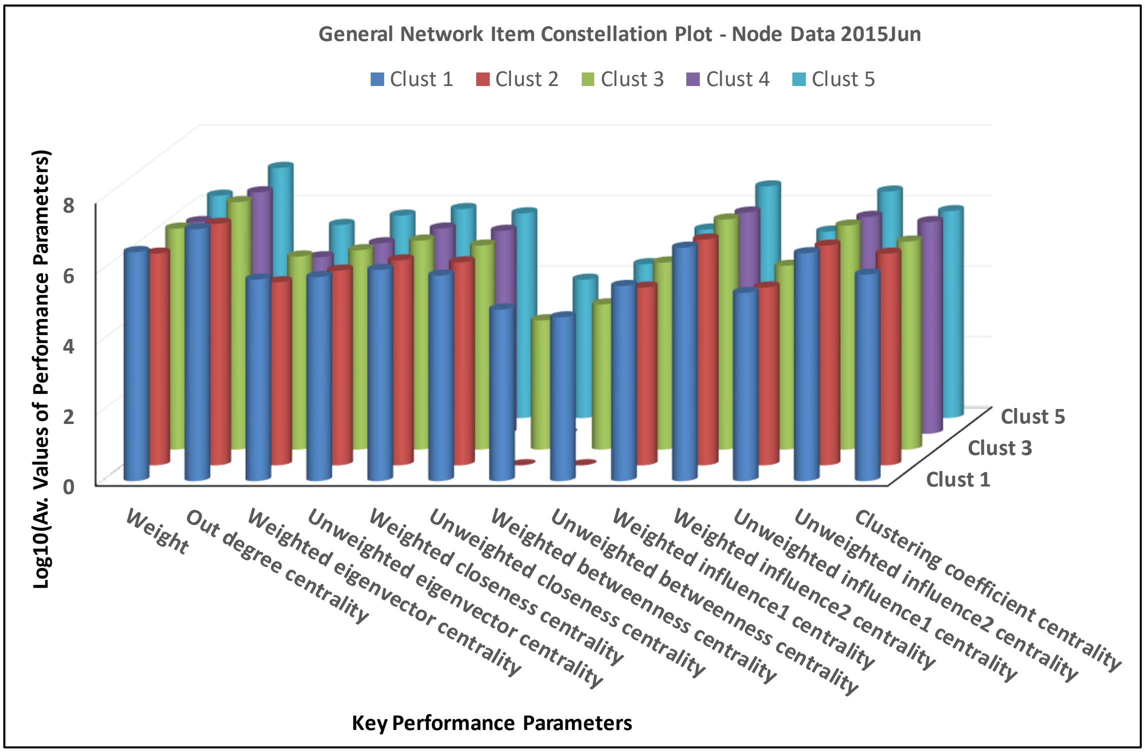


Figure 48: Key Performance Parameters General Network Item Constellation Plot - Node Data 2015Jun: showing Log10(av. values) of the various key performance parameters for cluster nodes 1, 2, 3, 4, and 5.

The percentage value distributions recorded in cluster node 1 include: Weight measure - 34.70%, Out degree centrality - 27.37%, Weighted eigenvector centrality – 37.83%, Unweighted eigenvector centrality – 27.79%, Weighted closeness centrality – 24.52%, Unweighted closeness centrality – 21.97%, Weighted betweenness centrality – 84.96%, Unweighted betweenness centrality – 54.91%, Weighted influence1 centrality 34.70%, Weighted influence2 centrality – 26.68%, Unweighted influence1 centrality – 27.37%, Unweighted influence2 centrality – 26.59%, and clustering coefficient centrality – 16.89%. These values are higher than the values for other clusters; and following next was the values for cluster nodes 5, 3, 2 and 4 in that order.

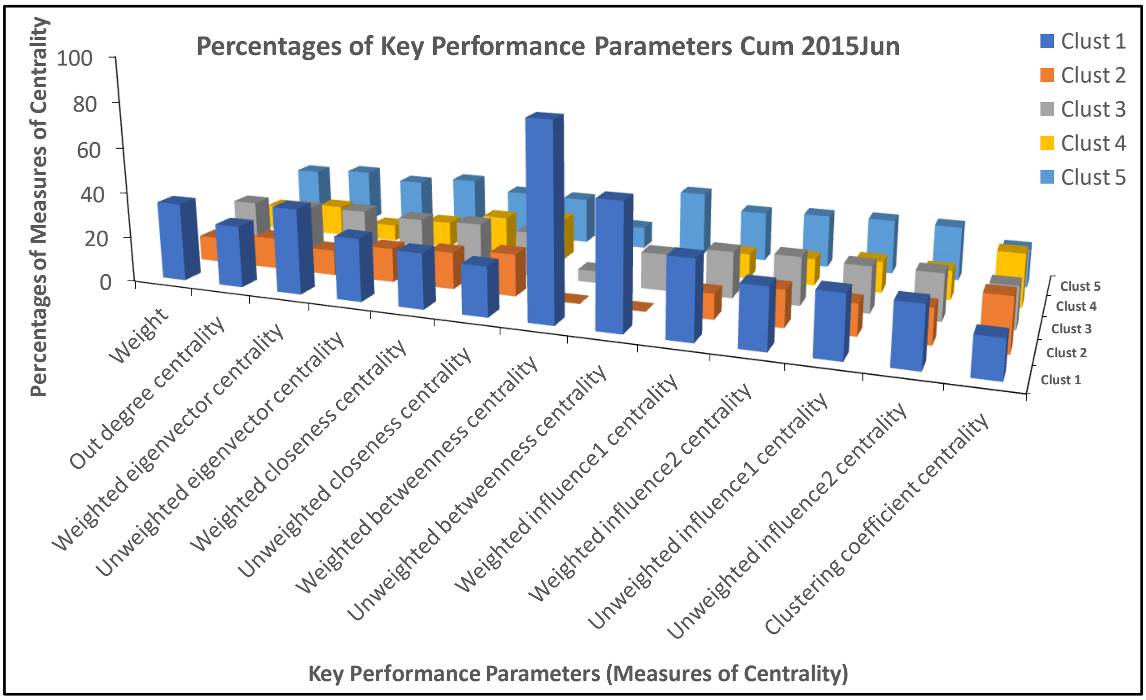


Figure 49: Percentage Values of Key Performance Parameters for Clusters 1 – 5 - General Network Item Constellation Plot - Node Data 2015Jun

The next stage of the ransomware events activities is represented in the cluster map for the cumulative half-year period ending 2015Dec (Figure 50a). A look at this plot (Figure 50a) shows a higher distribution density of the links than they were in Figure 47a. There is a higher distribution density of the links in all the measures of centrality resulting in reduced percentages ranging from 0.73 – 26.48% (Figure 51, Figure 52 and Appendix 6) in cluster node 1 compared to 16.89 – 84.96% that was recorded in Figure 47a, Figure 48, Figure 49 and Appendix 5. The loss of cluster node 5 in this period resulted in higher density distribution in cluster nodes 3 and 4 by percentage increases ranging from 21.82 – 99.04% and 0.06 – 26.67% respectively (Figure 52 and Appendix 6). This suggests that clusters 3 and 4 might have absorbed cluster node 5 based on pattern similarity. Unlike the node-links connections and density distribution for the period 2015Jun (Figure 47a), the links (connections) for the period 2015Dec (Figure 50a) appear to be more densely distributed in about 90% of the nodes, cluster nodes 1, 3 and 4, while a few 10% of the nodes, node 2, appear to be sparsely connected (Figure 50a).

Subsequently, the number of clusters formed (Figure 50a) for the second cumulative period ending 2015Dec is seen to be four (Figure 50b). This suggests that a slight increase in the number of events between Jul2015 and Dec2015 (Figure 29), increased the cluster average weight, thereby resulting in an increased distribution (cluster) density. Consequently, there was a resulting decrease in the number of clusters for the second cumulative period of the network given the slightly higher weight, weighted closeness centrality, and short distance betweenness centrality of the clustering items (Figure 51).

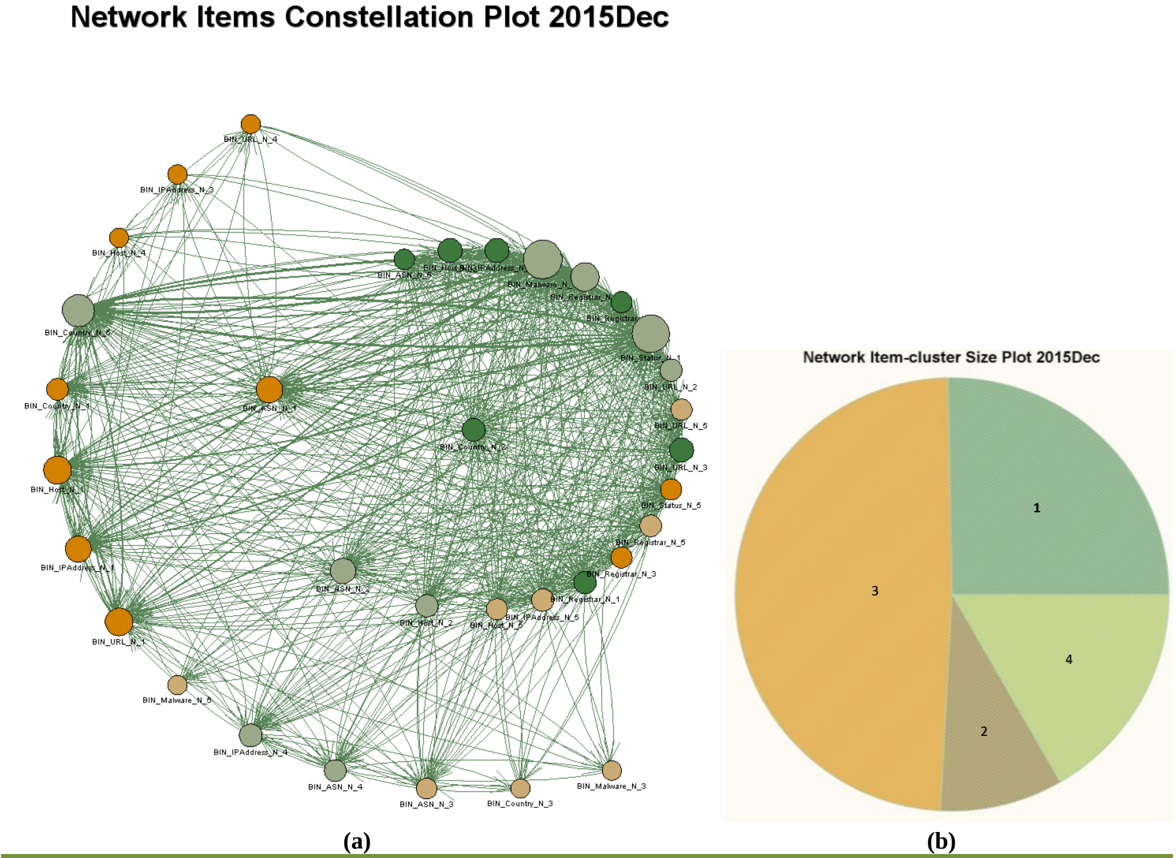


Figure 50: Network cluster graph showing (a) the nodes-links connections and density distribution, (b) Number and size of clusters for cumulative half-yearly period 2015Dec

The 47.33% and 20.88% increase in weight in clusters 3 and 4, respectively, affected their corresponding centrality measures: Out degree 31.99% and 25.94%, Weighted eigenvector

46.54% and 20.15%, Unweighted eigenvector 30.58% and 26.62%, weighted closeness 31.90% and 26.67%, Unweighted closeness 27.90% and 24.98%, Weighted betweenness 99.04% and 0.06%, Unweighted betweenness 54.02% and 16.72%, Weighted influence1 47.33% and 20.88%, Weighted influence2 28.76% and 26.14%, Unweighted influence1 31.99% and 25.94%, Unweighted influence2 30.31% and 26.57%, and Clustering coefficient 21.82% and 24.12% and short distance betweenness centrality of the clustering items (Figure 52 and Appendix 6).

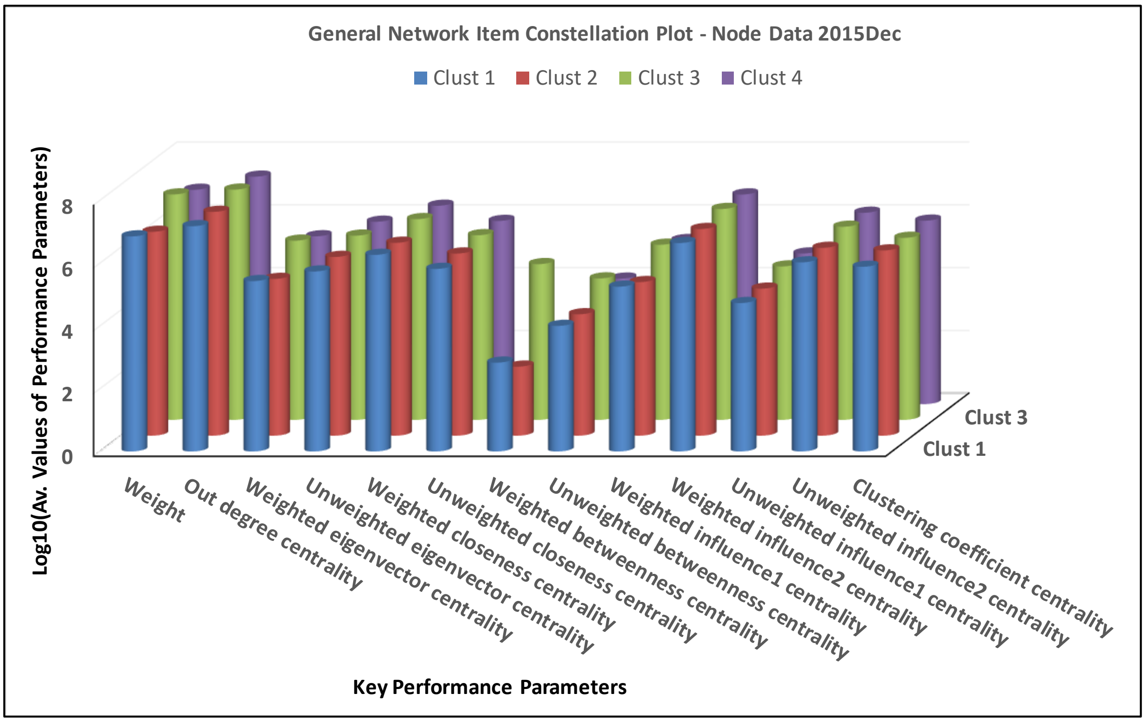


Figure 51: Key Performance Parameters General Network Item Constellation Plot - Node Data 2015Dec: showing Log10(av. values) of the various key performance parameters for cluster nodes 1, 2, 3, and 4.

The suggestion of higher distribution density is corroborated where the key performance parameters show slightly higher average values of 17.92 – 31.90% in closeness centrality among all the clusters. This contrasts with the values of 16.36 – 24.52% in closeness centrality (the key performance parameters) in the preceding period ending 2015Jun (Figure 48, Figure 49 and Appendix 5). It is also evident in the higher average weight values of 9.72

– 47.33% for each of the clusters in Figure 51 and Figure 52 compared with the values of 11.19 – 34.70% in Figure 48 and Figure 49. Furthermore, the percentage values of the average betweenness centrality of 0.06 – 99.04% and average clustering coefficient centrality of 21.82 – 27.58% show similar higher values in the 2015Dec period as against the respective values of 0.00 – 84.96 and 16.89 – 23.64% in 2015Jun period. This indicates the similarity of clustering-objects and distance between (betweenness centrality) clusters appear to be close.

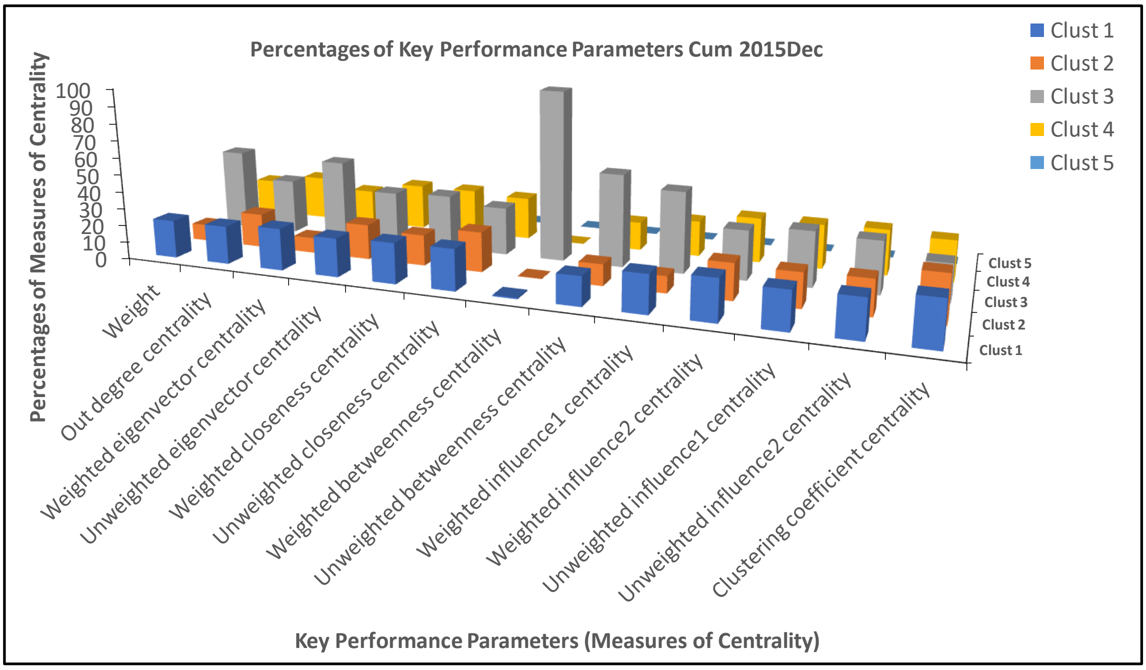


Figure 52: Percentages of Key Performance Parameters Cum 2015Dec Showing Percentage Distribution of the Link Connections

A more relaxed and sparsely distributed cluster is shown in the Network cluster graph showing (a) the node-links connections and density distribution (Figure 53a) for the third cumulative half-yearly time series ending 2016Jun. Still, fewer nodes are seen to be densely connected while others are more lightly and evenly connected. Consequently, the "Network cluster graph showing (b) Number and size of clusters for cumulative half-yearly period 2016Jun" (Figure 53b) shows a further decrease in the number of clusters to three as the number of events (observations) increased. The explanation for the decrease in the number of clusters could be seen in the "Key Performance Parameters General Network Item Constellation Plot - Node Data 2016Jun: showing Log10(av. values) of the various key performance parameters for cluster nodes 1, 2, and 3" (Figure 54, Appendix 7) for 2016Jun cumulative period.

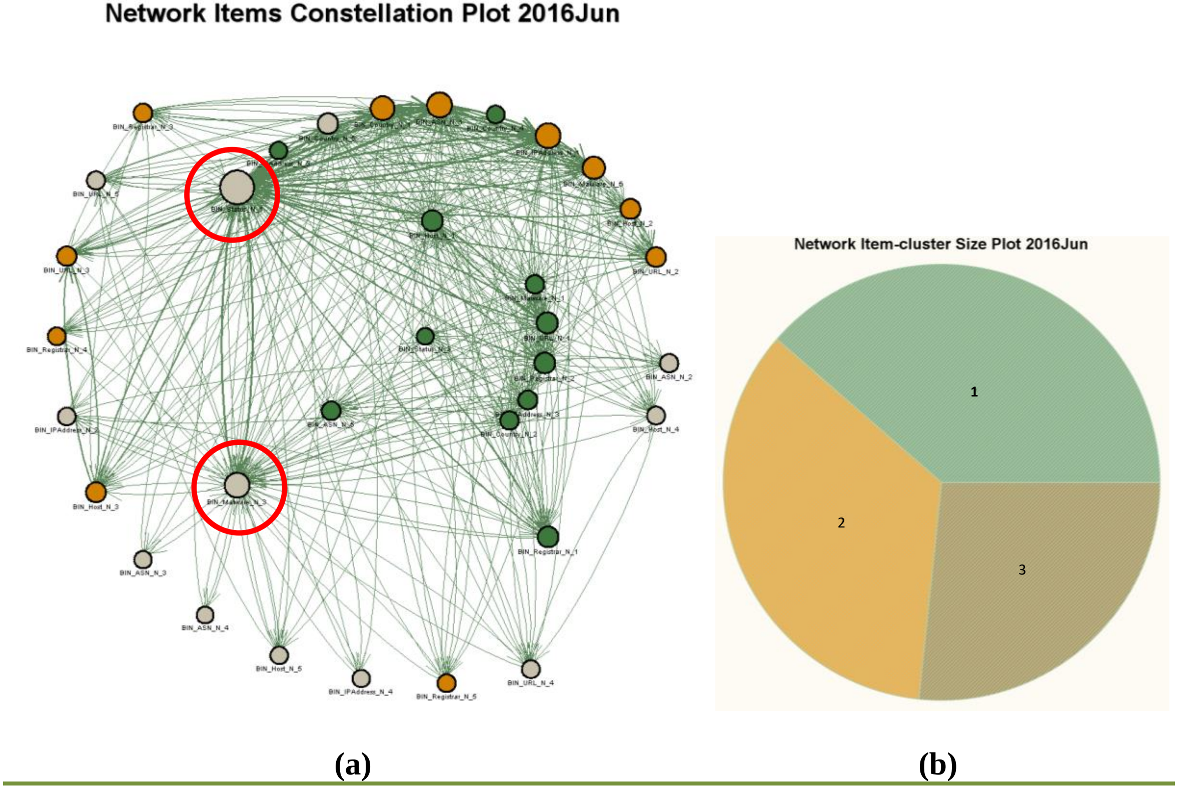


Figure 53: Network cluster graph showing (a) the nodes-links connections and density distribution, (b) Number and size of clusters for cumulative half-yearly period 2016Jun

The measures of centrality recorded high average weight of 27.58 – 39.88% and weighted closeness centrality 30.30 – 37.76% values for all the clustering nodes. The cumulative period 2016Jun recorded a clustering weighted/unweighted betweenness centrality of 0.00 – 94.40% and clustering coefficient centrality 32.22 – 34.08% compared to clustering weighted/unweighted betweenness centrality of 0.17 – 99.04% and clustering coefficient centrality of 21.82 – 27.58% in the previous period 2015Dec. These values are higher than those in Figure 48 and Figure 49 for 2015Jun and Figure 51 and Figure 52 for 2015Dec cumulative periods respectively. Another parameter that could explain the decrease in the number of clusters is the high average values of the clustering coefficient centrality, which ranks from 32.22% to 34.08% across the clusters. Notable are the values of the weighted closeness centrality (30.30 – 37.76%) and the unweighted closeness centrality (32.61 – 33.83%) in Figure 55. The values suggest that as the number of clusters decreases with the increased number of observations, the weighted closeness centrality increases while the unweighted closeness centrality appears to remain within the same range of values, indicating the build-up of threat and prompting action.

These results are evident in the minor spike of events activity in the time series between Jan2016 and Jun2016 in the plot of the “Multiple Time Series for all Variable Events in Ransomware Network” (Figure 29) and “Percentages of Time Series Events Activities Jan2015 – May2018” (Figure 30 and Appendix 4). The plots show the events activities increased from the range of 0.12 – 0.50% in Jan2016 to 8.67 - 12.91% in May2016, suggesting an average increase from 0.34% to 10.19% among all the network events. The point (2016Jun) where the number of clusters is seen to be at a minimum (Figure 29, Figure 30, Figure 53, Figure 55 and Appendix 4) suggests a change in the activity level of events within the ransomware network. Close observation shows a significant change, with an average increase of 10.19% activities in all the variables, especially in the connections to the BIN_Status and BIN_Malware nodes as highlighted in the cluster graph 2016Jun suggesting an onset of threat (Figure 30, Figure 53, and Appendix 4). Pre-judging the spike in the number of clusters for the cumulative half-year period 2016Dec (Figure 56a, b) it could be concluded that the minimum cluster threshold marks the onset of ransomware threat. However, conclusion can be drawn only after analysing the cluster formation and development for the period 2016Dec, when a pattern (trend) must have been established using three or more data points.

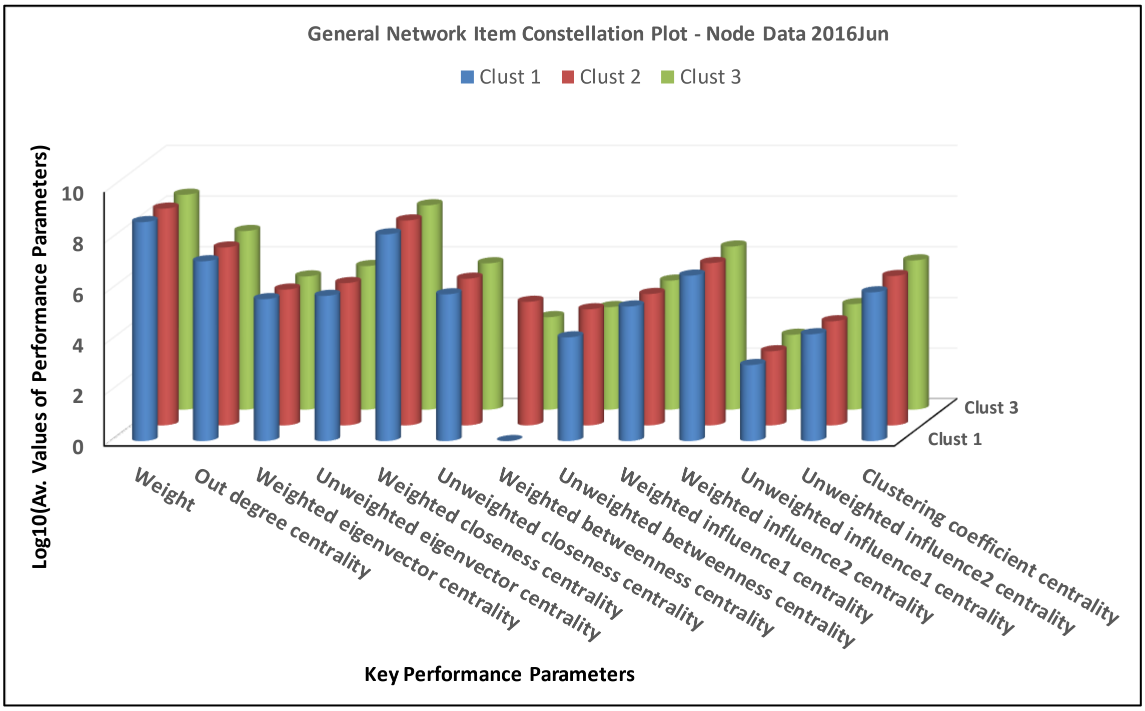


Figure 54: Key Performance Parameters General Network Item Constellation Plot - Node Data 2016Jun: showing Log10(av. values) of the various key performance parameters for cluster nodes 1, 2, and 3.

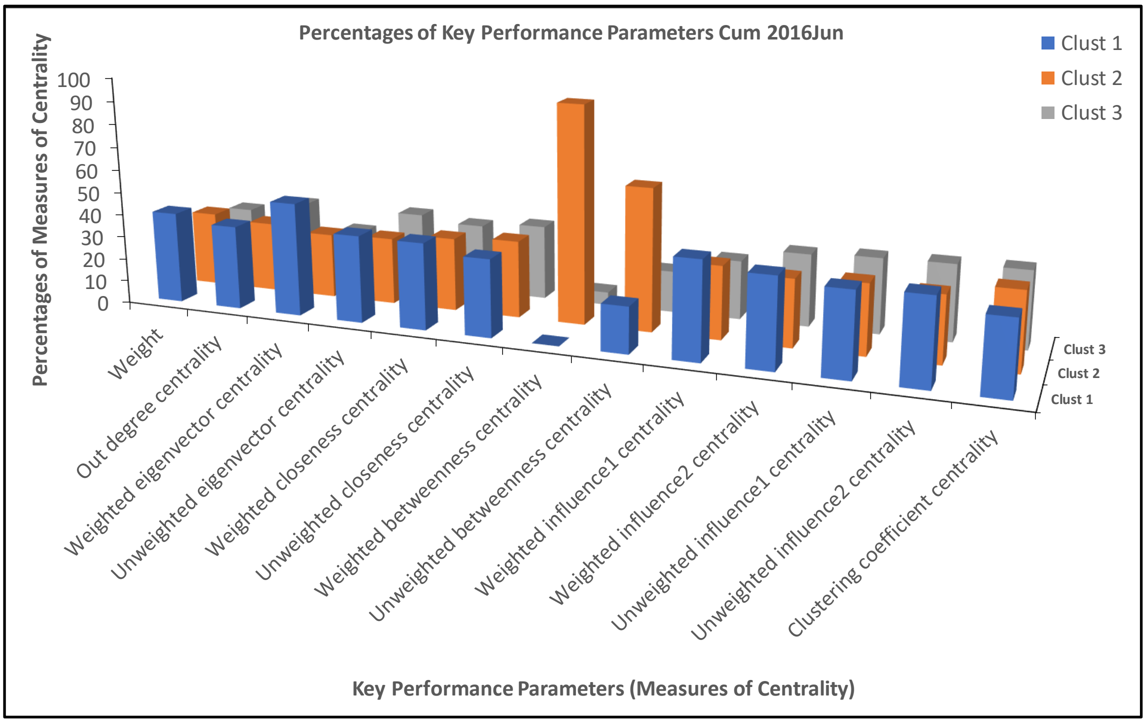
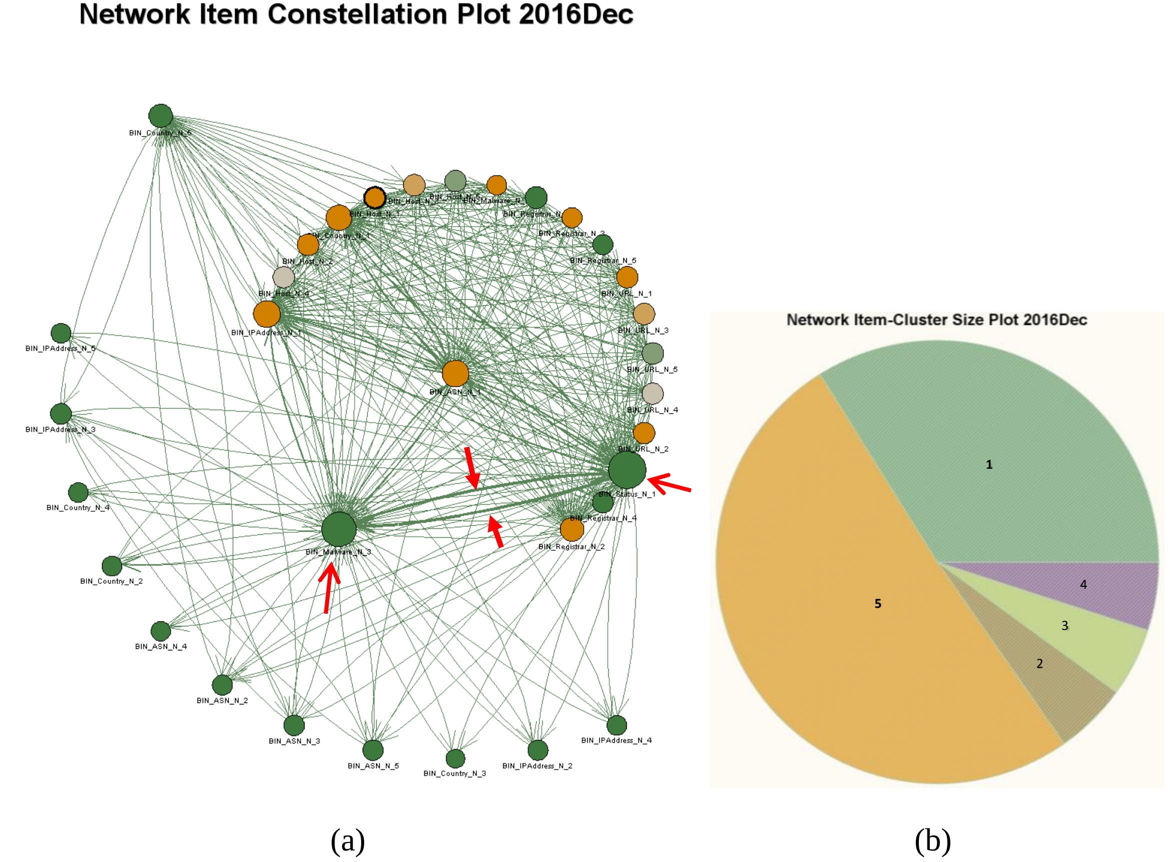


Figure 55: Percentages of Key Performance Parameters Cum 2016Jun Showing Percentage Distribution of the Link Connections

Exploring the graph of the fourth cumulative period of 2016Dec, the node-links connections and density distribution show further formation and development in the network clustering (Figure 56a, b) with an increase from 3 to 5 in the number of clusters. The network cluster graph (a) for 2016Dec shows a generally lower measures of centrality in the network than was observed in 2016Jun, except in cluster node 5 where the weighted and unweighted betweenness centrality recorded 100.00% and 69.76%, respectively. However, there is a higher density of links in a few of the nodes than there are in others.



(a) (b)

Figure 56: Network cluster graph showing (a) the nodes-links connections and density distribution, (b) Number and size of clusters for cumulative half-yearly period 2016Dec

The data for Sep2016 and Jan2017 in Figure 30 and Appendix 4 show a 24.96 - 29.17% and 51.00 - 56.77% increase in the link connections to all the variables, suggesting an average of 27.38% and 53.53% increase in network activities, especially, BIN_Malware (ransomware) and (ransomware) BIN_Status nodes among the few nodes that were most densely populated (connected) (Figure 56a). The two thick (bold) lines connecting BIN_Malware and BIN_Status suggest a strong link relationship between them. The high intensity connection strongly suggests an imminent active ransomware threat as shown earlier by the 53.53% increase in all the variable activities of the time series. The increases in the number of connections is evident in the distribution density reflected in the increase from 0.34% to 10.19% and 10.19% to 53.53% in the weight, out-degree centrality, weighted closeness centrality, clustering coefficient centrality, and other key performance parameters (Figure 57, Figure 58 and Appendix 8). It is interesting to note the same increase in the values of the weighted betweenness centrality in two cluster nodes relative to BIN_Malware and BIN_Status. At the period of active malware threat, it is important to note that all transmitting cluster nodes 1 (originating node), 2, 3, and 4 has 0.00% values for weighted betweenness centrality. This active threat period shows that terminal node 5 recorded 100.00% value for weighted betweenness centrality suggesting active transmission of threat unto network connected devices. These values show a great difference when compared with their corresponding values for the preceding cumulative period 2016Jun (the onset period of ransomware threat) where cluster nodes 1, 2 and 3 had 0.00%, 94.40% and 5.60% respectively, indicating active ransomware threat. This event activity corroborates the spike in events of the ransomware time-series plot (Figure 29, Figure 30 and Appendix 4). The increase in the number of clusters at this cumulative period could be an intended action to distribute ransomware through many URLs and IP addresses thereby infecting as many devices as possible.

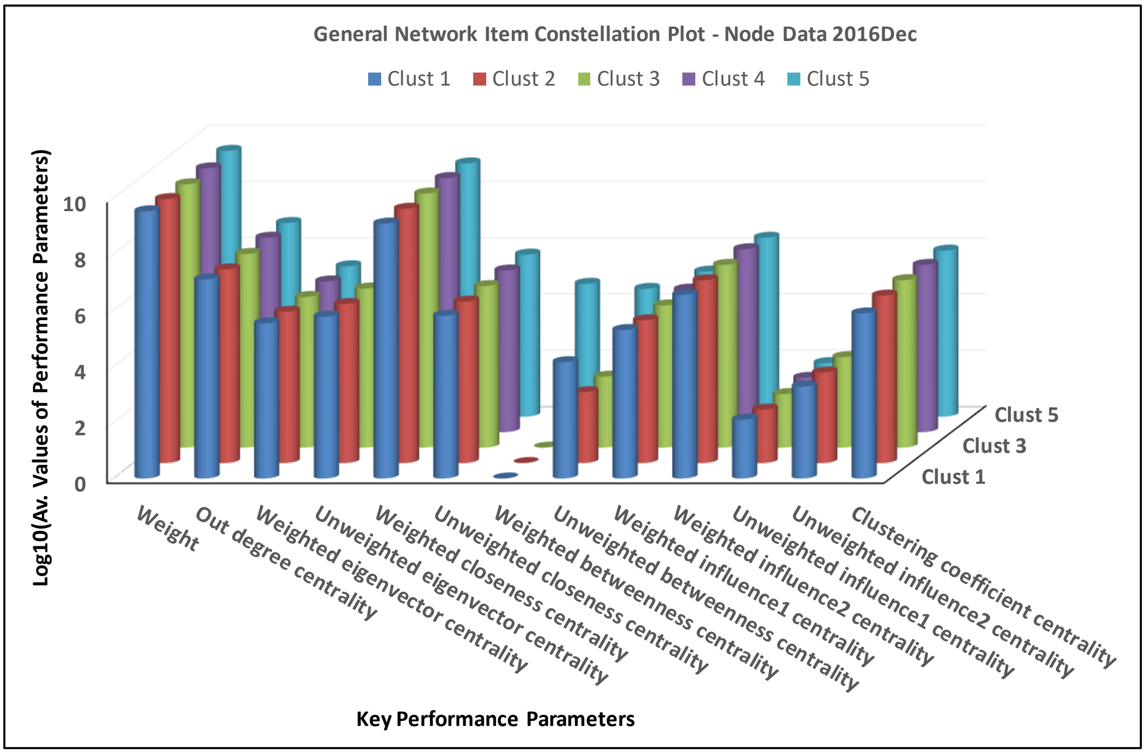


Figure 57: Key Performance Parameters General Network Item Constellation Plot - Node Data 2016Dec: showing Log10(av. values) of the various key performance parameters for cluster nodes 1, 2, 3, 4, and 5.

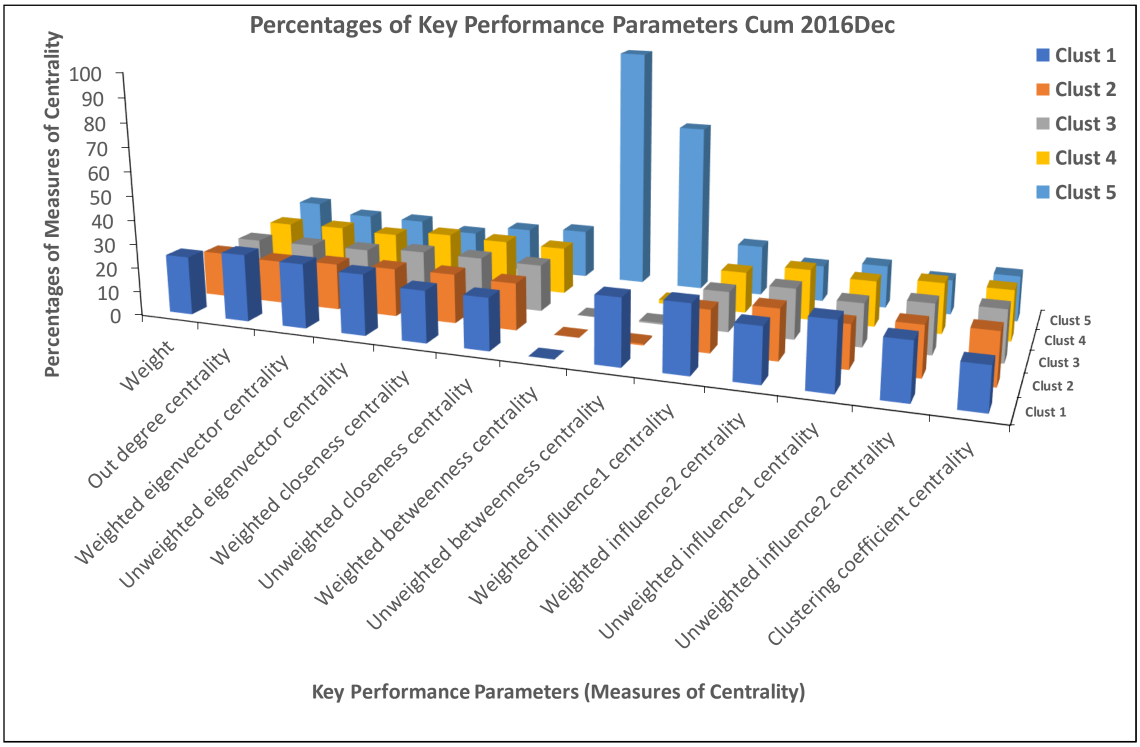
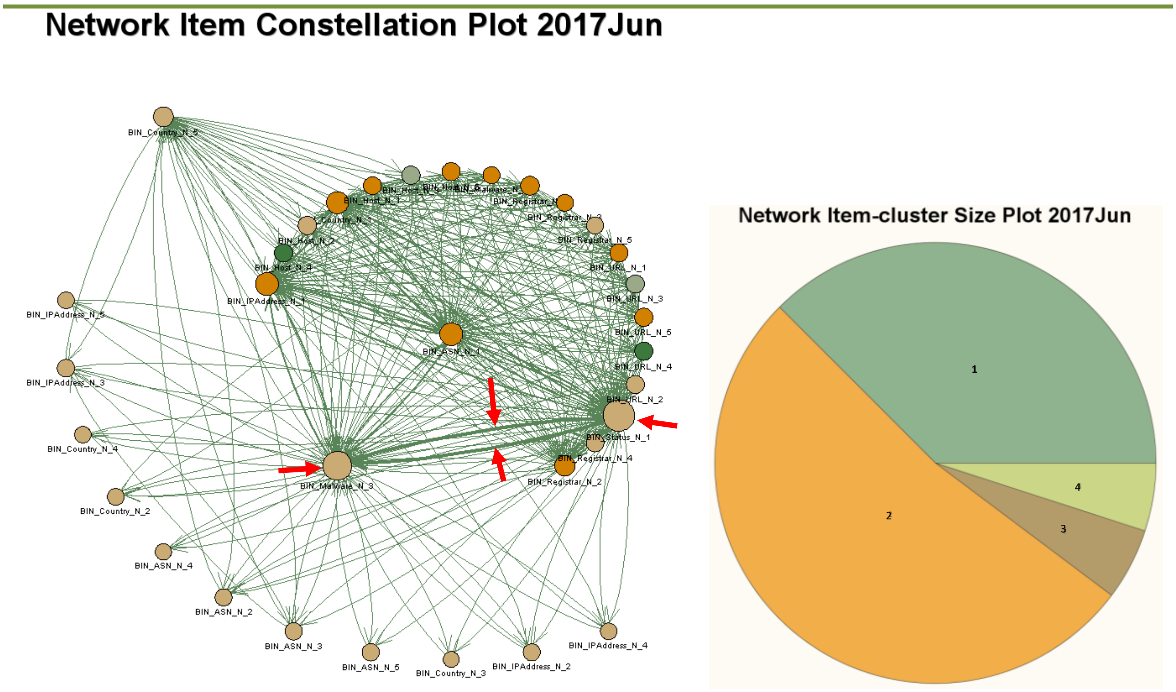


Figure 58: Percentages of Key Performance Parameters Cum 2016Dec Showing Percentage Distribution of the Link Connections

Analysis of further cumulative periods beyond the threat period 2016Dec reveals later event activities in the ransomware network. Principally, this shows whether the ransomware threat repeats after the initial attack. The 2017Jun cumulative period (Figure 59a, b) show a decrease in the number of clusters to four. Additionally, it shows an evenly connected nodes with about 50% of the cluster nodes more densely linked compared with others.



(a) (b)

Figure 59: Network cluster graph showing (a) the nodes-links connections and density distribution, (b) Number and size of clusters for cumulative half-yearly period 2017Jun

This is evident in Figure 60, Figure 61 and Appendix 9Appendix 9: Key Performance Parameters - General Network Item Constellation Plot - Node Data 2017Jun The number of clusters still decreased to four despite the increase in the number of events in the network. The increase in network activity is evident in the respective average percentage increase to 27.33%, 31.75%, 20.56% and 20.37% for clusters 1, 2, 3 and 4 (Figure 60, Figure 61 and Appendix 8) from 22.53%, 16.43%, 16.17%, 16.60% and 28.27% for clusters 1, 2, 3, 4 and 5 for the period 2016Dec (Figure 57, Figure 58 and Appendix 8). This means that clusters

1, 2, 3 and 4 of 2017Jun absorbed cluster 5 of 2016Dec because of closer pattern similarity, which could suggest the onset of another ransomware threat. However, like the node-links connections and density-distribution in the preceding cumulative period 2016Dec, there is still high weight and strong connections on the BIN_Malware (54.16%) and BIN_Status (51.00%) cluster nodes (Figure 59a). The connection is corroborated by the high weight of the cluster nodes recorded in the key performance parameters and Percentage Distribution of the Link Connections (Figure 60, Figure 61 and Appendix 9). The presence of such strong connections suggests there is a sustained threat or repeat of threat of ransomware after the initial attack. This suggestion is evident in the continued high out-degree centrality, which indicates the clusters have tightly connected nodes and ready to release threat. The out-degree centrality equally reveals the high importance score assigned to each of these nodes (BIN_Malware and BIN_Status) based on the number of links they hold (Figure 30 and Appendix 4). This shows that each of the two cluster nodes holds many direct connections to other nodes in the network showing active presence and distribution of ransomware threat through these nodes. The threat is also evident in the events percentage value of 100.00% recorded by the weighted betweenness centrality in cluster 2 while clusters 1, 3 and 4 recorded 0.00% each (Appendix 9). Like the out-degree centrality, weighted and unweighted eigenvector centralities show that these two nodes (BIN_Malware and BIN_Status) have significant influence on the network based on the number of links (54% and 51% respectively) they have with other nodes in the network (Figure 30 and Appendix 4). The eigenvector centrality identifies the nodes that have an influence on the entire network by determining the extended connections of the nodes. In the present network, the percentage values of weighted eigenvector centrality (32.01%) and unweighted eigenvector centrality (31.35%) show that cluster node 1 has the greatest influence in the network (Appendix 9).

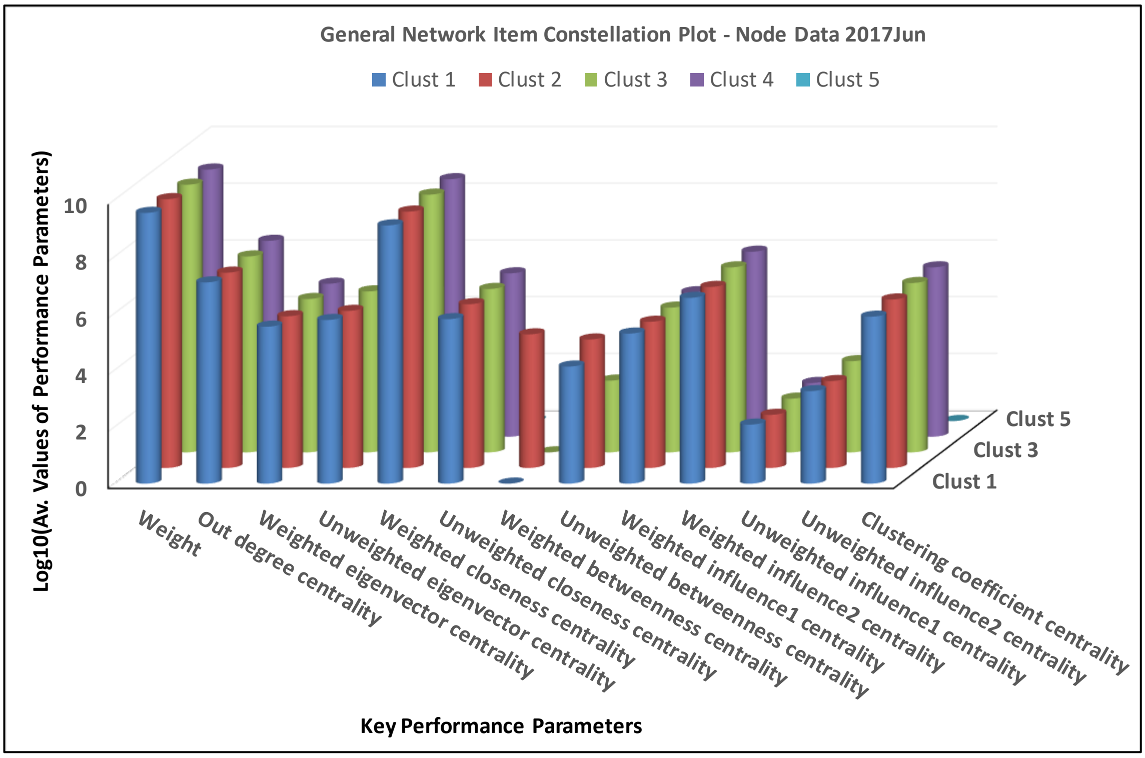


Figure 60: Key Performance Parameters for General Network Item Constellation Plot - Node Data 2017Jun: showing Log10(av. values) of the various key performance parameters for cluster nodes 1, 2, 3, 4, and 5.

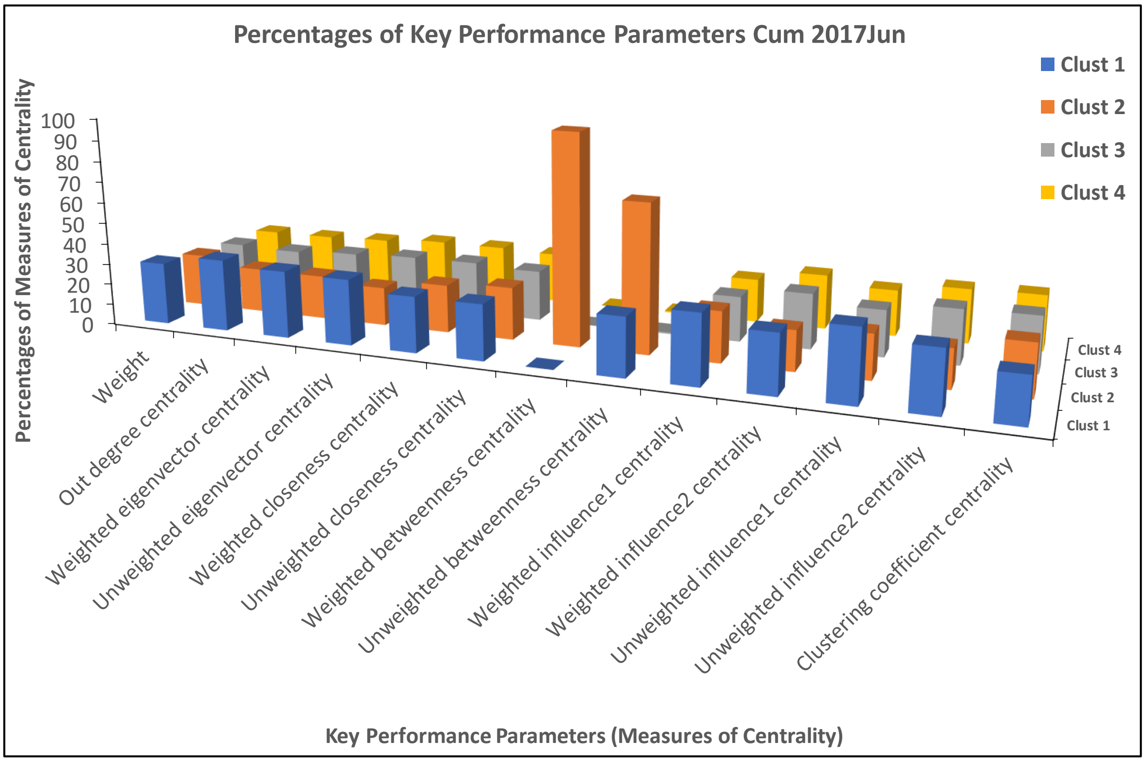
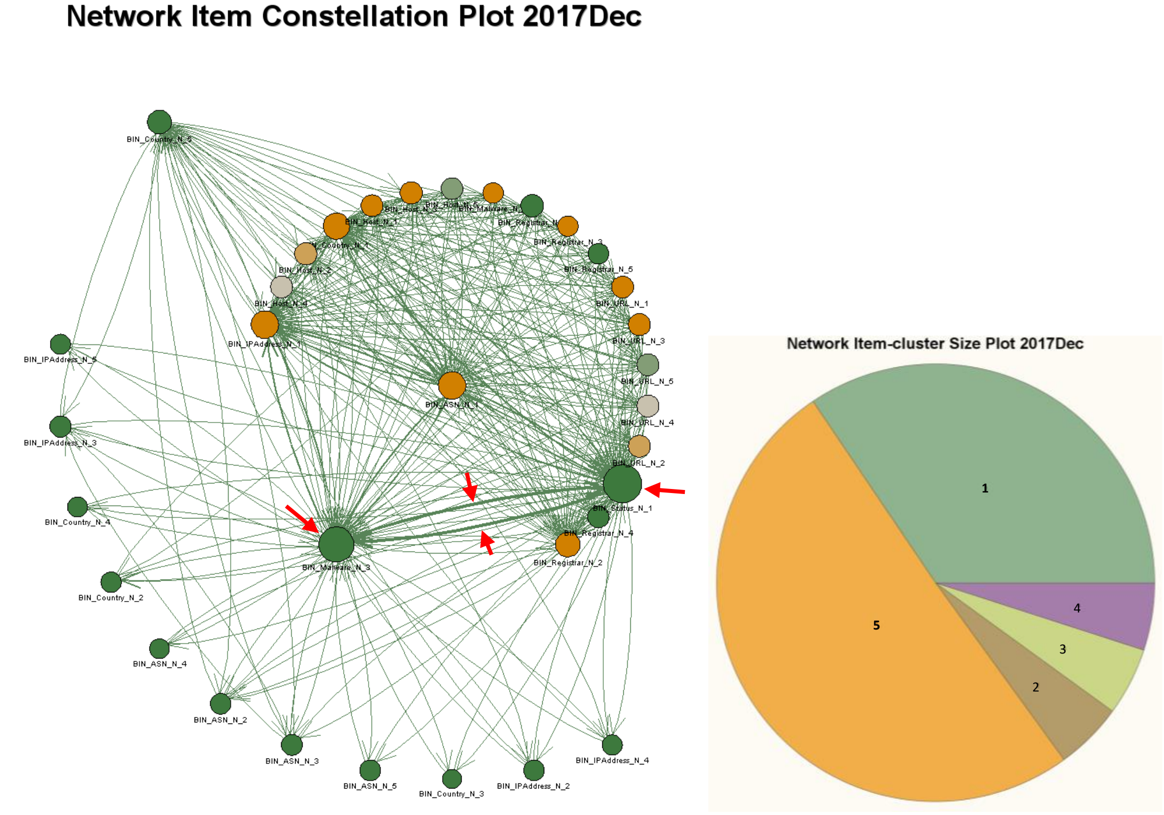


Figure 61: Percentages of Key Performance Parameters Cum 2017Jun Showing Percentage Distribution of the Link Connections

The high percentage of weighted (26.66%) and unweighted (26.59%) closeness centrality show the closeness of these nodes to other nodes based on the shortest path. This means that these nodes are well placed to influence the network and, therefore, be a major factor in the propagation of the ransomware threat. The clustering coefficient centrality demonstrates that the clustering nodes are densely connected. These values are of importance when determining the active cluster overlap in later sections. However, the drop in the number of clusters suggests the control of the spread of malware threats by direct intervention of some affected industries. The apparent repeat attack at this cumulative period indicates that any system that could effectively control the ransomware threat should have a self-healing mechanism and should run in a continuous cycle, and without this, there is significant risk of perpetuation of threats for undetermined and unconstrained time frames.

In the Network Item Constellation Plot for the cumulative period ending 2017Dec (Figure 62a), it is evident that the network is more sparsely distributed (Figure 63, Figure 64 and Appendix 10). The percentage average of all measures of centralities are 22.77%, 16.52%, 16.48%, 16.12% and 28.11% for clusters 1, 2, 3, 4 and 5, respectively (Figure 63, Figure 64 and Appendix 10). These values are lower than the respective values of 27.33%, 31.75%, 20.56% and 20.37% recorded for clusters 1, 2, 3 and 4 for the previous period ending 2017Jun (Figure 60, Figure 61 and Appendix 9). The connections are more even at the range of 16.12 - 16.52% with a few more densely connected nodes at a percentage range of 22.77 – 28.11%, BIN_Malware and BIN_Status (Figure 62a). The node-links connections and density distribution graph show a high volume of connections around a few cluster nodes, predominantly, around the same BIN_Malware and BIN_Status. Although there is a dispersion in the weight and density distribution of many cluster nodes, the number of clusters increased to five (Figure 62b). This indicates increased traffic in the activity level of the network, perhaps the entry and attack of a new ransomware into the network space.



(a) (b)

Figure 62: Network cluster graph showing (a) the nodes-links connections and density distribution, (b) Number and size of clusters for cumulative half-yearly period 2017Dec

Looking at the key performance parameters and their percentage values, except the increase in the number of clusters, all the measures of centrality (Figure 63, Figure 64 and Appendix 10) appear to have lower values compared with the preceding cumulative period 2017Jun. This indicates that similar ransomware activities are taking place in the network. Consequently, the ransomware threat appears to be persistent and self-propagating resulting in lasting impact and damage.

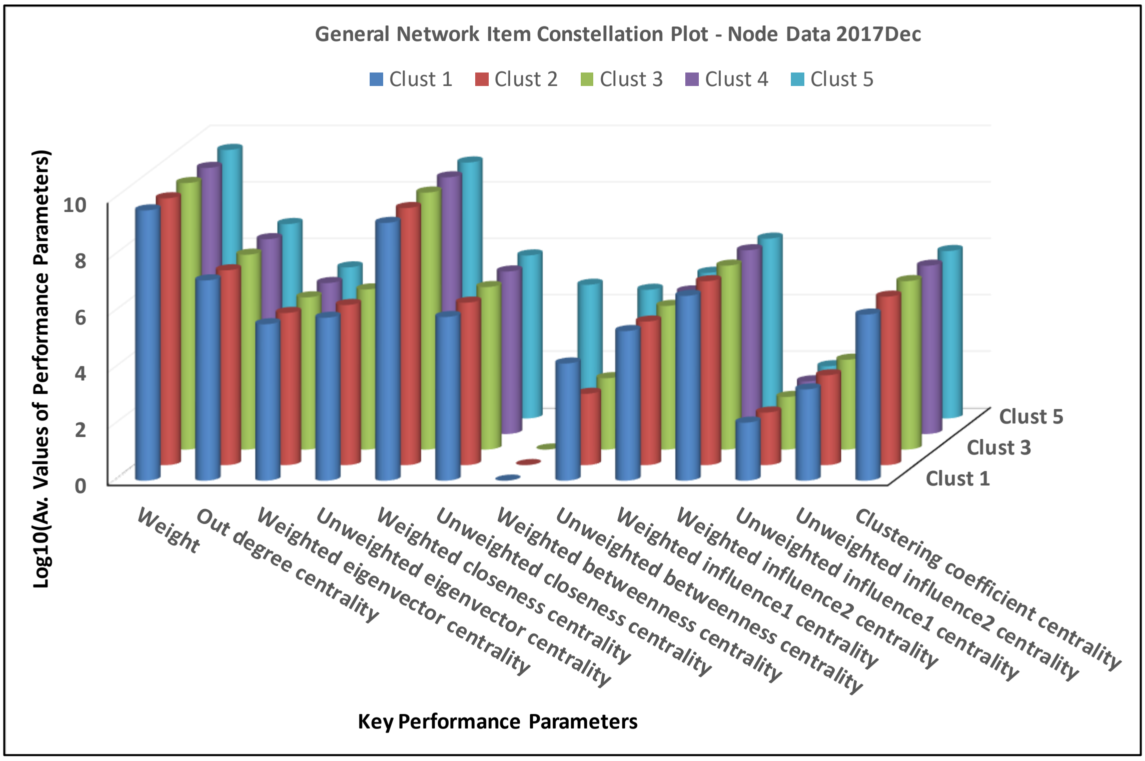


Figure 63: Key Performance Parameters for General Network Item Constellation Plot - Node Data 2017Dec: showing Log10(av. values) of the various key performance parameters for cluster nodes 1, 2, 3, 4, and 5.

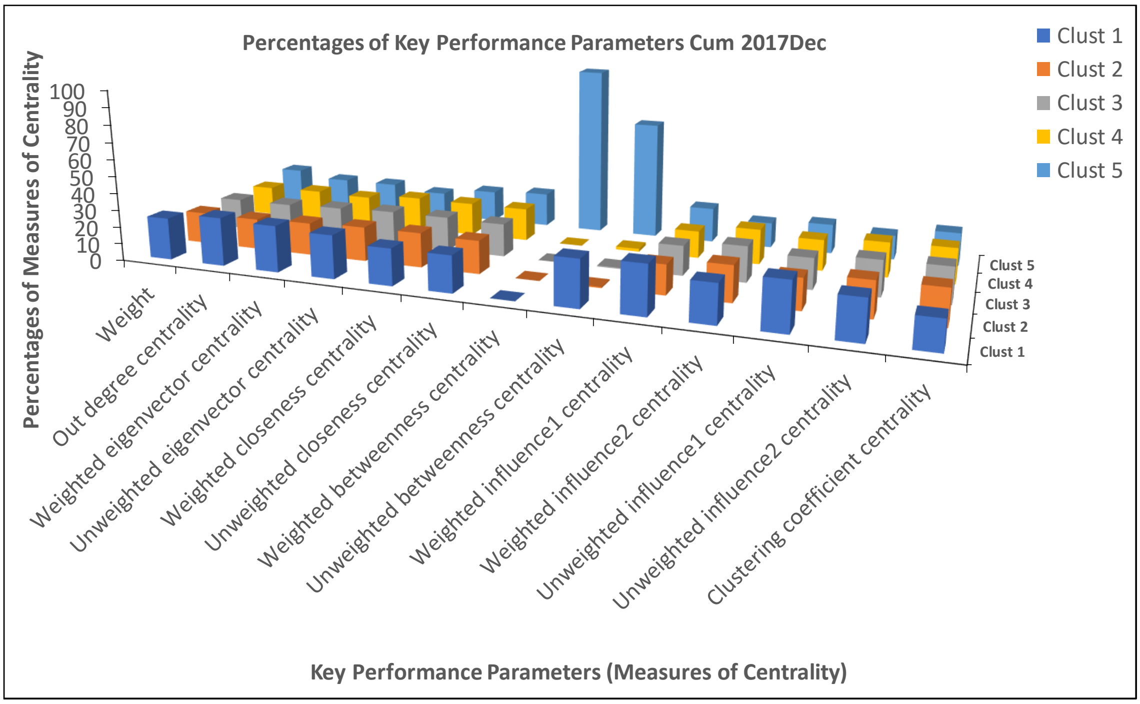


Figure 64: Percentages of Key Performance Parameters Cum 2017Dec Showing Percentage Distribution of the Link Connections

The last cumulative time series period 2018Jun (Figure 65a, b) does not show any visible topological difference with the preceding period 2017Dec (Figure 62a, b). The network structures appear the same in terms of node-link connections and density distribution, and the number of clusters. Clusters are distributed evenly with very few showing higher connection density. The number of clusters remains five as in the preceding period. The key performance parameters (Figure 66, Figure 67 and Appendix 11), also, look similar in all the measures of centralities. This suggests a stable and consistent state.

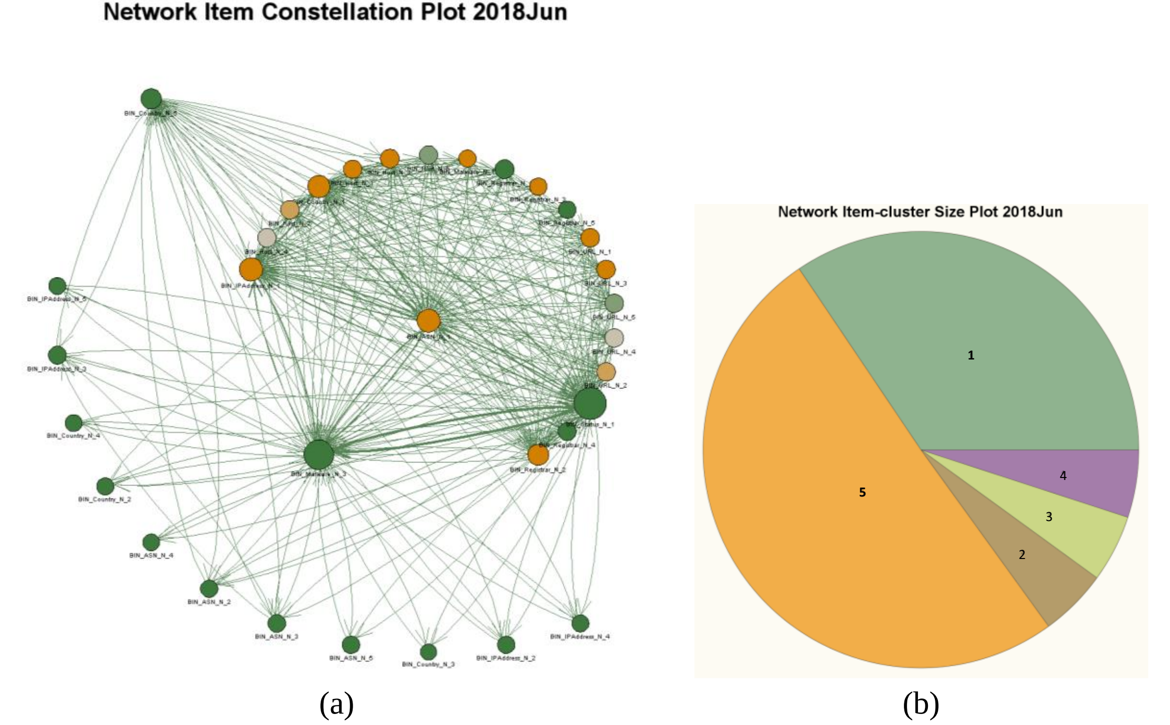


Figure 65: Network cluster graph showing (a) the nodes-links connections and density distribution, (b) Number and size of clusters for cumulative half-yearly period 2018Jun

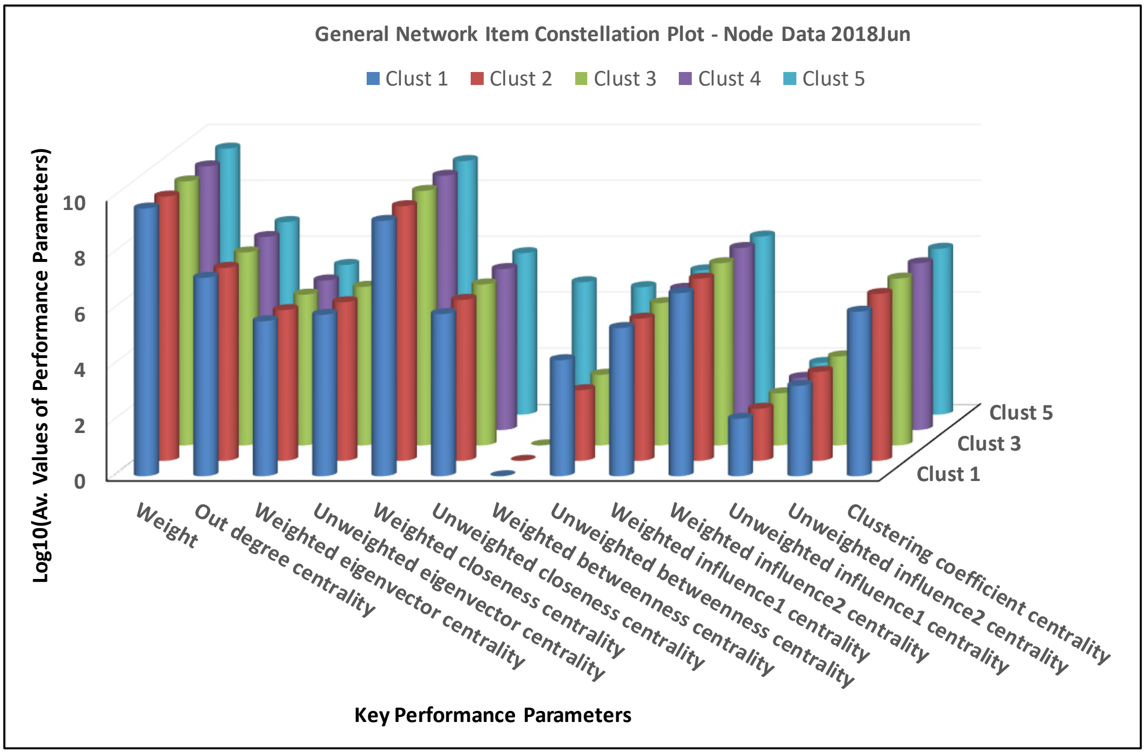


Figure 66: Key Performance Parameters for General Network Item Constellation Plot - Node Data 2018Jun: showing Log10(av. values) of the various key performance parameters for cluster nodes 1, 2, 3, 4, and 5.

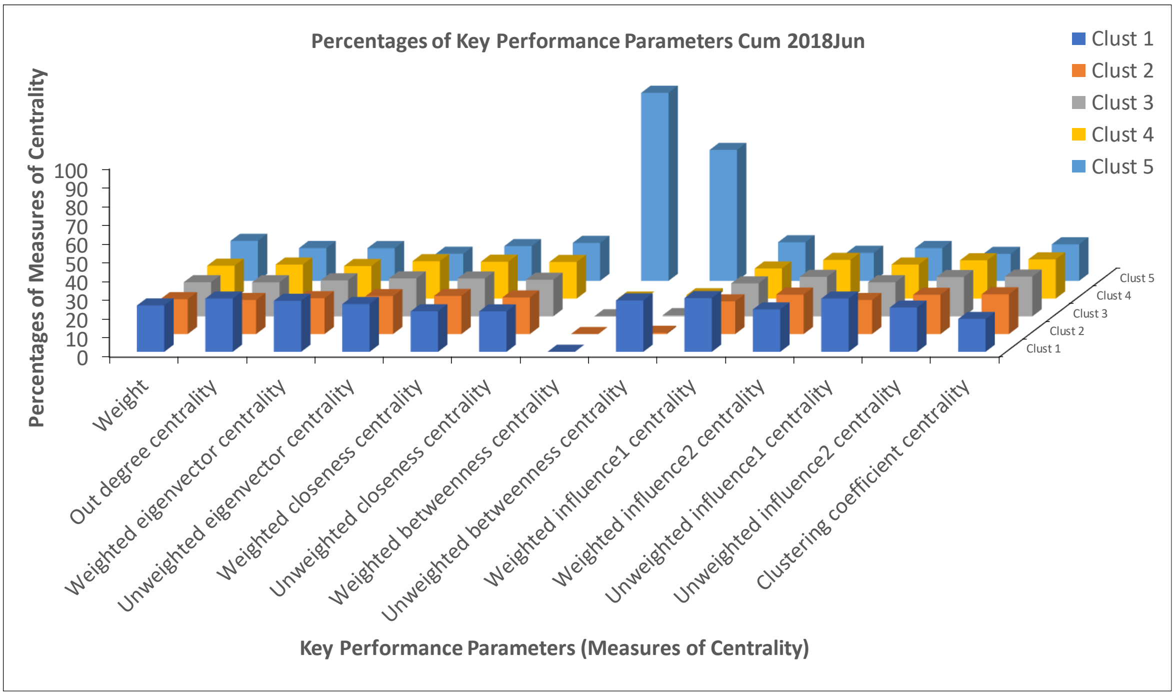


Figure 67: Percentages of Key Performance Parameters Cum 2018Jun Showing Percentage Distribution of the Link Connections

This section has established an understanding of the various stages in the formation and development of clusters in the ransomware network. The topological formation and development of clusters show that the network clustering nodes can disperse and form new clusters, onward propagating a persistent threat. Consequently, to have an effective control, a dynamic and self-healing (ransomware) control system that works in cycles to track the active cluster nodes each time it forms is essential. This research proceeds to investigate the clustering overlaps to identify the active cluster overlap to remove to dislodge any emerging ransomware threat.

6.2 Visualizing and Exploring Overlapping Clusters to Identify Active Cluster to Dislodge Ransomware Threat.

The preceding section established the pattern of the formation and development of clusters in a ransomware network. It reveals the role of key performance parameters of the links and contents of the network in the execution of ransomware threats, and identification of the attack profile. It established the onset of ransomware threat and the full-grown threat in the time series and the temporal pattern of the attack impact. The structure of the network reveals that the removal of the dominant cluster overlap is the key to controlling ransomware threat. This is because the dominant (active cluster) is the traffic generating cluster. It sends out all the outward transmissions and does not receive any incoming link. The structure of the network also established the need to track overlapping clusters and develop knowledge of the requirements of a system that can remove the active cluster overlap to dislodge any threat.

The current section investigates the dynamics of overlapping clusters in the network at different progressive and cumulative (period) time series (Figure 68 and section 6.1) to identify the most active cluster node that could be removed to dislodge a ransomware threat network and prevent the attack. Previous discussions have shown the clustering of these time series (ransomware) network to ensure the identified active cluster node to remove to prevent an attack or dislodge the network is the same at all cumulative periods of the time series. This means the active cluster overlap must be consistent across all the cumulative periods based on the cluster nodes intensities and other key performance parameters.

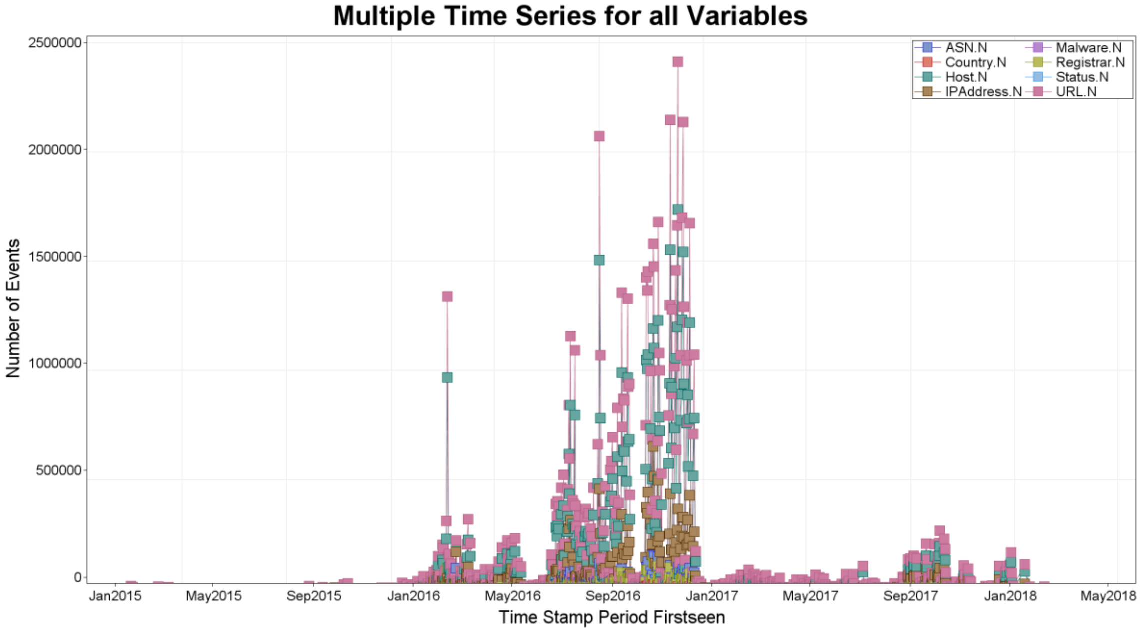


Figure 68: Multiple Time Series for all Variable Observations

The Number of Clusters for each Cumulative Half-yearly Time Series (Figure 69) gives a clear visual understanding of the time effect (or several observations) on the number of clusters within the cumulative half-yearly time series. The time series plot reveals a pattern of rise and fall in the event's activities. The plot of the number of clusters for each (Figure 69) reflects the same pattern exhibited in the time series plot (Figure 68). This suggests a decreasing number of clusters as the initial cumulative number of events increased. However, there was a sudden spike in the number of clusters between Jun2016 and Jan2017 (Figure 69) as was observed earlier in the plot of the time series (Figure 68). The sudden spike appears to suggest an event activity or an attack within that period. The number of clusters in the time series network appears to remain the same after this period, possibly because the input data for the next period (terminal period) did not change or there was not enough input event (data) in that period to cause structural change. There is a repeat of the same observation with the graphing of half-yearly (non-cumulative) events (Figure 70). Briefly, one of the challenges presented in the time series and number of clusters (Figure 68 and Figure 69) is to understand what happened along the entire time series, especially for the period Jan2016 to Jan2017; paying specific attention to Sep2016 to Jan2017. Section 6.1 has given an understanding of the events that took place along the time series in the ransomware network.

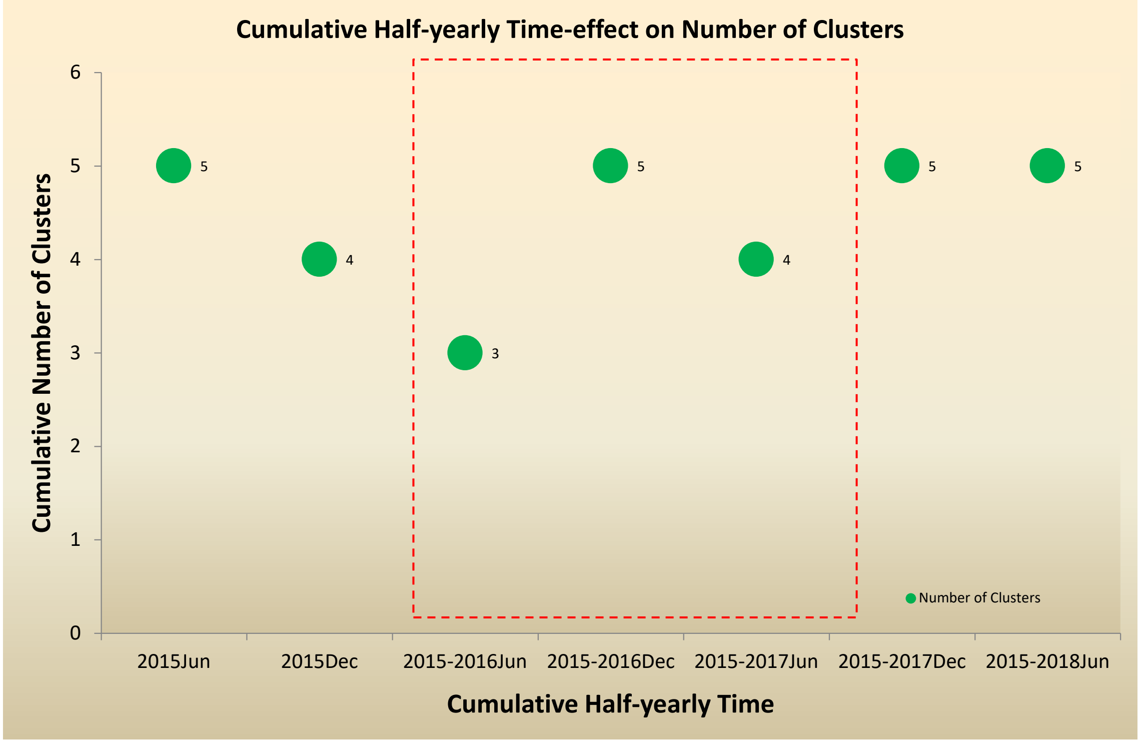


Figure 69: Number of Clusters for each Cumulative Half-yearly Time Series

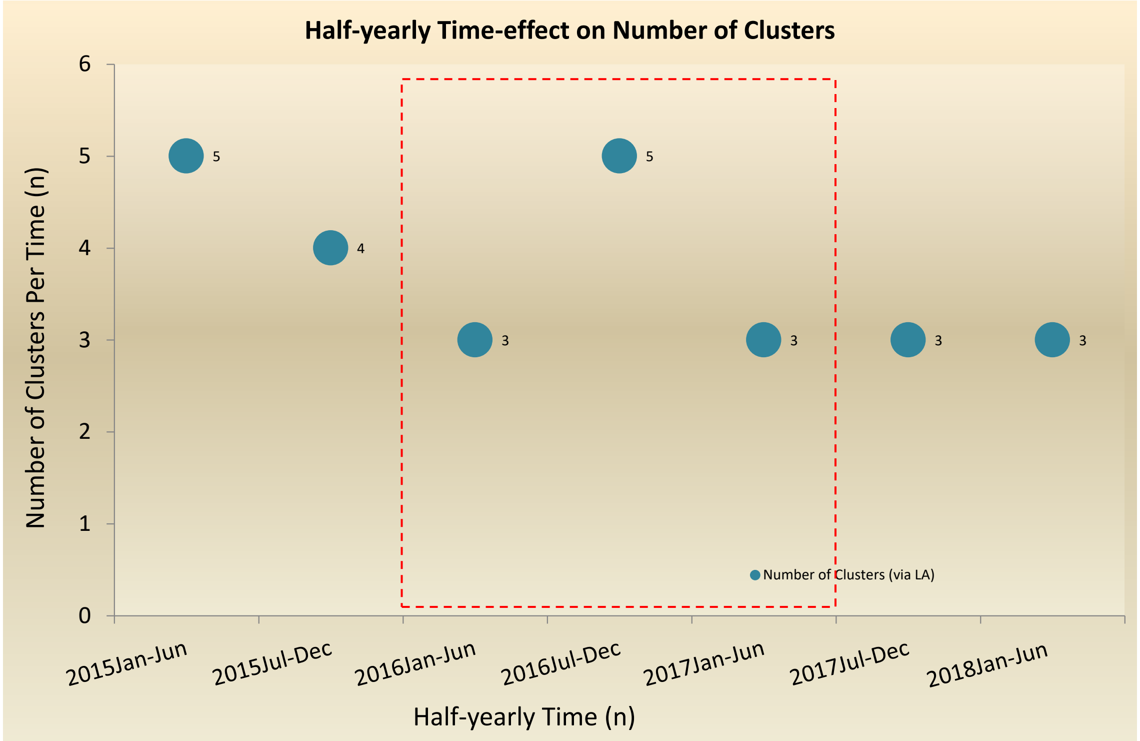
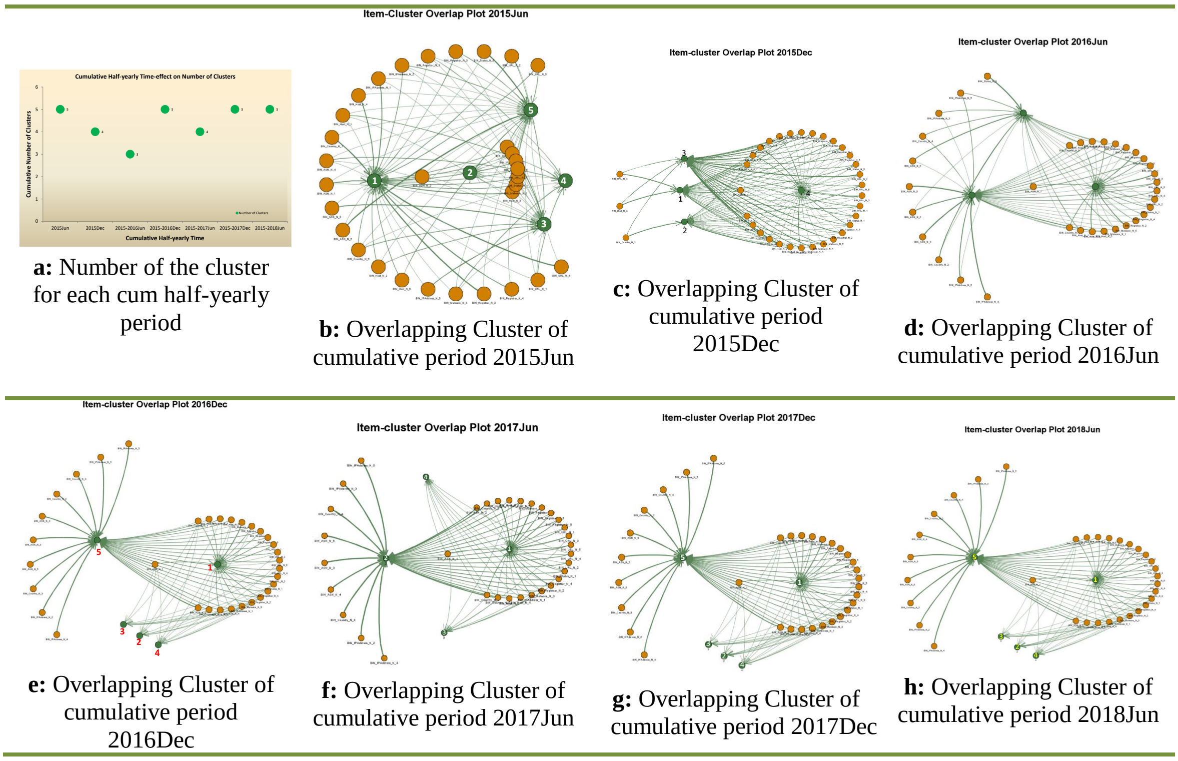


Figure 70: Number of Clusters for each Half-yearly Time Series

For further analysis to detect the active cluster overlap along the time series, cluster overlap maps were created for different cumulative half-year time-series. For ease of reference, a series map of the cluster overlaps for the various stages of the cumulative periods is presented (Figure 71b-h). The series map shows each (cumulative) periodic network cluster overlap, which was taken at the start of the time series January 2015 to the end June 2018. The map reveals that the periodic network of the overlapping cluster taken at the start of the time series, January to June 2015 has five overlaps (Figure 71b). Subsequent overlapping clusters (Figure 71c-h) show four, three, five, four, five and five cluster overlap, respectively, which are explained later.


Item-Cluster Overlap Plot 2015Jun
Cumulative Half-yearly Time-effect on Number of Clusters
Cumulative Number of Clusters
Cumulative Half-yearly Time
2015Jun
2015Dec
2015-2016Jun
2015-2016Dec
2015-2017Jun
2015-2017Dec
2015-2018Jun
Number of Clusters
Item-cluster Overlap Plot 2015Dec
Item-cluster Overlap Plot 2016Jun
Item-cluster Overlap Plot 2016Dec
Item-cluster Overlap Plot 2017Jun
Item-cluster Overlap Plot 2017Dec
Item-cluster Overlap Plot 2018Jun

a: Number of the cluster for each cum half-yearly period

b: Overlapping Cluster of cumulative period 2015Jun

c: Overlapping Cluster of cumulative period 2015Dec

d: Overlapping Cluster of cumulative period 2016Jun

e: Overlapping Cluster of cumulative period 2016Dec

f: Overlapping Cluster of cumulative period 2017Jun

g: Overlapping Cluster of cumulative period 2017Dec

h: Overlapping Cluster of cumulative period 2018Jun

Figure 71: Number of Overlapping cluster formation and development in various cumulative time series of the Network

Apparently, the overview of the plots shows the active clusters overlap, however, this will be clearer when the individual high-resolution view of each is analysed. The Network Item-Cluster overlaps plot (Figure 72) for the first cumulative period of the network shows five identifiable overlapping clusters (1, 2... 5) from the plot. The plot clearly shows the network has directed connections. However, the size, magnitude, and direction of the connecting links cannot be determined from this plot by mere visual observation. Therefore, the Cluster Overlap Map (Figure 73) shows the connecting links (edges) of the overlaps. This shows the feeder (traffic generating) cluster nodes and the nodes of the receiving cluster. This research has built the knowledge to understand and identify the active cluster overlap node and how its removal can help to dislodge the network and control the ransomware attack.

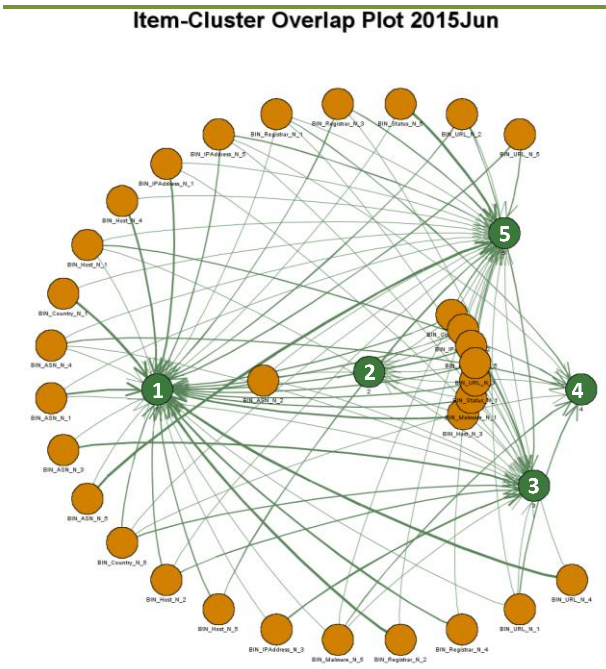


Figure 72: Network Item-cluster Overlap Plot 2015Jun

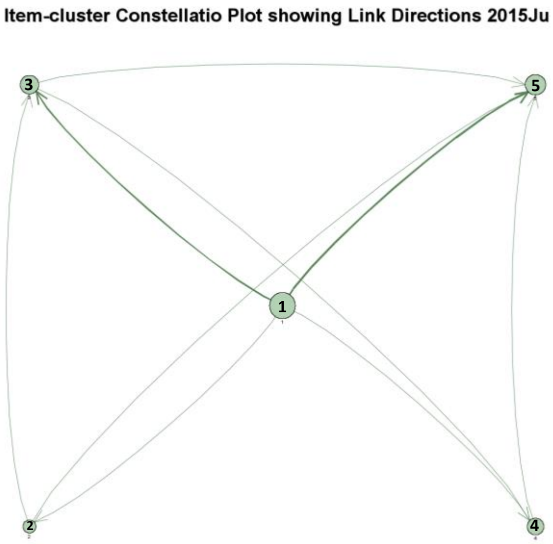

Item-cluster Constellatio Plot showing Link Directions 2015Ju
1
2
3
4
5

Figure 73: Cluster Overlap Map Showing Connecting Link Directions 2015Jun

The cluster-overlap map (Figure 73) reveals that cluster overlap node 1 has four outward connecting links and does not have any incoming links. Cluster overlap node 2 receives one in-coming connection from node 1 and sends two outward connections, one each, to nodes 3 and 5. In all, cluster overlap node 3 receives two in-ward connecting links from nodes 1 and 2, and sends two connecting links, one each, to cluster overlap nodes 4 and 5. Cluster overlap node 4 receives two in-ward links from nodes 1 and 3 and gives out one connecting link to cluster overlap node 5. At this period of the time series, node 5 can be seen to be a receiver-only node. It does not send any out-ward link. At this first period in the time series, it is correct to suggest that overlapping cluster node 1 is the dominant (active) cluster node or the events originating node, and consequently, the source of attack. The intensities of the overlapping cluster nodes 2015Jun (Figure 74) confirm the above understanding. The intensity values clearly show an apparent regular difference between the intensities of one clustering element (variable) and another. These intensities are related closely in values, and it can be seen too, that cluster node 1 appears to have a fully developed cluster with a very high clustering weight value (34.70%) and other measures of centrality ranging from 21.97-84.96%, indicating a potentially active source of attack (Figure 48, Figure 49 and Appendix 5). This argument is not true for the measures of centrality of clusters number 2, 3, 4, and 5 where the values range from 0.00-23.64%, 5.35-21.74%, 0.00-23.64% and 9.69-28.75% respectively.

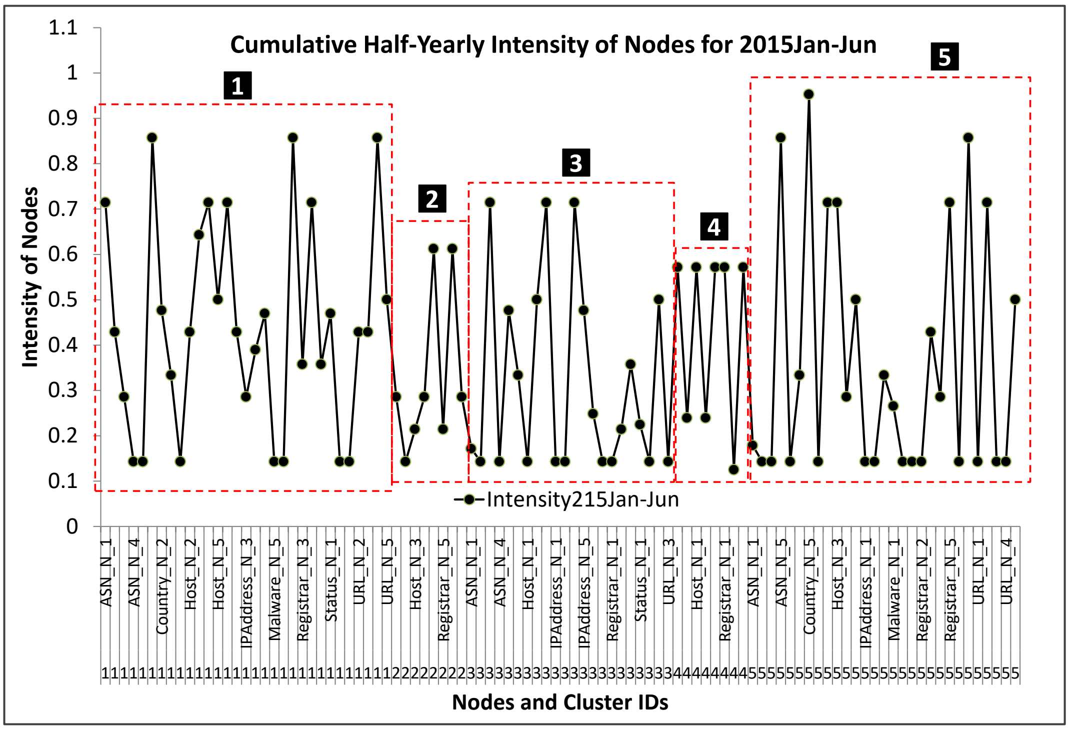


Figure 74: The Intensities of overlapping cluster nodes 2015June:
Showing the intensities of objects in cluster overlap nodes 1, 2, 3, 4, and 5

Figure 75 corroborates the above conclusion by showing that cluster 1 has the highest average intensity value of 43.97% for the cumulative period 2015Jun, while the average intensity values of clusters number 2, 3, 4, and 5 are 5.00%, 17.17%, 11.08% and 22.78% respectively with their individual intensities distributed wide apart. More importantly, the clusters have low cluster weights and weighted betweenness centrality and appear not to be fully developed (Figure 49 and Appendix 5). It is important to note here that cluster 5 has a clustering element with an intensity value equal to 1 suggesting the element was an outlier (alone by itself), with the potential to form or join a new cluster when there are objects with close similarities. In addition, it is important to observe that cluster node 5 is a terminal node with many incoming links and has the second largest average intensity value of 22.78%.

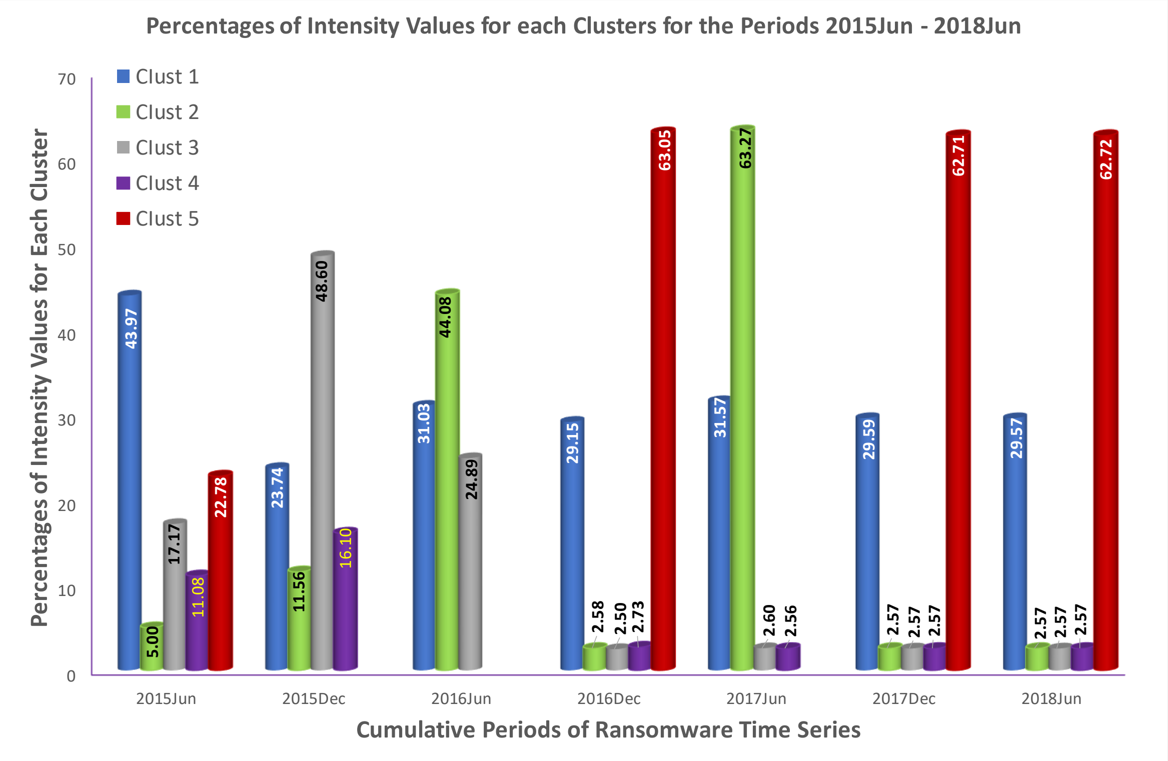


Figure 75: Percentages of Intensity Values for each Cluster for the Cumulative Periods 2015Jun - 2018Jun; the cylinders represent the Clusters as follows, Blue = Cluster 1, Green = Cluster 2, Grey = Cluster 3, Purple = Cluster 4 and Red = Cluster 5

The most active cluster node, cluster 1, has four outward links that feed all the other cluster nodes. It is the traffic originating link, which feeds other nodes, and receives from none. The thick dark arrows suggest that the major traffic routes originate from cluster node (links) 1 to 3 and 1 to 5 (Figure 73). It becomes apparent that its removal terminates traffic to other clusters as shown in the item-cluster overlap 2015Jun when the link and cluster node 1 is disconnected (Figure 76 and Figure 77), thereby effectively and immediately terminating the impact of the attack.

Item-cluster Overlap when Link Node 1 is Disconnected

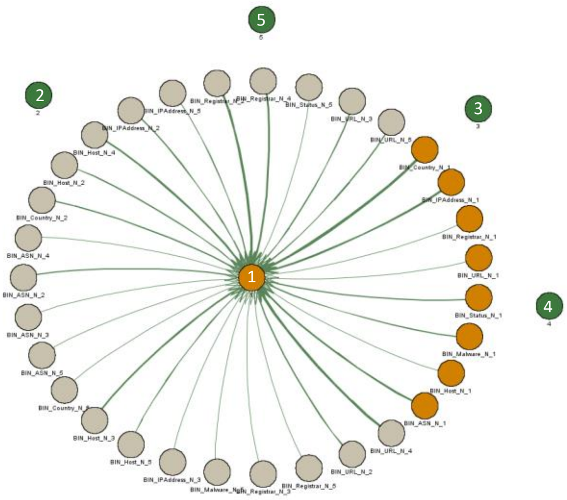


Figure 76: Item-cluster Overlap 2015Jun when Link Node 1 is disconnected

Item-cluster Overlap Plot when Cluster Node 1 is Disconnected

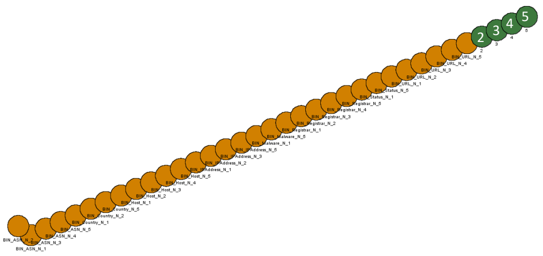
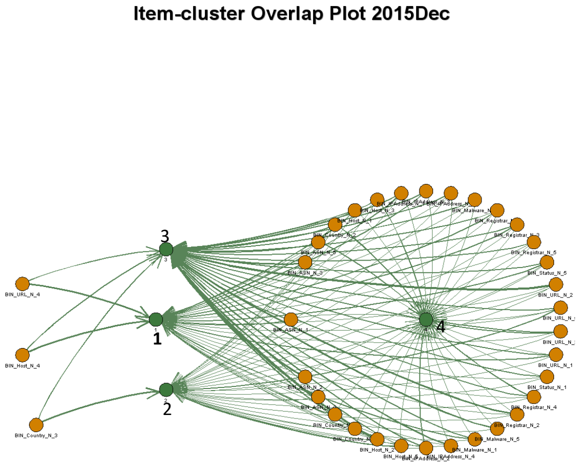


Figure 77: Item-cluster Overlap 2015Jun when cluster Node 1 is disconnected

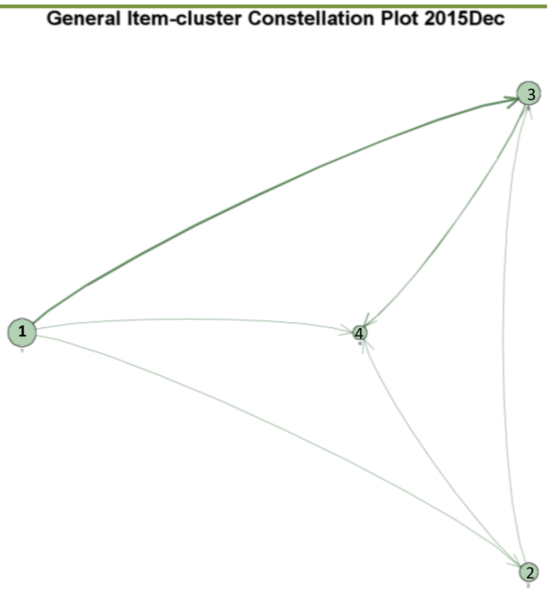
To establish consistency, there is a need to track this cluster node 1 to ensure it is the same cluster (1) node at different levels of the time series. This is done using the most consistent cluster overlap intensity parameter. Without repeating the graphs of the dislodged network for subsequent cumulative periods, the following passages present and discuss the plots of subsequent cumulative overlaps.

To track cluster node 1 at different periods of the time series, cluster maps were plotted for subsequent cumulative periods to show different consisting variable nodes for all different clusters. Subsequent cluster graphs presented for this discussion focus on the Item-cluster overlap plots. Hence, the Events Item-cluster overlap and links maps for the Cumulative Half-Year Periods 2015Dec are presented in Figure 78 and Figure 79 respectively. Unlike the previous cumulative period 2015Jun, the current period has four overlapping clusters (Figure 78), indicating the movement of clustering objects to clusters with stronger membership, and a topological transition towards impending threat. This is evident in the average percentage values of the intensities of the clusters (Figure 75). The values reveal that while the average percentage value of the intensities of the traffic originating node (cluster 1) decreased from 43.97% to 23.74%, the average percentage values of the intensity of cluster nodes 2, 3 and 4 increased from 5.00% to 11.56%; from 17.17% to 48.60%; and from 11.08% to 16.10% respectively. This also explains the disappearance of cluster node 5 suggesting that it was absorbed by other nodes based on close similarity values.



Item-cluster Overlap Plot 2015Dec

Figure 78: Network Item-cluster Overlap Plot 2015Dec



General Item-cluster Constellation Plot 2015Dec

Figure 79: Cluster Overlap Map Showing Connecting Link Directions 2015Dec

Visual observation shows a more evenly (normally) distributed number of edges among the nodes (vertices) in the overlapping cluster nodes 2015Dec. However, the cluster overlap map for the period 2015Dec (Figure 79) shows one (concentrated) major link represented by a thick dark arrow line, which has an outward connection from node 1 to 3. The intense connection and concentration of activity on this major link (nodes 1 – 3) is visually clearer in Figure 75 and Figure 80. The values show that cluster node 3 recorded the highest average percentage intensity value of 48.60% and cluster node 1 recorded the second highest value of 23.74%. This shows an increase from 17.17% and a decrease from 43.97%, respectively, in the preceding period. A less thick line connects node 3 (intensity = 48.60%) to node 4 (intensity = 16.10%), suggesting a less intense relationship between the two nodes. The lines show the order of magnitude of the links relationship and suggests the major traffic flows from cluster 1 through to cluster 3 and cluster 4. The remaining link lines are shown to be the same (or even). The connecting links (Figure 79) reveal that cluster node 1 originates three outward links, one link each, to nodes 2, 3, and 4. Conversely, node 2 sends outward link connections with one each to nodes 3 and 4, while node 3 sends one outward link connection to node 4. Clearly, the connecting links map shows that cluster node 1 generates and gives three outward links and does not receive any incoming connection, suggesting it is the attack-originating and distributing node. Nodes 2 and 3 receive incoming connections and send out-going connections while the overlapping cluster node 4 is a receiver-only node.

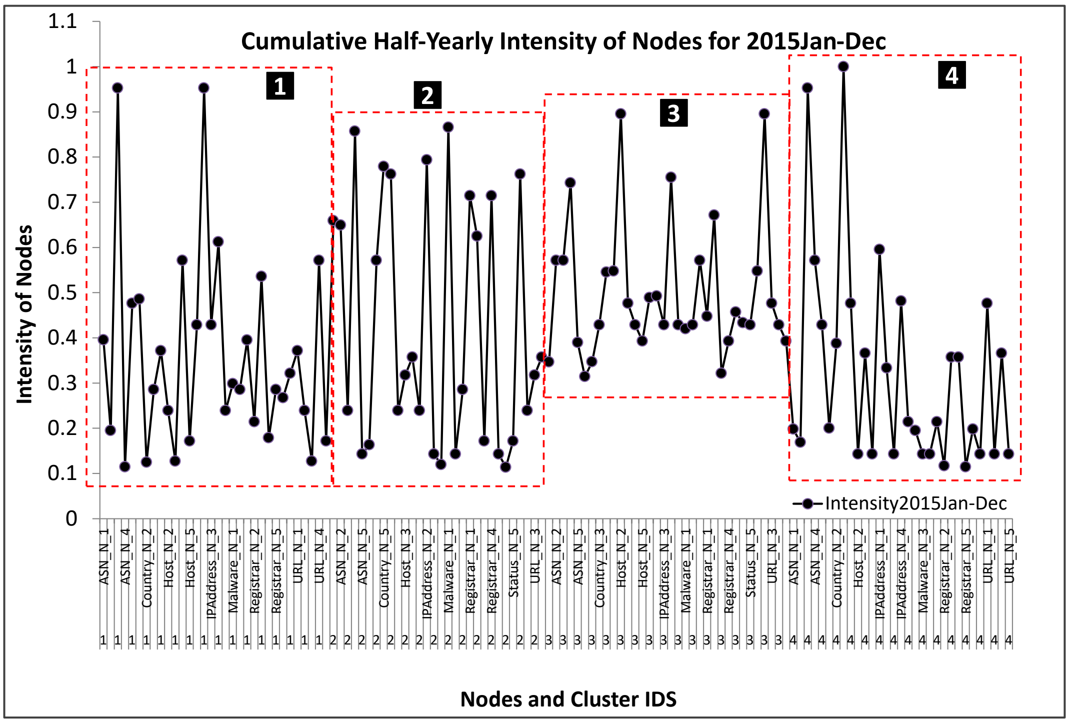


Figure 80: The Intensities of overlapping cluster nodes 2015Dec: Showing the intensities of objects in cluster overlap nodes 1, 2, 3, and 4

Like the preceding cumulative period 2015Jun, the intensity values of the elements show the same regular pattern around cluster node 1 such that the percentage of clustering elements are close together (2015Ju = 31.63%, 2015Dec = 25.20%) (Table 1 and Figure 81). Cluster node 1 also shows that it gave out all the outward links and did not receive any link from other nodes, therefore, corroborating the conclusion that it is the threat originating node, and consequently, the active cluster node (Figure 78, Figure 79 and Figure 80).

The average intensity of 11.56% for cluster node 2 for the cumulative period 2015Dec reveal a significant difference compared to the average intensity of 5.00% for the same node in the period ending 2015Jun. While the number of clustering elements increased significantly to 30 (23.62%) in the period 2015Dec from 8 (8.16%) in the 2015Jun period (Table 1 and Figure 81), the intensities varied in sizes for each individual values (Figure 74 and Figure 80). Cluster node 3 has a similar increase in the number of clustering elements to 34 (26.77%) for the period 2015Dec from 22 (22.45%) in 2015Jun with significant variations in the intensity values of 48.60% against the previous period 2015Jun (17.17%). There is an increase to 31 (24.41%) from 8 (8.16%) in the number of clustering elements in cluster node 4 for 2015Dec in contrast to 2015Jun. In addition, the intensity values vary with great irregularity suggesting that cluster node 4 (2015Dec) absorbed some of the closely relating clustering objects from the disappeared previous cluster node 5 (2015Jun).

Table 1: Number of Clustering Elements and Percentages

Period	Number of Clustering Elements						Percentage of Clustering Elements				
	Clust 1	Clust 2	Clust 3	Clust 4	Clust 5	Total	Clust 1	Clust 2	Clust 3	Clust 4	Clust 5
2015Jun	31	8	22	8	29	98	31.63	8.16	22.45	8.16	29.59
2015Dec	32	30	34	31		127	25.20	23.62	26.77	24.41	
2016Jun	23	33	23			79	29.11	41.77	29.11		
2016Dec	23	9	9	10	33	84	27.38	10.71	10.71	11.90	39.29
2017Jun	24	33	9	9		75	32	44	12	12	
2017Dec	23	9	9	10	33	84	27.38	10.71	10.71	11.90	39.29
2018Jun	23	9	9	10	33	84	27.38	10.71	10.71	11.90	39.29

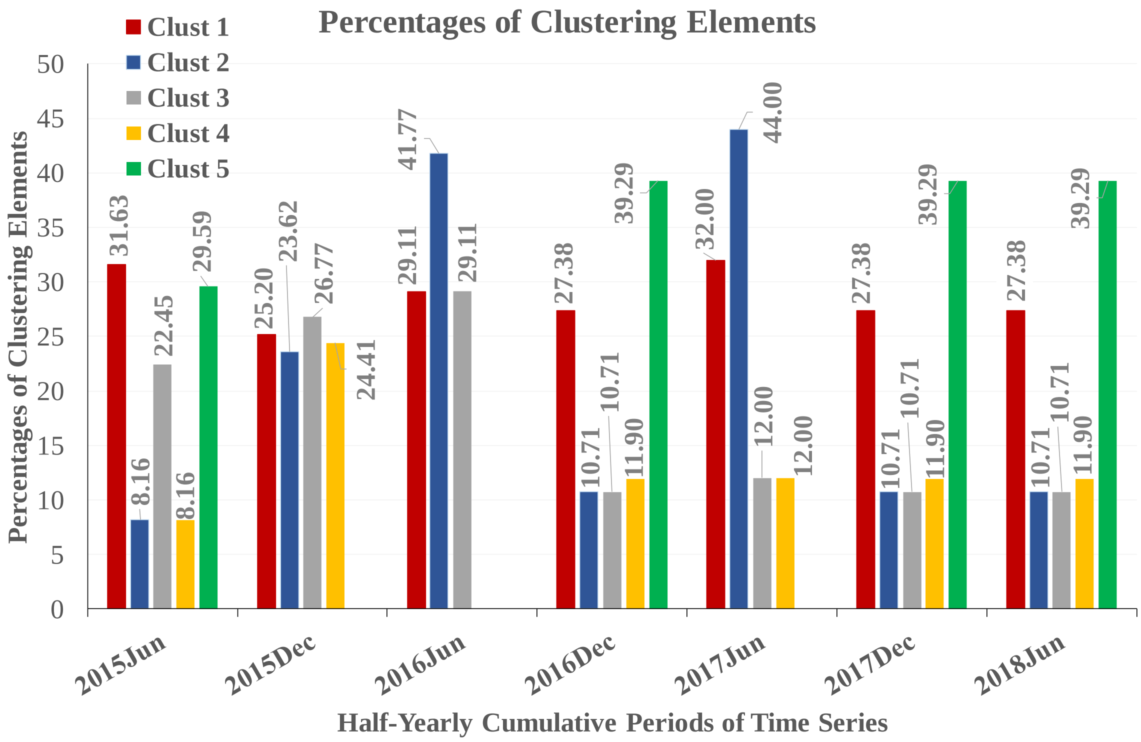


Figure 81: Percentages of Clustering Elements for Different Cluster and Cumulative Periods

Like cluster 5 for the period ending 2015Jun, cluster 4 has one clustering element that has an intensity value equal to 1, suggesting that it has an outlier element. It is worthy to note that while clusters 2, 3 and 4 expanded in the number of clustering elements for the period 2015Dec, cluster node 5 which was seen in period 2015Jun disappeared (Table 1 and Figure 81), indicating it was absorbed by other cluster nodes. Apparently, similarities developed with the entry of new events activities that resulted in the of absorption of cluster node 5. The inconsistencies in clusters 2, 3, and 4 suggest that like the period ending 2015Jun, cluster node 1 remains the most consistent, hence the active cluster overlap node, and the active source of ransomware attack.

Further discussions focus on subsequent cumulative periods, starting with analysis of the cumulative period ending 2016Jun. This period is represented by the Network Item-cluster Overlap Plot 2016Jun and Cluster Overlap Map Showing Connecting Link Directions 2016Jun (Figure 82 and Figure 83).

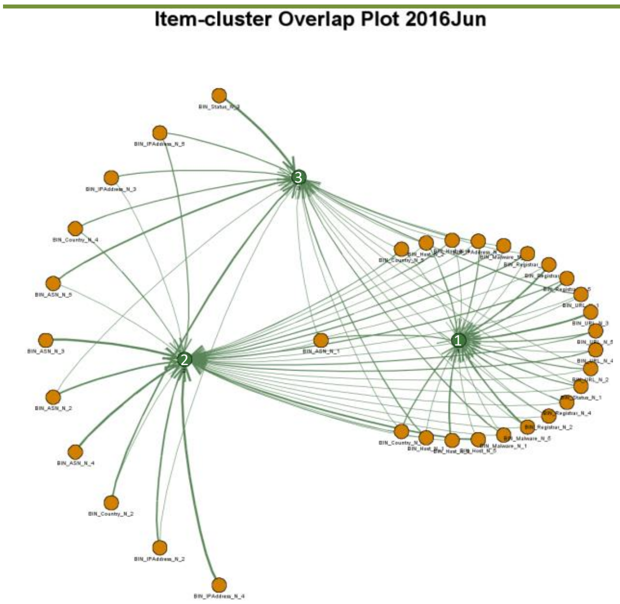


Figure 82: Network Item-cluster Overlap Plot 2016Jun

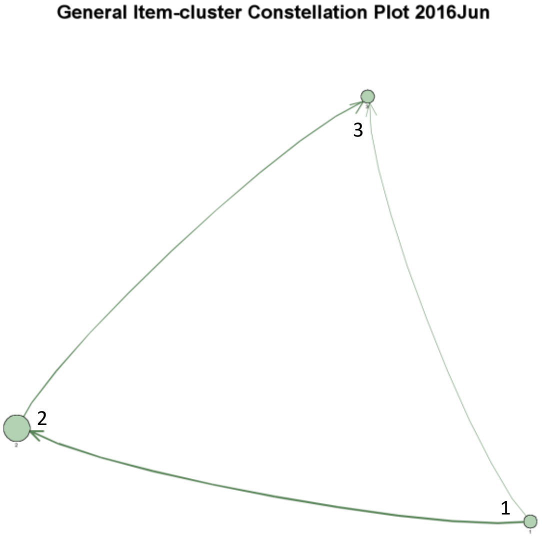


Figure 83: Cluster Overlap Map Showing Connecting Link Directions 2016Jun

The period has three overlapping clusters, which is one cluster less than the preceding period, suggesting stronger relationship in the intensities and similarities of objects within the clusters. This indicates the onset of ransomware threat. A Look at the network item-cluster overlap plot 2016 (Figure 82) reveals that the three overlapping clusters appear to be evenly connected. However, the Cluster Overlap Map Showing Connecting Link Directions 2016Jun (Figure 83) reveals the true connectivity of the overlaps. It is obvious that cluster overlap node 1 originates two outward links that connect cluster overlap nodes 2 and 3 suggesting that these two cluster nodes are the direct distribution nodes of ransomware threats to network connected devices. The cluster overlap map 2016Jun shows the major distribution route to be from cluster node 1 to node 2 and subsequently to the terminal node 3. The degree of the thick lines indicates this traffic route. Cluster overlap node 2 gives an outward link to the terminal cluster overlap node 3. While cluster overlap node 2 receives one incoming link, cluster overlap node 3 receives two incoming links. It is important to restate that the major link axis is from cluster node 1 to cluster node 2, and from cluster node

2 to cluster node 3. The direction of the link arrows on the cluster map (Figure 83) shows that the cluster overlap node 1 generated all outward links and did not receive any incoming links (connection) from other overlapping cluster nodes. The all-outward link direction of cluster overlap node 1 establishes it is the active node; it becomes clear that cluster overlap node 1 is the active node. Therefore, like the previous periods, the removal of the active cluster node 1 will dislodge the ransomware network and control any threat before it attacks.

The intensity values of the overlapping cluster node elements 2016Jun (Figure 84) corroborate these conclusions.

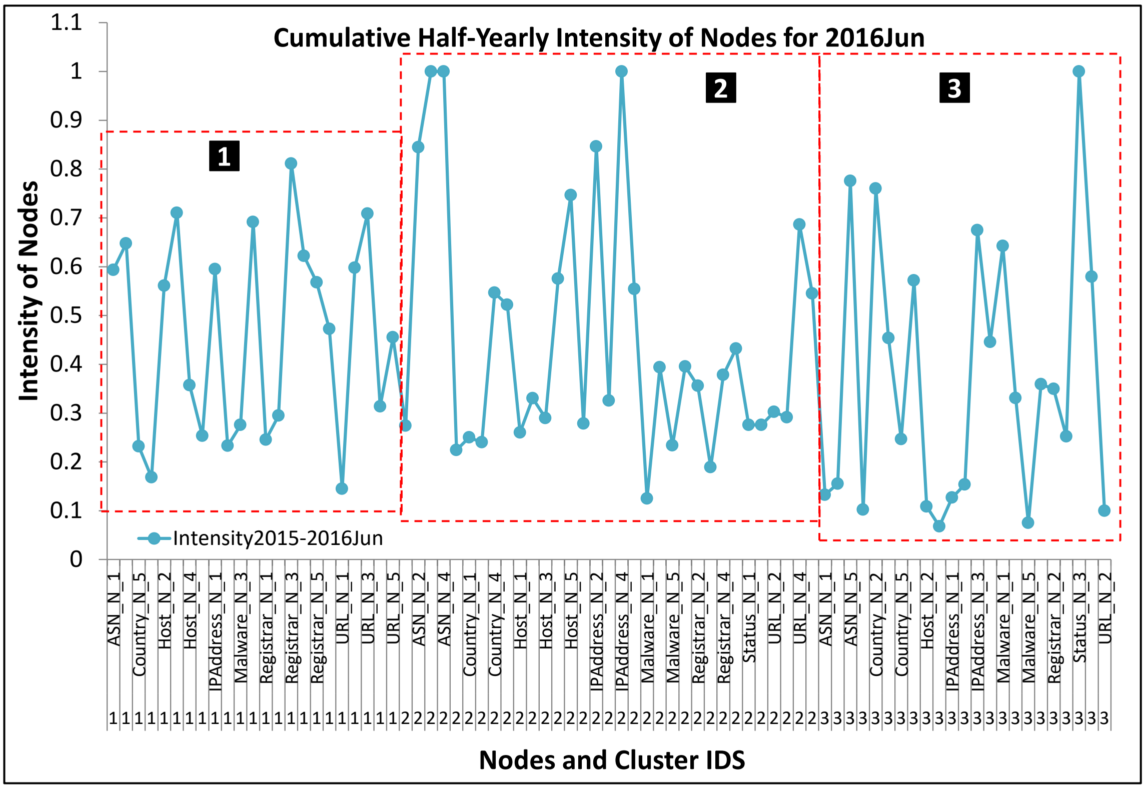


Figure 84: The Intensities of overlapping cluster nodes 2016Jun: Showing the intensities of objects in cluster overlap nodes 1, 2, and 3

It is obvious that the intensity values of the overlapping cluster node 1 are consistent with the values of the previous half-yearly cumulative periods (see also Table 1, Figure 75 and Figure 81). The table and figures show that the percentage values, number and percentage values of the clustering items for 2015Jun (43.97%, 31, 31.63%), 2015Dec (23.74%, 32, 25.20%) and 2016Jun (31.03%, 23, 29.11%) are closer and show consistent pattern than they are in other cluster nodes. In addition, the intensities of cluster nodes 2 and 3 are irregular as was the case in the preceding periods (Figure 74, Figure 80 and Figure 84). Figure 84 also shows that cluster overlap nodes 2 and 3 have some elements with intensity values that are equal to 1, indicating that they are outliers and stand-alone. While the overlapping clusters of later cumulative periods of the time series are examined, it is correct to propose at this point that cluster node 1 remains the active cluster node because of its highest and consistent percentage values in all measures of centrality.

Parameters for Cluster 1	**2015Jun**	**2015Dec**	**2016Jun**	**2016Dec**
Percentage of measures of centrality	27.41%	22.59%	39.26%	22.53%
Number and percentage of clustering items	31 (31.63%)	32 (25.20%)	23 (29.11%)	23 (27.38%)
Percentages of intensities of cluster objects	43.97%	23.74%	31.03%	29.15%

Consequently, its removal results in dislodging the ransomware network and controls ransomware threats before they attack. The removal could equally be undertaken at an earlier stage to minimize attack impact.

The investigation further examined the overlapping clusters for the period of the time series ending 2016Dec (Figure 85 and Figure 86). It is remarkable that after the network attained a minimum number of three clusters in the cumulative period 2016Jun, there was an increase in the number of overlapping clusters to five for the half-yearly cumulative period 2016Dec (Figure 86). This increase confirms the spike in the events' activities seen earlier in the multiple time series for all the variable events in the ransomware network (Figure 29 and Appendix 3). Furthermore, Table 1 and Figure 81 show there was an active attack with cluster 1 and cluster 5 recording a 27.38% 39.29% in the number of clustering events. In addition, Figure 75 shows a wide difference between the percentage intensity values of the originating cluster 1 (29.15%) and the terminal cluster 5 (63.05%) on the one hand and cluster 2 (2.58%), cluster 3 (2.50%) and cluster 4 (2.73%) conversely. This indicates an increase in the events' activities suggesting there was a full-grown ransomware threat (attack) that was highly active currently.

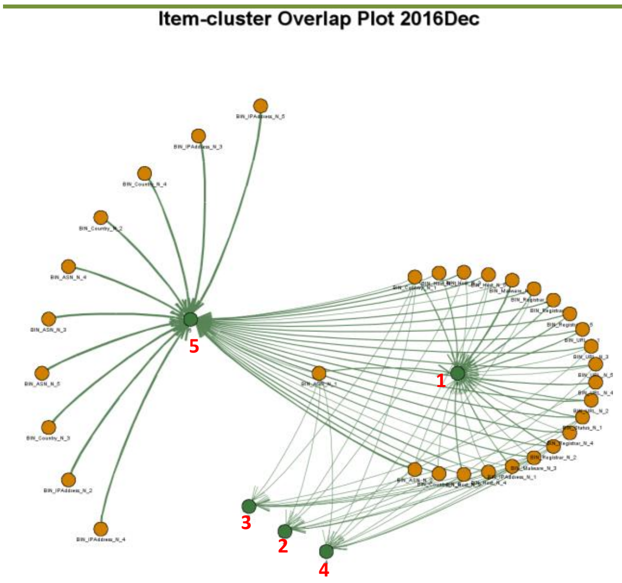


Figure 85: Network Item-cluster Overlap Plot 2016Dec

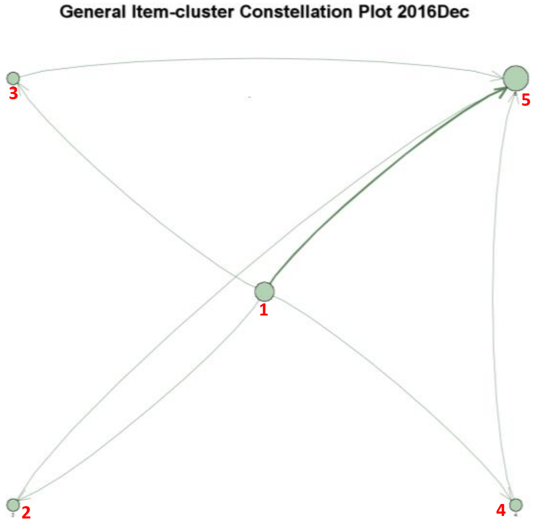


Figure 86: Cluster Overlap Map Showing Connecting Link Directions 2016Dec

However, a look at the Network Item-cluster Overlap Plot 2016Dec (Figure 85) shows that cluster nodes 1 and 5 are more densely connected at 29.15% and 63.05% respectively, while cluster nodes 2, 3, and 4 are more evenly connected and less densely connected at 2.58%, 2.50% and 2.73% (Figure 75, Figure 81 and Table 1). Nevertheless, the Cluster Overlap Map Showing Connecting Link Directions 2016Dec (Figure 86) gives a clearer understanding of the magnitude and direction of the connecting links. The origin and distribution links of the ransomware reveal that cluster node 1 is the link-generating node while the terminal cluster node 5 is a receiving node. Cluster overlap node 1 generated four outward links with one link each to the cluster nodes 2, 3, 4, and 5 respectively. Conversely, the overlapping cluster nodes 2, 3 and 4 send one outward link each to cluster node 5. The link direction map shows the major traffic (distribution) axis is from cluster overlap node 1 to node 5. The map evidently shows that cluster node 5 is a receiving only node and a terminal node. This means it receives incoming connections without sending any out-going connections suggesting that cluster node 5 is the major gateway for the distribution of attack. Conversely, cluster node 1 generates and sends outgoing connections without receiving an incoming connection. It is interesting to observe that all the BIN elements (BIN_Malware, BIN_Registrar, BIN_Host, BIN_URL, BIN_country, BIN_Status, and BIN_IPAddress) particularly the BIN_Malware are clustered in cluster node 1 in contrast to other cluster nodes. Consequently, like in the preceding cumulative periods, cluster node 1 is the dominant and active cluster overlap in the time series for the cumulative period ending 2016Dec. The active nature of the cluster node 1 suggests that its removal will dislodge the activities of the ransomware network and control any imminent ransomware attack.

The confirmation of the active nature of the overlapping cluster node 1 is validated by the intensity values of the clustering elements within each cluster (Figure 87). While the intensity values of the elements in the cluster overlap node 1 continue to remain consistent with preceding cumulative periods (2016Dec = 29.15%; 2016Jun = 31.03%), the intensity values of clusters node 2, 3, 4, and 5 continue to be inconsistent and irregular (2016Dec – clust 2 = 2.58, clust 3 = 2.50%, clust 4 = 2.73% and clust 5 = 63.05%; 2016Jun – clust 2 = 44.08%, clust 3 = 24.89%) (Figure 75, Figure 81 and Table 1). It is obvious that the number of clustering elements decreased greatly in clusters 2, 3, and 4 with the following percentage values:

- cluster 2 decreased from 41.77% down to 10.71%;
- cluster 3 decreased from 29.11% down to 10.71%;
- new cluster 4 recorded 11.90% and
- new cluster 5 recorded 39.29% respectively,

while the number of clustering elements for cluster 1 appears to remain unchanged from the previous period (from 29.11% to 27.38%).

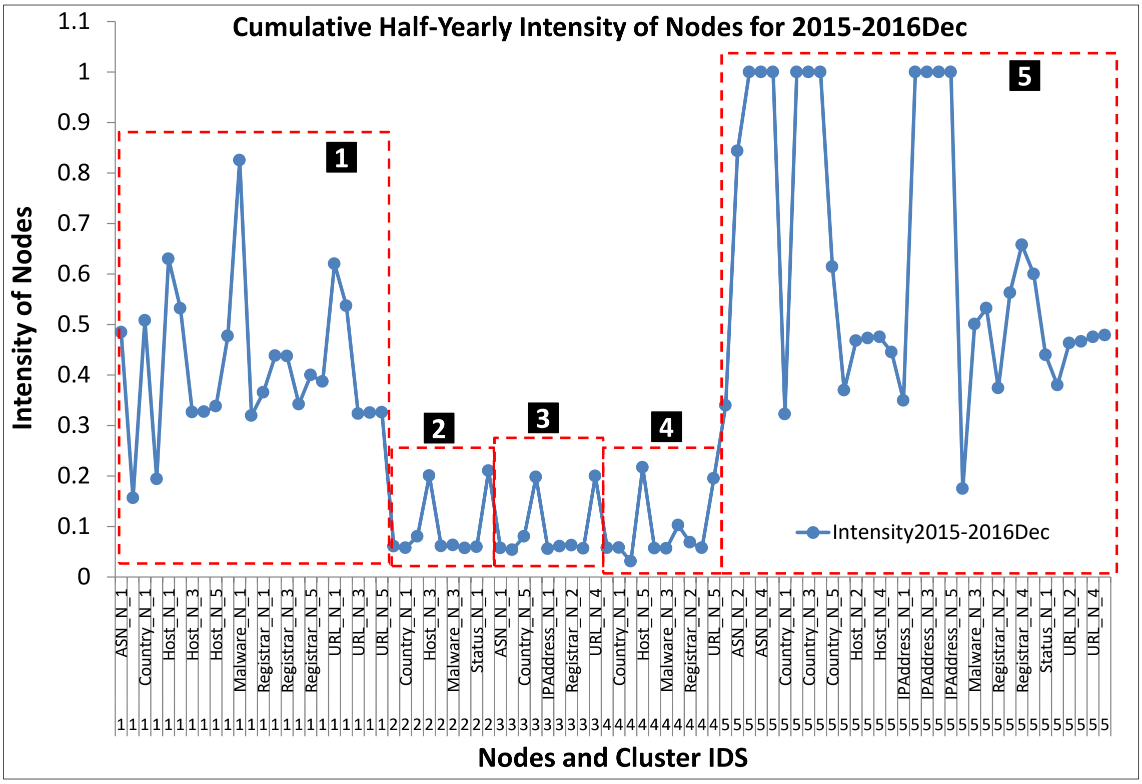


Figure 87: The Intensities of overlapping cluster nodes 2016Dec:
Showing the intensities of objects in cluster overlap nodes 1, 2, 3, 4, and 5

However, while the intensity values of cluster nodes 2, 3, and 4 decreased greatly, the percentage intensity value (39.29%) of cluster 5 is significantly high. Figure 87 shows that many of the intensity values in cluster 5 are equal to 1 resulting in the high percentage intensity value of 39.29% recorded by cluster 5 in 2016Dec. This suggests there are many (10) outlier elements (Figure 85), which are stand-alone in the cluster node 5. It also reveals a high intense activity suggesting there was an active attack. Therefore, the consistency of the intensity values of the clustering elements validates the conclusion that cluster node 1 remains consistent and continues to be the active cluster node that initiates and executes attacks, and its removal dislodges the threat.

In continuation, the formation and development of overlapping clusters and link connections plot for the cumulative period 2017Jun of the time series is presented in Figure 88 and Figure 89. Unlike the preceding period 2016Dec, the 2017Jun cumulative period has only 4 overlapping clusters (Figure 88). This period has two terminal nodes 3 and 4. Cluster overlap node 1 appears to have more edges connecting to it with the highest percentage values of measures of centrality than any other overlapping cluster node (Figure 60, Figure 61 and Appendix 9).

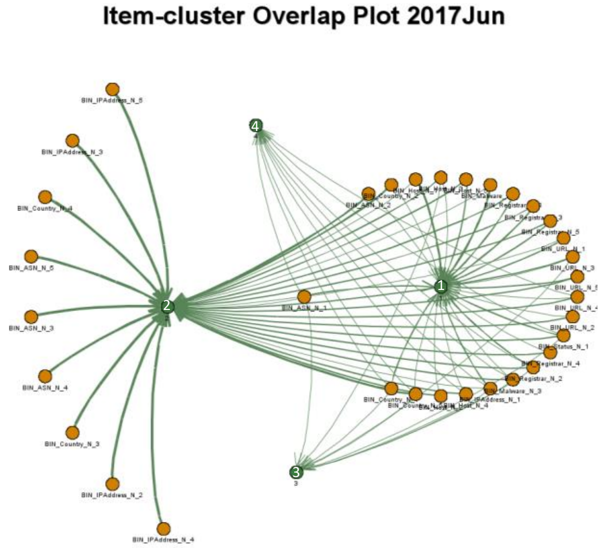


Figure 88: Network Item-cluster Overlap Plot 2017Jun

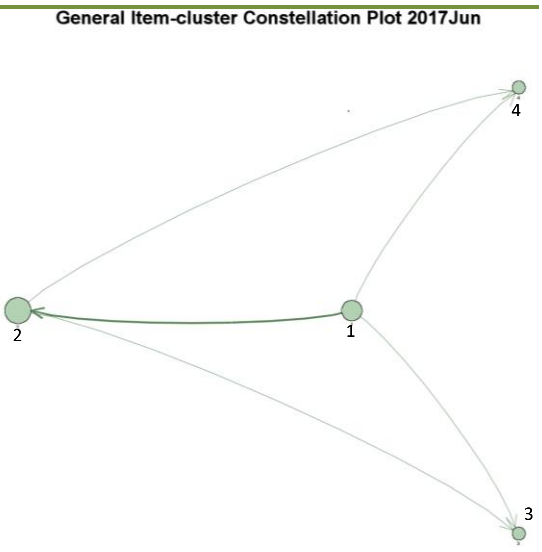


Figure 89: Cluster Overlap Map Showing Connecting Link Directions 2017Jun

Although Cluster 2 has many incoming connections, it has many (9) stand-alone (outlier) links with the intensity values for each clustering element equal to 1 resulting in in a very high percentage intensity value of 63.27% (Figure 90 and Figure 75). The network item-cluster overlap plot (Figure 88) also shows that the overlapping cluster nodes 3 and 4 are receiver-only (terminal) nodes; they received incoming connections and never gave any out-going connections. The explanation above is clearer from the Cluster Overlap Map Showing Connecting Link Directions 2017Jun (Figure 89). This plot reveals that the overlapping cluster node 1 generated and distributed three out-ward connections and did not receive any inward connections. Each of the three out-ward links connects to cluster node 2, 3, and 4, respectively. Figure 89 equally shows that the cumulative period 2017Jun has two terminal nodes namely cluster 3 and 4. Clearly, the main distribution axis is from cluster node 1 to cluster node 2 as indicated by the thick arrow-line that links them. This understanding is further corroborated by the high percentage intensity value of 63.27% and high percentage value of clustering elements (44%) recorded by cluster node 2 (Figure 75, Figure 81 and Table 1). The dominant role of cluster overlap node 1 leaves no doubt that it is the active node as it was in the preceding cumulative periods. Consequently, the removal of cluster overlap node 1, as in the preceding periods, dislodges the ransomware distribution network and controls potential threats.

The intensities of the overlapping cluster nodes 2017Jun (Figure 90) further demonstrates the consistency of cluster overlap node 1 as the active cluster overlap. Clearly, the percentage of the measures of centrality, percentages of clustering elements and the values and percentages of the intensities of cluster node 1 are consistent with those of the node of the preceding cluster 1 (Appendix 5, Figure 81, Table 1 and Figure 75,). Note that cluster node 1 has not recorded any outlier element throughout the time series, implying that it had not recorded any intensity values that are equal to 1, meaning that all the objects are strongly linked for attack. All the intensity values of cluster node 1 consistently fall within the range of 0.1 to 0.9.

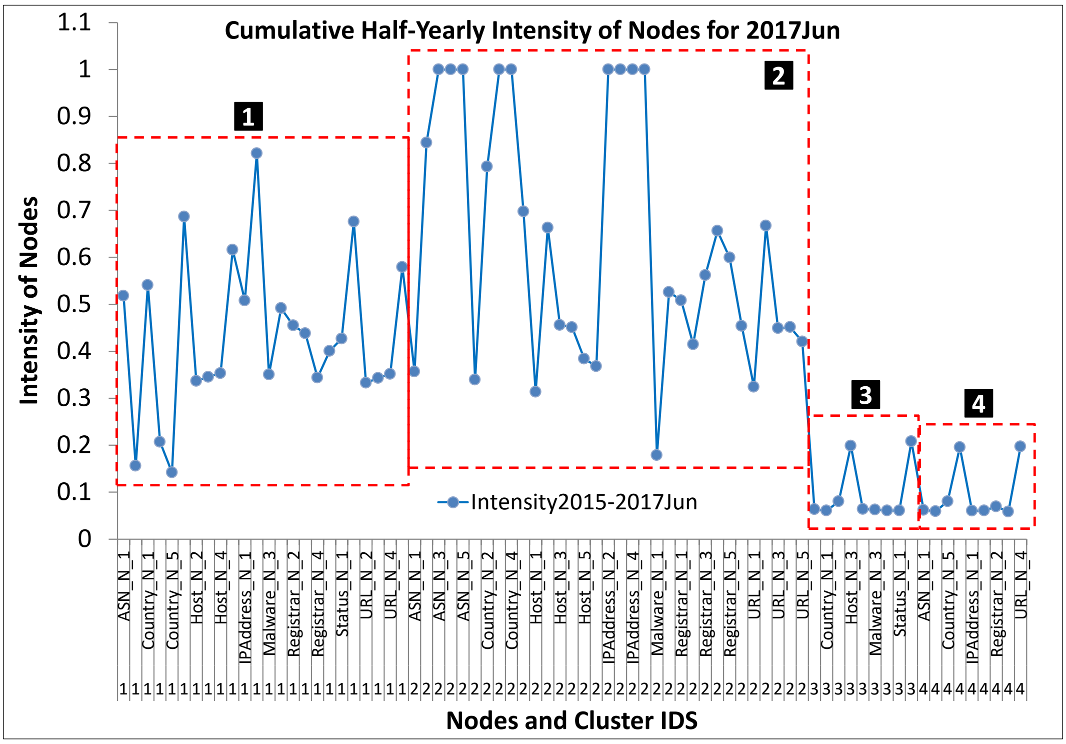


Figure 90: The Intensities of overlapping cluster nodes 2017Jun: Showing the intensities of objects in cluster overlap nodes 1, 2, 3, and 4

Moreover, each successive percentage values of the cluster intensities at different stages of the time series appear to be consistent and similar when compare with other clusters, i.e. 2015Jun – 43.97%; 2015Dec – 23.74%; 2016Jun – 31.03%; 2016Dec – 29.15% and 2017Jun – 31.57% (Figure 75). In addition, cluster node 1 shows to have the same number of clustering objects (elements) at every cumulative stage of the time series, 2015Jun – 31.63%; 2015Dec – 25.20%; 2016Jun – 29.11%; 2016Dec – 27.38% and 2017Jun – 32.00% (Table 1 and Figure 81). Conversely, cluster overlap node 2 continues to show irregular values of intensity in the clustering objects. It is observed also that cluster node 2 recorded about nine (9) outlier elements with each outlying element having an intensity value equal to 1 (Figure 90). On the other hand, cluster nodes 3 and 4 have few clustering objects and lack most of the ransomware clustering elements (variables). It is worthy to note that cluster nodes 3 and 4 have very low-intensity values that are$\leq$ 0.2, i.e. 2.60% and 2.56% respectively (Figure 75). These values validate the earlier suggestions that cluster node 1 continues to be the active cluster node of the ransomware network. Therefore, its removal dislodges the ransomware network and controls ransomware threat before it attacks.

The next cumulative network item-cluster overlap plot 2017Dec and cluster overlap map showing connecting link directions 2017Dec (Figure 91and Figure 92) respectively, reveal that the period 2017Dec has 5 cluster overlaps. Looking at the network item-cluster overlap plot 2017Dec (Figure 91) suggests that the cluster node 1 is more densely connected with percentage clustering elements of 27.38% than the other cluster nodes, except the terminal cluster node 5, which has percentage clustering elements value of 39.29%. Cluster overlap nodes 2, 3, and 4 appear to be more evenly connected with few links distribution (10.71%, 10.71% and 11.90% respectively).

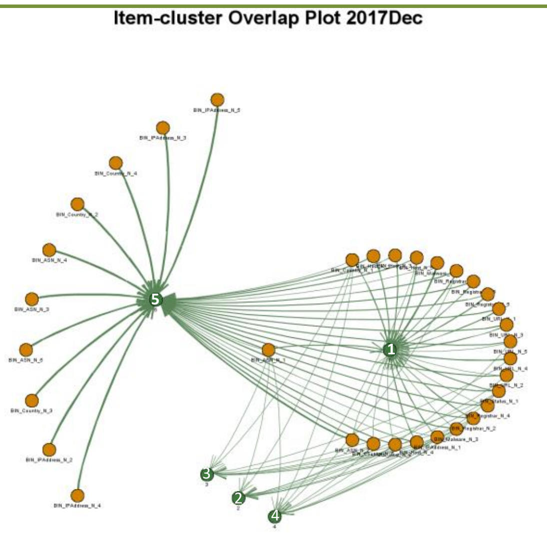


Figure 91: Network Item-cluster Overlap Plot 2017Dec

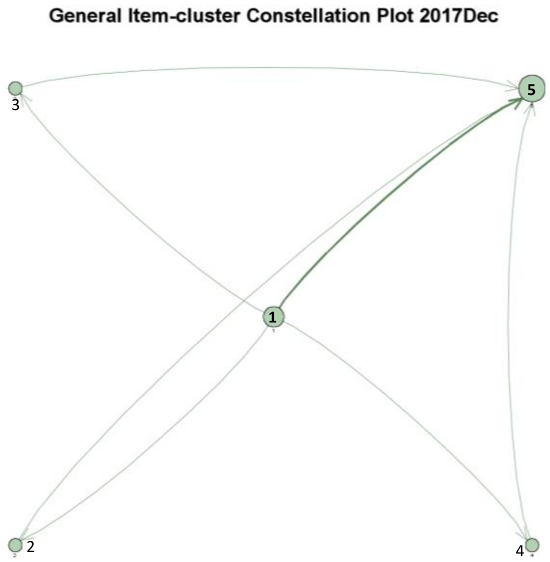


Figure 92: Cluster Overlap Map Showing Connecting Link Directions 2017Dec

While cluster node 5 shows many connections with high percentage number of clustering elements of 39.29% (Table 1 and Figure 81), it has as many as 10 outlier (stand-alone) elements with each outlier recording absolute intensity value that equals 1.00 (Figure 93). Looking at the “cluster overlap map showing connecting link directions 2017Dec” (Figure

92), cluster node 1, as seen in the previous periods, generates four outward connections. Each of these connections is connected to node 2, 3, 4, and 5 respectively. However, the major link distribution axis originates from cluster overlap node 1 to cluster overlap node 5 as shown by the connecting thick arrow line. The percentage values of the intensities (cluster 1 = 29.59% and cluster 5 = 62.71%) and clustering elements (cluster 1 = 27.38% and cluster 5 = 39.29%) confirm the observation (Figure 75, Table 1 and Figure 81). While cluster node 2 receives one incoming connection, it sends an out-ward connection to cluster node 5. Similarly, cluster nodes 3 and 4 receive one inward connection from cluster node 1 and send one outward connection to cluster node 5 respectively. Generally, cluster node 5 receives inward connections only and does not send out an outward connection. It is evident that cluster node 1 generates all the outward connections and does not receive an inward connection. It is also true that while further cumulative events in the network altered the structure of other overlapping nodes, it did not alter the structure of cluster node 1. This understanding is demonstrated in the percentage values of the measures of centrality of the various cumulative period of the time series, the percentage values of clustering intensity (Figure 75) and the percentage values of clustering objects (Table 1 and Figure 81). Therefore, the active nature of this node makes it the appropriate cluster overlap node to remove to dislodge the ransomware network and control threats.

Further analysis of “the intensities of the (network item-cluster) overlapping cluster objects 2017Dec” (Figure 93), shows the intensities of the overlapping objects of cluster node 1 are consistent with the intensities of the preceding cumulative periods of the time series. This shows a similar number of clustering objects as observed in the previous cumulative periods where the intensity values fall within the range 0.1≥Intensity≤0.9. This contrasts with the irregular values of intensity recorded in cluster nodes 2, 3, 4 and 5 (Figure 75, Table 1 and Figure 81).

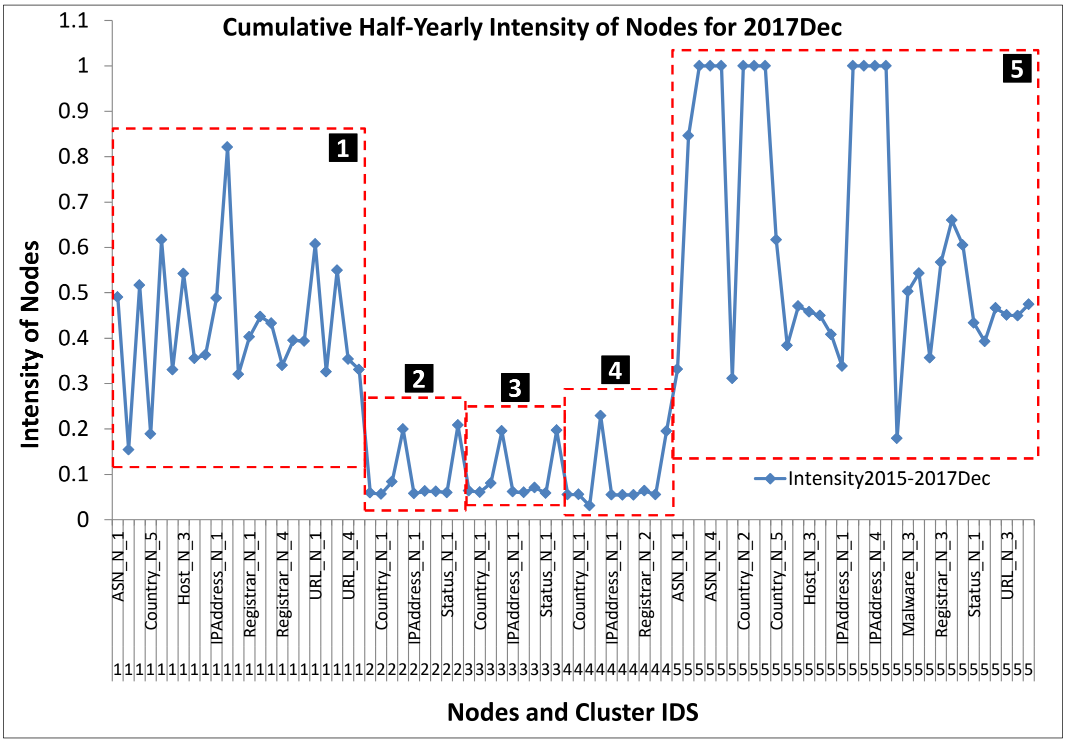


Figure 93: The Intensities of overlapping cluster nodes 2017Dec:
Showing the intensities of objects in cluster overlap nodes 1, 2, 3, 4, and 5

Evidently, nodes 2, 3, and 4 recorded intensity values, which are less than 0.2, i.e. 2.57%, 2.57% and 2.57% and percentage values of clustering elements of 10.71%, 10.71% and 11.90% respectively. This appears to confirm the non-link relationship between nodes 2 and 3, and nodes 2 and 4, respectively (Figure 92). Conversely, the intensities of clustering elements in cluster overlap node 5 show very divergent values ranging from 0.1 to 1.00 confirming the presence of many (10) outlier elements (Figure 93). This results in non-representative centroid and very high error sum of squares (SSE), making the cluster node a dominant dispersion outlet for ransomware threat. In addition, Figure 75 shows that cluster node 5 recorded the highest intensity value of 62.71% for the cumulative period 2017Dec. This observation is evident in Table 1 and Figure 81 where cluster node 5 also recorded the highest percentage value of 39.29% for the number of clustering objects. So far, the discussions have shown that cluster node 1 has remained the active cluster overlap node. Consequently, its removal effectively dislodges the entire ransomware network and prevents ransomware threat.

The "Network item-cluster overlap plot 2018Jun" (Figure 94), "Cluster overlap map showing connecting link directions 2018Jun" (Figure 95), and "The Intensities of overlapping cluster nodes 2018Jun" (Figure 96) for cumulative period 2018Jun are technically the same as the cumulative period ending 2017Dec. It is right to conclude that the same discussions for the period 2017Dec apply also to the current cumulative period ending 2018Jun.

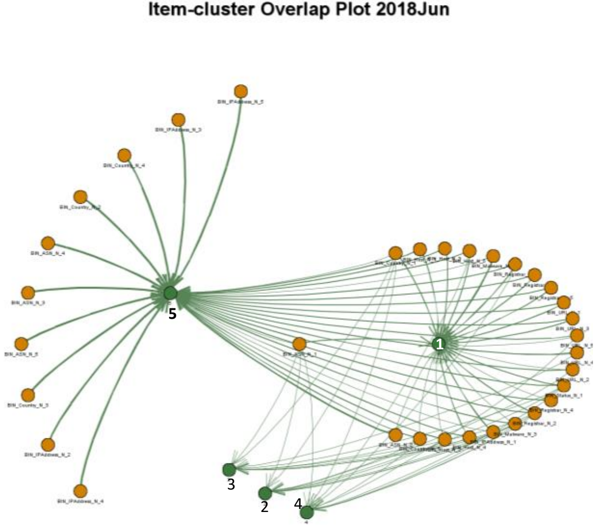


Figure 94: Network Item-cluster Overlap Plot 2018Jun

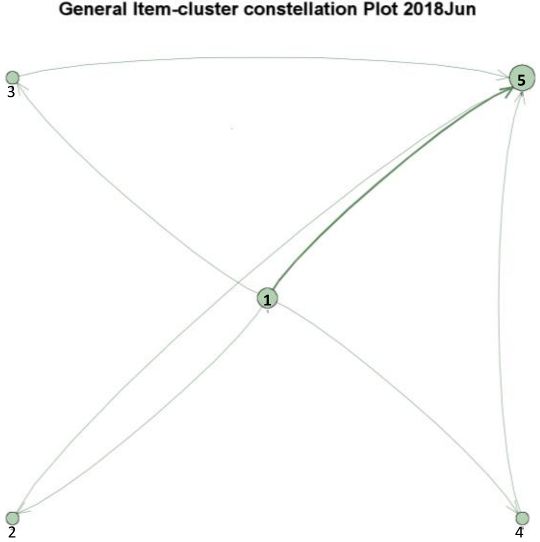


Figure 95: Cluster Overlap Map Showing Connecting Link Directions 2018Jun

There are many links connections to cluster node 1 and 5 shown in the network item-cluster overlap plot 2018Jun (Figure 94). However, many (10) of the clustering elements in cluster node 5 are outliers having intensity values equal to 1.0 each (Figure 96). Cluster node 1 shows to be the link-generating (outward links) node, while node 5 shows to be the link-receiving (inward links) node. Cluster nodes 2, 3, and 4 are shown to be evenly connected with intensity values 0.1 ≤ Intensity ≤0.2; i.e. 2.57%, 2.57% and 2.57% (Figure 75). This is further demonstrated with the percentage values of 10.71%, 10.71% and 11.90% for the clustering elements (Table 1 and Figure 81). Clearly, the “Cluster Overlap Map Showing Connecting Link Directions 2018Jun” (Figure 95) reveal the directions and magnitude of the connections. The plot shows that cluster overlap node 1 generated four outward connections with one connection each to nodes 2, 3, 4, and 5. It is shown also that the major connection axis is the link from cluster node 1 to node 5 suggesting it is the major transmission link for ransomware traffic. This is confirmed by the high percentage intensity values of 29.57% for cluster 1 and 62.72% for cluster 5 (Figure 75); and high percentage values of number of clustering elements of 27.38% for cluster 1 and 39.29% for cluster 5 (Table 1 and Figure 81). While cluster node 2 receives one inward connection from node 1, it sends an outward connection to cluster node 5. Similarly, cluster nodes 3 and 4 receive one inward connection from cluster node 1 and send outward connections each to node 5. In other words, all ransomware traffic is directed to cluster node 5. This is explained by the even percentage distribution of the intensity values of 2.57%, 2.57% and 2.57% (Figure 75) and number of clustering elements of 10.71%, 10.71% and 11.90% (Table 1 and Figure 81). There are no connections between cluster nodes 2 and 3, and cluster nodes 2 and 4 suggesting there is no ransomware traffic between those nodes.

The preceding conclusion is validated by the regularity of the intensities of the overlapping cluster nodes 2018Jun (Figure 96). The consistency of the values of the intensities of cluster overlap node 1 at the various cumulative periods of the ransomware time series (network) agrees with the various topological dynamics (changes) and other key performance parameters in the network. Consistent with the intensity values of the previous periods the clustering elements in cluster node 1 maintains the intensity values within the range of 0.1 and 0.9, i.e. an average of 29.57% (Figure 75), indicating close membership (association) of the objects resulting in a smaller error sum of squares (SSE), and an active source of attack. Cluster nodes 2, 3, and 4 have low-intensity values of 2.57%, 2.57% and 2.57% suggesting low ransomware traffic.

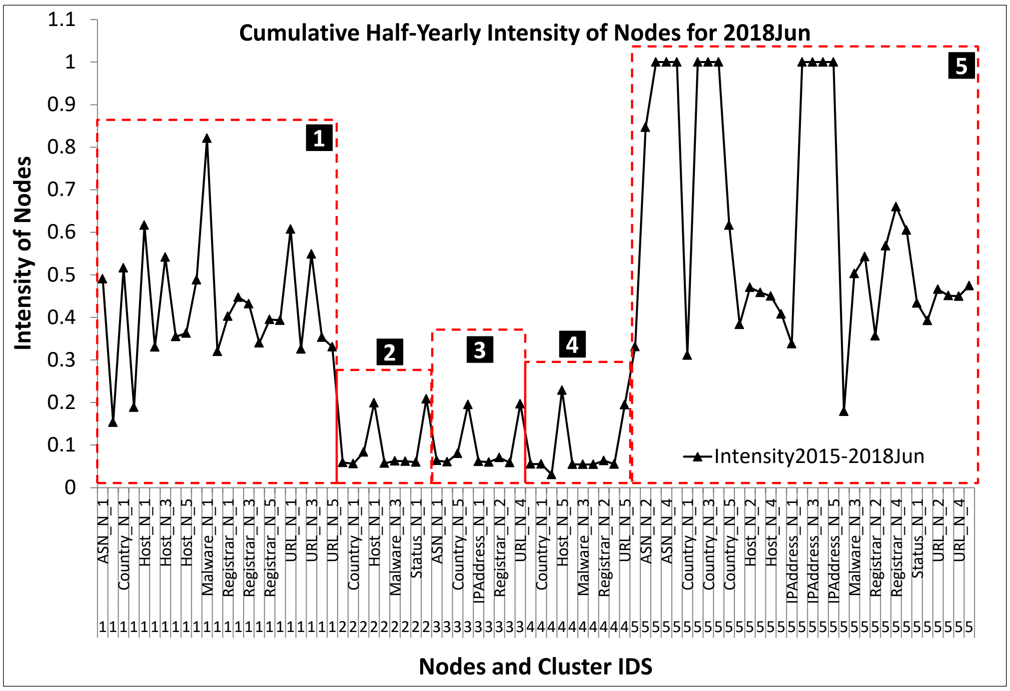


Figure 96: The intensities of overlapping cluster nodes 2018Jun

Conversely, cluster node 5 appears to have many connections but most (10) of them are outliers with intensity values equal to 1.0, thereby increasing the error sum of squares (SSE) and making the centroid non-representative. In addition, it received inward connections and did not give outward connections resulting in the highest percentage intensity value of 62.72% thereby making it an outlet for the distribution of threat into network connected devices. Therefore, cluster overlap node 5 is a terminal link node.

6.3 The Profiling and Tracking of Active Cluster Overlap Using Intensity Maps

The intensities map for the various cumulative periods of the time series (Figure 97 and Figure 98) are presented to show the intensities cluster maps and their respective average weights for different cluster overlaps. The intensities cluster map was configured to show the weight or degree of the cluster intensities by the size of each node circle and its color intensity (Figure 97). The Cluster (Intensities) Average Weight for Half Yearly Cumulative Periods 2015Jun - 2018Jun is to show their corresponding numerical values (Figure 98). With the largest node circle and darkest colour shading, cluster overlap node 1 proves to have the highest intensities degree and weight of 212 (26.40%) validating earlier observations.

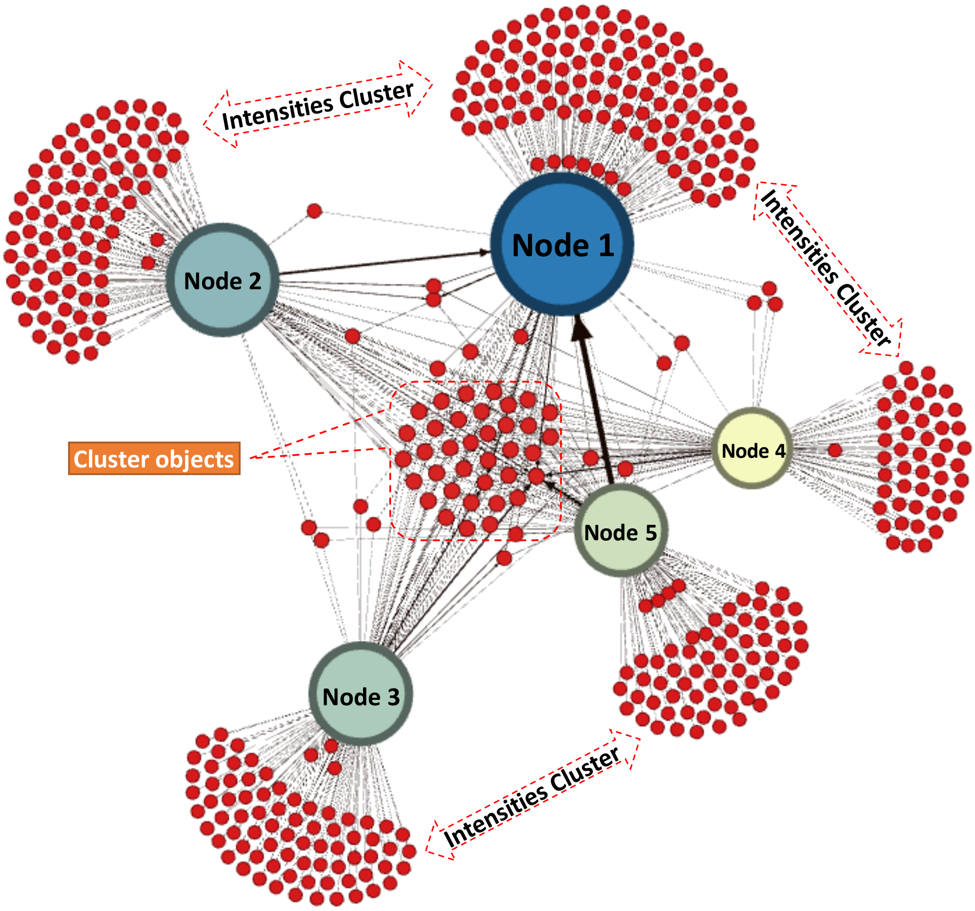


Figure 97: Intensity map for the various cumulative periods of the time series: showing the cluster of intensities of objects cluster nodes 1, 2, 3, 5, and 4 in order of magnitude

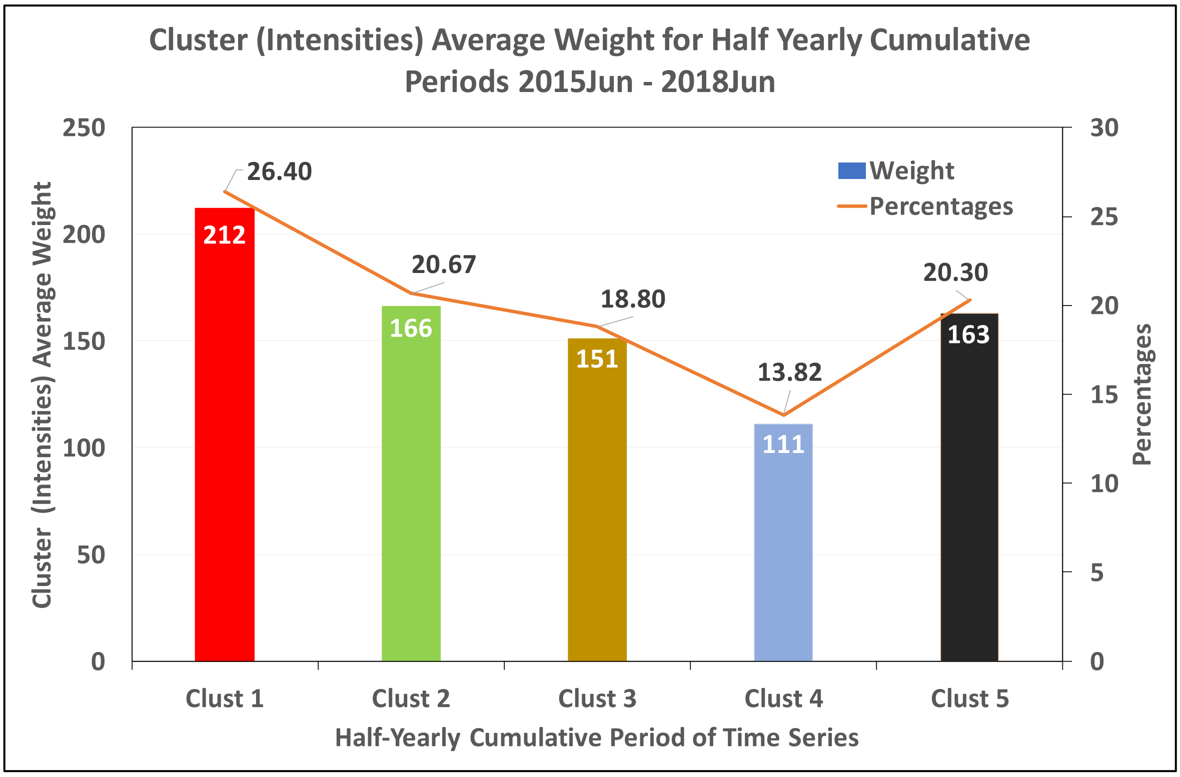


Figure 98: Cluster (Intensities) Average Weight for Half Yearly Cumulative Periods 2015Jun - 2018Jun

The cluster map node 1 shows there are more clustering objects in the cluster overlap node 1 for the entire time series. Similarly, intensities cluster nodes 2, 3, 5, and 4 follows in the same order of degree (weight) 166 (20.67%), 151 (18.80%), 163 (20.30%) and 111 (13.82%) respectively (Figure 98). A further look at the map reveals that the major distribution link is between cluster node 1 and cluster node 5, and between cluster node 1 and cluster node 2, respectively. The cluster objects (elements); including Malware, URL, Registrar, IP Address, ASN, Country, and Status could be seen, as indicated, at the middle of the intensities map. In all representations in this investigation, cluster overlap node 1 proves to be consistent in its structure and pattern, and therefore, shows to be the active cluster overlap node and source of threat, and its removal results in dislodging ransomware attack.

Further validation of the consistency of cluster overlap node 1, as the active overlap node, is the single-plot presentation of the intensity values (for cluster node 1) for all the cumulative periods of the ransomware time series (Figure 99). The intensity values show intense similarities and closeness amongst them, this suggests a very close association thus making them cluster together. The intensities of the ransomware (malware) elements in cluster node 1 consistently peaked to approximately 0.83 during the 2016Dec, 2017Jun, 2017Dec and 2018Jun, which were the periods when the time series (Figure 29) showed a phenomenal spike in the event's activities.

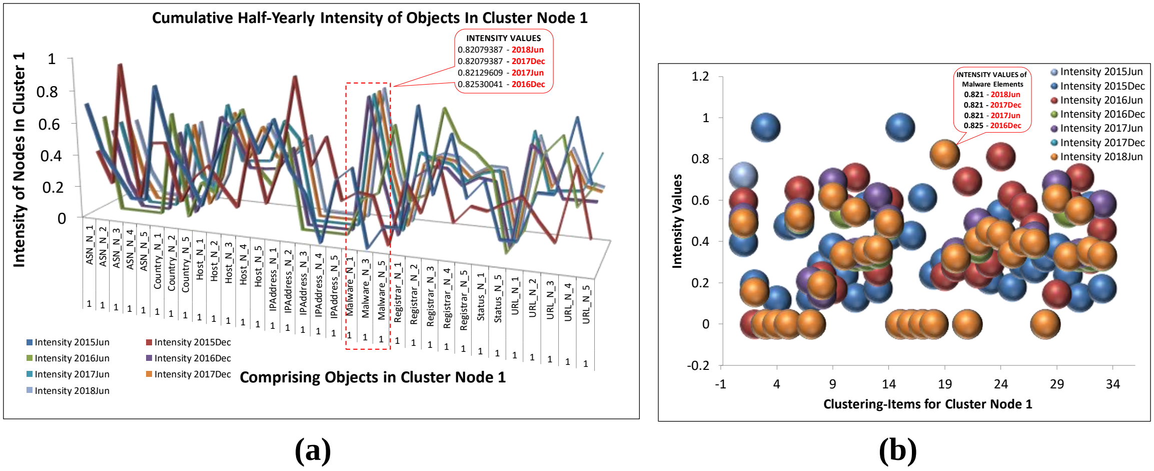
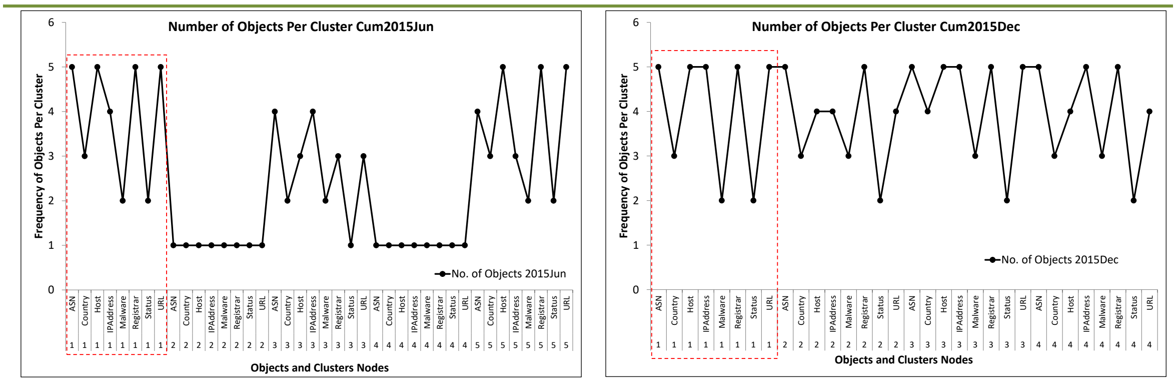


Figure 99: (a) Intensity values for cluster overlap node 1 for all cumulative periods of the ransomware time series, (b) Scatter plot showing the degree of closeness of the intensity values across all cumulative periods

The scatter plot showing the degree of closeness of the intensity values across all cumulative periods (Figure 99) further validates the consistency and similarity of the cluster overlap node 1 as the active cluster overlap node. Therefore, the nature and consistency of cluster overlap node 1 as the only ransomware traffic-generating hub, in the ransomware time series, make it the active cluster overlap node. Hence, the removal of the active cluster overlap node 1 effectively dislodges the ransomware networks and effectively disrupts any potential threat.

6.4 The Profiling and Tracking of Active Cluster Overlap Using the Objects (Variables) Counts

The topological transformations of the various cumulative ransomware networks and cluster overlaps were a function of the activities of the network contents (clustering objects and variables). Therefore, an understanding of the impact of these contents (variables) on the structure of the overlaps is sought to further validate cluster node 1 as the active node. To achieve this objective, further tracking of cluster node 1 is conducted using the clustering objects (components) counts (Figure 100 to Figure 105). The frequency of the objects per cluster for each of the cumulative period shows different frequencies. The frequency plot for Cumulative Periods (a) 2015Jun and (b) 2015Dec (Figure 100) show a high and consistent frequency count of consisting clustering objects in cluster node 1. It consists of high-frequency counts of complete components of the network, ASN, Country, Host, IPAddress, Malware, Registrar, Status, and URL. This indicates that cluster overlap node 1 contains all the network elements to launch a ransomware threat. These values are confirmed in Table 1 and Figure 81 where the number of clustering elements and their percentages are 31 (31.63%) and 32 (25.20%) for the cumulative periods ending 2015Jun and 2015Dec.



(a) (b)

Figure 100: Frequency of Objects per Cluster for Cumulative Periods (a) 2015Jun and (b) 2015Dec: showing the consistency of cluster overlap node 1

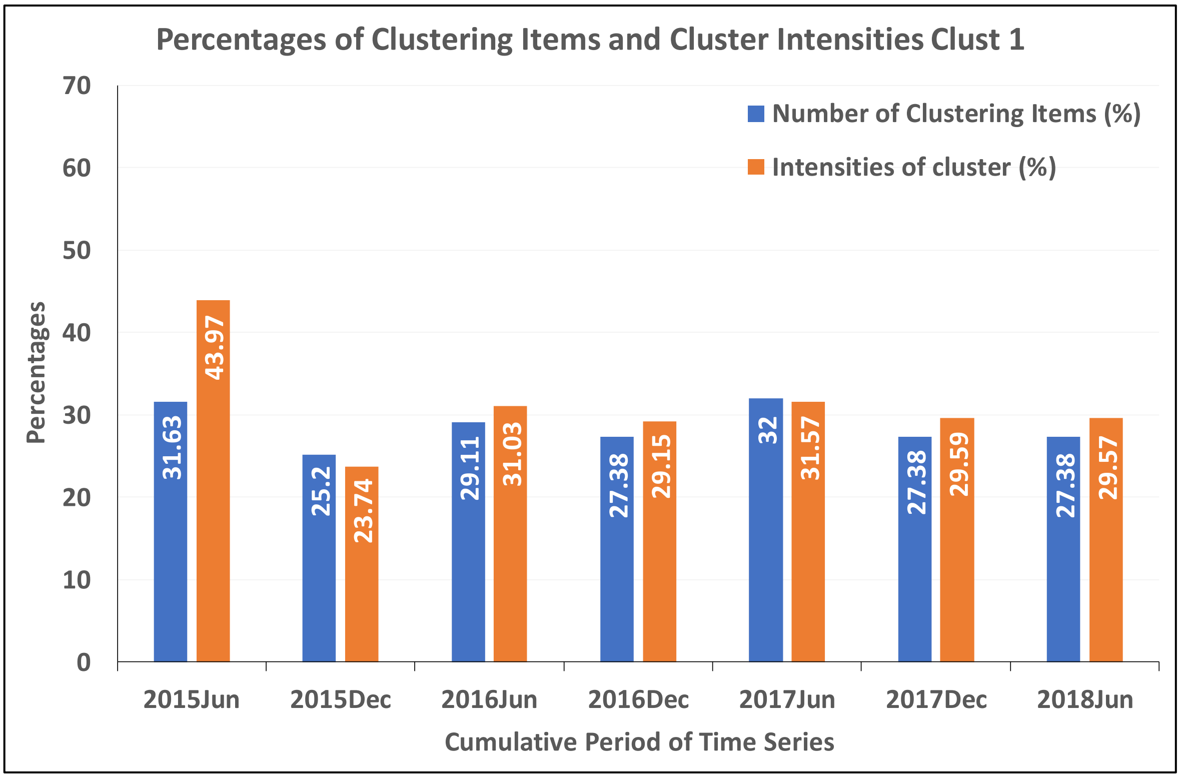
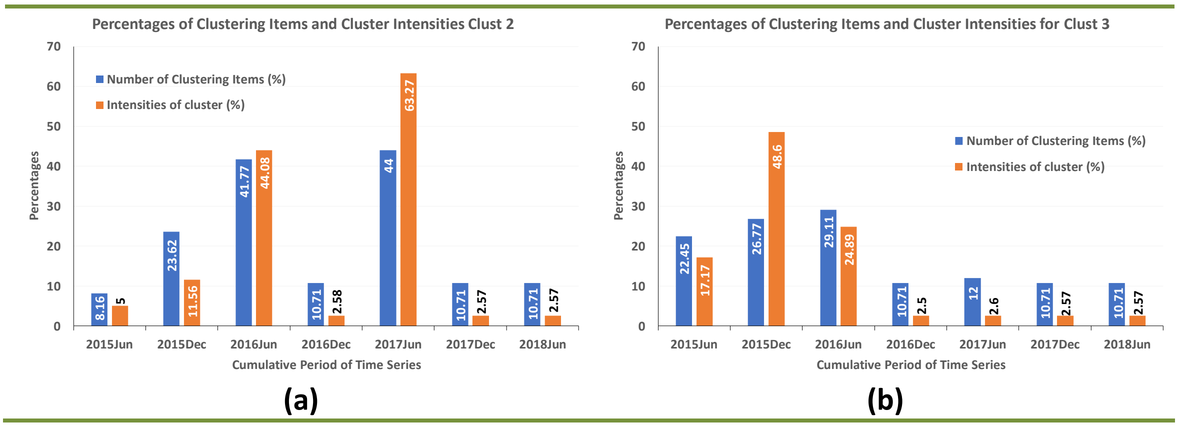
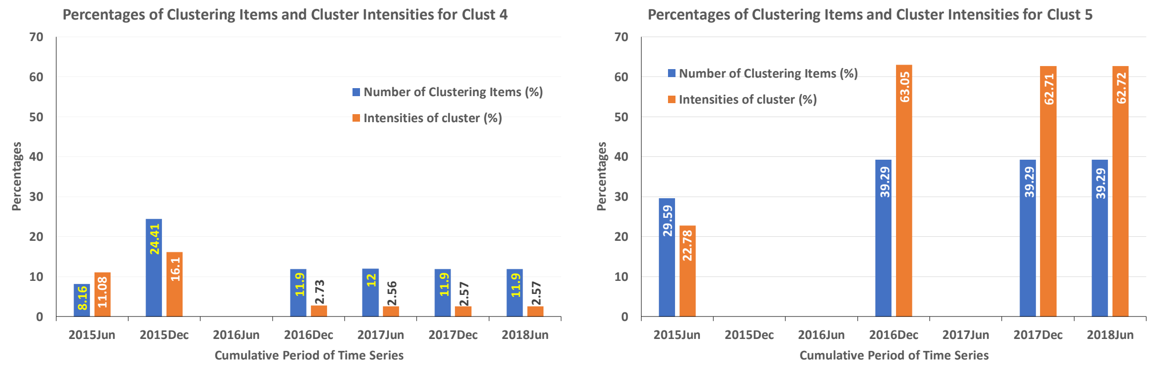


Figure 101: Percentages of Clustering Items and Cluster Intensities of Cluster 1 for the Cumulative Periods 2015Jun – 2018Jun

The percentages of numbers of clustering elements stated above are confirmed in the plot of "Percentages of Clustering Items and Cluster Intensities of Cluster 1 for the Cumulative Periods 2015Jun – 2018Jun" (Figure 101). In contrast, cluster overlap nodes 2, 3, 4 and 5 recorded irregular frequency of objects counts for the same period (a) 2015Jun and (b) 2015Dec {(Figure 100) and (Figure 102 a-b)}.



(a) **(b)**

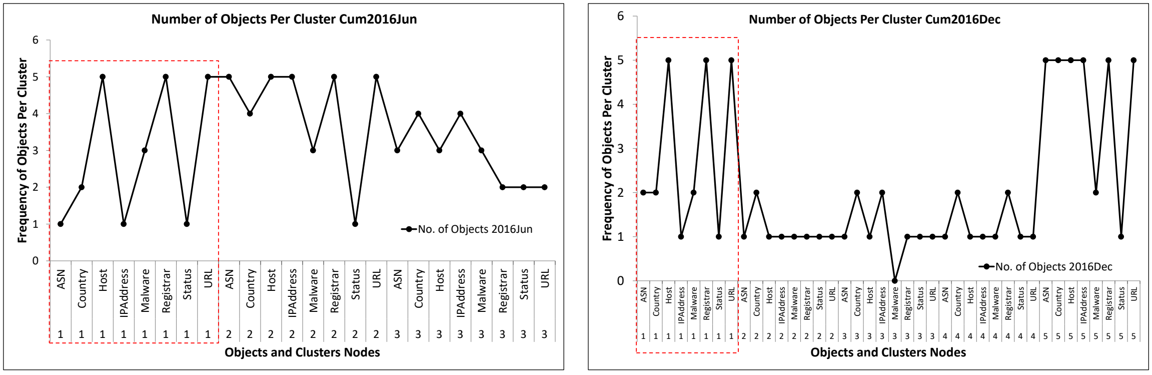


(c) (d)

Figure 102: Percentages of Clustering Items and Cluster Intensities of Clusters 2, 3, 4 and 5 for the Cumulative Periods 2015Jun – 2018Jun

While cluster node 2 in 2015Jun recorded a very low frequency and percentage of objects counts (8, 8.16%), the corresponding cluster node 2 in 2015Dec recorded higher frequencies (30, 23.62%) (Table 1 and Figure 81). This suggests a comparative difference in the formation, development, and weight of cluster node 2 for the two periods. It is concluded that the links traffic to cluster node 2 in 2015Jun was not as high as it was in 2015Dec, as cluster node 2 in 2015Jun was still forming (Figure 100a, b and Figure 102a, b). However, as the ransomware events activities increased, cluster node 2 in 2015Dec had more object similarities that clustered together resulting in higher frequency and percentage of clustering objects. This phenomenon was proven earlier in the preceding discussions (Figure 46a, b and Figure 71b, c). This is also evident in the intensity values of the overlapping cluster nodes 2015Jun and 2015Dec (Figure 74, Figure 75 and Figure 80). Similarly, cluster node 3 had irregularity in the frequency and percentage counts between the two cumulative periods (22.45% and 26.77%) except there is a higher frequency counts in cluster node 3 in the cumulative period (b) 2015Dec (Figure 100a, b and Figure 102b). Like cluster node 2, the frequency and percentages of objects (counts) in the cumulative period (a) 2015Jun (22, 22.45%) is lower than it was in the period (b) 2015Dec (34, 26.77%) (Table 1 and Figure 81). The objects in Nodes 2 and 4 for the period 2015June show that there are outliers (single) and stand-alone, while the objects for the same nodes for the period (b) 2015Dec are multiple in clusters. Finally, there is a drop to four in the number of clusters in (b) 2015Dec suggesting there are more similarities created in most of the clustering elements as the number events activities increased in the time series of the ransomware network. Conclusively, the main objective here is to show that the cluster overlap node 1 has been very consistent as the active cluster overlap node even when investigated using the frequency counts of the clustering objects.

Looking further at the frequency of objects counts brings focus to the cumulative period (a) 2016Jun and (b) 2016Dec (Figure 103a, b). As seen in the previous cumulative periods, the frequencies and percentages of the objects count for cluster overlap nodes 1 for the periods (a) 2016Jun (23, 29.11%) and (b) 2016Dec (23, 27.38%) are regular and consistent (Table 1 and Figure 81). This means that cluster node 1 is a fully formed and developed (active) ransomware transmission hub. Similar irregularities recorded in the previous cumulative periods are also recorded in the frequency of objects counts for cluster nodes 2 and 3 for the period (a) 2016June (Figure 103a). It is obvious that the increased number of clustering objects in these nodes reduced the number of clusters from the previously recorded four to three.



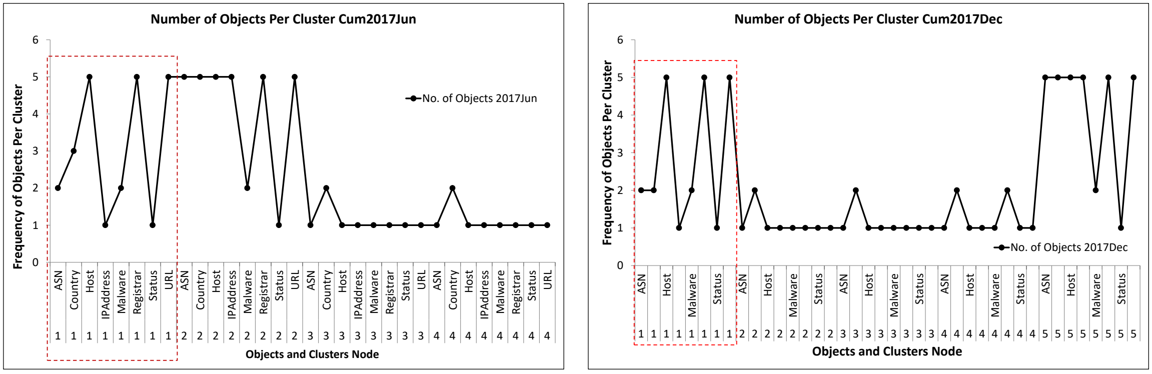
(a) (b)

Figure 103: Frequency of Objects per Cluster for Cumulative Periods (a) 2016Jun and (b) 2016Dec: showing the consistency of cluster overlap node 1

The increased number of clustering objects suggests increased ransomware traffic through those nodes. After what seemed to be an increased (intense) activity in the period ending 2016Jun where all the clusters appear to be fully formed, there was a release or dispersion of traffic (threat) to network devices in subsequent period 2016Dec (Figure 103b). The result was an increase in the number of clusters to 5 for the period 2016Dec. The increase to 5 in the number of clusters is reflected in the decrease in percentage intensities of cluster 1 = 29.15%, cluster 2 = 2.58% and cluster 3 = 2.50%; and the appearance of cluster 4 = 2.73% and a high value cluster 5 = 63.05% compared to their values in 2016Jun (Figure 75). The same decrease in distribution pattern is shown in the number and percentages of clustering elements for cluster 1 = 23, 27.28%; cluster 2 = 9, 10.71%; cluster 3 = 9, 10.71%; cluster 4 = 10, 11.90% and cluster 5 = 33, 39.29% (Table 1 and Figure 81) This agrees with the spike in ransomware activities recorded earlier in the time series of the ransomware network (Figure 29). While cluster overlap node 1 remains stable and consistent in this period, cluster nodes 2, 3, 4, and 5 show inconsistencies and irregular topological patterns (Figure 101 and Figure 102a-d). Most of these nodes not only have reduced frequency of objects counts, they show to have outlier (stand-alone) objects. An exception is cluster node 5, which increased its frequency of cluster objects counts (63.05%) and had proven earlier to be a receiver node with major traffic from cluster node 1. Worthy of note is the frequency transition between 2015Dec and 2016Jun (Figure 69, Figure 100b and Figure 103a) where the number of clusters decreased from four to a minimum of three, which suggests the onset of ransomware threats and full threat, respectively. Notably, the percentage values of clustering elements increased from cluster 1 = 25.20%; cluster 2 = 23.62%; cluster 3 = 26.77 and cluster 4 = 24.41% in 2015Dec to cluster 1 = 29.11%; cluster 2 = 41.77% and cluster 3 = 29.11% in 2016Jun. These percentage decreased in values for cluster 1 = 27.38%; cluster 2 = 10.71%; cluster 3 = 10.71% and gained cluster 4 = 11.90% and cluster 5 = 39.29% in 2016Dec (Table

1 and Figure 81). At the point of minimum threshold of the number of clusters, onset of threat, the percentage values of intensity distribution shows that cluster 1 recorded an average intensity for 2016Jun = 31.03%; 2016Dec = 29.15%; 2017Jun = 31.57%; 2017Dec = 29.59% and 2018Jun = 29.57%. Conversely, at the point of active threat, cluster 5 recorded percentage intensity values as follows: 2016Dec = 63.05%, 2017Dec = 62.71% and 2018Jun = 62.72% while clusters 2, 3 and 4 recorded approximate value of 2.58% each. The exception here is cluster 2 in 2017Jun which recorded 63.27% (Figure 75). However, the consistency of node 1 as the traffic generating and active cluster overlap node remains valid. Therefore, its removal, especially at the point where the number of clusters is lowest (the onset), would disrupt and prevent the threat at the proceeding maximum number of clusters (at the spike).

The results of subsequent cumulative periods are like those of the preceding periods. It is obvious that the frequency of the object counts of cluster overlap node 1 for the periods 2017Jun and 2017Dec (Figure 104a, b) remains structurally regular and consistent. This corroborates the results of the structures of the various cumulative network clusters, the key performance parameters and the intensity values of the objects of the cluster overlaps. Therefore, this observation shows that cluster overlap node 1 continues to be the active cluster overlap node. The irregularities and topological differences in cluster nodes 2, 3, 4, and 5 are clearly visible in the frequency (counts) of the clustering objects (Figure 104a, b).

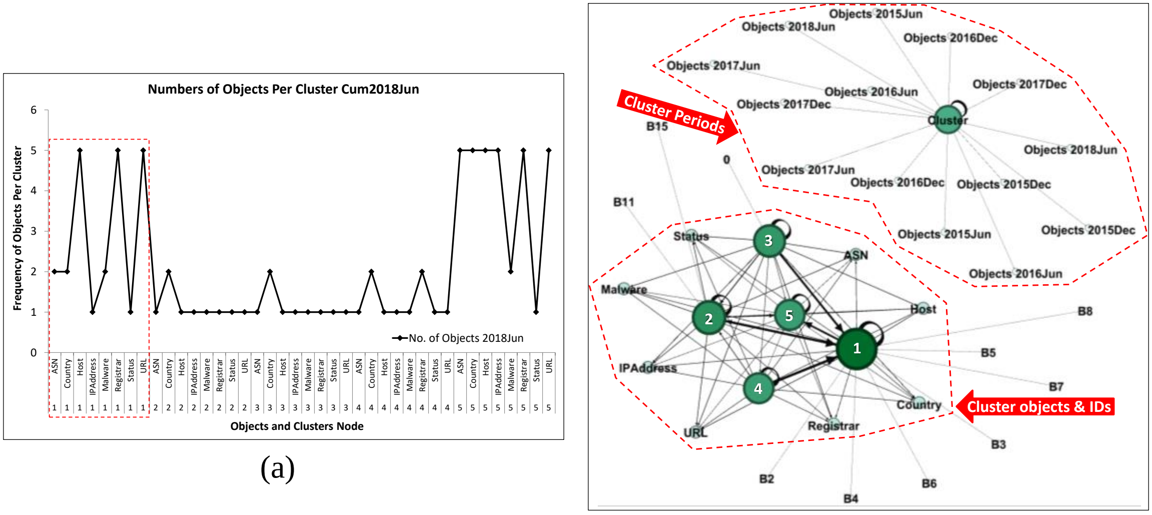


(a) (b)

Figure 104: Frequency of Objects per Cluster for Cumulative Periods (a) 2017Jun and (b) 2017Dec: showing the consistency of cluster overlap node 1

While cluster node 2 in the cumulative period 2017Jun (Table 1, Figure 81 and Figure 104a) recorded an increased frequency count (44.00%) in the clustering objects, it contrasts with the frequency of objects (10.71%) for the corresponding cluster node 2 in the cumulative period 2017Dec. Clusters nodes 3 and 4, have similar objects frequency values of 12.00% and 12.00% for 2017Jun; 10.71% and 11.90% for 2017Dec and 39.29% new cluster 5 in 2017Dec. Like the frequency of objects for cluster node 5 during the period of threat 2016Dec (Figure 103b), the frequency of the clustering objects increased (39.29%) in the same pattern during the period 2017Dec. This phenomenon tends to suggest that the ransomware threat repeated, which implies that the system required to control the threat would be self-healing and cyclic (repeating in cycles). However, the interesting fact in this case is that the active cluster overlap will still have the same attributes, which remains cluster overlap node 1 making it easy for the system to target and select it for removal.

Arguably, the frequencies of clustering objects for all cluster nodes in the cumulative period 2018Jun (Figure 105a) are the same as it was in 2017Dec. Hence, without repetition, the discussion held for 2017Dec is adopted for 2018Jun.



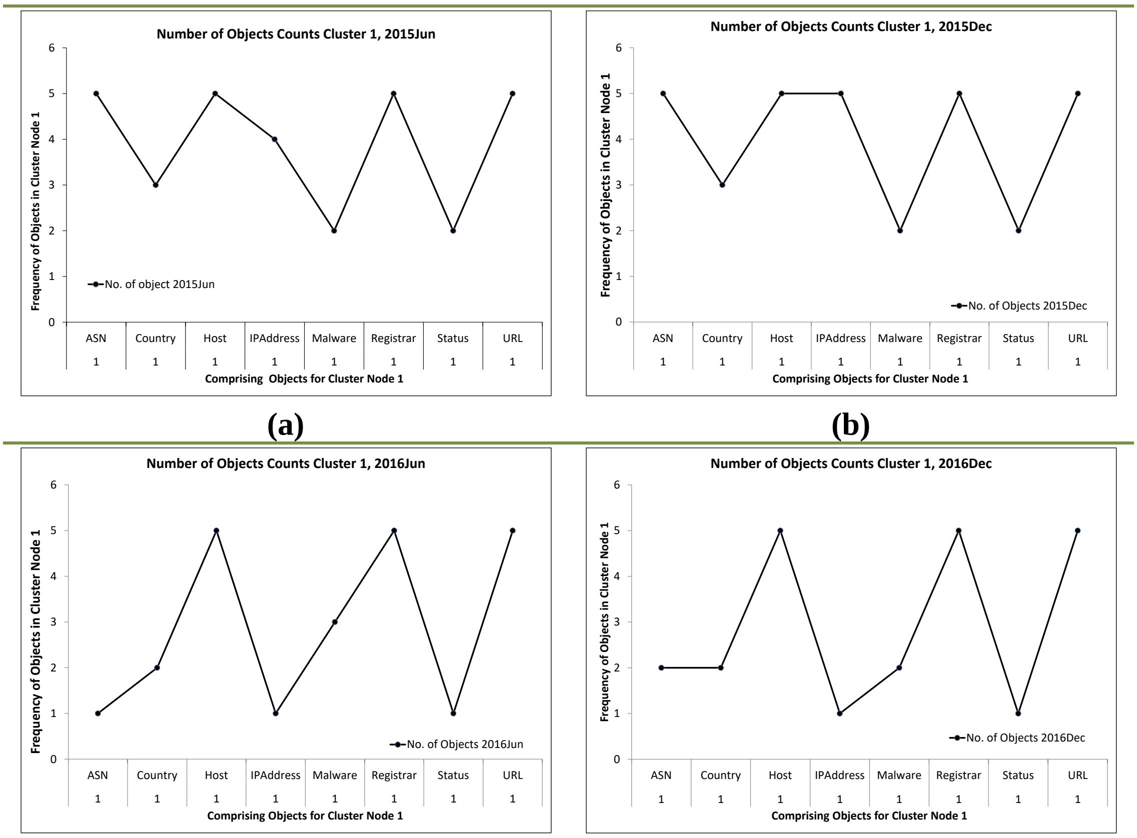
(a)

(b)

Figure 105: Frequency of Objects per Cluster for (a) Cumulative Periods 2018Jun, (b) Cumulative Periods 2015-2018 showing various cluster periods on top, and degree of cluster objects for various cluster overlaps IDs 1, 2, 3, 4, 5, at bottom

Generally, the single plot for the frequency of objects for the clusters in all the cumulative periods is represented in the cumulative periods 2015-2018 (Figure 105b). The upper part of the plot shows the cluster of the various cumulative periods of the time series. The corresponding distribution frequency of cluster of objects amongst the clusters is highlighted in the lower part of the plot. The cluster objects frequency distribution shows the degree of object frequency for each node by the size and color intensity of the node circle. Like in the earlier observations cluster overlap node 1 has the highest frequency of cluster objects suggesting that it has the highest number of clustering objects and is more densely linked than the other nodes. This degree of connection is followed by node 2, node 3, node 4, and node 5, respectively. In conclusion, node 1 remains the active cluster overlap node as observed in earlier discussions. These results validate the understanding that cluster overlap node 1 is the active cluster overlap node. Consequently, its removal dislodges the ransomware network and controls any ransomware threat.

The graphical representation of the tracking of cluster overlap node 1 using frequency of objects counts for cluster 1, Cum2015Jan-2018Jun [(a) – (h)] (Figure 106) gives clarity on the consistency, pattern and structural formation of cluster overlap node 1. Looking at the plots (a) – (g), there is no doubt that cluster node 1 is consistently the active cluster node. Placing all the frequency counts for cluster node 1 in one plot (Figure 106h) shows a very high degree of consistency in the frequency counts for all the cumulative periods. The slight difference in the frequency value is seen in 2015Jun and 2015Dec prior to the onset of ransomware threat where the object (ASN) in the plots recorded higher frequencies. However, at the onset of threat in 2016Jun and 2017Jun (Figure 106c, e) the same object (ASN) recorded a very low frequency. Subsequently, the frequency increased slightly and remained at the same values when there was a full-grown threat in the period 2016Dec, 2017Dec and 2018Jun.

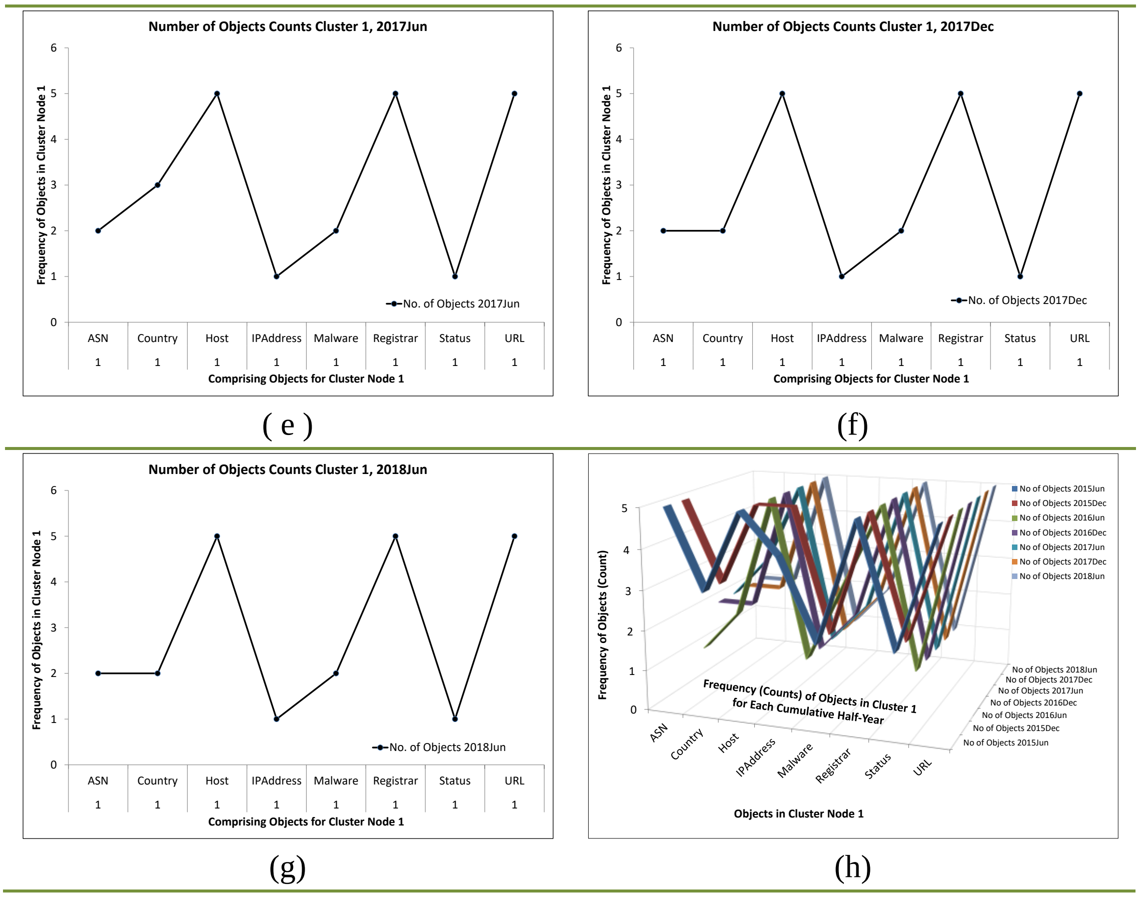


(a)

(b)

(c)

(d)



(e)

(f)

(g)

(h)

Figure 106: Tracking Cluster 1 using Number of Objects Counts for Cluster 1, Cum2015Jan-2018Jun [(a) – (h)]: Showing the consistency of cluster overlap node 1

Convincingly, the results have proved that cluster overlap node 1 is the active cluster node in the entire time series of the ransomware network. The results have been validated in various ways using the visualization of the ransomware networks, the key performance parameters of the network, the overlaps clustering objects intensities, and the frequencies (counts) of overlaps clustering objects. Therefore, it can be concluded that the removal of the active cluster overlap node (1) would consequently dislodge the ransomware network and control ransomware threat.

6.5 Summary of Key Findings

Various methodological approaches have been used to investigate and explore the clustering overlaps in a ransomware network based on the link structure and content relevance to build a knowledge and understanding of cluster evolution and development. The approaches include visualization of the clusters; exploration of the measures of centrality (key performance parameters), intensities of the clusters and the profiling of the clustering items to understand the topological evolution and development of the clusters. These approached were carried out to detect the dominant (active) cluster that can be removed to dislodge the ransomware network and prevent threat before it attacks. To give a summary of the findings, it is important to revisit the data in the following Figure 107, Table 2, Figure 108 and Figure 109.

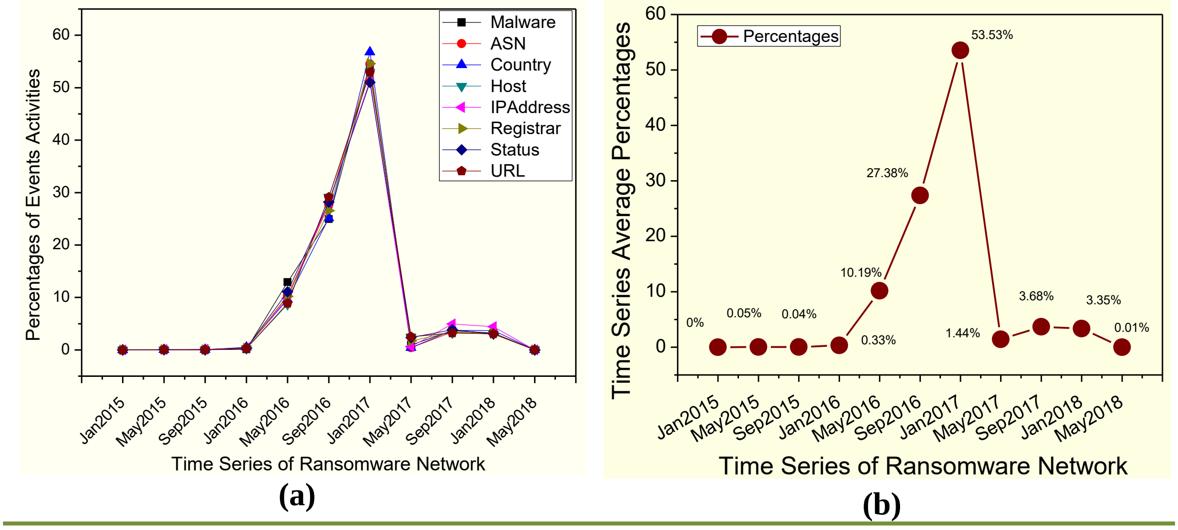
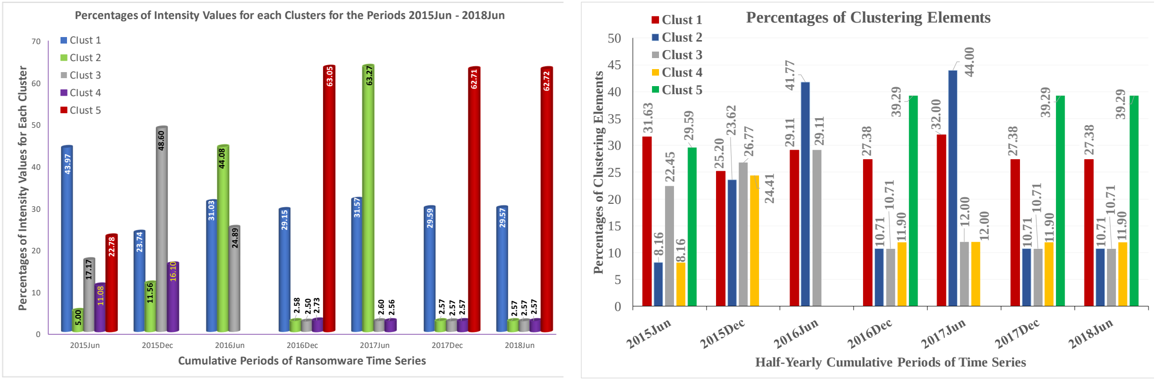


Figure 107: The (a) Number of Events Activities (b) Percentages of Number of Events Activities of Multiple time series for all variable

Table 2: Number of Clustering Elements and Percentages

Period	Number of Clustering Elements						Percentage of Clustering Elements				
	Clust 1	Clust 2	Clust 3	Clust 4	Clust 5	Total	Clust 1	Clust 2	Clust 3	Clust 4	Clust 5
2015Jun	31	8	22	8	29	98	31.63	8.16	22.45	8.16	29.59
2015Dec	32	30	34	31		127	25.20	23.62	26.77	24.41	
2016Jun	23	33	23			79	29.11	41.77	29.11		
2016Dec	23	9	9	10	33	84	27.38	10.71	10.71	11.90	39.29
2017Jun	24	33	9	9		75	32	44	12	12	
2017Dec	23	9	9	10	33	84	27.38	10.71	10.71	11.90	39.29
2018Jun	23	9	9	10	33	84	27.38	10.71	10.71	11.90	39.29



(a) (b)

Figure 108: Percentages of (a) Number of Clustering Elements and (b) Intensity Values for each Cluster for the Periods 2015Jun - 2018Jun

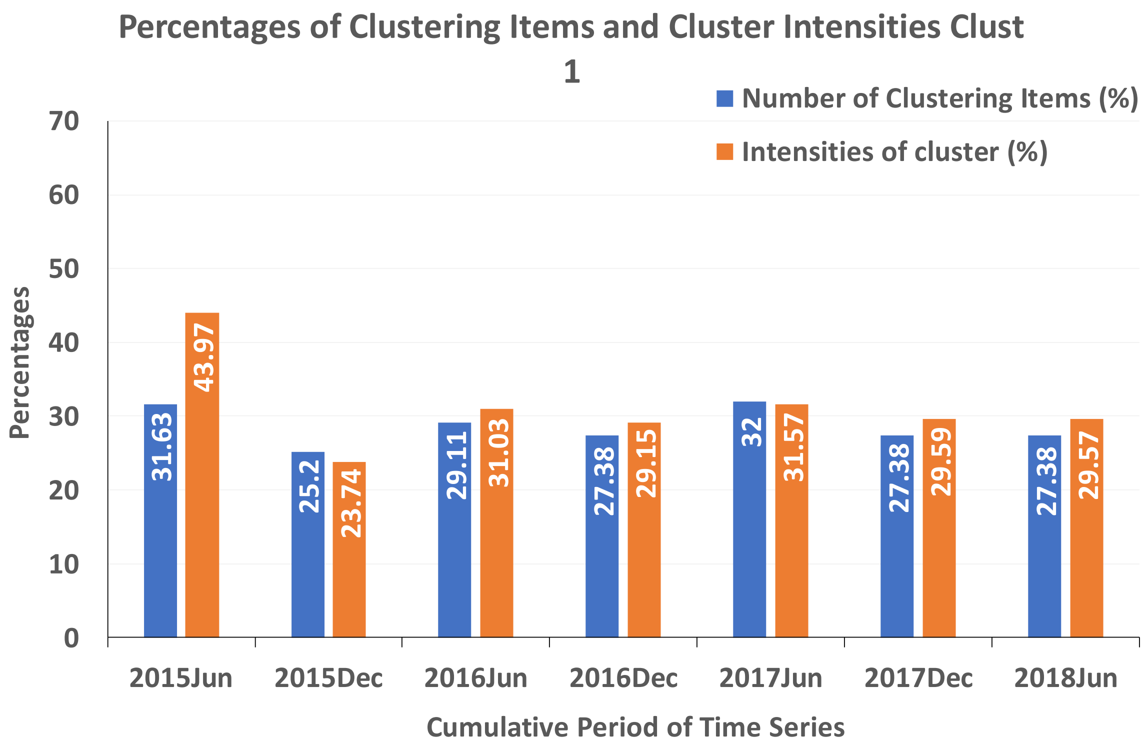


Figure 109: Percentages of (blue) Number of Clustering Elements and (orange) Intensity Values for the Active Cluster (Cluster 1) for the Periods 2015Jun - 2018Jun

Reading the data from the figures and table above, the investigation records the following findings regarding the formation and development of the overlapping clusters and the active overlapping cluster node in the ransomware network. The following findings are recorded:

1) The percentage distribution of the events variables remained stationary at **0.00% - 0.33%** between 2015Jun and 2016Jun. The events variables increased to **10.19%, 27.38%** and reached a climax at **53.53%** in 2016Dec before the values decreased drastically to **1.44%, 3.68%, 3.35% and 0.01%** (Figure 107a, b)
2) Prior to the onset of threat, the number of clusters decreased progressively until the onset of the threat at the lowest minimum threshold (2016Jun) where the clustering elements and percentage values are close to one another (**29.11%,**

41.77% and 29.11%) and the differences between them are **-12.11% and 12.11%** (Table 2 and Figure 108a).

3) At the onset of threat (2016Jun), the percentage values of the clusters are close to one another and the differences between the percentages of the intensity of originating cluster **(31.03%)**, the cluster with largest intensity **(44.08%)** and the intensity of the terminal cluster **(24.89%)** are **-13.05% and 14.19%** respectively (Table 2 and Figure 108b).

4) At the onset of threat (2016Jun), the number of clustering objects are close to each other and the differences between the percentage values of the number of clustering objects of the originating cluster **(29.11%)**, the cluster node with the largest percentage value **(41.77%)** and the terminal cluster node **(29.11%)** are **-12.66% and 12.66%** respectively (Table 2 and Figure 108b).

5) Proceeding the onset of threat is the active threat (2016Dec) where the percentage values of the active/originating cluster node is **27.38%**, the terminal cluster node is **39.29%**, and other cluster nodes are **10.71%, 10.71% and 11.90%** (Table 2 and Figure 108b).

6) The active overlapping cluster is the cluster that has the most regular, consistent and closely distributed number of clustering objects, measures of centrality and intensity values in all the cumulative periods of the ransomware network time series (Table 2, Figure 108 and Figure 109).

Therefore, any self-healing or cyclic system automated with the above findings will automatically detect and dislodge a ransomware threat before an attack occurs. The above findings will help to reach a summary, conclusion, recommendations and projected future work in the next chapter.

CHAPTER 7

SUMMARY, CONCLUSION, AND RECOMMENDATION FOR FUTURE WORK

The aim of this research was to study the topological evolution of clusters in a half-yearly cumulative period of the time series of a ransomware network. Besides the visual representation and understanding of the structural evolution of the ransomware network clusters, the study investigated the formation and development of overlapping clusters. Understanding the dynamics of the cluster overlaps was important to inform the creation of tools to control ransomware threats. The study sought to identify the active cluster overlap node to remove that would consequently dislodge the ransomware network and prevent threats before they occurred. The investigation applied different exploratory approaches to understand the dynamics of an attack profile, investigate, establish and validate the results, which found that the removal of the active cluster overlap node effectively dislodged the ransomware network and controlled the threats. However, the effectiveness of the practical application of any system based on this principle is beyond the scope of this project and should be investigated in future studies.

7.1 Visualization of Network Cluster and Key Performance Parameters

Visualization was very important to reveal hidden patterns in the network that would make it possible to establish the dynamics of the clustering entities. In visualizing the topological changes within the clusters of the ransomware network at different cumulative periods of the time series, it was possible to identify the key performance parameters that influenced the structural changes in the network clusters. Consequently, the pattern of the formation and development of these clusters prior to the onset and full-grown threat of the ransomware was established. The network clusters revealed that while there was a period of gradual increase in the density of the distribution of links (connections) around the cluster nodes, as the activities of the ransomware increased, there was more concentration of these links on a specific cluster overlap node. This specific node showed to be generating all the outward links and never received an inward link (connection). Additionally, it was revealed there was a corresponding decrease in the number of clusters of the ransomware network during the same periods as the events' activities and link distribution density increased. This suggested that the onset of the ransomware threat was at the point when the number of clusters was at its minimum. The full-grown threat proceeded immediately after the minimum number of cluster threshold by recording a maximum number of clusters. This indicates a period when the threat was dispersed to different target networks and devices. The structural dynamics of the network clusters were characterized by the values of the key performance parameters. In such characterizations, a high weight value revealed the magnitude of links (traffic) to a cluster node. Out degree centrality, weighted eigenvector centrality, and unweighted eigenvector centrality showed that cluster overlap node 1 had the highest degree of outward links and generated and distributed all the traffic to other nodes. Other parameters confirming the dominance of cluster overlap node 1 included the consistently high values of weighted and unweighted closeness centrality, and clustering coefficient centrality.

7.2 Visualization of Network Cluster Overlaps and the Intensities of the Objects of Cluster Overlaps

The visualization of ransomware network created an understanding of the clustering dynamics to motivate the analysis of the overlapping clusters of the network. Further investigation used the cluster overlaps and the corresponding intensities of the clustering objects of the overlapping clusters to identify and track the active cluster overlap node at the different cumulative periods. This was to establish and ensure the consistency of the identified active cluster node at the various cumulative periods of the time series of the ransomware network. Like the network clusters, the cluster overlaps showed a period of gradual decrease in the number of clusters and increased connections within a few cluster nodes. The decrease in the number of clusters reached a minimum value (threat onset) prior to a sudden increase (full-grown threat) in the number of clusters. These connections showed that cluster node 1 was both the originator and distributor of the outward traffic and never received an inward traffic, thereby becoming the active source of threat. The intensity values of the clustering objects revealed that the weight, density, and strength of the closeness (association) between the intensities of the clustering objects were higher in the cluster overlap node 1 than it was in the other cluster nodes. This implies that the differences between the intensity values of the clustering objects were small and evenly distributed. It was also seen that the range of the values fell in the thresholds of 0.1≥Intensity≤0.9 for each object. Unlike the other cluster overlap nodes, cluster overlap node 1 did not have any outlier objects or objects with a single intensity value equal to 1.00 or less than 0.1. The intensities of the malware (ransomware) object was seen to record a high and consistent peak value of 0.82 for the cumulative periods 2016Dec, 2017Jun, 2017Dec, and 2018Jun, which were the periods when the time series of the network recorded the highest events activities (Figure 29 and Figure 99a, b). In other words, it was the period when the analysis showed the onset and full-grown ransomware threat. The analysis using intensity values proved and validated cluster overlap node 1 as the active cluster overlap node as well as the links or events generating node. Consequently, its removal effectively dislodged the network and apparently controlled ransomware threat. The experimental and exploratory procedures in this research shows that the concept of active cluster overlap node is potentially an effective/useful and new identifier that both enables an automated response and pro-actively controls the attack at an earlier time-point than was evidently achieved by manual processes.

7.3 Frequencies (counts) of Objects of Cluster Overlaps

To understand the formation and development of the cluster overlaps, further analysis and validation was conducted using the frequencies of each of the clustering objects. The results gave insight into how the objects were entering or leaving a cluster. Consistent with the cluster weight or distribution (link) density, the frequency counts explain why the number of clusters increased or decreased in the successive cumulative periods. In all the cumulative periods, the values of the frequencies of cluster overlap node 1 proved to be consistent and regular in contents than the other cluster overlaps. Like the other analysis and validation parameters, the frequency counts showed that cluster overlap node 1 was the active cluster overlap node. Therefore, the removal of the active cluster overlap node 1 dislodged the ransomware network and consequently controlled the ransomware threats.

7.4 Characteristics of active cluster overlap node and Recommendations

Recommendations for future work is built on the following characteristics of the active cluster overlap node in a ransomware network:

a) The onset of threat is detected when an apparent zero (%) stationary distribution of events (network objects) suddenly increased to ≥8.67%≤12.91% and ≥24.96%≤30.00%; and the active threat is detected when the values reach the maximum of ≥51.00%≤56.77% prior to near zero % drop in activities.

b) At the point of active threat, the weighted betweenness centrality for the terminal cluster node recorded a value of 100% while other clusters recorded 0%. The values of all other measures of centrality appear to be within the same range.

c) The active cluster overlap has no outlier objects; therefore, it does not have any object intensity values that are equal to 1.00.

d) The intensity values of clustering objects in the active cluster lies between ≥0.1≤0.9.

e) The intensity value of the ransomware object at the period of threat is maximum at 0.82.

f) The frequency counts explain why the number of clusters increased or decreased in the successive cumulative periods.

g) In all the cumulative periods, the values of the frequencies of the active cluster overlap node 1 proved to be consistent and regular in contents than the other cluster overlaps.

7.5 Review of research aim, methodology, and contributions

This research aimed to develop detailed knowledge and understanding of the dynamic emergence of network overlap clusters and how this might be detected and evaluated by automated processes grounded in data science principles. The automated active cluster overlap detection and removal model (Exploratory Machine-Learning Cluster Overlap System – **EMCOS**) to be built from this principle seeks to help industries to take a quick decision to counter and control ransomware threats before they attack. To achieve this and demonstrate the effectiveness of the research a prototype tool would be created in a future work. The identification of the characteristics of the dynamic active overlapping cluster in a ransomware network is the precursor to the creation of the prototype **EMCOS.** These characteristics are identified by profiling the ransomware network using:

- Visualization of the emergence and development of clusters and cluster overlaps to reveal hidden patterns
- Measures of centrality of the clustering objects to identify key performance parameters and how they influence the distribution of the objects
- Intensity values of the clustering objects to understand how the objects of the overlapping clusters fit within the clusters
- Frequency of the clustering Objects to understand the distribution events (activities) within each of the clusters in the network.

To this end, the methodology adopted in this investigation was effective in achieving the aim of the investigation, therefore developing the knowledge and understanding of the characteristics of overlapping cluster that could be applied in future work to create an automated EMCOS to detect and control ransomware threat.

7.6 Conclusion and Future Works

The aim of this investigation was to explore and identify the characteristics of the active cluster overlap to remove from a ransomware network to respond timely in the control of ransomware threat. In pursuit of this aim, the exploratory machine learning approach was applied to investigate and analyse the topological formation and evolution of the network clusters and cluster overlap at different cumulative periods of the ransomware time series. This was to identify the characteristics of the active cluster overlap to remove and at what stage to remove it to control a threat. The parameters that influenced the dynamics and structural changes were analysed and discussed. These were related to the various dynamics of the ransomware threat at different stages of development ranging from a period of no-threat, on-set of threat, to full-grown threat. All the parameters used to investigate and validate the consistency of the active cluster overlap node proved that the onset of the threat was at the cumulative period in the time series when the number of clusters was at a minimum prior to a phenomenal increase in the number of clusters, which was at the full-grown threat. The threat was found to continue to persistently function within the threshold of this minimum and maximum number of clusters. Therefore, the removal of the active cluster overlap node 1 effectively dislodged the ransomware network and apparently controlled any intending ransomware threat. For emphasis, the experimental and exploratory procedures in this research shows that the concept of active cluster overlap node is potentially a reliable and new identifier that both enables an automated response and pro-actively controls the attack at an earlier time-point than was evidently achieved by manual processes.

In conclusion, the investigation has proved that a ransomware control tool built on the principle of the dynamics of the cluster overlap could help to bring security and relief to all who carry out transactions on the network.

The present investigation has developed the knowledge and understanding of the formation and development of active overlapping cluster. The identification of the characteristics of the active overlapping cluster is key to the future work of developing an automated system to help industries to make timely decision to counter and control threat before they attack. Future work would focus on building an automated active cluster overlap detection and removal model (Exploratory Machine-Learning Cluster Overlap System – **EMCOS**) using the principles developed in the present work.

However, the effectiveness of the practical application of any system that is developed based on this principle is beyond the scope of the present study because of social, ethical, and need for moral network, etc.; and is left for future studies to characterize. In addition, a more detailed understanding of how the ransomware events join a given cluster overlap node or leave it to form a new one would be the focus of a future study.

References

[1] Cyberark Labs. Analyzing Ransomware and Potential Mitigation Strategies. 2016;10. Available from: http://lp.cyberark.com/rs/316-CZP-275/images/CyberArk Lab Research - Ransomware-073116-web.pdf

[2] Jang-jaccard J. A survey of emerging threats in cybersecurity. J Comput Syst Sci [Internet]. Elsevier Inc.; 2014;80(5):973–93. Available from: http://dx.doi.org/10.1016/j.jcss.2014.02.005

[3] Analysis A, Ransomeware L. CryptoWall CryptoWall Version.

[4] Symantec. An ISTR Special Report: Ransomware and Businesses 2016. Ransomware and Businesses. 2016.

[5] Accenture. Cost of Cyber Crime Study 2017 Insights on the Security Investments that Make a Difference. 2017;

[6] Morgan S. Global Ransomware Damage Costs Predicted To Hit $11.5 Billion By 2019 [Internet]. 2017. Available from: https://cybersecurityventures.com/ransomware-damage-report-2017-part-2/

[7] Hornby A. Advanced Learner’s Dictionary. In: Turnbul J, Hornby, editors. oxford. 8th ed. 2010. p. 1732.

[8] News SW. LeChiffre Ransomware Hits Indian Banks , Pharma Company. 2016. p. 3–4.

[9] What Is Ransomware ? How Do I Protect My Networks ? How Do I Respond To Ransomware ? :2.

[10] Symantec. The evolution of ransomware. 2015;57.

[11] O’Brien D. Special Report : Ransomware and Businesses 2016. Symantec Corp. 2016;1–30.

[12] Kent SL. Bitcoin and Money Laundering - Mining for an Effective Solution.

2005;1.

[13] Pathak PB, Nanded YM. A Dangerous Trend of Cybercrime: Ransomware Growing Challenge. Int J Adv Res Comput Eng Technol [Internet]. 2016;5(2):371–3. Available from: http://ijarcet.org/wp-content/uploads/IJARCET-VOL-5-ISSUE-2-371-373.pdf

[14] Böhme R, Christin N, Edelman B, Moore T. Bitcoin: Economics, Technology, and Governance. J Econ Perspect [Internet]. 2015;29(2):213–38. Available from: http://pubs.aeaweb.org/doi/10.1257/jep.29.2.213

[15] Kharraz A, Robertson W, Balzarotti D, Bilge L, Kirda E. Cutting the gordian knot: A look under the hood of ransomware attacks. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). 2015;9148:3–24.

[16] Hampton N, Baig ZA. Ransomware: Emergence of the cyber-extortion menace. Aust Inf Secur Manag Conf [Internet]. 2015;13:47–56. Available from: http://ro.ecu.edu.au/ism/180

[17] Savage K, Coogan P, Lau H. The evolution of ransomware. 2015;

[18] Richardson R, North M. Ransomware : Evolution , Mitigation and Prevention. 2017;13(1):10–21.

[19] Paracha AM. Analysis of Increasing Malwares and Cyber Crimes Using Economic Approach. 2013;3(3):487–91.

[20] Bojanc R, Jerman-Blažič B. An economic modelling approach to information security risk management. Int J Inf Manage. 2008;28(5):413–22.

[21] Lelarge M, Bolot J. Economic incentives to increase security in the internet: The case for insurance. Proc - IEEE INFOCOM. 2009;1494–502.

[22] Lelarge M. Economics of malware: Epidemic risks model, network externalities and incentives. 2009 47th Annu Allert Conf Commun Control Comput Allert 2009.

2009;2009:1353–60.

[23] Talha KA, Alper DI, Aydin C. APK Auditor: Permission-based Android malware detection system. Digit Investig [Internet]. Elsevier Ltd; 2015;13:1–14. Available from: http://dx.doi.org/10.1016/j.diin.2015.01.001

[24] Ren J, Yang X, Yang L-X, Xu Y, Yang F. A delayed computer virus propagation model and its dynamics. Chaos, Solitons & Fractals. Elsevier Ltd; 2012;45(1):74–9.

[25] Chen Z, Ji C. Spatial-temporal modeling of malware propagation in networks. IEEE Trans Neural Networks. 2005;16(5):1291–303.

[26] Hofmeyr S, Moore T, Forrest S, Edwards B, Stelle G. Modeling Internet-Scale Policies for Cleaning up Malware. Econ Inf Secur Priv III [Internet]. 2013;149–70. Available from: http://link.springer.com/10.1007/978-1-4614-1981-5_7

[27] Hofmeyr S, Moore T, Forrest S, Edwards B, Stelle G. Modeling Internet-Scale Policies for Cleaning up Malware. In: Bruce Schneier, editor. Economics of Information Security and Privacy III. Springer; 2013. p. 149–70.

[28] Wang P, Wang Y-S. Malware behavioural detection and vaccine development by using a support vector model classifier. J Comput Syst Sci [Internet]. 2015;81(6):1012–26. Available from: http://www.sciencedirect.com/science/article/pii/S0022000014001780

[29] Wang P, Wang Y-S. Malware behavioural detection and vaccine development by using a support vector model classifier. J Comput Syst Sci. Elsevier Inc.; 2014;1(Ml):1–15.

[30] Ross RS. Guide for Conducting Risk Assessments. Spec Publ (NIST SP) - 800-30 Rev 1 [Internet]. 2012;(September):95. Available from: http://dx.doi.org/10.6028/NIST.SP.800-30r1

[31] ISO. ISO/IEC 27000:2018(en), Information technology — Security techniques —

Information security management systems — Overview and vocabulary [Internet]. ISO Online Browsing Platform (OBP). 2018 [cited 2018 Sep 1]. Available from: https://www.iso.org/obp/ui/#iso:std:iso-iec:27000:ed-5:v1:en

[32] Swanson M, Bowen P, Phillips AW, Gallup D, Lynes D. Contingency Planning Guide for Federal Information Systems [Internet]. [cited 2018 Sep 1]. Available from: http://csrc.nist.gov/publications.

[33] Memisevic R. Machine Learning in MATLAB Last time. Control. 2007;1–4.

[34] Choosing the right estimator — scikit-learn 0.19.1 documentation.

[35] Hu L, Chan KCC. Fuzzy Clustering in a Complex Network Based on Content Relevance and Link Structures. IEEE Trans Fuzzy Syst. 2016;24(2):456–70.

[36] Bora DJ, Gupta DAK. A Comparative study Between Fuzzy Clustering Algorithm and Hard Clustering Algorithm. Int J Comput Trends Technol [Internet]. 2014;10(2):108–13. Available from: http://www.ijcttjournal.org/archives/ijctt-v10p119

[37] Schaeffer SE. Graph clustering. Comput Sci Rev. 2007;1(1):27–64.

[38] Kovács IA, Palotai R, Szalay MS, Csermely P. Community landscapes: An integrative approach to determine overlapping network module hierarchy, identify key nodes and predict network dynamics. PLoS One. 2010;5(9):1–14.

[39] Barabási A-L. NETWORK SCIENCE. 10. Spreading phenomena. Netw Sci. 2015;1–63.

[40] Rogers J. Presbyterian guidelines for biblical interpretation: Their origin and application to homosexuality. Biblic Theol Bull. 2007;37(4):174–83.

[41] Kirik AM, Arslan A, Çetİnkaya A, Gül M. A Quantitative Research on the Level of Social Media Addiction among Young People in Turkey Türkiye ' deki Gençlerin Sosyal Medya Bağımlılık Düzeylerine Yönelik Nicel Bir Araştırma.

2015;3(September):108–22.

[42] Fogel J, Nehmad E. Computers in Human Behavior Internet social network communities : Risk taking , trust , and privacy concerns. Comput Human Behav [Internet]. Elsevier Ltd; 2009;25(1):153–60. Available from: http://dx.doi.org/10.1016/j.chb.2008.08.006

[43] Brown JO, Broderick AJ, Lee N, Brown JO, Broderick AJ. ONLINE COMMUNITIES : CONCEPTUALIZING THE ONLINE SOCIAL NETWORK. 2007;21(3):2–20.

[44] Donetti L. Detecting Network Communities: a new systematic and efficient algorithm. 2008;1–8.

[45] Atzori L. From “ Smart Objects ” to “ Social Objects ”: The Next Evolutionary Step of the Internet of Things. 2014;(January):97–105.

[46] Shirkhorshidi AS, Aghabozorgi S, Wah TY. A Comparison Study on Similarity and Dissimilarity Measures in Clustering Continuous Data. Dalby AR, editor. PLoS One. Public Library of Science; 2015 Dec;10(12):e0144059.

[47] Wang X, Zhan ZQ, Heipke C. an Efficient Method To Detect Mutual Overlap of a Large Set of Unordered Images for Structure-From-Motion. ISPRS Ann Photogramm Remote Sens Spat Inf Sci. 2017;IV-1/W1(June):191–8.

[48] What is “Wanna Decryptor”? A look at the ransomware that brought down the NHS - Mirror Online.

[49] What is WannaCry ransomware and why is it attacking global computers? | Technology | The Guardian.

[50] Ransomware TW, Switch K, Saved T, Pcs U, Harm F. The WannaCry Ransomware “Kill Switch” That Saved Untold PCs From Harm |. 2017;1–7.

[51] WannaCry ransomware cyber-attacks slow but fears remain - BBC News.

[52] Center for Strategic and International Studies. Net Losses: Estimating the Global Cost of Cybercrime. Mcafee. 2014.

[53] Greisiger M. Cyber Liability & Data Breach Insurance Claims Cyber Liability & Data Breach Insurance Claims. 2012;(October).

[54] Ponemon Institute. Cost of Data Breach. Ponemon Inst. 2016;(May):1–30.

[55] Ponemon, Ponemon Instititution. 2013 Cost of Data Breach Study : Global Analysis. Ponemon Instititution Benchmark Res Spons by Symantec. 2013;(May):1–25.

[56] Ponemon. 2013 Cost of Data Breach Study : Global Analysis Benchmark research sponsored by Symantec. 2013;(May).

[57] Symantec. Symantec internet security threat report trends for 2014. Symantec internet Secur Threat Rep trends 2014. 2015;20(v2.0):119.

[58] Ogazi-Onyemaechi BC, Dehghantanha A, Choo K-KR. Performance of Android Forensics Data Recovery Tools. In: Contemporary Digital Forensic Investigations of Cloud and Mobile Applications. Syngress Media Inc; 2016. p. 91–110.

[59] Bennett D. The Challenges Facing Computer Forensics Investigators in Obtaining Information from Mobile Devices for Use in Criminal Investigations. Inf Secur J. 2012;21(3):159–68.

[60] Netstar A. Research says shipments of smartphones grew 74 percent in 2010. Vol. 6. 2015. p. 5–9.

[61] Berg I, Insight B. Research says shipments of smartphones grew 74 percent in 2010 . News Archive. 2011;

[62] Fuente JD La, Santiago J, Román A, Dumitrache C, Casasanto D. When you think about it, your past is in front of you: How culture shapes spatial conceptions of time. Psychol Sci. 2014;25(9):1682–90.

[63] Data I, Idc T, Quarterly W, Phone M. IDC : Smartphone shipments to overtake

feature phones worldwide in 2013. 2013;(Idc):1–7.

[64] Simão A, Sícoli F, Melo L, Deus F, Sousa Júnior R. Acquisition and Analysis of Digital Evidence in Android Smartphones. Int J Forensic Comput Sci. 2011;6(1):28–43.

[65] Dilworth BD. Worldwide Smartphone Sales to Hit 1 . 5 billion in 2017 : IHS Report. 2017;

[66] Articles L, Webinars L. Smartphone Users Worldwide Will Total Mobile users pick up smartphones as they become more Best Practices in Digital Video Advertising Go beyond the articles : Hear from our clients : Want to learn more ? 2014.

[67] Neeraj M. Smartphone Sales 2015 – 2017 : India Will Surpass The US [REPORT] the engine of. 2015.

[68] Gartner says worldwide device shipments to grow 1.5 percent, to reach 2.5 billion units in 2015. 2015.

[69] Gozalvez J. Advances in Wireless Power Transfer. IEEE Veh Technol Mag. 2015;(december):14–32.

[70] Roland T, Trend B. Roland Berger Trend Compendium About the Roland Berger Trend Compendium 2030 What is it ? Use it ! 2014;(March):1–29.

[71] Woodgate P, Coppa I, Hart N. Global Outlook 2014 : Spatial Information Industry. CRC for Spatial Information; 2014.

[72] Reynolds R. Trends influencing the growth of digital textbooks in US higher education. Publ Res Q. 2011;27(2):178–87.

[73] Aouad L, Kechadi T, Trentesaux J. An Open Framework for Smartphone. Adv Digit Forensics VIII, IFIP AICT 383. 2012;159–66.

[74] - FND, - AD, - RM, - NFBMS, - S bin S. A Data-centric Model for Smartphone Security. Int J Adv Comput Technol [Internet]. 2013;5(9):9–17. Available from:

http://www.aicit.org/ijact/global/paper_detail.html?jname=IJACT&q=2934

[75] Mohtasebi S H, Dehghantanha A. Towards a Unified Forensic Investigation Framework of Smartphones. Int J Comput Theory Eng. 2013;5(2):351–5.

[76] Hoog A. Android and mobile forensics. Android Forensics [Internet]. 2011;(October 2008):1–40. Available from: http://linkinghub.elsevier.com/retrieve/pii/B9781597496513100019

[77] CBC News M. Cybercrime Moving to smartphones and tablets, say experts. 2013;0–2. Available from: http://www.cbc.ca/news/canada/manitoba/cybercrime-moving-to-smartphones-and-tablets-say-experts-1.1877058

[78] Savoldi A., Gubian P., Echizen I. A comparison between windows mobile and Symbian S60 embedded forensics. In: IIH-MSP 2009 - 2009 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing. 2009. p. 546–50.

[79] Ablon L, Libicki MC, Golay AA. Markets for Cybercrime Tools and Stolen Data: Hacker’s Bazaar. Natl Secur Res Div. 2014;

[80] Hoog A. Android Forensics: Investigation, Analysis and Mobile Security for Google Android. Security. 2011. 432 p.

[81] Mohtasebi S, Dehghantanha A. A Mitigation Approach to the Privacy and Malware Threats of Social Network Services. 2011;448–59.

[82] Sabena F, Dehghantanha A, Andrew PS. A review of vulnerabilities in identity management using biometrics. In: 2nd International Conference on Future Networks, ICFN 2010. 2010. p. 42–9.

[83] Kuntze N, Rudolph C, Alva A, Endicott-popovsky B, Christiansen J, Kemmerich T. On the Creation of Reliable Digital Evidence. In: Peterson G, Shenoi S, editors. Advances in Digital Forensics VIII, IFIP AICT 383. VIII. IFIP International

Federation for Information Processing 2012; 2012. p. 3–17.

[84] Ayers R, Jansen W, Brothers S. Guidelines on mobile device forensics (NIST Special Publication 800-101 Revision 1). NIST Spec Publ [Internet]. 2014;1(1):85. Available from: http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-101r1.pdf

[85] Damshenas M, Dehghantanha A, Mahmoud R. A Survey on Digital Forensics Trends. Int J Cyber-Security Digit Forensics. 2014;3(4):209–34.

[86] Dehghantanha A, Franke K. Privacy-respecting digital investigation. 2014 12th Annu Conf Privacy, Secur Trust PST 2014. 2014;129–38.

[87] Aminnezhad a, Dehghantanha a, Abdullah M. A Survey on Privacy Issues in Digital Forensics. Int J Cyber-Security Digit Forensics. 2012;1(4):311–23.

[88] Shaerpour K, Dehghantanha A, Mahmod R. Trends in Android Malware Detection. J Digit Forensics, Secur Law. 2013;8(3):21–40.

[89] Dezfouli Farhood N, Dehghantanha A, Mahmod R, Mohd Sani Binti NF, Shamsuddin S bin, Daryabar F. A Survey on Malware Analysis and Detection Techniques. 2013;5(October):42–51.

[90] Computer Viruses and Malware [Internet]. Vol. 22. 2006. Available from: http://link.springer.com/10.1007/0-387-34188-9

[91] Barmpatsalou K, Damopoulos D, Kambourakis G, Katos V. A critical review of 7 years of Mobile Device Forensics. Digit Investig [Internet]. Elsevier Ltd; 2013;10(4):323–49. Available from: http://dx.doi.org/10.1016/j.diin.2013.10.003

[92] Buchanan-Wollaston J, Storer T, Glisson W. Comparison of the Data Recovery Function of Forensic Tools. Adv Digit Forensics IX SE - 22 [Internet]. 2013;410:331–47. Available from: http://dx.doi.org/10.1007/978-3-642-41148-9_22%5Cnhttp://link.springer.com/10.1007/978-3-642-41148-9_22

[93] Grispos G, Storer T, Glisson WB. A comparison of forensic evidence recovery techniques for a windows mobile smart phone. Digit Investig [Internet]. Elsevier Ltd; 2011;8(1):23–36. Available from: http://dx.doi.org/10.1016/j.diin.2011.05.016

[94] Damshenas M, Dehghantanha A, Mahmoud R. A Survey on Malware propagation, analysis and detection. Int J Cyber-Security Digit Forensics. 2013;2(4):10–29.

[95] Tassone C, Martini B, Choo K-KR, Slay J. Mobile device forensics: A snapshot. Trends issues crime Crim justice. 2013;(460):1–7.

[96] Glisson WB, Storer T, Buchanan-Wollaston J. An empirical comparison of data recovered from mobile forensic toolkits. Digit Investig. 2013;10(1):44–55.

[97] Sidheeq M, Dehghantanha A, Kananparan G. Utilizing trusted platform module to mitigate botnet attacks. ICCAIE 2010 - 2010 Int Conf Comput Appl Ind Electron. 2010;(Iccaie):245–9.

[98] Shaerpour K, Dehghantanha A, Mahmod R, Technology I. Virtualized Honeynet Intrusion Prevention System in Scada. :11–5.

[99] Campbell-kelly M. Computing. 2009. p. 62–9.

[100] Shea B. The History Of The Internet (1969-2012). Internet Soc [Internet]. 2013; Available from: http://www.adweek.com/socialtimes/internet-history/475846

[101] Cerpa N, Bardeen M, Kitchenham B, Verner J. Evaluating logistic regression models to estimate software project outcomes. Inf Softw Technol [Internet]. Elsevier B.V.; 2010;52(9):934–44. Available from: http://dx.doi.org/10.1016/j.infsof.2010.03.011

[102] Microsoft. Evolution of Malware. 2016.

[103] Thonnard, O et al. Targeted Attacks: Definition and Common Traits. In: Research in Attacks, Intrusion, and Defenses; Industrial Espionage and Targeted Attacks. 2012. p. 66–85.

[104] Allodi L, Kotov V, Massacci F. MalwareLab: Experimentation with Cybercrime Attack Tools. Usenix Work Cyber Secur Exp Test, CSET. 2013;1–8.

[105] Mathew J. Cyber Crime costs Global Economy $500bn Annually. International Business Times. 2013.

[106] Williams R. Cyber Crime costs Global Economy $445bn Annually. The Telegraph. 2014.

[107] CHEONG C. Cyber crime costs global economy $445bn a year. TheActuary, Mag Actuar Prof [Internet]. 2015;87:1–5. Available from: http://www.theactuary.com/news/2015/09/cyber-crime-costs-global-economy-445bn-a-year/

[108] Hawes J. 2013 An Epic Year For Data Breaches With Over 800 Million Records Lost. Naked Secur. 2014;

[109] Farid Daryabar, Ali Dehghantanha, Nur Izura Udzir, Nor Fazlida binti Mohd Sani SBS. A Survey about Impacts of Cloud Computing on Digital Forensics. Int J Cyber-Security Digit Forensics [Internet]. 2013;2(2):77–94. Available from: http://sdiwc.net/security-journal/Browse-Archive.php?ptid=1&ptsid=66&vnum=2&inum=2

[110] Ru DVZM, Smith DL. Melissa (computer virus). 2016. p. 7–8.

[111] Eric C. VBS.LoveLetter.Var. 2014.

[112] Edmundson HP. Theory of self-reproducing automata. Vol. 5, Information Storage and Retrieval. 1969. p. 151.

[113] THOMPSON K. Reflections on trusting trust revisited. Commun ACM. 2003;46(6):112.

[114] Qrwhzruwk RI, Yluxvhv F, Zrupv F, Kruvhv U, Pdozduh V, Uhvhdufk U, et al. Timeline of computer viruses and worms. 1949;

[115] Russell D, Gangemi Sr. GT. Computer Security Basics. Sebastopol, CA, USA: O'Reilly &amp; Associates, Inc.; 1991.

[116] Wikipedia. Denial-of-service attack - From Wikipedia, the free encyclopedia. Wikipedia. 2016;1–5.

[117] Wlph D, Zklfk Y, Zkhwkhu F, Qrw RU, Shuydghg LW, Qdphg ZD V, et al. ANIMAL Source Code. 1974;

[118] John W. The animal. 1985. p. 27–8.

[119] Mrayati M, Alam YM, At-Tayyan MH. Series on Arabic Origins of Cryptology al-Kindi's Treatise on Cryptanalysis. 2003. 206 p.

[120] Systems C. Spl_211 © 2001,. 2001;

[121] Stanley J, Steinhardt B. Bigger Monster, Weaker Chains: The Growth of an American Surveillance Society. Civ Lib vs Natl Secur a Post 9/11 World. 2003;(January):53–79.

[122] Sg ITUD. Teletraffic Engineering Handbook Contact : 2001.

[123] Iversen VB, Others. Teletraffic engineering handbook. Vol. 2, Itu-D Sg. 2002. 16 p.

[124] Course COM. Teletraffic Engineering and Network Planning. 2007.

[125] Wood JA. The Darknet: A Digital Copyright Revolution. Richmond J Law Technol. 2009;16(4):1.

[126] Cavallaro D. Cyberpunk and cyberculture. 2000;282.

[127] Fallis A. The Shockwave Rider. J Chem Inf Model [Internet]. 2013;53(9):1689–99. Available from: https://en.wikipedia.org/wiki/The_Shockwave_Rider

[128] Skrenta R, Warner T, Cloner E, Cloner E. First virus hatched as a practical joke. 2007;6–8.

[129] Cohen F. Computer viruses. Theory and experiments. Comput Secur. 1987;6(1):22–35.

[130] Lq SW, Swrjudsklf FU. Backdoor (computing).

[131] Karger PA, Schell RR. MULTICS security evaluation, volume II: vulnerability analysis. Electron Syst Div Air Force Syst Command. 1974;

[132] Karger PA, Schell RR. Thirty years later: Lessons from the Multics security evaluation. Proc - Annu Comput Secur Appl Conf ACSAC. 2002;2002–Janua:119–26.

[133] Wysopal C, Shields T. Static Detection of Application Backdoors Detecting both malicious software behavior and malicious. Info.

[134] Szor BP. Techniques and Computer Virus Generator Kits. 2005;

[135] Symantec. Symantec Security Update – November 2 0 0 5 Symantec ^{TM} Security Update - November 2005 Worldwide and EMEA Monthly report examining recent high-severity vulnerabilities , and spam activity . 2005.

[136] Leyden J. PC virus celebrates 20th birthday. Regist. 2006;6–8.

[137] Szor P. The Art of Computer Virus Research and Defense. Vol. 43, The Art of Computer Virus Research and Defense. 2005. 744 p.

[138] Ferbrache D. A Pathology of Computer Viruses. 1992. 1-299 p.

[139] Yluxv V, Rxwqxpehuhg X, Yluxv EWKH. Ping-Pong virus. 2014;7–8.

[140] Capek PG, Chess DM, White SR, Fedeli A. Merry Christma: An early network worm. IEEE Secur Priv. 2003;99(5):26–34.

[141] Spafford EH, Heaphy KA, Ferbrache DJ. A Computer Virus Primer. Department of Computer Science. 1989.

[142] Popkin RH. The History of.

[143] Chokoshvili D. The Role of the Internet in Democratic Transition: Case Study of the Arab Spring. North. 2011;

[144] Coffman KG, Odlyzko AM. The size and growth rate of the Internet. First Monday.

1998;3(10):l-25.

[145] Iru MV, Dq L V, Xvhg V, Glvuxsw WR, Rshudwlrqv F, Vhqvlwlyh J, et al. Fdxvh kdup riwhq dv vderwdjh h j 6wx[qhw ru wr h[wruw sd\phqw &u\swr/rfnhu.

[146] Ssoh R, Dqg SS, Ri O. 7kh ³)ruelgghq´ $ssoh $ss 6wruhv dqg wkh ,ooxvlrq ri &rqwuro 3duw ,. 78.

[147] Wkh Q, Vhfwruv E, Iorss RI, Yluxv DIL, Glvn DU, Dv V, et al. 6suhdglqj d iloh lqihfwlqj yluxv. 5:7–10.

[148] Wang W. Steal This Computer Book 4.0: What They Won’t Tell You About the Internet. 2006. 384 p.

[149] Nachenberg C. Understanding and managing polymorphic viruses. Symantec Enterp Pap. 1996;30:16.

[150] Cd CB. Injury Research. 2012;107.

[151] Rosenberger R. Michelangelo Fiasco : a Historical Timeline. 1992;

[152] Ponemon Institute LLC. 2016 Cost of Data Breach Study : Global Analysis. 2016 Cost Data Breach Study Glob Anal [Internet]. 2016;(June):31. Available from: https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=SEL03094WWEN

[153] Walker J. ANIMAL Source Code. 2016;1–29.

[154] Chen T, Robert J-M. Evolution of Viruses and Worms. Stat Methods Comput Secur [Internet]. 2004;1–16. Available from: http://vx.netlux.org/lib/atc01.html

[155] Nathoo K. A case study solaris sadmind exploitation. 2005.

[156] Voyiatzis AG, Serpanos DN. Pulse: A class of super-worms against network infrastructure. Proc - 23rd Int Conf Distrib Comput Syst Work ICDCSW 2003. 2003;(March):28–33.

[157] Chen T. Trends in viruses and worms. Internet Protoc J. 2003;23–33.

[158] Office. USGA. Information Security: Code Red, Code Red II, and SirCam Attacks

Highlight Need for Proactive Measures . United States. General Accounting Office.; 2001.

[159] Berghel H. The Code Red Worm: Malicious software knows no bounds. Commun ACM. 2001;44(12):15–9.

[160] Bilar D. Opcodes as predictor for malware. Int J Electron Secur Digit Forensics [Internet]. 2007;1(2):156. Available from: http://www.inderscience.com/link.php?id=16865

[161] Rusch JJ. The Social Psychology of Computer Viruses and Worms *. America (NY). 2002;

[162] Okon BE, Umunnah RA. Identification-Of-Threats-To-Communication-Security-Effects-On-Victims-And-Ways-Of-Mitigating-The-Threats.docx. 2015;6(5).

[163] Baecher P, Koetter M, Holz T, Dornseif M, Freling F. The nepenthes platform: An efficient approach to collect malware. Recent Adv intrusion Detect. 2006;165–84.

[164] Moore D, Paxson V, Savage S, Shannon C, Staniford S, Weaver N. Inside the slammer worm. IEEE Secur Priv. 2003;1(4):33–9.

[165] Weapox CY, Guard ONY, Divided ANE. Virus. Computer (Long Beach Calif). 2005;(July).

[166] Berghel H. Malware Month. 2003;46(12):15–9.

[167] Mancoridis S. Software Analysis for Security. 2008;109–18.

[168] Sweigart C. The Welchia Worm. 2003.

[169] Bailey M, Cooke E, Jahanian F, Watson D, Nazario J. The Blaster Worm: Then and Now. IEEE Secur Priv Mag. 2005 Jul;3(4):26–31.

[170] Schultz EE. The Sobig Worm Variants.

[171] Rasheed MF. Modelling Virus Propagation in P2P Networks. Int J Comput Sci Issues. 2012;9(2):580–7.

[172] Sober worm returns as largest outbreak of 2005. Netw Secur. 2005;2005(12):2.

[173] Peltomäki M, Ovaska M, Alava M. Worm spreading with immunization : An interplay of spreading and immunity time scales. Physica A. Elsevier B.V.; 2011;390(23–24):4152–9.

[174] Netsky &amp; Bagle dominate virus top 10 in March. Vol. 2004, Computer Fraud & Security. 2004.

[175] Caswell B, Orebaugh A. Phishing Exposed [Internet]. Database. 2005. 750 p. Available from: http://books.google.com/books?id=UyktqN6GnWEC&pgis=1

[176] Detection M. Malware Detection. Vol. 53, Journal of Chemical Information and Modeling. 1989. 160 p.

[177] Stewin P, Bystrov I. Detection of Intrusions and Malware, and Vulnerability Assessment. Vol. 7591, Dimva. 2016. 21-41 p.

[178] Peng S, Yu S, Yang A. Smartphone malware and its propagation modeling: A survey. IEEE Commun Surv Tutorials. 2014;16(2):925–41.

[179] Sood AK, Enbody R. Chapter 6 - Maintaining Control and Lateral Movement. Target Cyber Attacks [Internet]. 2014;95–111. Available from: http://www.sciencedirect.com/science/article/pii/B9780128006047000061

[180] Islam R, Tian R, Batten LM, Versteeg S. Classification of malware based on integrated static and dynamic features. J Netw Comput Appl [Internet]. Elsevier; 2013;36(2):646–56. Available from: http://dx.doi.org/10.1016/j.jnca.2012.10.004

[181] Bash E. Computer Viruses and Malware. Vol. 1, Advances in Information Security. Springer; 2006. 25-29 p.

[182] Li C, Jiang W, Zou X. Botnet : Survey and Case Study. 2009;1184–7.

[183] Slides P. Malware : Malicious Software Viruses , Worms , Trojans , Rootkits Backdoors. 2010;

[184] Symantec. Internet Security Threat Report. Vol. 19, Symantec 2013 Trends. 2014.

[185] Security HPE. 2014 Global Report on the Cost of Cyber Crime Sponsored by HP Enterprise Security. Ponemon Institue Res Rep. 2014;(October).

[186] (PonemonInstitute). 2015 Cost of Cyber Crime Study: Global. Hewlett Packard Enterp. 2015;(October).

[187] Ponemon Institute. 2015 Cost of Data BreachSstudy: Global Analysis. 2015;(May):1–30.

[188] Advanced Malware Detection - Signatures vs. Behavior Analysis - Infosecurity Magazine.

[189] Analysis of Signature-Based and Behavior-Based Anti-Malware Approaches. Int J Adv Res Comput Eng Technol. 2013;2(6):2278–1323.

[190] Shabtai A, Kanonov U, Elovici Y, Glezer C, Weiss Y. " Andromaly ": a behavioral malware detection framework for android devices. 2012;161–90.

[191] Whang JJ, Dhillon I, Gleich D. Non-exhaustive , Overlapping k -means. Proc SIAM Int Conf Data Min. 2015;936–44.

[192] Objectives L. Cluster Analysis. 2011.

[193] Lauwers O, De Moor B. A time series distance measure for efficient clustering of input output signals by their underlying dynamics. 2017 Mar;

[194] SAS Enterprise Miner ^{TM} High-Performance Data Mining Node Reference for.

[195] Antonenko PD, Toy S, Niederhauser DS. Using cluster analysis for data mining in educational technology research. Educ Technol Res Dev. 2012;60(3):383–98.

[196] Mandalapu S, Wang Y, Ni XS. Building Neural Network Model in Base SAS ® (From Scratch) 2 . FEED FORWARD AND BACKWARD PROPAGATION NEURAL NETWORK. 2018;(Figure 1):1–20.

[197] Dostál P, Pokorný P. Cluster analysis and neural network. Tech Comput Prague

2009, 17th Annu Conf Proc [Internet]. 2008; Available from: http://dsp.vscht.cz/konference_matlab/MATLAB08/prispevky/025_dostal.pdf

[198] Karim A, Loqman C, Boumhidi J. Determining the number of clusters using neural network and max stable set problem. Procedia Comput Sci [Internet]. Elsevier B.V.; 2018;127:16–25. Available from: https://doi.org/10.1016/j.procs.2018.01.093

[199] Pang H, Zhao H. Building pathway clusters from Random Forests classification using class votes. 2008;12(1):1–12.

[200] Pang H, Lin A, Holford M, Enerson BE, Lu B, Lawton MP, et al. Pathway analysis using random forests classification and regression. 2006;22(16):2028–36.

[201] Muñoz-González L, Biggio B, Demontis A, Paudice A, Wongrassamee V, Lupu EC, et al. Towards Poisoning of Deep Learning Algorithms with Back-gradient Optimization. 2017; Available from: http://arxiv.org/abs/1708.08689

[202] Cabaj K, Gregorczyk M, Mazurczyk W. Software-defined networking-based crypto ransomware detection using HTTP traffic characteristics. Comput Electr Eng. 2018;66:353–68.

[203] Cusack G, Michel O, Keller E. Machine Learning-Based Detection of Ransomware Using SDN. 2018;1–6.

[204] Zanero S, Continella A, Zingaro G, Guagnelli A, Maggi F, De Pasquale G, et al. ShieldFS: A Self-healing, Ransomware-aware Filesystem. 2016;336–47.

[205] Nath H V, Mehtre BM. Static Malware Analysis. 2014;440–50.

[206] Shaukat SK, Ribeiro VJ. RansomWall : A Layered Defense System against Cryptographic Ransomware Attacks using Machine Learning. 2018;

[207] Faruki P, Laxmi V, Gaur MS, Vinod P. Mining control flow graph as API call-grams to detect portable executable malware . 2012;(September 2015):130–7.

[208] Park Y, Reeves DS, Stamp M. Deriving common malware behavior through graph

clustering. Comput Secur [Internet]. Elsevier Ltd; 2013;39:419–30. Available from: http://dx.doi.org/10.1016/j.cose.2013.09.006

[209] Park YH, Reeves DS. Deriving common malware behavior through graph clustering (Short Paper). 2011;497–502.

[210] Vatamanu C, Cosovan D, Gavrilu D, Luchian H. A Comparative Study of Malware Detection Techniques Using Machine Learning Methods. 2015;9(5):1115–22.

[211] Khan MD, Shaikh MT, Ansari R, Suriya M, Suryawanshi S. Malware detection using Machine Learning Algorithms. 2017;6(9):195–9.

[212] Gandotra E, Bansal D, Sofat S. Malware Analysis and Classification. 2014;(April):56–64.

[213] Malware H, Perdisci R, Ariu D, Giacinto G. Scalable Fine-Grained Behavioral Clustering of HTTP-Based Malware.

[214] Bayer U, Milani-Comparetti P, Hlauscheck C, Kruegel C, Kirda E. Scalable, Behavior-Based Malware Clustering. 16th Symp Netw Distrib Syst Secur. 2009;120–9.

[215] Ulrich B, Imam H, Davide B, Engin K, Christopher K. A View on Current Malware Behaviors. AVAR Conf [Internet]. 2011; Available from: http://static1.esetstatic.com/us/resources/white-papers/AVAR-EICAR-2010.pdf

[216] Zolkipli MF, Jantan A. An approach for malware behavior identification and classification. ICCRD2011 - 2011 3rd Int Conf Comput Res Dev. 2011;1(December 2015):191–4.

[217] Rieck K, Trinius P, Willems C, Holz_aff2n3 T. Automatic Analysis of Malware Behavior Using Machine Learning. J Comput Secur [Internet]. 2011;19(4):639–68. Available from: http://dl.acm.org/citation.cfm?id=2011216.2011217

[218] Ye Y, Li T, Chen Y, Jiang Q. Automatic malware categorization using cluster

ensemble. 2010;95.

[219] Guofei Gu1, Roberto Perdisci, Junjie Zhang and WL. BotMiner: Clustering Analysis of Network Traffic for Protocol- and Structure-Independent Botnet Dete.pdf. USENIX Secur Symp. 2008;139–54.

[220] Schultz MG, Eskin E, Zadok F, Stolfo SJ. Data mining methods for detection of new malicious executables. 2002;38–49.

[221] Avenue M, Hill M, Cohen WW, Of C, Pruning R. Fast Effective Rule Induction. Learning. 1994;

[222] Marpaung JAP, Sain M, Lee H. Survey on malware evasion techniques : state of the art and challenges. 2012;(Mic):744–9.

[223] Afianian A, Niksefat S, Sadeghiyan B, Baptiste D. Malware Dynamic Analysis Evasion Techniques: A Survey. 2018;1–33. Available from: http://arxiv.org/abs/1811.01190

[224] Types C, Attacks C, Techniques H. Common Types of Cybersecurity Attacks A look at the various types of hacking techniques At a Glance : 2019;1–7.

[225] Padmavathi G, Divya S. A Survey on Various Security Threats and Classification of Malware Attacks , Vulnerabilities and Detection Techniques. 2013;(June).

[226] Khan I. An introduction to computer viruses: Problems and solutions. Libr Hi Tech News. 2012;29(7):8–12.

[227] Seberry JHJ. Computer Viruses An Introduction. 2000;19(1):122–31.

[228] Smith C, Matrawy A. Computer worms: Architectures, evasion strategies, and detection mechanisms. J Inf … [Internet]. 2008;4:69–83. Available from: http://www.softcomputing.net/jias/smith.pdf

[229] Moffie M, Cheng W, Kaeli D, Zhao Q. Hunting Trojan Horses. 2007;12–7.

[230] Gupta S, Das S, Pal PK. Predictability in Viral Computing. Procedia Technol

[Internet]. Elsevier B.V.; 2012;4:530–5. Available from: http://www.sciencedirect.com/science/article/pii/S2212017312003635

[231] Deng W, Liu Q, Cheng H, Qin Z. A Malware Detection Framework Based on Kolmogorov Complexity. J Comput Inf Syst [Internet]. 2011;7(8):2687–94. Available from: http://www.jofcis.com

[232] Gavriluţ D, Cimpoeşu M, Anton D, Ciortuz L. Malware detection using machine learning. Proc Int Multiconference Comput Sci Inf Technol IMCSIT '09. 2009;4(May 2014):735–41.

[233] Chandrasekaran M, Sankaranarayanan V, J. Upadhyaya S. SpyCon: Emulating User Activities to Detect Evasive Spyware. In 2007. p. 502–9.

[234] Chuvakin A. An overview of unix rootkits. Megasecurity Labs [Internet]. 2003;1–27. Available from: http://www.thehackademy.net/madchat/vxdevl/avtech/An Overview of Unix Rootkits.pdf

[235] Keylogger G, Ladakis E, Koromilas L, Vasiliadis G, Polychronakis M, Ioannidis S. You Can Type , but You Can ' t Hide : A Stealthy. 2013;

[236] Risks CS, Threats C. Common threats to be aware of. 2019;1–10.

[237] Sultan H, Khalique A, Alam SI, Tanweer S. a Survey on Ransomeware: Evolution, Growth, and Impact. Int J Adv Res Comput Sci [Internet]. 2018;9(2):802–10. Available from: http://www.ijarcs.info/index.php/Ijarcs/article/view/5858%0Ahttp://dx.doi.org/10.26483/ijarcs.v9i2.5858

[238] Registered, England, Wales. Ransomware White Paper. VAT Reg Number [Internet]. 2016;880(5589479):8618–88. Available from: https://swgfl.org.uk/Uploads/4f/4fccce1e-134b-4872-b965-7c7a23dc3244.pdf

[239] Savage K, Coogan P, Lau H. Information Resources. Res Manag [Internet].

2011;54(5):59–63. Available from: http://openurl.ingenta.com/content/xref?genre=article&issn=0895-6308&volume=54&issue=5&spage=59

[240] 3/29/2019 Ransomware attacks: Radical menace for cloud computing - IEEE Conference Publication. 2019;8300039.

[241] 3/29/2019 Ransomware at X-Rays - IEEE Conference Publication. 2019;8276776.

[242] Chen Q, Bridges RA. Automated behavioral analysis of malware: A case study of wannacry ransomware. Proc - 16th IEEE Int Conf Mach Learn Appl ICMLA 2017. 2018;2018–Janua(July):454–60.

[243] 3/29/2019 Detection and prevention of crypto-ransomware - IEEE Conference Publication. 2019;8249052.

[244] Morgan J. Five Things You Need To Know About Cryptolocker. Forbes [Internet]. 2015; Available from: http://www.forbes.com/sites/jacobmorgan/2015/05/04/5-things-you-need-to-know-about-telecommuting/#70fa5aa212a0

[245] Wyke BJ, Researcher ST, Emerging S, Team T. The current state of ransomware. 2015;(December):1–3. Available from: papers2://publication/uuid/016593B9-DCF8-468B-B3A7-CE9C9CC71F85

[246] Ramesh V, Glass RL, Vessey I. Research in computer science: an empirical study.

[247] Wainer J, Novoa Barsottini CG, Lacerda D, Magalhães de Marco LR. Empirical evaluation in Computer Science research published by ACM. Inf Softw Technol. Elsevier; 2009 Jun;51(6):1081–5.

[248] Tichy WF, Lukowicz P, Prechelt L, Heinz EA. Experimental Evaluation in Computer Science: A Quantitative Study.

[249] Valizadegan H, Jin R. Generalized Maximum Margin Clustering and Unsupervised Kernel Learning.

[250] Dunn† JC. Well-Separated Clusters and Optimal Fuzzy Partitions. J Cybern. Taylor & Francis; 1974 Jan;4(1):95–104.

[251] Virmani D, Jain S. Clustering Based Topology Control Protocol for Data Delivery in Wireless Sensor Networks.

[252] Knight M, Nunes M, Nason G. Modelling, Detrending and Decorrelation of Network Time Series. 2016;

[253] Sharp A, McDermott P. Workflow modeling : tools for process improvement and applications development. Artech House; 2009. 449 p.

[254] Bohte SM, La Poutre H, Kok JN. Unsupervised clustering with spiking neurons by sparse temporal coding and multilayer RBF networks. IEEE Trans Neural Networks. 2002 Mar;13(2):426–35.

[255] Ferreira LN, Zhao L. Time Series Clustering via Community Detection in Networks. 2015 Aug;

[256] Rosvall M, Bergstrom CT. Mapping change in large networks. 2010;

[257] Waltman L, Van Eck NJ, Noyons ECM. A unified approach to mapping and clustering of bibliometric networks.

[258] Buteikis A. 03 Time series with trend and seasonality components.

[259] SAS/ETS ® 13.2 User's Guide The TIMESERIES Procedure.

[260] OECD Glossary of Statistical Terms - Trend-cycle Definition.

[261] Statistical Analysis Handbook 2018 edition - Dr M J de Smith.

[262] The Use of Butterworth Filters for Trend and Cycle Estimation in Economic Time Series.

[263] Hindrayanto I, Jacobs JPAM, Osborn DR. On Trend-Cycle-Seasonal Interactions. SSRN Electron J. 2014;(417).

[264] Approach to an irregular time series on the basis of the fractal theory.

[265] Sevtap Kestel A. Time Series Analysis Classical Time Series. 2013;

[266] Zhang Y. Time Series Analysis 时间序列分析.

[267] Seasonally-Adjusted Data: What it Really Means | Bureau of Transportation Statistics.

[268] Hood CC. Comparison of Time Series Characteristics for Seasonal Adjustments from SEATS and X-12-ARIM A.

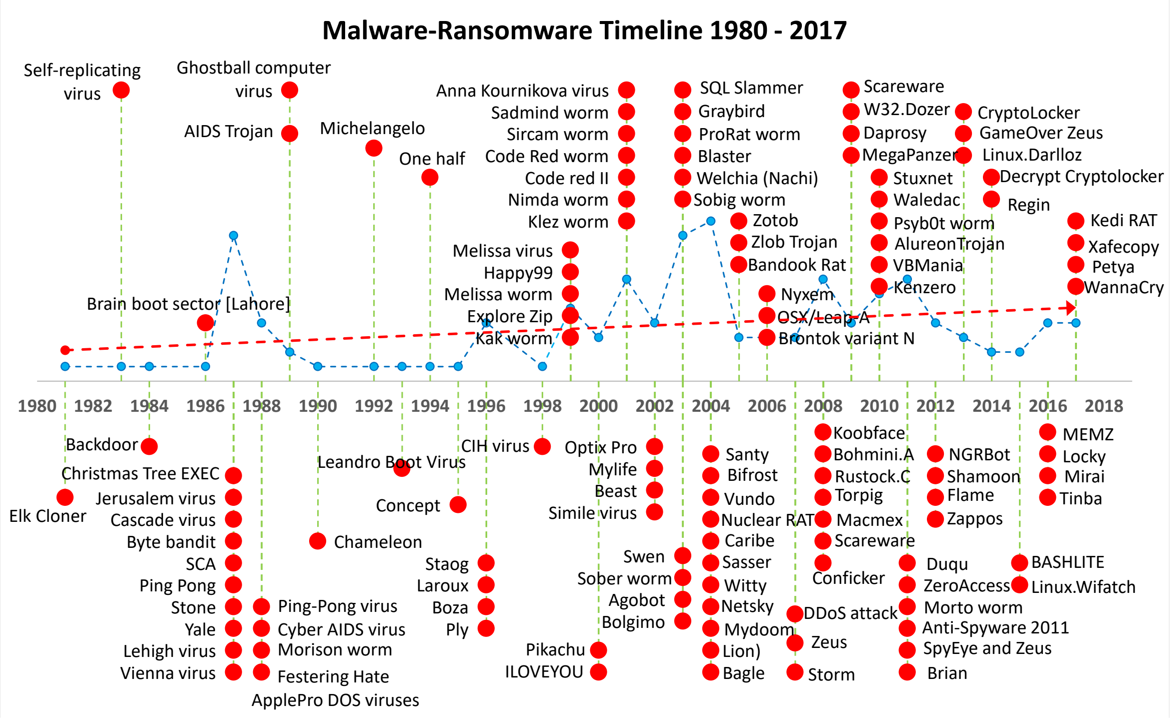
[269] Beveridge S, Nelson CR. A New Approach to Decomposition of Economic Time Series Into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle. J Monet Econ. North-Holland Pubhshing Company; 1981;7:151–74.

[270] De Livera AM, Hyndman RJ, Snyder RD. Forecasting time series with complex seasonal patterns using exponential smoothing. 2010;

[271] SAS/ETS ® 13.2 User’s Guide The SIMILARITY Procedure.

APPENDICES

Appendix 1: The Malware-Ransomware Timeline for the Period 1980-2017 showing names of malware against date of release, the dotted blue line shows the frequency for each year, while the dotted red line is the trend showing increased malware release over the years



Appendix 3: Time Series Events Activities Jan2015 - May2018

Period	Time Series Events Activities Jan2015 - May2018							
	Malware_N	ASN_N	Country_N	Host_N	IPAddress_N	Registrar_N	Status_N	URL_N
Jan2015	0	0	0	0	0	0	0	0
May2015	14	2214	306	35921	13488	2919	8	49467
Sep2015	15	3647	504	14026	18915	1066	6	18433
Jan2016	103	20598	3270	191934	82907	15368	51	258659
May2016	11475	574265	62471	6052799	2155983	472868	1566	8487426
Sep2016	22185	1538818	163316	20357444	6173588	1228925	3987	27907747
Jan2017	48136	3033579	368912	36978444	11228061	2525873	7212	50683484
May2017	834	24325	2697	1735947	140946	79562	334	2391980
Sep2017	3260	189397	24273	2268704	1080715	156407	543	3079716
Jan2018	2849	177578	23973	2132827	970512	144786	432	2918581
May2018	14	558	102	11766	2355	445	2	16146
Total	**88885**	**5564979**	**649824**	**69779812**	**21867470**	**4628219**	**14141**	**95811639**

Appendix 4: Percentages of Time Series Events Activities Jan2015 – May2018

Period	Percentages of Time Series Events Activities Jan2015 – May2018							
	Malware_N	ASN_N	Country_N	Host_N	IPAddress_N	Registrar_N	Status_N	URL_N
Jan2015	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
May2015	0.02	0.04	0.05	0.05	0.06	0.06	0.06	0.05
Sep2015	0.02	0.07	0.08	0.02	0.09	0.02	0.04	0.02
Jan2016	0.12	0.37	0.50	0.28	0.38	0.33	0.36	0.27
May2016	12.91	10.32	9.61	8.67	9.86	10.22	11.07	8.86
Sep2016	24.96	27.65	25.13	29.17	28.23	26.55	28.19	29.13
Jan2017	54.16	54.51	56.77	52.99	51.35	54.58	51.00	52.90
May2017	0.94	0.44	0.42	2.49	0.64	1.72	2.36	2.50
Sep2017	3.67	3.40	3.74	3.25	4.94	3.38	3.84	3.21
Jan2018	3.21	3.19	3.69	3.06	4.44	3.13	3.05	3.05
May2018	0.02	0.01	0.02	0.02	0.01	0.01	0.01	0.02

Appendix 5: Key Performance Parameters - General Network Item Constellation Plot - Node Data 2015Jun

General Network Item Constellation Plot - Node Data 2015Jun (Values and Percentages)

Parameters	Clust 1	Clust 2	Clust 3	Clust 4	Clust 5	Total	Percentage (Ratio*100)				
							Clust 1	Clust 2	Clust 3	Clust 4	Clust 5
Weight	3.10	1.00	1.83	1.00	2.00	**8.93**	34.70	11.19	20.52	11.19	22.39
Out degree centrality	13.80	7.00	10.33	7.00	12.29	**50.42**	27.37	13.88	20.49	13.88	24.37
Weighted eigenvector centrality	0.52	0.15	0.30	0.10	0.30	**1.36**	37.83	11.32	21.74	7.37	21.73
Unweighted eigenvector centrality	0.61	0.33	0.45	0.25	0.54	**2.18**	27.79	15.13	20.57	11.53	24.98
Weighted closeness centrality	0.97	0.65	0.84	0.66	0.84	**3.95**	24.52	16.36	21.27	16.60	21.24
Unweighted closeness centrality	0.67	0.57	0.61	0.57	0.63	**3.04**	21.97	18.62	19.95	18.62	20.83
Weighted betweenness centrality	0.07	0.00	0.00	0.00	0.01	**0.09**	84.96	0.00	5.35	0.00	9.69
Unweighted betweenness centrality	0.04	0.00	0.01	0.00	0.02	**0.08**	54.91	0.00	16.34	0.00	28.75
Weighted influence1 centrality	0.34	0.11	0.20	0.11	0.22	**0.98**	34.70	11.19	20.52	11.19	22.39
Weighted influence2 centrality	4.13	2.52	3.32	1.86	3.67	**15.50**	26.68	16.23	21.41	12.00	23.69
Unweighted influence1 centrality	0.22	0.11	0.16	0.11	0.19	**0.79**	27.37	13.88	20.49	13.88	24.37
Unweighted influence2 centrality	2.88	1.69	2.24	1.41	2.62	**10.83**	26.59	15.58	20.68	12.98	24.17
Clustering coefficient centrality	0.71	1.00	0.79	1.00	0.73	**4.23**	16.89	23.64	18.60	23.64	17.24

Appendix 6: Key Performance Parameters - General Network Item Constellation Plot - Node Data 2015Dec

General Network Item Constellation Plot - Node Data 2015Dec (Values and Percentages)

Parameters	Clust 1	Clust 2	Clust 3	Clust 4	Total	Percentages (Ratio*100)			
						Clust 1	Clust 2	Clust 3	Clust 4
Weight	7.10	3.13	15.22	6.71	**32.16**	22.08	9.72	47.33	20.88
Out degree centrality	15.00	13.50	21.67	17.57	**67.74**	22.14	19.93	31.99	25.94
Weighted eigenvector centrality	0.27	0.10	0.51	0.22	**1.10**	24.35	8.97	46.54	20.15
Unweighted eigenvector centrality	0.54	0.49	0.74	0.65	**2.42**	22.38	20.42	30.58	26.62
Weighted closeness centrality	1.83	1.40	2.49	2.08	**7.80**	23.51	17.92	31.90	26.67
Unweighted closeness centrality	0.66	0.63	0.76	0.68	**2.74**	23.94	23.19	27.90	24.98
Weighted betweenness centrality	0.00	0.00	0.09	0.00	**0.09**	0.73	0.17	99.04	0.06
Unweighted betweenness centrality	0.01	0.01	0.03	0.01	**0.06**	16.84	12.42	54.02	16.72
Weighted influence1 centrality	0.18	0.08	0.38	0.17	**0.80**	22.08	9.72	47.33	20.88
Weighted influence2 centrality	4.38	3.83	5.24	4.76	**18.22**	24.05	21.04	28.76	26.14
Unweighted influence1 centrality	0.05	0.05	0.08	0.06	**0.24**	22.14	19.93	31.99	25.94
Unweighted influence2 centrality	1.06	0.96	1.42	1.25	**4.69**	22.56	20.55	30.31	26.57
Clustering coefficient centrality	0.77	0.80	0.63	0.70	**2.89**	26.48	27.58	21.82	24.12

Appendix 7: Key Performance Parameters - General Network Item Constellation Plot - Node Data 2016Jun

General Network Item Constellation Plot - Node Data Cum 2016Jun (Values and Percentages)

Parameters	Clust 1	Clust 2	Clust 3	Total	Percentage (Ratio*100)		
					Clust 1	Clust 2	Clust 3
Weight	440.45	359.33	304.64	**1104.42**	39.88	32.54	27.58
Out degree centrality	12.55	10.50	11.27	**34.32**	36.56	30.60	32.85
Weighted eigenvector centrality	0.40	0.23	0.18	**0.81**	49.36	28.24	22.39
Unweighted eigenvector centrality	0.54	0.41	0.46	**1.42**	38.01	29.16	32.82
Weighted closeness centrality	144.84	122.52	116.22	**383.58**	37.76	31.94	30.30
Unweighted closeness centrality	0.62	0.62	0.60	**1.83**	33.83	33.56	32.61
Weighted betweenness centrality	0.00	0.08	0.00	**0.08**	0.00	94.40	5.60
Unweighted betweenness centrality	0.01	0.04	0.01	**0.06**	20.15	61.77	18.08
Weighted influence1 centrality	0.21	0.15	0.12	**0.48**	42.57	31.77	25.66
Weighted influence2 centrality	3.39	2.55	2.74	**8.68**	39.06	29.38	31.56
Unweighted influence1 centrality	0.00	0.00	0.00	**0.00**	36.56	30.60	32.85
Unweighted influence2 centrality	0.02	0.01	0.01	**0.04**	37.53	29.18	33.29
Clustering coefficient centrality	0.74	0.79	0.78	**2.31**	32.22	34.08	33.70

Appendix 8: Key Performance Parameters - General Network Item Constellation Plot - Node Data 2016Dec

General Network Item Constellation Plot - Node Data 2016Dec (Values and Percentages)

Parameters	Clust 1	Clust 2	Clust 3	Clust 4	Clust 5	Total	Percentage (Ratio*100)				
							Clust 1	Clust 2	Clust 3	Clust 4	Clust 5
Weight	3360.70	2574.50	2428.50	2526.50	2984.18	**13874.38**	24.22	18.56	17.50	18.21	21.51
Out degree centrality	12.50	8.00	8.00	8.50	8.00	**45.00**	27.78	17.78	17.78	18.89	17.78
Weighted eigenvector centrality	0.34	0.25	0.23	0.23	0.23	**1.29**	26.59	19.19	18.22	18.24	17.76
Unweighted eigenvector centrality	0.59	0.47	0.47	0.48	0.34	**2.35**	25.12	19.88	19.88	20.54	14.58
Weighted closeness centrality	1224.71	1163.85	1137.13	1144.35	1064.65	**5734.70**	21.36	20.29	19.83	19.95	18.57
Unweighted closeness centrality	0.63	0.57	0.57	0.58	0.60	**2.95**	21.40	19.40	19.40	19.57	20.23
Weighted betweenness centrality	0.00	0.00	0.00	0.00	0.05	**0.05**	0.00	0.00	0.00	0.00	100.00
Unweighted betweenness centrality	0.01	0.00	0.00	0.00	0.04	**0.05**	27.28	0.65	0.65	1.65	69.76
Weighted influence1 centrality	0.20	0.12	0.12	0.12	0.15	**0.70**	27.98	17.41	16.59	17.04	20.98
Weighted influence2 centrality	3.52	3.27	3.27	3.29	2.37	**15.72**	22.39	20.81	20.77	20.93	15.10
Unweighted influence1 centrality	0.00	0.00	0.00	0.00	0.00	**0.00**	27.78	17.78	17.78	18.89	17.78
Unweighted influence2 centrality	0.00	0.00	0.00	0.00	0.00	**0.01**	23.34	20.59	20.59	20.96	14.51
Clustering coefficient centrality	0.77	0.93	0.93	0.91	0.83	**4.36**	17.61	21.28	21.28	20.92	18.92

Appendix 9: Key Performance Parameters - General Network Item Constellation Plot - Node Data 2017Jun

General Network Item Constellation Plot - Node Data 2017Jun (Values and Percentages)

Parameters	Clust 1	Clust 2	Clust 3	Clust 4	Total	Percentage (Ratio*100)			
						Clust 1	Clust 2	Clust 3	Clust 4
Weight	3490.73	2978.28	2706.00	2601.00	**11776.01**	29.64	25.29	22.98	22.09
Out degree centrality	12.36	7.67	8.00	8.00	**36.03**	34.31	21.28	22.20	22.20
Weighted eigenvector centrality	0.34	0.22	0.25	0.25	**1.06**	32.01	21.00	23.84	23.15
Unweighted eigenvector centrality	0.59	0.34	0.47	0.47	**1.87**	31.35	18.42	25.12	25.12
Weighted closeness centrality	1264.08	1085.62	1205.06	1187.57	**4742.33**	26.66	22.89	25.41	25.04
Unweighted closeness centrality	0.63	0.59	0.57	0.57	**2.36**	26.59	25.03	24.19	24.19
Weighted betweenness centrality	0.00	0.05	0.00	0.00	**0.05**	0.00	100.00	0.00	0.00
Unweighted betweenness centrality	0.01	0.03	0.00	0.00	**0.05**	27.95	70.64	0.70	0.70
Weighted influence1 centrality	0.19	0.14	0.13	0.12	**0.59**	33.21	24.29	21.52	20.98
Weighted influence2 centrality	3.52	2.40	3.30	3.30	**12.53**	28.12	19.18	26.36	26.34
Unweighted influence1 centrality	0.00	0.00	0.00	0.00	**0.00**	34.31	21.28	22.20	22.20
Unweighted influence2 centrality	0.00	0.00	0.00	0.00	**0.01**	29.20	18.57	26.11	26.11
Clustering coefficient centrality	0.76	0.87	0.93	0.93	**3.49**	21.91	24.84	26.62	26.62

Appendix 10: Key Performance Parameters - General Network Item Constellation Plot - Node Data 2017Dec

General Network Item Constellation Plot - Node Data 2017Dec (Values and Percentages)

Parameters	Clust 1	Clust 2	Clust 3	Clust 4	Clust 5	Total	Percentage (Ratio*100)				
							Clust 1	Clust 2	Clust 3	Clust 4	Clust 5
Weight	3774.80	2842.50	2775.00	2679.50	3256.88	**15328.68**	24.63	18.54	18.10	17.48	21.25
Out degree centrality	12.50	8.00	8.00	8.00	7.71	**44.21**	28.28	18.10	18.10	18.10	17.43
Weighted eigenvector centrality	0.35	0.25	0.25	0.22	0.22	**1.29**	27.13	19.08	19.11	17.30	17.38
Unweighted eigenvector centrality	0.59	0.47	0.47	0.46	0.34	**2.34**	25.37	20.17	20.17	19.88	14.41
Weighted closeness centrality	1363.20	1285.83	1273.39	1241.29	1169.74	**6333.45**	21.52	20.30	20.11	19.60	18.47
Unweighted closeness centrality	0.63	0.57	0.57	0.57	0.59	**2.94**	21.46	19.45	19.45	19.46	20.18
Weighted betweenness centrality	0.00	0.00	0.00	0.00	0.05	**0.05**	0.00	0.00	0.00	0.00	100.00
Unweighted betweenness centrality	0.01	0.00	0.00	0.00	0.04	**0.05**	27.33	0.65	0.65	1.77	69.61
Weighted influence1 centrality	0.20	0.12	0.12	0.11	0.14	**0.70**	28.57	17.40	17.51	16.01	20.51
Weighted influence2 centrality	3.53	3.29	3.29	3.23	2.34	**15.67**	22.53	20.97	20.97	20.61	14.92
Unweighted influence1 centrality	0.00	0.00	0.00	0.00	0.00	**0.00**	28.28	18.10	18.10	18.10	17.43
Unweighted influence2 centrality	0.00	0.00	0.00	0.00	0.00	**0.01**	23.49	20.88	20.88	20.38	14.36
Clustering coefficient centrality	0.77	0.93	0.93	0.92	0.85	**4.39**	17.45	21.13	21.13	20.86	19.43

Appendix 11: Key Performance Parameters - General Network Item Constellation Plot - Node Data 2018Jun

General Network Item Constellation Plot - Node Data 2018Jun (Values and Percentages)

Parameters	Clust 1	Clust 2	Clust 3	Clust 4	Clust 5	Total	Percentage (Ratio*100)				
							Clust 1	Clust 2	Clust 3	Clust 4	Clust 5
Weight	3778.30	2843.50	2781.50	2682.50	3261.12	**15346.92**	24.62	18.53	18.12	17.48	21.25
Out degree centrality	12.50	8.00	8.00	8.00	7.71	**44.21**	28.28	18.10	18.10	18.10	17.43
Weighted eigenvector centrality	0.35	0.25	0.25	0.22	0.22	**1.29**	27.11	19.06	19.13	17.30	17.39
Unweighted eigenvector centrality	0.59	0.47	0.47	0.46	0.34	**2.34**	25.37	20.17	20.17	19.88	14.41
Weighted closeness centrality	1364.85	1286.86	1275.52	1242.71	1171.08	**6341.01**	21.52	20.29	20.12	19.60	18.47
Unweighted closeness centrality	0.63	0.57	0.57	0.57	0.59	**2.94**	21.46	19.45	19.45	19.46	20.18
Weighted betweenness centrality	0.00	0.00	0.00	0.00	0.05	**0.05**	0.00	0.00	0.00	0.00	100.00
Unweighted betweenness centrality	0.01	0.00	0.00	0.00	0.04	**0.05**	27.33	0.65	0.65	1.77	69.61
Weighted influence1 centrality	0.20	0.12	0.12	0.11	0.14	**0.70**	28.56	17.39	17.52	16.01	20.52
Weighted influence2 centrality	3.53	3.28	3.29	3.23	2.34	**15.67**	22.53	20.97	20.97	20.61	14.92
Unweighted influence1 centrality	0.00	0.00	0.00	0.00	0.00	**0.00**	28.28	18.10	18.10	18.10	17.43
Unweighted influence2 centrality	0.00	0.00	0.00	0.00	0.00	**0.01**	23.49	20.88	20.88	20.38	14.36
Clustering coefficient centrality	0.77	0.93	0.93	0.92	0.85	**4.39**	17.45	21.13	21.13	20.86	19.43

Appendix 12: Key Performance Parameters General Network Item Constellation Plot - Node Data 2015Jun

General Network Item Constellation Plot - Node Data 2015Jun					
	Clusters				
Parameters (Av. Values)	**Clust 1**	**Clust 2**	**Clust 3**	**Clust 4**	**Clust 5**
Weight	3.1000	1.0000	1.8333	1.0000	2.0000
Out degree centrality	13.8000	7.0000	10.3333	7.0000	12.2857
Weighted eigenvector centrality	0.5159	0.1544	0.2965	0.1005	0.2964
Unweighted eigenvector centrality	0.6056	0.3297	0.4483	0.2512	0.5443
Weighted closeness centrality	0.9686	0.6463	0.8400	0.6557	0.8389
Unweighted closeness centrality	0.6676	0.5660	0.6065	0.5660	0.6332
Weighted betweenness centrality	0.0730	0.0000	0.0046	0.0000	0.0083
Unweighted betweenness centrality	0.0437	0.0000	0.0130	0.0000	0.0229
Weighted influence1 centrality	0.3391	0.1094	0.2005	0.1094	0.2188
Weighted influence2 centrality	4.1344	2.5156	3.3177	1.8594	3.6719
Unweighted influence1 centrality	0.2156	0.1094	0.1615	0.1094	0.1920
Unweighted influence2 centrality	2.8797	1.6875	2.2396	1.4063	2.6183
Clustering coefficient centrality	0.7144	1.0000	0.7869	1.0000	0.7293

Appendix 13: Key Performance Parameters General Network Item Constellation Plot - Node Data 2015Dec

General Network Item Constellation Plot - Node Data 2015Dec					
	Clusters				
Parameters	**Clust 1**	**Clust 2**	**Clust 3**	**Clust 4**	**Clust 5**
Weight	7.1000	3.1250	15.2222	6.7143	
Out degree centrality	15.0000	13.5000	21.6667	17.5714	
Weighted eigenvector centrality	0.2679	0.0986	0.5121	0.2217	
Unweighted eigenvector centrality	0.5424	0.4948	0.7409	0.6451	
Weighted closeness centrality	1.8334	1.3978	2.4876	2.0804	
Unweighted closeness centrality	0.6554	0.6348	0.7638	0.6840	
Weighted betweenness centrality	0.0007	0.0002	0.0918	0.0001	
Unweighted betweenness centrality	0.0100	0.0074	0.0321	0.0099	
Weighted influence1 centrality	0.1775	0.0781	0.3806	0.1679	
Weighted influence2 centrality	4.3825	3.8344	5.2417	4.7643	
Unweighted influence1 centrality	0.0536	0.0482	0.0774	0.0628	
Unweighted influence2 centrality	1.0582	0.9638	1.4218	1.2464	
Clustering coefficient centrality	0.7651	0.7967	0.6304	0.6967	

Appendix 14: Key Performance Parameters General Network Item Constellation Plot - Node Data 2016Jun

General Network Item Constellation Plot - Node Data 2016Jun			
	Clusters		
Parameters	Clust 1	Clust 2	Clust 3
Weight	440.4545	359.3333	304.6364
Out degree centrality	12.5455	10.5000	11.2727
Weighted eigenvector centrality	0.3998	0.2287	0.1813
Unweighted eigenvector centrality	0.5386	0.4132	0.4650
Weighted closeness centrality	144.8393	122.5194	116.2239
Unweighted closeness centrality	0.6207	0.6157	0.5983
Weighted betweenness centrality	0.0000	0.0754	0.0045
Unweighted betweenness centrality	0.0125	0.0382	0.0112
Weighted influence1 centrality	0.2055	0.1534	0.1239
Weighted influence2 centrality	3.3921	2.5515	2.7408
Unweighted influence1 centrality	0.0010	0.0008	0.0009
Unweighted influence2 centrality	0.0165	0.0128	0.0146
Clustering coefficient centrality	0.7444	0.7873	0.7786

Appendix 15: Key Performance Parameters General Network Item Constellation Plot - Node Data 2016Dec

General Network Item Constellation Plot - Node Data 2016Dec					
	Clusters				
Parameters	Clust 1	Clust 2	Clust 3	Clust 4	Clust 5
Weight	3360.7000	2574.5000	2428.5000	2526.5000	2984.1765
Out degree centrality	12.5000	8.0000	8.0000	8.5000	8.0000
Weighted eigenvector centrality	0.3422	0.2471	0.2346	0.2348	0.2287
Unweighted eigenvector centrality	0.5899	0.4669	0.4669	0.4823	0.3423
Weighted closeness centrality	1224.7076	1163.8538	1137.1322	1144.3490	1064.6544
Unweighted closeness centrality	0.6305	0.5714	0.5714	0.5766	0.5958
Weighted betweenness centrality	0.0000	0.0000	0.0000	0.0000	0.0543
Unweighted betweenness centrality	0.0140	0.0003	0.0003	0.0008	0.0358
Weighted influence1 centrality	0.1966	0.1223	0.1165	0.1197	0.1474
Weighted influence2 centrality	3.5210	3.2717	3.2659	3.2909	2.3741
Unweighted influence1 centrality	0.0001	0.0001	0.0001	0.0001	0.0001
Unweighted influence2 centrality	0.0019	0.0017	0.0017	0.0017	0.0012
Clustering coefficient centrality	0.7683	0.9286	0.9286	0.9127	0.8257

Appendix 16: Key Performance Parameters General Network Item Constellation Plot - Node Data 2017Jun

General Network Item Constellation Plot - Node Data 2017Jun				
	Clusters			
Parameters	Clust 1	Clust 2	Clust 3	Clust 4
Weight	3490.7273	2978.2778	2706.0000	2601.0000
Out degree centrality	12.3636	7.6667	8.0000	8.0000
Weighted eigenvector centrality	0.3399	0.2229	0.2532	0.2458
Unweighted eigenvector centrality	0.5860	0.3443	0.4695	0.4695
Weighted closeness centrality	1264.0788	1085.6217	1205.0628	1187.5655
Unweighted closeness centrality	0.6280	0.5912	0.5714	0.5714
Weighted betweenness centrality	0.0000	0.0512	0.0000	0.0000
Unweighted betweenness centrality	0.0134	0.0338	0.0003	0.0003
Weighted influence1 centrality	0.1944	0.1422	0.1260	0.1228
Weighted influence2 centrality	3.5249	2.4037	3.3039	3.3008
Unweighted influence1 centrality	0.0001	0.0001	0.0001	0.0001
Unweighted influence2 centrality	0.0018	0.0012	0.0016	0.0016
Clustering coefficient centrality	0.7643	0.8665	0.9286	0.9286

Appendix 17: General Network Item Constellation Plot - Node Data 2017Dec

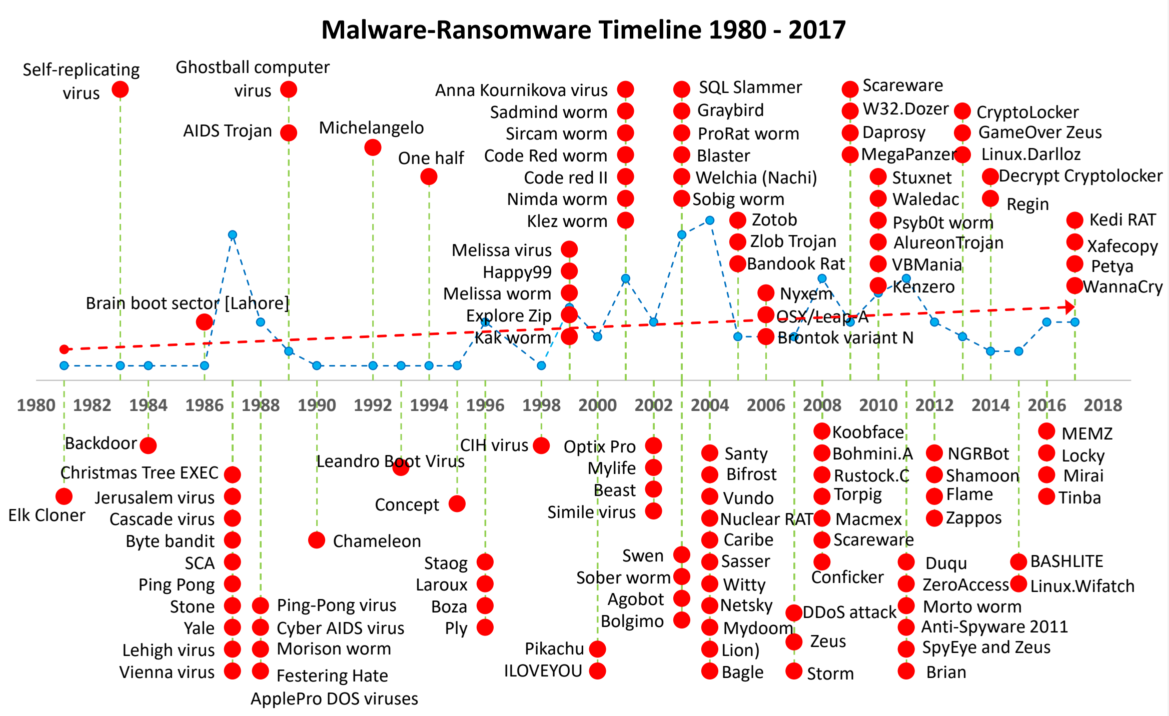
General Network Item Constellation Plot - Node Data 2017Dec					
	Clusters				
Parameters	Clust 1	Clust 2	Clust 3	Clust 4	Clust 5
Weight	3774.8000	2842.5000	2775.0000	2679.5000	3256.8824
Out degree centrality	12.5000	8.0000	8.0000	8.0000	7.7059
Weighted eigenvector centrality	0.3499	0.2461	0.2465	0.2232	0.2243
Unweighted eigenvector centrality	0.5930	0.4714	0.4714	0.4648	0.3368
Weighted closeness centrality	1363.1984	1285.8315	1273.3897	1241.2872	1169.7447
Unweighted closeness centrality	0.6305	0.5714	0.5714	0.5716	0.5930
Weighted betweenness centrality	0.0000	0.0000	0.0000	0.0000	0.0542
Unweighted betweenness centrality	0.0142	0.0003	0.0003	0.0009	0.0361
Weighted influence1 centrality	0.2004	0.1221	0.1228	0.1123	0.1439
Weighted influence2 centrality	3.5297	3.2853	3.2860	3.2289	2.3375
Unweighted influence1 centrality	0.0001	0.0001	0.0001	0.0001	0.0001
Unweighted influence2 centrality	0.0017	0.0015	0.0015	0.0015	0.0010
Clustering coefficient centrality	0.7668	0.9286	0.9286	0.9167	0.8538

Appendix 18: General Network Item Constellation Plot - Node Data 2018Jun

General Network Item Constellation Plot - Node Data 2018Jun

	Clusters				
Parameters	Clust 1	Clust 2	Clust 3	Clust 4	Clust 5
Weight	3778.3000	2843.5000	2781.5000	2682.5000	3261.1176
Out degree centrality	12.5000	8.0000	8.0000	8.0000	7.7059
Weighted eigenvector centrality	0.3498	0.2459	0.2468	0.2232	0.2243
Unweighted eigenvector centrality	0.5930	0.4714	0.4714	0.4648	0.3368
Weighted closeness centrality	1364.8501	1286.8553	1275.5178	1242.7077	1171.0776
Unweighted closeness centrality	0.6305	0.5714	0.5714	0.5716	0.5930
Weighted betweenness centrality	0.0000	0.0000	0.0000	0.0000	0.0542
Unweighted betweenness centrality	0.0142	0.0003	0.0003	0.0009	0.0361
Weighted influence1 centrality	0.2003	0.1220	0.1229	0.1123	0.1439
Weighted influence2 centrality	3.5292	3.2849	3.2859	3.2285	2.3377
Unweighted influence1 centrality	0.0001	0.0001	0.0001	0.0001	0.0001
Unweighted influence2 centrality	0.0017	0.0015	0.0015	0.0015	0.0010
Clustering coefficient centrality	0.7668	0.9286	0.9286	0.9167	0.8538

Appendix 19: The Malware-Ransomware Timeline for the Period 1980-2017 showing names of malware against date of release, the dotted blue line shows the frequency for each year, while the dotted red line is the trend showing increased malware release over the years



Appendix 20: Performance of Android Forensics Data Recovery Tools

PERFORMANCE OF ANDROID FORENSICS DATA RECOVERY TOOLS

CHAPTER
7

Performance of Android Forensics Data Recovery Tools

B.C. Ogazi-Onyemaechi, A. Dehghantanha*, K.-K.R. Choo*[†,‡]

***University of Salford, Salford, United Kingdom [†]University of Texas at San Antonio, San Antonio, TX, United States [‡]University of South Australia, Adelaide, SA, Australia**

1. INTRODUCTION

Smart mobile devices, particularly smartphones, are increasingly popular in today's Internet-connected society [1–4]. For example, few years ago in 2010, shipments of smartphone grew by 74% to 295 million units [3,4]. Unsurprisingly, sales of smartphones have been increasing since then [5,6], and it has been estimated that 1.5 billion smartphones will be sold by 2017 and 1 billion mobile subscribers by 2022 [7–15].

Such devices are generally used to make phone calls, send SMS messages, web browsing, locate places of interests, map navigation, image and video capture, entertainment (e.g., gaming and lifestyle), business and economic transactions (e.g., internet banking), take notes, create and view documents, etc. [6,16–18]. Due to their widespread adoption in corporate businesses, these devices are a rich source of information (e.g., corporate data and intellectual property) [19–22]. The potential to target such devices for criminal activities (e.g., malware such as banking Trojans) or be used as an attack launch pad (e.g., used to gain unauthorized access to corporate data) [19,23–29], makes it important to ensure that we have the capability to conduct a thorough investigation of such devices [22,30,31,18,32–35].

While there are a small number of forensic tools that can be used in the forensic investigation of smart mobile devices [36], the extent to which data can be recovered varies, particularly given the wide range of mobile devices and the constant evolution of mobile operating systems and hardware [37,38]. For example, recovering data from the internal memory of a smartphone remains a challenge [34,39,40]. Further to these challenges is the requirement to create forensically sound and effective tools and procedures [36,20,41].

Therefore, it is essential that the forensic community keeps pace with forensic solutions for smart mobile devices [42,43,2–4]. This is the focus of this chapter. Specifically, we study the effectiveness of five popular mobile forensics tools, namely: Phone Image Carver, AccessData FTK (Forensic Tools Kit), Foremost, Recover My Files, and DiskDigger, in recovering evidential data from a factory-restored Samsung Galaxy Note 3 running Android Jelly Bean version 4.3.

The structure of this chapter is as follows. Section 2 reviews related work. Section 3 outlines the methodology and our experiment setup. Section 4 presents our findings, and Section 5 concludes this chapter.

2. RELATED WORK

The present investigation is conducted based on the current acquisition method on smartphones and the extent of available forensic techniques and tools on the analysis of evidence. Smartphones have many profound sections of where and how evidence is collected when it is dealing with a crime occurrence [44]. Therefore with the ever-increasing features and utility, it becomes much more complicated to collect evidences from a smartphone [43,45]. Consequently, there are different acquisition methods on different architectures and software that a smartphone operates on [18,39]. These differences in the architecture make it extremely difficult to perform similar acquisition method on different devices and operating systems [23]. However, NIST created guidelines for Computer Forensics Tools Testing (CFTT) to provide for the differences in architecture [45]. Therefore mobile device forensics has been defined as the science of recovering digital evidence from a mobile device under forensically sound conditions using accepted methods. Notwithstanding, mobile device forensics is considered an evolving specialty in the field of digital forensics [23,25]. The procedures for the validation, preservation, acquisition, examination, analysis, and reporting of digital information had been discussed [25]. Although there had been an established program and guidelines, mobile device forensics, like any other evolving field, still has its own forensics challenges [26]. Evolvement on different areas of usage for mobile phones provides even further challenges in their investigation. For example, investigation of cloud applications on mobile phones [44,46–48], malwares on smartphones, [49–51], and investigating mobile phones as part of botnets [52] and SCADA [53] systems are all challenging forensics research areas. In view of the evolving nature of mobile device forensics, it is suggested that forensic practitioners who rely primarily on general-purpose mobile forensic toolkits might find that no single forensic tool could recover all relevant evidence data from a device [6]. Therefore researchers are working to establish the best forensic tools and procedures that are reliable for mobile device's investigation [27,31,33,46,54].

Investigations conducted in the field of mobile device forensics still show variations in research opinions on the effectiveness and reliability of different forensic tools when applied to different mobile device architectures [23]. The removal of internal memories from a mobile device or their mirroring procedure is evasive and complex because of difficulties in having direct hardware access [16]. To resolve such challenges, five mobile device forensic scenarios were studied; and a method to perform data acquisition of an Android smartphone regardless of the architecture was proposed. The method was validated using Motorola Milestone II A953, Sony Ericson Xperia X10 miniPro, Motorola Defy, Samsung Galaxy S 9000a, Motorola II, and Motorola Milestone A853 [16]. These architecture-based difficulties were confirmed by investigating Symbian- and Windows-based mobile devices [23]; which revealed that tools for forensic investigation of smartphone mobile devices always would pose challenges in forensic investigation because of the continual evolving nature of the technology [23,40]. The data recovery capabilities of EnCase, FTK, Recuva, R-Studio, and Stellar Phoenix from a desktop Windows XP were compared [36]. The comparison revealed that EnCase, FTK, Recuva, and R-Studio performed identically when recovering marker files from most images [36,53]. The experiment showed also that Stellar Phoenix corrupted two bytes in each of two text files (even when these files had not been deleted) and added a padding of zeroes at the end of another file that had not been deleted [36]. Nonetheless, the study concluded that no two tools produced identical results [36]. Further to this, a different study conducted on Samsung Star 3G phone used TK file Explorer 2.2, MOBILedit 4, and Samsung PC studio for logical acquisition of different data files for analysis, and this was to create a framework for forensic investigation of Samsung Star 3G device [17]. Evidence collection and analysis conducted on a Nexus 4 phone discovered a flaw that allowed access to all data on the device without a device wipe that occurs when the bootloader is unlocked [55]. The challenges of smartphone forensics continue to expand even with the emergence of Linux-based Firefox OS, which has no procedure yet for forensic investigation [56].

To ensure a sound forensic investigation, care should be taken to preserve and retrieve the volatile data inside the memory of the mobile device [47]; hence, a backup and acquisition process was proposed to work on windows mobile phones, android mobile phones, and iPhones [47]. The recovery of information held on a Windows Mobile smartphone was investigated using different approaches to acquisition and decoding, accepting AccessData FTK and DD imagers. The investigation concluded that no one technique recovers all information of potential forensic interest from the device [35]. Further work on mobile device forensics shows that a diverse collection of smartphone forensic tools has been introduced; however, these studies do not guarantee data integrity, which is required for digital forensic investigation [48]. Therefore, Android device acquisition utilizing Recovery Mode was investigated to analyze the Android device Recovery Mode variables that compromise data integrity at the time of acquisition. Consequently, an Android data acquisition tool that ensures integrity of acquired data was developed [48]. It was noted there was not yet specific procedures or rules to collect evidence from a smartphone [7]. Therefore, forensic investigators use existing procedures in the acquisition of digital evidence [7]. Furthermore, it was suggested that the relative amount of important evidence that could be gathered from the smartphone differ based on different versions of software system that runs on the smartphone. However, nothing was done on the acquisition of evidence on a formatted android device—where it was claimed that all data and applications were erased [7].

The challenges of data recovery for forensic investigation extend to cloud computing environment [46]. Therefore, utilizing TPM in hypervisor [42], implementing multifactor authentication and updating the cloud service provider policy to provide persistent storage devices are proposed to overcome the difficulties in Cloud forensics [51,57]; which include limited access to obtain evidence from the cloud, seizure of physical evidence for integrity validation, evidence presentation, or rulings on data saved in different locations [46]. In a cloud-related forensic investigation, Guidance Encase and AccessData forensic Toolkit were evaluated; and the tools show that they can successfully return volatile and nonvolatile data from the cloud [50]. Thus, a foundation is laid for the development of new acquisition methods for the cloud that will be trustworthy and forensically sound [50]. To ensure a reliable cloud-based forensic investigation, a step-by-step technique for evidence data collection was proposed [55]. There was a review of 7 years of research into forensic investigation of various smartphone mobile device platforms, data acquisition scheme and information recovery methods in order to provide comprehensive reference material to enhance future research [33]. Prior to the advancement of forensic tools, the traditional method of memory acquisition focused on the physical memory. This procedure most often requires the removal of the memory chip from the chipboard. These methods put valuable evidence at risk because during the removal process there might be loss or damage of essential evidence.

Emphasizing on the need for accurate and reliable forensic tools and procedures, there was a warning against unfair application of wrong forensic techniques and evidence to secure conviction in fervour of the prosecution [52]. Notably, some challenging areas include "erroneous allegations of knowledgeable possession, misuse of time stamps and metadata, control and observation of the discovery process." Other challenging areas include "authentication issues, deficiencies and the lack of verification for proprietary software tools, deliberate omission or obfuscation of exculpatory evidence and inadvertent risks resulting from the use of legitimate services" [52].

The foregoing review shows there is need to investigate the recovery performance of Phone Image Carver, AccessData FTK, Foremost, Recover My Files, and DiskDigger from FTK and DD images acquired from Android mobile smartphones.

3. EXPERIMENT SETUP

In our experiments, we used the popular Samsung Galaxy S2 i9100 as the case study device. The device has an internal memory of 16 GB, random access memory (RAM) of 1 GB, running Android Gingerbread version 2.3.4 operating system (OS) (Android OS, Ice Cream Sandwich version 4.0.3). The focus of our study was on the internal memory; thus, no memory card was inserted in the phone. Prior to the experiments, the phone was preloaded with Enron dataset [58–64], which is considered similar to data collected for fraud detection.

Therefore, it is a good dataset for the present investigation. The device was subsequently factory reset before taking images on the phone (see Fig. 1). The reason for factory resetting the phone was to wipe all preloaded data and investigate the effectiveness of the forensic tools in recovering the data erased from the device.

Fig. 1 shows that two different image acquisition processes, logical and physical image acquisition, were conducted after the device was restored to default factory state. AFLogical forensic tool was used for logical Image acquisition on the formatted disk (device). The tool captures the call-log calls, contacts phones, MMS, MMS-Parts, and SMS, which were contained in the preloaded Enron dataset. It stores this information in a zip folder named forensics.zip within the device itself. The .zip folder contains .csv files, which hold the logs of the device.

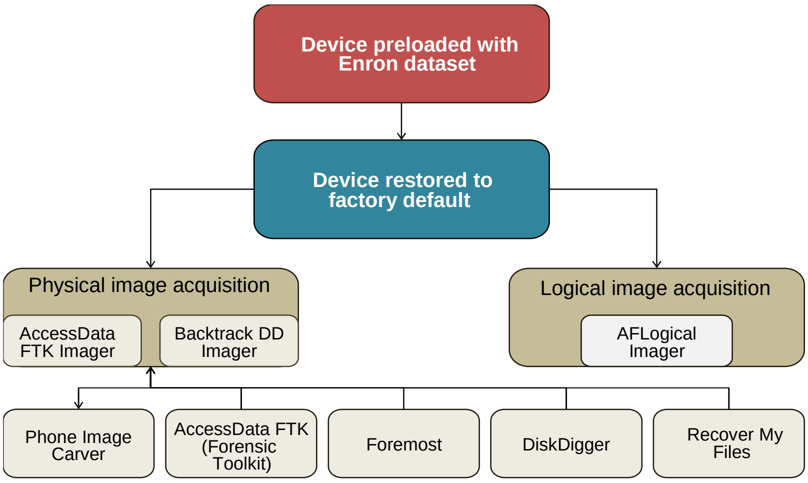

Device preloaded with Enron dataset
Device restored to factory default
Physical image acquisition
AccessData FTK Imager
Backtrack DD Imager
Logical image acquisition
AFLogical Imager
Phone Image Carver
AccessData FTK (Forensic Toolkit)
Foremost
DiskDigger
Recover My Files

FIG. 1 Schematic representation of processes conducted on the case study device.

However, when opened .cvs files show blank suggesting that logical acquisition of data is not executable in a reformatted device. Comparing to another investigation, analysis of logical image, bb file, from Blackberry PlayBook device did not produce direct data files of user activity. However, it produced some key files that can assist to further trace device usage [33,53,65]. Related studies confirm also that Encase and FTK forensic tools could not recover all data from NTFS-formatted logical disk partitions [34,52].

On the other hand, AccessData FTK Imager 3.1.3.2 and Backtrack dd Imager were used for physical acquisition of images. Subsequently, Phone Image (Carver v1.6.0), AccessData FTK, Foremost, DiskDigger, and Recover My Files (v4.7.2) were employed to analyze the two different physical images.

During the image creation process using AccessData FTK Imager 3.1.3.2, the backup option was not selected. This was to ensure that no backup data was available. The physical memory of the device was imaged and analyzed using several tools. The resulting image revealed that the tool created only 2.227 GB image file compared to the 16 GB physical memory capacity of the device. This means that AccessData FTK imager recovered only

14% of the device memory capacity. The imager separated the physical drive into eight images. The size of the first seven images is 1.46 GB and the eighth image is 767 MB. This experiment compares with another study where AccessData FTK Imager recovered a higher average of 86.4% of the physical disk capacity from the various image segments [34].

Another image was acquired from the case study device using .dd imager in Backtrack. This tool converts the memory into a disk dump. Similar to the acquisition of image using AccessData, the dd imager used command line prompt in Backtrack dd Imager without any filter. The image size converted by this tool shows only 11 GB of the 16 GB memory capacity of the device, which translates to 68.75% recovery of the device capacity. The remaining memory was ignored due to slack spaces, which is the disk space between the end of the file content and the end of the last cluster in which the file is saved [34]. Two types of disk dumps were created after imaging the device. One of the disk dumps uses the access data FTK images, while the other uses backtrack dd image. Subsequently, several recovery tools were used to perform analysis on the images to determine the effectiveness of these tools. Prior to analysis, validation checks were conducted on the images using hash calculation for integrity checking [48] as shown in Table 1. Image numbers S2.001–S2.008 in column one of Table 1 are the validation and integrity check identification numbers for each of the eight images acquired by AccessData FTK imager. DD, on the other hand is the validation and integrity check ID number for the DD image that was acquired as a single image. The corresponding MD5 sums and SHA1 values are given in columns two and three of Table 1.

TABLE 1 validation and Integrity Check of Images

Image	MD5	SHA1
S2.001	df14a97ed884e959f79d718a4a3e8de0	838f1291740988d86e2b2b22625646e20e9e535a
S2.002	334bc971671ad78a09abadd81aa2419f	aa42e0269f4cb2fd53b065f0aea56823fb770d88
S2.003	53393c41f197b08a693db24600b2eab1	35cc22977e11e6c966c569260812fde04471401e
S2.004	4396a40fb1825166db005e39d211b5a8	fecb8eabff34ea01ad4a84cc283625af5ca0319f
S2.005	6e9a8fc2cc1235da1d33a51d73a53c30	1b03158408ae5e567d6a7e27993ab46ecd8fe686
S2.006	Ab49d0350af243d6d3d6df1791adb58f	Eb7a722cc1b14fc8bdf06c81390df45ab34f5a50
S2.007	35d8479c1dc29edacb33ad9dcaa07d5b	Eac0ba1a3e4e8ef8b0195ab682081c4d12d9d36e
S2.008	01a5c72710d2223d012e8e7b71e9055c	C782fb92564990314de7baefa2db748ac186aa7b
DD	1eeac023329e6d70ffcc78e7230c1ca7	76ae66d29894ae6b21f73bac87578c9dd1202c77

4. RESULTS AND DISCUSSIONS

Two types of image acquisition namely logical image and physical image acquisition were conducted. AFLogical was used to acquire the logical disk image. The tool captures the calllog calls, contacts phones, MMS, MMS-Parts, and SMS that were contained in the preloaded Enron dataset. It stores this information in a zip folder named forensics.zip within the device itself. The .zip folder contains .csv files that holds the logs of the device. However, when opened the .cvs files show blank suggesting that logical acquisition of data is not executable in a reformatted device. This revelation agrees with particularly the respective findings of Buchanan-Wollaston and Mercuri. The blank .cvs files confirm that Encase and FTK were unable to recover some data from NTFS-formatted logical disk partitions, except by further procedure where the data acquired must be decoded [33–35,52,66].

Earlier on AccessData image and backtrack dd image were acquired for the physical disk of the device. The content of the FTK Image and BackTrack dd image of the target Samsung Mobile device were loaded on different forensic tools. Three different tools namely, Phone Image Carver v. 1.6.0, AccessData FTK (Forensic ToolKit 1.81.3), and dd Image FTK were used to analyze the images.

TABLE 2 Recovered Files Format using Phone Image Carver

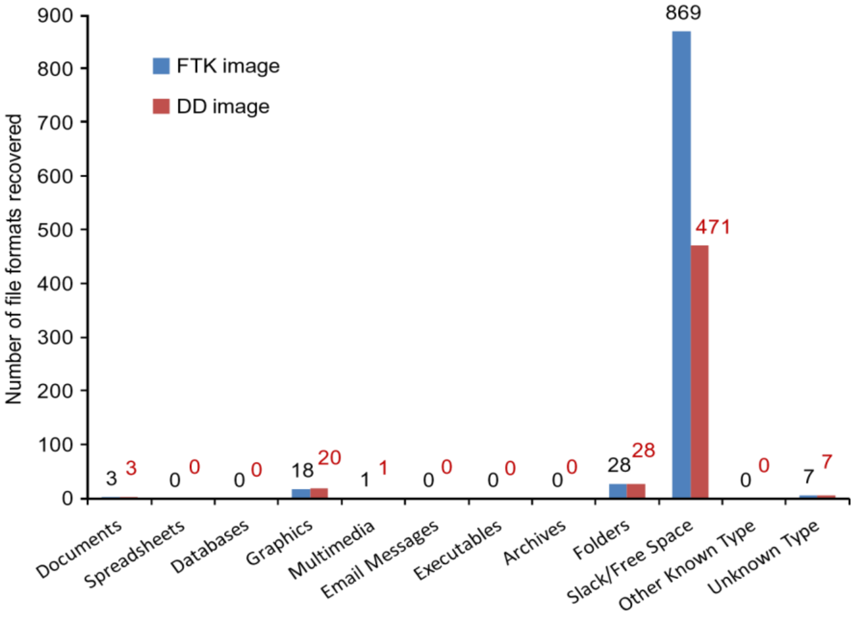
	File Format										
FTK image	DocX	HTML	JPG	MP3	SQLite	SWF Flash	Text	Text UTF-16	Text-Shift-JIS	Zip
Backtrack dd image	DocX	HTML	JPG	MP3	SQLite	SWF Flash	Text	Text UTF-16	Text-Shift-JIS	Zip

Table 2 shows the various files format recovered from the images using Phone Image carver. Phone Image Carver recovered evidence of many different formats of data, confirming that Phone Image Carver supports over 300 file types [35]. The present experiment supports the finding which suggests that .docx files were the only Office documents detected. However, it disagrees that Phone Image Carver did not detect .jpg files [35]. This study reveals that utilizing Phone Image Carver tool is extremely time-consuming and not efficient; suggesting that it is not suitable for real investigation. Phone Image Carver tool does not list out the deleted files according to file formats, however the information still exists within the image. It was noted that Phone Image Carver does not permit addition of further file types to its database, meaning that a number of file types from the data set would not be detected [35]. An additional feature of Phone Image Carver tool is that the activities performed in it are recorded in a log file. This offers a great opportunity for the analyst to go back and review his steps each time he uses Phone Image Carver tool.

The FTK and dd images acquired in Section 3 were analyzed using AccessData FTK. Analysis revealed that 12 [12] categories of files were recovered from each of the two images as shown in Fig. 2.

Fig. 2 shows minor differences in the recovery function of both FTK and dd images from the same restored device. The recovered files in Fig. 2 reveals that different images from different acquisition tools give different depth of evidence recovered in the analysis [34]. While FTK Image recovered total file items of 926, dd Image recovered far less number of 530 when the same FTK tool was used on both images (Tables 3 and 4). It is clear that entries in some of the file items are the same, some with zero entries (no recovery); however, few others show significant differences between both images. Worthy of note are the entries under FTK Image for “Unchecked items (926), Other thumbnails (18) and Filtered in (926).”

The analysis shows the corresponding entries for dd Image are lower, except for graphics and other thumbnails where dd Image and FTK Image recovered virtually equal numbers of files. This suggests that dd images provide negligently better recovery for graphic and thumbnail files. However, both images recorded similar values under File Status and File Category, except for Slack/free space where FTK Image had significantly higher value (869) than dd Image (471). The analysis suggests that in the event of recoverable evidences from slack spaces, FTK image gives more recovery files than dd image. In other words, slack spaces could contain deleted files, deleted file fragments, and hidden data [54,56]. The negative side of it is that recovering from it could result in a waste of time if there were no recoverable evidence in the slack spaces. It is worthy to note that files were not recovered from FTK Image 8, which contains the Slack/free spaces. On the percentage file recovery of FTK Image, Mercuri noted that CFTT Program tests revealed defects in the hard disk image preparation process on Windows XP OS [34,52]. Taken into account, the analysis revealed that defect might still have some drawback on the performance of FTK as an imager in the present study on smartphone.



File formats recovered from FTK & DD images using AccessData FTK

FIG. 2 Recovery and analysis of data from both FTK and dd images using FTK forensic toolkit.

TABLE 3 FTk Image Result using FTk

File Items	No	File Status	No	File Category	No
Total file items	926	KFF Alert Files	0	Documents	3
Checked items	0	Bookmarked items	0	Spreadsheets	0
Unchecked items	926	Bad extension	18	Databases	0

Flagged thumbnails	0	Encrypted files	0	Graphics	18
Other thumbnails	18	From email	0	Multimedia	1
Filtered in	926	Deleted files	2	Email messages	0
Filtered out	0	From recycle bin	0	Executables	0
		Duplicate items	2	Archives	0
		OLE subitems	0	Folders	28
		Flagged ignore	0	Slack/free space	869
		KFF ignorable	0	Other known type	0
		Data carved files	0	Unknown type	7

TABLE 4 dd Image Result using FTk

File Items	No	File Status	No	File Category	No
Total file items	530	KFF alert files	0	Documents	3
Checked items	0	Bookmarked items	0	Spreadsheets	0
Unchecked items	530	Bad extension	20	Databases	0
Flagged thumbnails	0	Encrypted files	0	Graphics	20
Other thumbnails	20	From email	0	Multimedia	1
Filtered in	530	Deleted files	2	Email messages	0
Filtered out	0	From recycle bin	0	Executables	0
		Duplicate items	2	Archives	0
		OLE subitems	0	Folders	28
		Flagged ignore	0	Slack/free space	471
		KFF ignorable	0	Other known type	0
		Data carved files	0	Unknown type	7

Further analysis was conducted on the FTK and Backtrack dd Images using Foremost in Backtrack forensic tool. The FTK Images (1–8) were analyzed in segments, the way they were imaged. Fig. 3 shows the numbers of various file formats recovered from the two images.

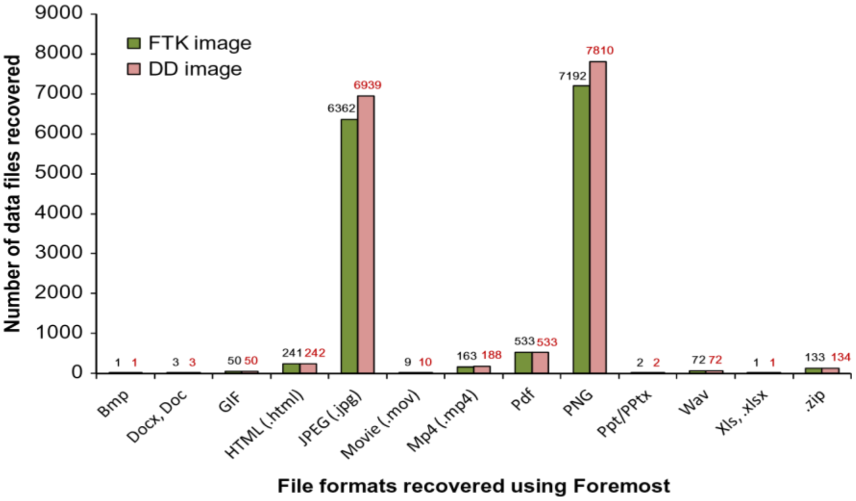
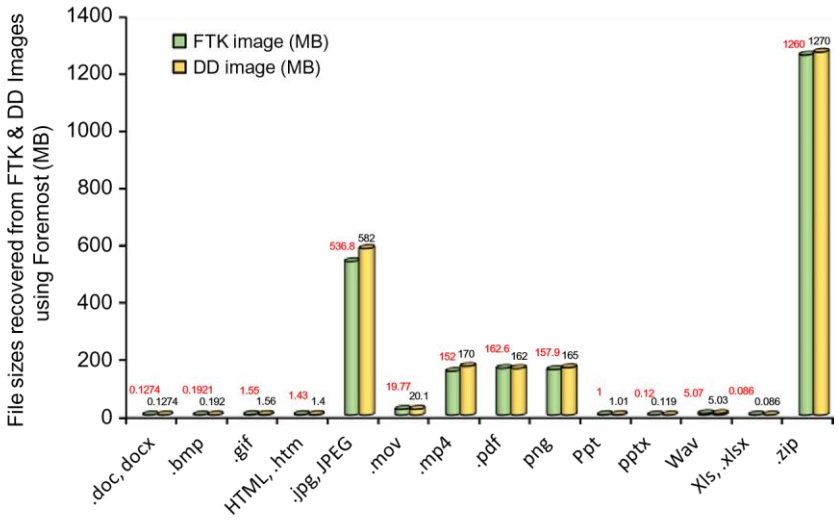


FIG. 3 Number of data files recovered from FTK and dd images using Foremost in Backtrack.

Unlike AccessData FTK, analysis conducted using Foremost showed format-specific files from both images. Apparently, Foremost recovered similar types of files from the two images, particularly, bmp, docx/doc, gif, html, movie, pdf, ppt/pptx, wav, xls/xlsx, and zip. The analysis showed that both images recovered the largest number of jpeg and png files. While analysis in Foremost showed that FTK image recovered 6362 jpeg and 7192 png files, dd image on the other hand recovered 6939 and 7810, respectively. The large number of graphics analyzed (recovered) suggests that the slack spaces contain more deleted graphic files, therefore foremost performed better in recovering them [54]. However, analysis in Foremost revealed a higher number of files from dd Image than it did from FTK image for jpeg/jpg, mp4, and png. Compared to the AccessData FTK, Foremost analyzed a higher number of file types. The recovery of high number of file types could be a result of high-performance of Foremost tool. The absence of slack/free spaces in the analysis suggests that Foremost performed better in recovering all deleted files, deleted file fragments, and hidden data files that were present in the slack/ free spaces of the images, which AccessData FTK could not recover as shown in Fig. 2 [54,56].



File formats recovered from FTK & DD Images using Foremost

FIG. 4 Data file sizes recovered from FTK and dd images using Foremost in Backtrack.

The large sizes of data file shown in Fig. 4 further corroborates the suggestion that Foremost recovered deleted files, deleted files fragments, and perhaps hidden files that might be contained in the slack/free spaces which AccessData FTK could not recover as shown in Fig. 2 [54]. Comparing the data file sizes with the number of files recovered in Fig. 3, it is evident that Foremost forensic tool recovered more data from Backtrack dd Image than it recovered from FTK Image. This suggests also that Foremost forensic tool performs better in analyzing data files from Backtrack dd Image than FTK Image. The analysis reveals also that the zip, jpeg, mp4, png, and pdf files were the files most recovered using Foremost forensic tool. This result shows that Foremost performs better than AccessData FTK in recovering files from FTK and dd images, particularly dd image of a restored Android mobile smartphone device.

The next forensic tool to use in the analysis of the two images (FTK and dd acquired images) was Recover My Files version 4.2.7. The result of the analysis is shown in Fig. 5. The number of data files recovered from the two images shows no differences in the performance of Recover My Files as a forensic tool. The tool recovered 16,756 files from the FTK image, which is comparable with 16,758 files recovered from Backtrack dd image.

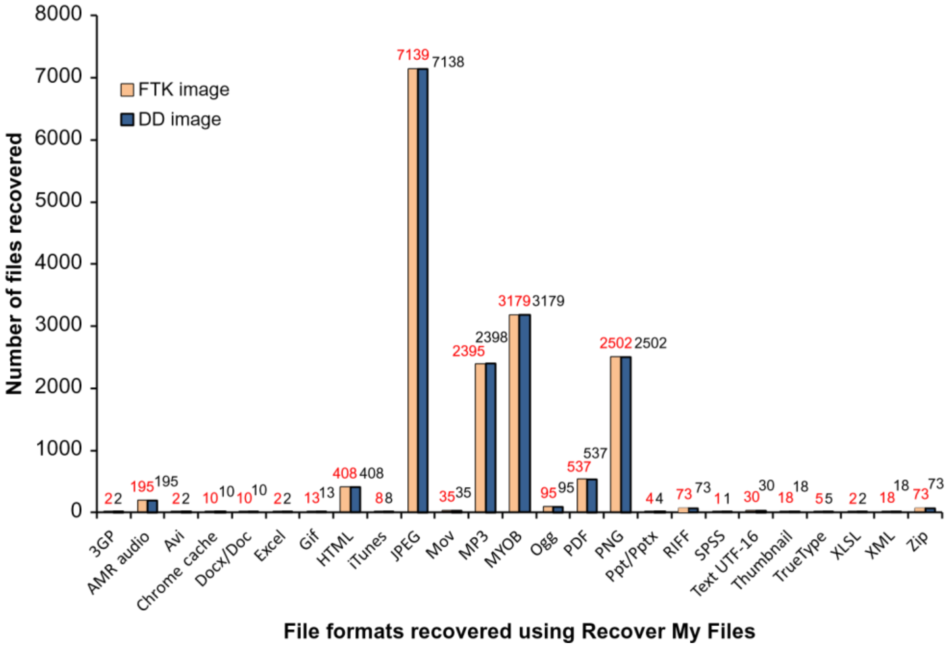
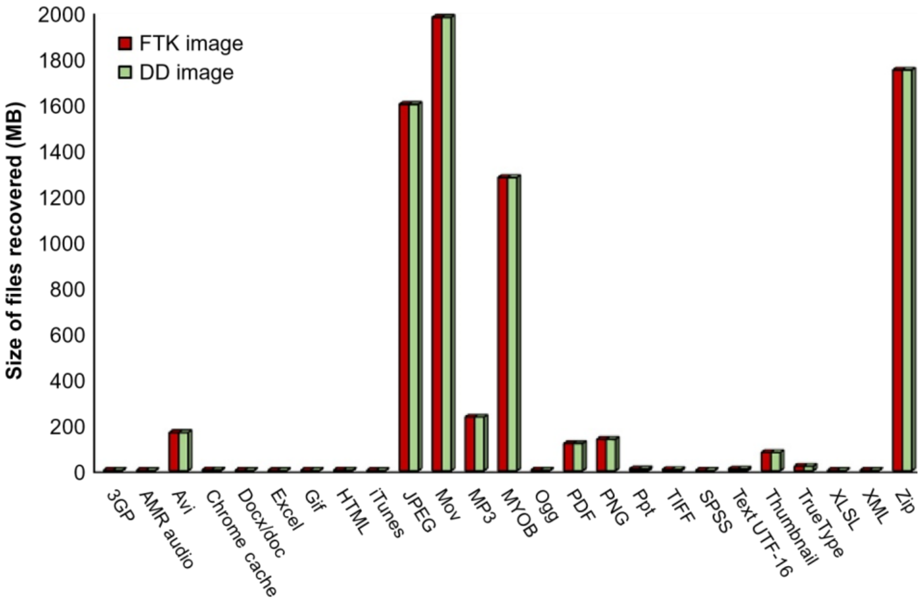


FIG. 5 Number of data files formats recovered from FTK and dd images using Recover My Files.

The results show also that jpeg, myob, png, mp3, pdf, html, and amr audio files were the most recovered in the order of listing. The tool has shown to recover more file formats and showed no slack/free space. The additional file formats it recovered include 3gp, amr audio, avi, chrome cache, itunes, mp3, myob, ogg, tif, spss, text UTF-16, thumbnails, and truetype. Similarity in the recovery function of Recover My Files on the two images is revealed further in Fig. 6. It is evident in Fig. 6 that the tool recovered equal sizes of data files from both FTK and dd images. The tool recovered 7.37 GB of data files for each file image confirming that mov, zip, jpeg, myob, mp3, avi, png, and pdf are among the most recovered files format.

Comparing Recover My File and Foremost in Figs. 7 and 8, the results show that Foremost recovered 14,762 and 15,985 data files from FTK and dd images respectively while Recover My File recovered 16,756 and 16,758 data files respectively from the same images. This recovery performance is repeated in the size of data files recovered, where Foremost recovered 2.35 and 2.38 GB of data files from FTK and dd images. On the other hand, Recover My File recovered 7.37 and 7.39 GB data files respectively from the same FTK and dd images.



Data file formats recovered using Recover My Files

FIG. 6 Sizes of data files recovered from FTK and dd images using Recover My Files.

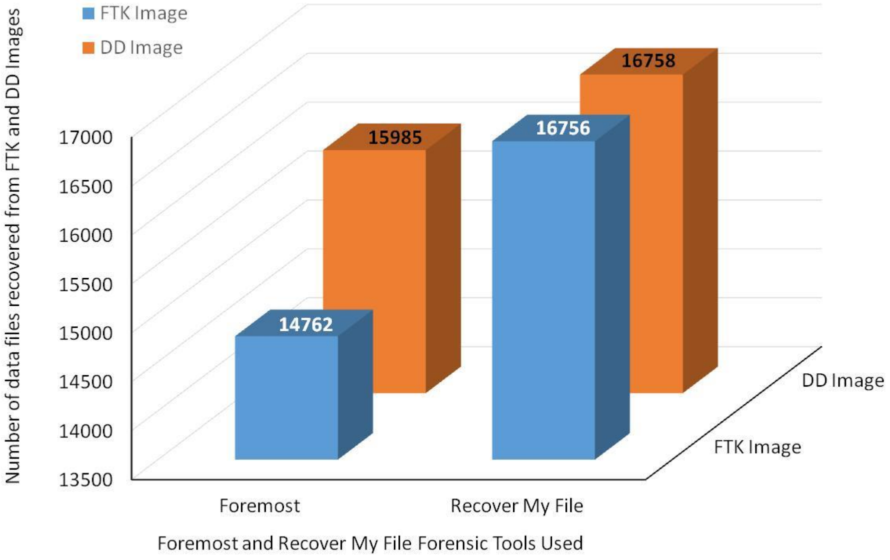


FIG. 7 Number of files recovered between Foremost and Recover My File.

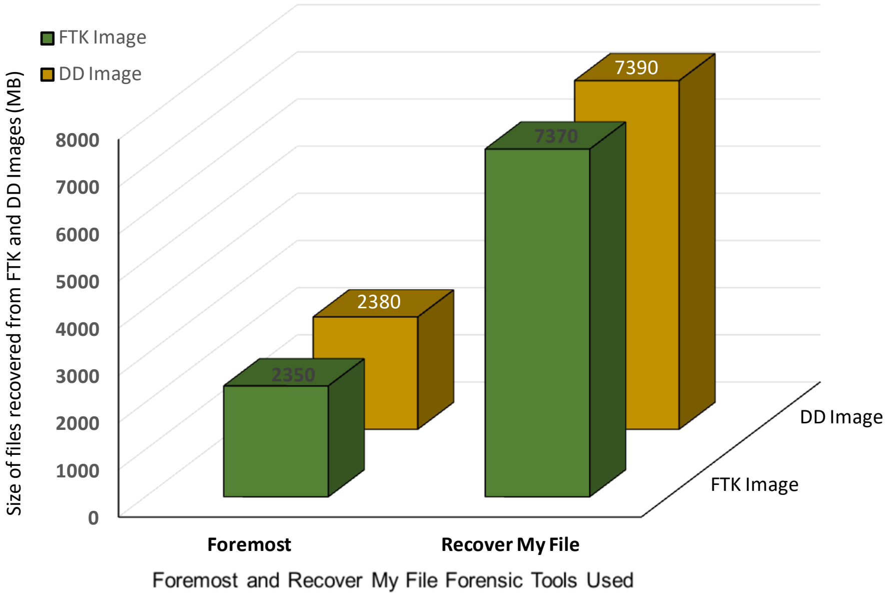


FIG. 8 Size of files recovered between Foremost and Recover My File.

Both the number and size of data files recovered could elaborate the huge difference in the recovery functions of the two forensic tools. While results show that foremost recovered large number of data files formats in jpeg (6939), png (7810), pdf (533), and hmtl (242) for FTK and dd Images, Recover my File recovered an average of jpeg (7139), myob (3179), png (2502), mp3 (2398), pdf (537), and html (408). Similarly, Foremost recovered an average size in jpeg (0.582 GB), zip (1.27 GB), mp4 (0.17 GB), pdf (0.162 GB), and png (0.165), while Recover My File recovered some phenomenal sizes in mov (1.98 GB), zip (1.75 GB), jpeg (1.6 GB), myob (1.28 GB), mp3 (0.2329 GB), and avi (0.1651 GB) for the two images. These analyzes reveal that Recover My File is more effective than Foremost in recovering data files from FTK and dd images of a restored device.

The superior performance of Recover My File over Phone Carver image, AccessData FTK and Foremost was exciting. Therefore, one more analysis was conducted on the acquired FTK and Backtrack dd Images using DiskDigger forensic tool. The analyzes are shown in Fig. 9.

Fig. 9 reveals that DiskDigger recovered also many different data files format. What is interesting here is that DiskDigger recovered equal number of data files from both FTK and DD Images. The figure shows that the largest numbers of files recovered from both images are from jpeg (7,178), tif (6939), png (2502), and mp3 audio (2398) files in the listed order. It is important to observe that DiskDigger tool did not record any slack/free space. This implies that the tool made deep recovery of all files in the two images. The distribution of the number of recovered data files (see Fig. 9) corresponds to the sizes of data files recovered from the images (see Fig. 10).

As shown in Fig. 10, the zip folder recorded the largest data file sizes, while jpeg, tif, mp3, avi, and mp3 followed in the same order.

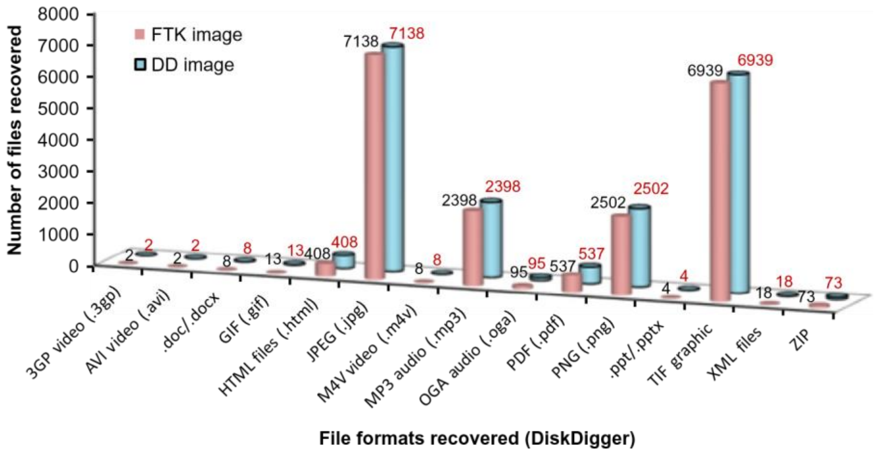


FIG. 9 FTK and DD image analysis of number of data files using DiskDigger tool.

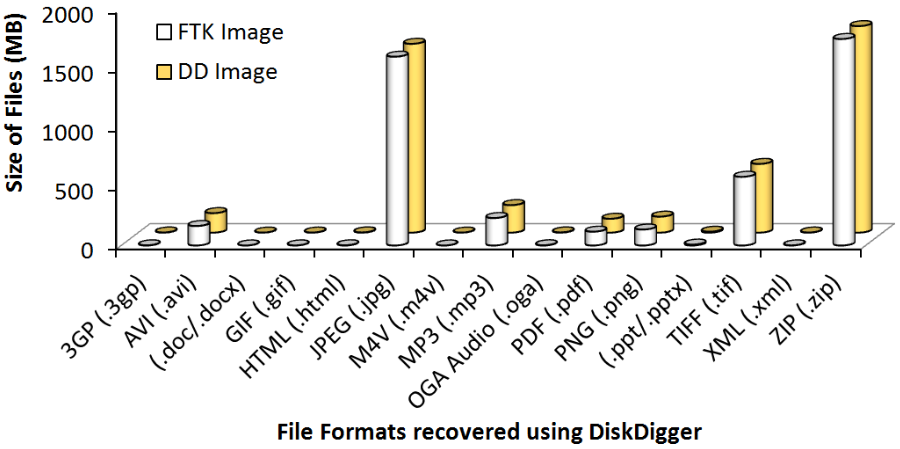


FIG. 10 FTK and DD image analysis of size of data files using DiskDigger tool.

The results in Fig. 10 suggest that Recover My File and DiskDigger show better recovery performance than Phone Image Carver and AccessData FTK. Although DiskDigger recovered 15 different file formats, Recover My File recovered 25 file formats, indicating a better deep recovery performance. However, DiskDigger recovered a greater total number of files (20,145) for both Images than the total number of files (16,758) recovered by Recover My File as shown in Fig. 11. The greater number of data files recovered from DiskDigger comes from the number of .tif files (6939) for both FTK and dd Images, which Recover My File did not recover.

Comparing Recover My File and DiskDigger in terms of data file size, it is clear that Recover My File recovered a larger file size (7.387 GB) than the data file size (4.596 GB) recovered by DiskDigger shown in Fig. 12. This difference is explained by the larger number of file formats recovered by Recover My File. These additional file formats and sizes (MB) include, amr audio (0.496 MB), chrome cache (2.6 MB), itunes (0.121 MB), mov (1980 MB),

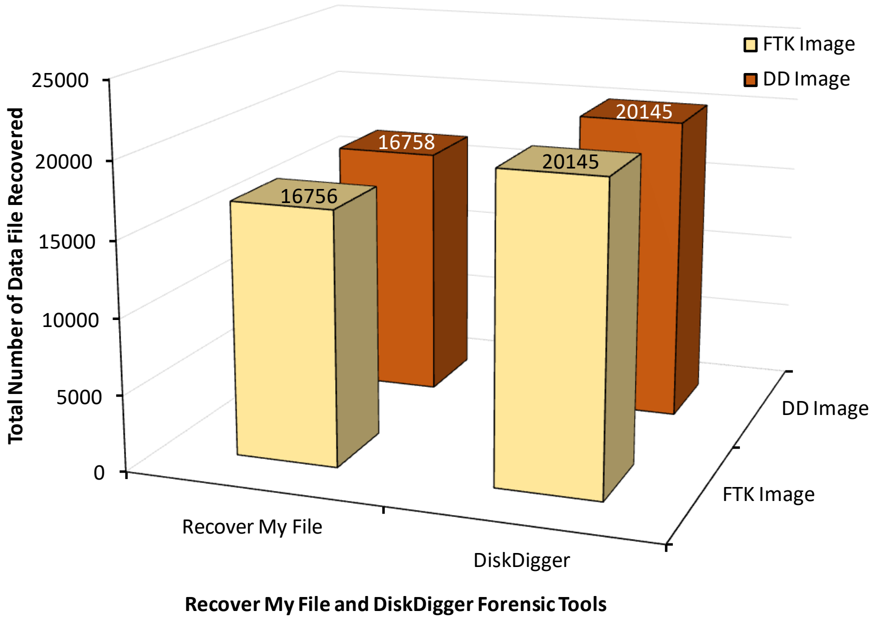


FIG. 11 Comparing data file number recovered using Recover My File and DiskDigger

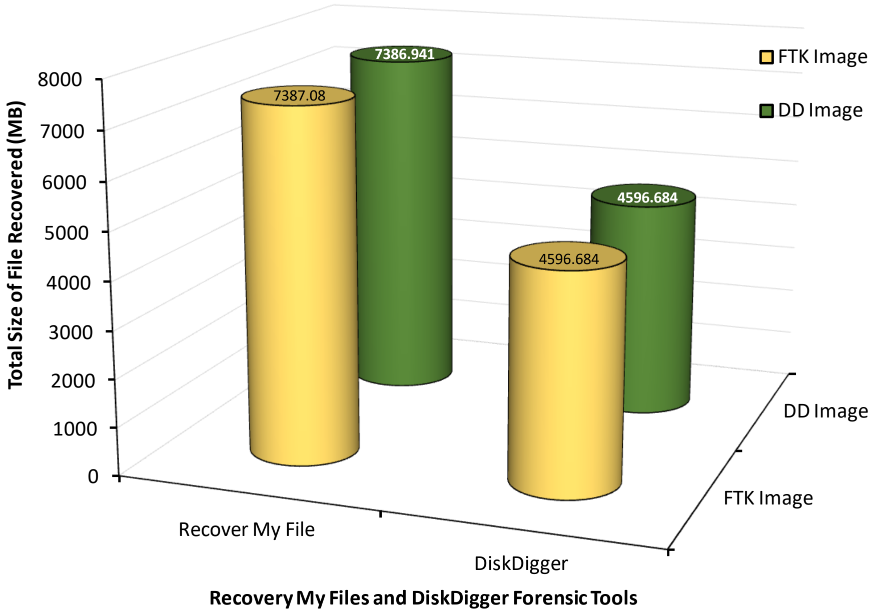


FIG. 12 Comparing data file size recovered using Recover My File and DiskDigger

myob (1280 MB), ogg (1.8 MB), spss (0.657 MB), text UTF-16 (7.3 MB), thumbnails (77.7 MB), and truetype (18.3 MB). It is evident that the two additional data file formats from Recover My File that made major contribution to the difference are .mov and .myob. The results suggest that DiskDigger performs better than Recover My File only in the area of the number of data files, particularly the .tif files (6939), recovered by each from the two images. Conversely, Recover My File performs better in terms of deep search for different data file formats, and the size of data files recovered.

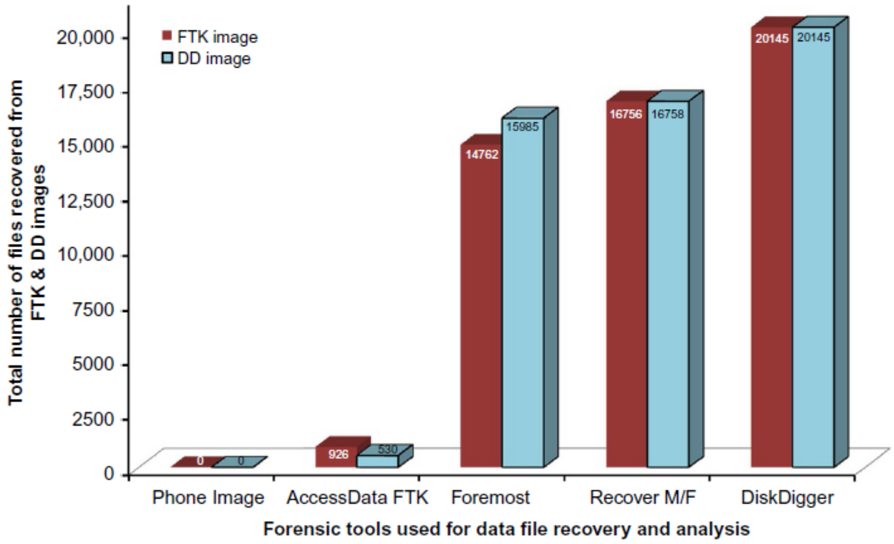


FIG. 13 Total number of files recovered from FTK & DD images using different forensic tools.

The summary of the recovery performance of the five forensic tools is shown in Figs. 13 and 14. The figures show that Phone Image Carver AccessData FTK did not perform well under the present experimental conditions as a forensic recovery tool. Foremost, on the hand recovered more file formats and appreciable large number of data files with corresponding data file size. However, Foremost shows to recover slightly higher number of data files (jpeg and png) from dd image than FTK image while the sizes of the data file recovered show no difference. DiskDigger appears to have performed well compared to Recover My File. It recovered many file formats that is still less than the number of data file formats recovered by Recover My File. DiskDigger proves to recover the highest number of data files but less than the size of data files recovered by Recover My File. Therefore, the study shows that Recover My File has the best recovery function as a forensic tool. It proved to have the deepest search penetration, recovered the highest number of data file formats, and recovered the largest size of data file formats, although it was less than DiskDigger in the number of data files recovered.

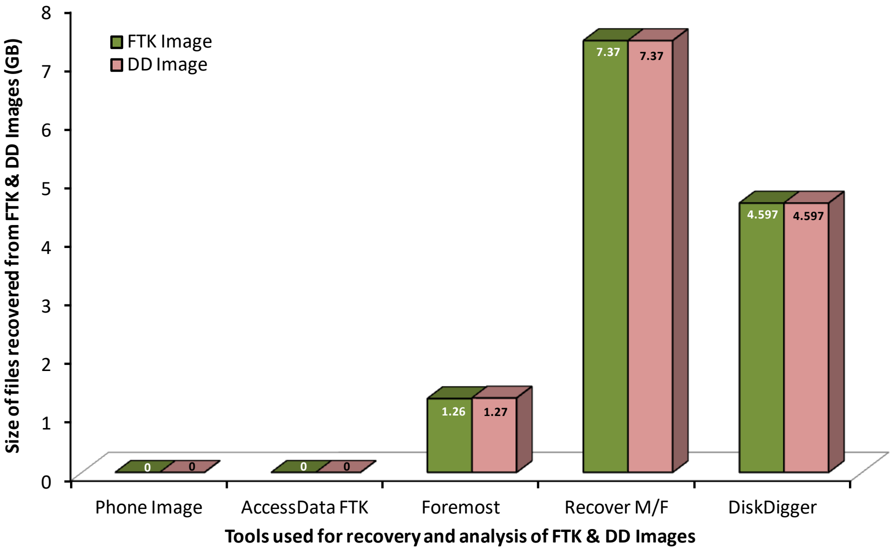


FIG. 14 Total size of data files recovered from FTK & DD images using different forensic tools.

5. CONCLUSION AND FUTURE WORKS

The increasing use of smartphone for various social-economic transactions led to a consequent increase in cybercrimes committed through smartphones. The nature of these devices and the variety of applications resided on them deemed challenging challenges to forensic investigators. To address these challenges, this paper investigated the use of different forensic tools for recovering data files from a restored Android mobile phone. The data was extracted using different forensic tools, namely AccessData FTK and Backtrack dd, on the physical image acquisition of the device. The focus of this paper was to investigate the data recovery functions of Photo Image Carver, AccessData FTK, Foremost, Recover My Files and DiskDigger in forensic investigation of FTK and DD images from smartphone mobile device. The study revealed that .dd images compare more favorably for android mobile forensic investigation than FTK images, judging from the size of evidence it holds. Moreover, the experiment revealed that Phone image carver recovered some file types including SQLite, SWF Flash, and Text-Shift JIS, which other tools could not recover. Phone Image Carver keeps log file record of activities performed in it, giving the analyst the opportunity to review his previous steps each time he uses the tool. Foremost proves to recover more number of files data than Phone Image Carver and AccessData FTK. On the other hand, Recover My Files and DiskDigger had greater percentage recovery performance than Foremost, Phone Image Carver, and AccessData FTK. Both Recover My Files and DiskDigger recovered many data file formats suggesting they had a deep penetration recovery capability. However, Recover My Files recovered more files of type mov, zip, JPEG, and MYOB than DiskDigger recovered. Recover My Files proves to recover the greatest percentage of evidence by recovering 3GP, AMR audio, avi, itunes, Myob, ogg, thumbnails, and truetype files, which no other tools recovered. Therefore, Recover My Files proves to be the best recovery tool in this study.

In conclusion, the analysis tools used in this experiment showed different levels of recovery performance. Most of the tools recovered major file formats that other tools did not recover, suggesting that no single forensic tool could recover all forensic evidences in a smartphone image. In future, further similar studies are suggested to be conducted on other mobile platforms such as iPhone and compare and contrast results with those presented in this paper.

References

[1] D. Bennett, The challenges facing computer forensics investigators in obtaining information from mobile devices for use in criminal investigations, Inf. Secur. J. Glob. Perspect. 21 (3) (2012) 159–168.

[2] A. Netstar, Research says shipments of smartphones grew 74 percent in 2010, European Communications, 10 March 2011, Available from: http://www.eurocomms.com/industry-news/7655-research-says-shipments-of-smarthttp://www.eurocomms.com/industry-news/7655-research-says-shipments-of-smartphones-grew-74-percent-in-2010phones-grew-74-percent-in-2010.

[3] Berg Insight, Research says shipments of smartphones grew 74 percent in 2010, News Archive, 2011.

[4] Berg Insight, Research says shipments of smartphones grew 74 percent in 2010, European Communications. Available from: http://www.eurocomms.com/industry-news/7655-research-says-shipments-of-smartphones-grew-74- percent-in-2010, 2011 (accessed 28.12.15).

[5] J.D. La Fuente, J. Santiago, A. Román, C. Dumitrache, D. Casasanto, When you think about it, your past is in front of you: how culture shapes spatial conceptions of time, Psychol. Sci. 25 (9) (2014) 1682–1690.

[6] Data I, Idc T, Quarterly W, Phone M, IDC: smartphone shipments to overtake feature phones worldwide in 2013, Available online at: http://thenextweb.com/insider/2013/03/04/idc-smartphone-shipments-to-overtake- feature-phones-worldwide-in-2013/#gref.

[7] A. Simão, F. Sícoli, L. Melo, F. Deus, R. Sousa Júnior, Acquisition and analysis of digital evidence in Android smartphones, Int. J. Forensic Comput. Sci. 6 (1) (2011) 28–43.

[8] B.D. Dilworth, Worldwide smartphone sales to Hit 1.5 billion in 2017: IHS report, 2017.

[9] Smartphone users worldwide will total mobile users pick up smartphones as they become more best practices in digital video advertising go beyond the articles: hear from our clients: want to learn more? Available from: http://www.emarketer.com/Article/Smartphone-Users-Worldwide-Will-Total-175-Billion-2014/1010536, 2014. (accessed 19.03.16).

[10] M. Neeraj, Smartphone sales 2015–2017: India will surpass the US [report], Available from: http://dazehttp://dazeinfo.com/2015/07/04/global-smartphone-sales-2015-2017-india-will-surpass-us-report/info.com/2015/07/04/global-smartphone-sales-2015-2017-india-will-surpass-us-report/, 2015 (accessed 19.03.16).

[11] Gartner says worldwide device shipments to grow 1.5 percent, to reach 2.5 billion units in 2015, Newsroom.
Available from: http://www.gartner.com/newsroom/id/3088221, 2015 (accessed 19.03.16).

[12] J. Gozalvez, Advances in wireless power transfer, IEEE Veh. Technol. Mag. 10 (2015) 14–32.

[13] Roland T, Trend B, Roland Berger Trend Compendium 2030 – Trend 2 Globalization and Future Markets, 2014, pp. 1–29, Retrieved July 27, 2016, from: https://www.rolandberger.com/publications/publication_pdf/ roland_berger_trend_compendium_2030_trend_2_globalization_and_future_markets_20140501.pdf.

[14] P. Woodgate, I. Coppa, N. Hart, Global Outlook 2014: Spatial Information Industry, Australia and New Zealand Cooperative Research Centre for Spatial Information, 2014, pp. 1–25.

[15] R. Reynolds, Trends influencing the growth of digital textbooks in US higher education, Publ. Res. Q. 27 (2) (2011) 178–187.

[16] L. Aouad, T. Kechadi, J. Trentesaux, An open framework for smartphone, in: G. Peterson, S. Shenoi (Eds.), Advances in Digital Forensics VIII, IFIP AICT, 383, 2012, pp. 159–166. IFIP International Conference on Digital Forensics, Springer, Berlin, Heidelberg.

[17] N. Dezfouli Farhood, A. Dehghantanha, R. Mahmod, N.F.B. Mohd S, S.B. Shamsuddin, A data-centric model for smartphone security, Int. J. Adv. Comput. Technol. 5 (9) (2013) 9–17.

[18] S.H. Mohtasebi, A. Dehghantanha, Towards a unified forensic investigation framework of smartphones, Int. J. Comput. Theory Eng. 5 (2) (2013) 351–355.

[19] A. Hoog, Android and mobile forensics, in: J. McCash (Ed.), Android Forensics: Investigation, Analysis and Mobile Security for Google Android, 1st ed., Elsevier, Amsterdam, 2011, pp. 1–40 (Chapter 1).

[20] CBC News, Cybercrime moving to smartphones and tablets, say experts [Internet]. CBC News, 2013, Available from: http://www.cbc.ca/news/canada/manitoba/cybercrime-moving-to-smartphones-and-tablets-say- experts-1.1877058 (accessed 28.12.2015).

[21] A. Savoldi, P. Gubian, I. Echizen, A comparison between windows mobile and Symbian S60 embedded forenhttp://refhub.elsevier.com/B978-0-12-805303-4.00007-1/rf0085sics, in: IIH-MSP 2009–2009 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2009, pp. 546–550.

[22] L. Ablon, M.C. Libicki, A.A. Golay, Markets for Cybercrime Tools and Stolen Data: Hacker's Bazaar, National Security Research Division, Santa Monica, CA, 2014.

[23] S. Mohtasebi, A. Dehghantanha, A mitigation approach to the privacy and malware threats of social network services, in: V. Snasel, J. Platos, E. El-Qawasmeh (Eds.), Digital Information Processing and Communications, Springer, Berlin, 2011, pp. 448–459.

[24] F. Sabena, A. Dehghantanha, P.S. Andrew, A review of vulnerabilities in identity management using biometrics, in: Proc. 2nd International Conference on Future Networks, ICFN 2010, 2010, pp. 42–49.

[25] K. Nicolai, R. Carsten, A. Aaron, E.-P. Barbara, C. John, K. Thomas, On the creation of reliable digital evidence, in: Advances in Digital Forensics VIII, IFIP AICT 383, 2012, pp. 3–17.

[26] A. Rick, B. Sam, J. Wayne, Guidelines on Mobile Device Forensics, National Institute of Standards and Technology, United States, 2014, pp. 15–43.

[27] M. Damshenas, A. Dehghantanha, R. Mahmoud, A survey on digital forensics trends, Int. J. Cyber-Secur. Digit. Forensic 3 (4) (2014), 209–234.

[28] A. Dehghantanha, K. Franke, Privacy-respecting digital investigation, in: 2014 12th Annual International Conference on Privacy, Security and Trust (PST), 2014, pp. 129–138.

[29] A. Aminnezhad, A. Dehghantanha, M. Abdullah, A survey on privacy issues in digital forensics, Int. J. CyberSecur. Digit. Forensic 1 (4) (2012) 311–323.

[30] K. Shaerpour, A. Dehghantanha, R. Mahmod, Trends in Android malware detection, J. Digit. Forensic Secur. Law 8 (3) (2013) 21–40.

[31] F.N. Dezfouli, A. Dehghantanha, R. Mahmod, N.F.B.M. Sani, S.B. Shamsuddin, F. Daryabar, A survey on malhttp://refhub.elsevier.com/B978-0-12-805303-4.00007-1/rf0130ware analysis and detection techniques, Int. J. Adv. Comput. Technol. 5 (2013) 42–51.

[32] J. Aycock, Computer viruses and malware, Advances in Information Security, vol. 22, Springer, US, 2006, pp. 25–29.

[33] K. Barmpatsalou, D. Damopoulos, G. Kambourakis, V. Katos, A critical review of 7 years of mobile device fohttp://refhub.elsevier.com/B978-0-12-805303-4.00007-1/rf0140rensics, Digit. Investig. 10 (4) (2013) 323–349.

[34] J. Buchanan-Wollaston, T. Storer, W. Glisson, Comparison of the data recovery function of forensic tools, in: G. Peterson, S. Shenoi (Eds.), Advances in Digital Forensics IX, 9th IFIP WG 11.9 International Conference on Digital Forensics, IFIP Advances in Information and Communication Technology, vol. 410, Springer, Berlin, 2013, pp. 331–347.

[35] G. Grispos, T. Storer, W.B. Glisson, A comparison of forensic EVIDENCE recovery techniques for a windows mobile smart phone, Digit. Investig. 8 (1) (2011) 23–36.

[36] M. Damshenas, A. Dehghantanha, R. Mahmoud, A survey on malware propagation, analysis and detection, Int. J. Cyber-Secur. Digit. Forensic 2 (4) (2013) 10–29.

[37] C. Tassone, B. Martini, K.-K.R. Choo, J. Slay, Mobile device forensics: a snapshot, Trends Issues Crime Crim. Justice 460 (2013) 1–7.

[38] W.B. Glisson, T. Storer, J. Buchanan-Wollaston, An empirical comparison of data recovered from mobile forensic toolkits, Digit. Investig. 10 (1) (2013) 44–55.

[39] NIST, Have your computer forensics tools been tested? Computer Forensics Tool Testing Handbook, National Institute of Standards and Technology (NIST), 2015.

[40] R. Ayers, W. Jansen, S. Brothers, Guidelines on mobile device forensics, (NIST Special Publication 800-101 Revision 1) 2014, 85 pp.

[41] B.-W. Joe, S. Tim, G. William, Comparison of the data recovery function of forensic tools, in: Advances in Digital Forensics IX, IFIP AICT 410, 2013, pp. 331–347.

[42] M. Sidheeq, A. Dehghantanha, G. Kananparan, Utilizing trusted platform module to mitigate botnet attacks, in: 2010 International Conference on Computer Applications and Industrial Electronics (ICCAIE), 2010, pp. 245–249.

[43] K. Shaerpour, A. Dehghantanha, R. Mahmod, Virtualized honeynet intrusion prevention system in Scada, in:
The International Conference on E-Technologies and Business on the Web (EBW2013), 2013, pp. 11–15.

[44] M.M.N. Umale, P.A.B. Deshmukh, P.M.D. Tambhakhe, Mobile phone forensics challenges and tools classificahttp://refhub.elsevier.com/B978-0-12-805303-4.00007-1/rf0190tion: a review, Int. J. Recent Innov. Trends Comput. Commun. 2 (2014) 622–626.

[45] M. Mulazzani, M. Huber, E. Weippl, Data visualization for social network forensics, in: G. Peterson, S. Shenoi (Eds.), Advances in Digital Forensics VIII, IFIP AICT, 383, Springer, Berlin, 2012, pp. 115–126. IFIP International Federation for Information Processing 2012.

[46] M. Damshenas, A. Dehghantanha, R. Mahmoud, S. Bin Shamsuddin, Forensics investigation challenges in cloud computing environments, in: Proc. 2012 Int. Conf. Cyber Secur. Cyber Warf. Digit. Forensic, CyberSec, 2012, pp. 190–194.

[47] F.N. Dezfouli, A. Dehghantanha, R. Mahmoud, N.F.B.M. Sani, S.bin. Shamsuddin, Volatile memory acquisihttp://refhub.elsevier.com/B978-0-12-805303-4.00007-1/rf0205tion using backup for forensic investigation, in: Proc. 2012 Int. Conf. Cyber Secur. Cyber Warf. Digit. Forensic, CyberSec, 2012, pp. 186–189.

[48] N. Son, Y. Lee, D. Kim, J.I. James, S. Lee, K. Lee, A study of user data integrity during acquisition of Android devices, Digit. Investig. 10 (Suppl.) (2013) S3–S11.

[49] N. Kuntze, C. Rudolph, A. Alva, B. Endicott-popovsky, J. Christiansen, T. Kemmerich, On the creation of reliable digital evidence, in: G. Peterson, S. Shenoi (Eds.), Advances in Digital Forensics VIII, IFIP AICT, 383, Springer, Berlin, 2012, pp. 3–17. IFIP International Federation for Information Processing 2012.

[50] J. Dykstra, A.T. Sherman, Acquiring forensic evidence from infrastructure-as-a-service cloud computing: exploring and evaluating tools, trust, and techniques, Digit. Investig. 9 (2012) S90–S98.

[51] N. Borhan, A. Dehghantanha, A Framework of TPM, SVM and boot control for securing forensic logs, Int. J. Comput. Appl. 50 (13) (2012) 15–19.

[52] R. Mercuri, Criminal defense challenges in computer forensics, in: S. Goel (Ed.), Digit Forensics and Cyber Crime, First International ICST Conference, ICDF2C 2009, LNICST, vol. 31, Springer, Berlin, 2010, pp. 132–138.

[53] M. Al Marzougy, I. Baggili, A. Marrington, BlackBerry PlayBook backup forensic analysis, in: M. Rogers, K.C. Seigfried-Spellar (Eds.), Digit Forensics and Cyber Crime, 4th International Conference, ICDF2C 2012, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 114, Springer, Berlin, 2013, pp. 239–252.

[54] S. Vandeven, B. Filkins, Forensic images: for your viewing pleasure, Inf. Secur. (2014) 1–38.

[55] Q. Do, B. Martini, K.-K.R. Choo, A cloud-focused mobile forensics methodology, IEEE Cloud Comput. 2 (4) (2015) 60–65.

[56] M. Mulazzani, S. Neuner, P. Kieseberg, M. Huber, S. Schrittwieser, E. Weippl, Quantifying windows file slack size and stability, in: G. Peterson, S. Shenoi (Eds.), Advances in Digital Forensics IX, 9th IFIP WG 11.9 International Conference on Digital Forensics, IFIP Advances in Information and Communication Technology, vol. 410, Springer, Berlin, 2013, pp. 183–193.

[57] M. Damshenas, A. Dehghantanha, R. Mahmoud, S. Shamsuddin, Cloud computing and conflicts with digital forensic investigation, Int. J. Digit. Content Technol. Appl. 7 (9) (2013) 543–553.

[58] J. Shetty, J. Adibi, The Enron Email Dataset—Database Schema and Brief Statistical Report, Inf. Sci. Inst. Tech. Report 2004.

[59] S.P. Hotel, N. Beach, Workshop on Link Analysis, Counterterrorism and Security at the SIAM International Conference on Data Mining Workshop on Link Analysis, Counterterrorism and Security, 2005.

[60] J. Shetty, J. Adibi, Discovering important nodes through graph entropy: the case of Enron email database, in: Proc. 3rd Int. Work Link Discov., 2005, pp. 74–81.

[61] J. Diesner, T.L. Frantz, K.M. Carley, Communication networks from the Enron email corpus "it's always about the people. Enron is no different", Comput. Math. Organ. Theory 11 (2006) 201–228 (abstract).

[62] B. Klimt, Y. Yang, The Enron Corpus: a new dataset for Email classification research, in: J.-F. Boulicaut, E. Floriana, G. Fosca, P. Dino (Eds.), European Conference on Machine Learning, Vol. 3201, Springer, Berlin, Heidelberg, 2004 September 20, pp. 217–226.

[63] J. Diesner, K.M. Carley, Exploration of communication networks from the Enron email corpus, in: Int. Conf. Data Mining Workshop Link Anal. Counterterrorism Secur., 2005.

[64] Y. Zhou, K.R. Fleischmann, W.A. Wallace, Automatic text analysis of values in the Enron email dataset: clusterhttp://refhub.elsevier.com/B978-0-12-805303-4.00007-1/rf0290ing a social network using the value patterns of actors, in: Proc. Annu. Hawaii Int. Conf. Syst. Sci., 2010.

[65] M. Al Marzougy, B. Ibrahim, M. Andrew, BlackBerry Playbook backup forensic analysis, in: M. Rogers, K.C. Seigfried-Spellar (Eds.), ICDF2C 2012, LNICST 114, 2013, pp. 239–252.

[66] F. Cruz, A. Moser, M. Cohen, A scalable file based data store for forensic analysis, Digit. Investig. 12 (2015)
S90–S101.

END