IJCSIS Vol. 11 No. 1, January 2013
ISSN 1947-5500

International Journal of Computer Science & Information Security

Cornell University Library

Cogprints

Google scholar

.docstoc
find and share professional documents

ScientificCommons

View my documents on
Scribd

BASE
Bielefeld Academic Search Engine

SCIRUS
search engine for science

SciRate.com

CiteSeerx beta

dblp.uni-trier.de
Computer Science
Bibliography

Q-Sensei BETA

DOAJ DIRECTORY OF OPEN ACCESS JOURNALS

EBSCO HOST

ProQuest

Editorial
Message from Managing Editor

*Since May 2009, the **International Journal of Computer Science and Information Security** (IJCSIS), has been promoting the dissemination of knowledge in research areas of computer applications and practices, and advances in information security. The themes focus mainly on innovative developments, research issues/solutions in computer science and related technologies.*

IJCSIS archives publications; abstracting/indexing, editorial board and other important information are available online on homepage. IJCSIS editorial board consisting of reputable experts solicits your contribution to the journal with your research papers, projects, surveying works and industrial experiences. IJCSIS appreciates all the insights and advice from authors and reviewers. Indexed by the following International agencies and institutions: Google Scholar, Bielefeld Academic Search Engine (BASE), CiteSeerX, SCIRUS, Cornell's University Library EI, Scopus, DBLP, DOI, ProQuest, EBSCO. Google Scholar reported a large amount of cited papers published in IJCSIS.

IJCSIS is currently accepting manuscripts for upcoming issues based on original qualitative or quantitative research, an innovative conceptual framework, or a substantial literature review that opens new areas of inquiry and investigation in Computer science. Case studies and works of literary analysis are also welcome.

*We look forward to your collaboration. For further questions please do not hesitate to contact us at **ijcsiseditor@gmail.com**.*

A complete list of journals can be found at:
http://sites.google.com/site/ijcsis/

IJCSIS Vol. 11, No. 1, January 2013 Edition

ISSN 1947-5500 © IJCSIS, USA.

Journal Indexed by (among others):

IJCSIS EDITORIAL BOARD

TABLE OF CONTENTS

Neetu Narwal, Asst. Prof., Maharaja Surajmal Institute, Affiliate College of GGSIP University & Research Scholar, Lingayas University
Dr. Mayank Singh, Ass. Prof., Research Guide, Lingayas University

Abstract — Internet comprises of huge volume of data but the information it contains, are highly unstructured. There exists a lot of algorithm for extraction and identification of main content of the web page, as along with the main content there exist an irrelevant data, which is of little significance to the user and often distracting in many cases. Web page with only relevant information is of greater significance to the user. We have proposed an algorithm to extract the main contents of web page and store the contents in XML file, which can be utilized for useful purpose.

Keywords- Web Data Extraction, DOM tree, Visual Cues, XML

Asia Samreen, Department of Computer Science, Bahria University, Karachi Campus, Karachi, Pakistan
Gulnaz Ahmed, Department of Computer Science, Bahria University, Karachi Campus, Karachi, Pakistan

Abstract — Wireless sensor Networks are very famous these days due to their coverage and enormous benefits. Cluster making efficiently and dynamically for such networks as to increase the life time of network nodes, is the question in front of researchers. Cooperation among Relay nodes and Edge nodes along with controlled energy consumption requires an efficient way such as LEACH for communication. We have revealed various factors to elaborate the issues relevant to WSNs for instance Cluster-head selection, low data reception rate, relay node placement and number of relay node assignment for optimal power usage.

Keywords-Cluster-head; Relay node placement; Relay node; Edge node; Power usage; LEACH

Neetu Narwal, Asst. Prof., Maharaja Surajmal Institute, Affiliate College of GGSIP University & Research Scholar, Lingayas University
Dr. Mayank Singh, Ass. Prof., Research Guide, Lingayas University

Abstract - Each day, the WWW grows by nearly a million electronic pages, adding to the hundreds of millions previously on-line. WWW is a platform for swapping numerous types of information, ranging from research papers, and learning contents, to audio-visual content and various software's [4]. Mining log data is a very comprehensive research area developed to solve the issues related to usage of computer and internet on single computer or on network of computers. This article provides a analysis of monitoring strategy for computer and internet usage of teenage computer.

Keywords: Usage Mining, Mining Technologies, Monitoring Strategy

4. Paper 31071240: Design and Implementation of Security Framework for Cognitive Radio Networks Resource Management (pp. 15-29)

Obeten O. Ekabua and Ifeoma U. Ohaeri
Department of Computer Science, North-West University, Mafikeng Campus, Private Bag X2046, Mmabatho 2735, South Africa

Abstract: Designing and implementing a secure communication for any network is an important issue for the optimal control of resource usage in a resource constrain network environment. Therefore, in this paper, we design and implement a joint authentication and authorization framework by transforming the framework requirement analysis. The framework is a security infrastructure capable of monitoring and controlling access to the limited spectrum resources, dynamically managing data and information in CRN, for a secured communication and quality of service (QOS). We explained how the various components in the framework interact to ensure a secured communication and effective access control.

Keywords: Network Management, Security, Authentication, Authorization, Access Control.

5. Paper 25121204: Cyber Crimes Analysis Based-On Open Source Digital Forensics Tools (pp. 30-43)

Victor O. Waziri PhD, Department of Cyber Security Science; School of Information and Communication Technology; Federal University of Technology, Minna-Nigeria
Okongwu N. O, Economic and Financial Crimes Commission, Nigeria
Audu Isah PhD, Department of Mathematics/ Statistics, School of Federal University of Technology, Minna-Nigeria
Olawale S. Adebayo, Department of Cyber Security Science; School of Information and Communication Technology; Federal University of Technology, Minna-Nigeria
Shafi'í Mohammed Abdulhamid, Department of Cyber Security Science; School of Information and Communication Technology; Federal University of Technology, Minna-Nigeria

Abstract - In this paper, we are present the digital forensic open source tools: Fiwalk, Bulk_Extractor, Foremost, Sleuth Kit, and Autopsy which are all Linux based forensic tools to extract evidences that could be presented in the court of law. Fiwalk reads a disk image and outputs a block of XML containing all the disk image of resident and deleted files. Foremost recovers files by using their headers, footers and data structures. The Sleuth Kit and Autopsy perform various aspects of file system analysis. The Autopsy Forensic Browser is a graphical web interface that presents the results generated by Sleuth Kit. This research project demonstrates the usefulness of the above-mentioned forensic tools for analysis and recovery of obliterated data from hard drives. This paper found that Sleuth Kit, Autopsy Forensic Browser, Fiwalk, Bulk_Extractor, and Foremost all provide effective file system analysis and recovery tool sets. The increasing complexity of storage devices requires that the investigator employs different forensic tool set to complement his arsenal of tools. No single digital forensic tool would be sufficient for an entire digital forensic investigation case. With this consideration, this paper employs various forensic tools. The demonstration of the effectiveness of these digital forensic tools utilized in this paper could serve as an alternative for investigators looking to expand their digital forensic tool set functionality in the court of law. Details of the experiments are fully given at the expense of bulkiness since this works is aim at enhancing the utilities of open source forensics tools applications.

Keywords: Digital Forensics, Fiwalk, Foremost, Sleut Kits Bulk_Extractor, Autopsy, Linux, Ontologies

Web Content Extraction A Heuristic Approach

Neetu Narwal
Asst. Prof., Maharaja Surajmal Institute
Affiliate College of GGSIP University
Research Scholar, Lingayas University
neetunarwal@gmail.com

Dr. Mayank Singh
Ass. Prof., Research Guide, Lingayas University
mayanksingh2005@gmail.com

Abstract— **Internet comprises of huge volume of data but the information it contains, are highly unstructured. There exists a lot of algorithm for extraction and identification of main content of the web page, as along with the main content there exist an irrelevant data, which is of little significance to the user and often distracting in many cases. Web page with only relevant information is of greater significance to the user. We have proposed an algorithm to extract the main contents of web page and store the contents in XML file, which can be utilized for useful purpose.**

Keywords- Web Data Extraction, DOM tree, Visual Cues, XML

I. Introduction

World Wide Web is considered as largest information repository presented in the form of web documents. With the advent of different programming languages for web development viz., XML, ASP, PHP etc. these are highly typed languages, but there still exist a high percentage of web sites designed in HTML. The HTML is the simplest markup language designed for data representation, but on the other hand it makes the web page highly unstructured. The HTML Web Page designer can design web page by ignoring syntactic rules (i.e. ignoring closing tags, wrong nested tags, wrong parameters and incorrect parameter values) thus making the web page highly unstructured.

A web page contains various types of information represented in different forms such as text, image, video or audio. The Business web sites are mostly dynamic, these are designed using technologies i.e., JSP, PHP, Javascript, VBScript, ASP which extracts information at runtime from the server databases. as, along with the useful information there exist lots of unwanted information [1], which is of little significance to the user.

A web page contains navigation bar, advertisements, contact information etc. that are not related to the content of the web page and sometimes it becomes distracting to the user. Advertisements are the source of revenue for most of the websites but it becomes overhead for the web user who has to scroll the web page to read the relevant content as most part of the web page are occupied by these advertisements and links to different pages. These links at different portion of a page usually contribute to the page rank or HITS.

The specific aim of the user while accessing the web site is to view the relevant information he/she wishes to acquire and the first look of the web page must satisfy the user with his/her intent so that user doesn't have to scroll much and jump from one link to another in order to collect the relevant information. Most Information extraction system consider web as the smallest and undividable unit, there exist varied Information Extraction System that incorporate approaches like manual, supervised and automatic learning. The manual method has high accuracy level but it is time consuming, whereas supervised learning is a semi-automatic approach, which involves human interaction to decide positive and negative result. Automatic approach [2] have less human interaction but at the same time these are lesser reliable. But, a trusted Automatic Information Extraction Technique is the need of the hour.

The web page segmentation is classified as Top down or Bottom up approach, and they consider various methods to extract the web page contents based on

§ DOM Tree method

§ Web Page Layout method

§ Visual Characteristics

We have proposed a Hybrid algorithm, which takes web page as input and by making use of DOM tree and Visual cues of the web page i.e, Alignment, Font size and Font color etc. extracts the contents of the web page. The extracted visual blocks are stored in the xml file inside node tag as objects, which can be utilized for further applications.

Extraction of useful information from the web pages has many applications like mobile phone adaptation, extraction of useful contents for visually impaired or text summarization.

We have implemented the algorithm in Javascript and PHP and the output is stored in the XML and text file. We have conducted the experiment on 30 website and found the result with high precision.

The rest of the paper is organized as follows. In Section 2 describes the background and the related work done in the area of web information extraction. In Section 3 we have described the approach of the proposed system for Information Extraction. In Section 4 we have displayed the algorithm of the proposed system. Section 5 demonstrates the application and experimentation of developed algorithm on the web pages. Finally the conclusion is mentioned in Section 6.

II. Previous Work And Background

Web data extraction system is a sequence of steps that extracts the content of the web page by incorporating different approaches like manual, supervised learning and automatic learning. Web page segmentation technique makes use of either top down or bottom up approach, by incorporating the use of various methods to extract the web contents i.e., DOM Tree method, Web Page Layout method and Visual Characteristics.

A. Top-down Web Page Segmentation Approach

This approach begins by considering the complete web page as a block and then partitions the block iteratively into smaller blocks using different features obtained from the web contents. Partitioning decision is based on the DOM Tree, Page Layout and Visual Characteristics.

DOM based web extraction method, make use of the Document Object Model (,http://www.w3.org/DOM/) described by W3 Consortium. DOM Document is a collection of nodes arranged in a hierarchy, which allows a developer to navigate through the tree to extract the information.

The author[4] proposed a Web page segmentation method of segmenting a Web page into logical blocks and then classifying the segmented blocks into informative blocks that contain the page's core contents and noise blocks that contain irrelevant information such as menus, advertisements, or copyright statements.

A vision-based page segmentation (VIPS) algorithm [5] make use of the visual characteristics of web page viz. font color, font name, top margin, left margin to identify the coherence of each segment and then used it to divide the web page into semantic blocks.

Vision based Extraction of data Record (VER) algorithm [2] used VIPS algorithm to extract the blocks from the web page and then analysed the data regions to extract the data records proposed. The author [13] presents an automatic approach to extract the main content of the web page using tag tree & heuristics to filter the clutter and display the main content.

The author[6] uses information measure, to dynamically select the entropy-threshold that partitions blocks into either informative or redundant. Informative content blocks are distinguished parts of the page, whereas redundant content blocks are common parts.

Author used [8] the entropy of contained terms in the DOM Tree for segmentation, [9] also used content size and entropy value to measures the strength of local patterns within the subtree and used it for web page segmentation..

B. Bottom-up Web Page Segmentation Approach

The Bottom up approach considers the leaf-node of DOM tree as atomic unit and then using certain heuristic rules to combine the atomic unit to form blocks.

Author [10] segmented the web pages into cohesive microunits and represented it as tag tree and based on heuristic rules aggregated the nodes into segment blocks.

Using Graph theoretic approach [11], DOM nodes are considered as complete weighted graph and edge weight estimates the cost required to combine the connected nodes into one block. The author used gestalt theory[3] to capture the visual layout, content and functional aspects and implemented with semantic web techniques. The author [12] used quantitative measure of text density and used it for iterative block fusion based on computer vision.

III. Proposed Methodology

The proposed system extracts the information from the web page by considering the DOM tree structure and Visual characteristics of each node. An overview of the proposed system is shown in Fig 1. Our approach is a hybrid technique that considers semantic information and the visual cues of the web page for extracting information.

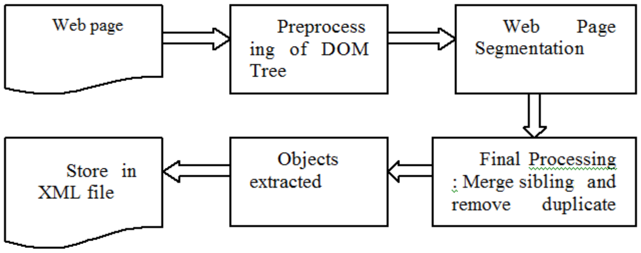
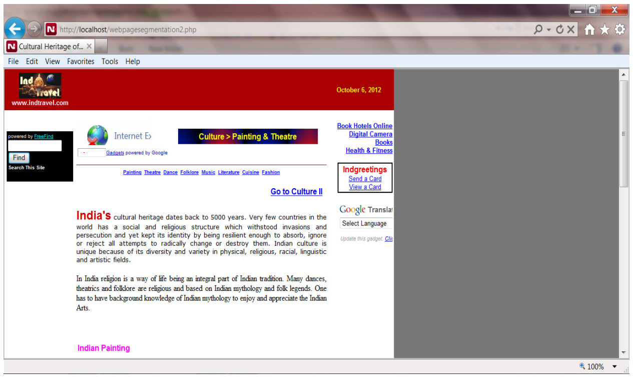

Web page
Preprocessing of DOM Tree
Web Page Segmentation
Final Processing : Merge sibling and remove duplicate
Objects extracted
Store in XML file

Fig 1: Proposed System Methodology.

A. Preprocessing (Remove Invalid Nodes)

Input to the preprocessing module is the Web page comprising of tags and texts. Some tags like <SCRIPT>, <BR>, <NOSCRIPT> etc. do not have any size dimensions so they do not contribute to the visual areas of the web page, but the tags like <SCRIPT> contains some preprocessing steps which are executed and provides some output. Preprocessing steps remove all the invalid nodes from the DOM Tree for further processing.



India's cultural heritage dates back to 5000 years. Very few countries in the world has a social and religious structure which withstood invasions and persecution and yet kept its identity by being resilient enough to absorb, ignore or reject all attempts to radically change or destroy them. Indian culture is unique because of its diversity and variety in physical, religious, racial, linguistic and artistic fields.

In India religion is a way of life being an integral part of Indian tradition. Many dances, theatrics and folklore are religious and based on Indian mythology and folk legends. One has to have background knowledge of Indian mythology to enjoy and appreciate the Indian Arts.

Fig 2: Sample Web page input to the Web Segmentation Algorithm

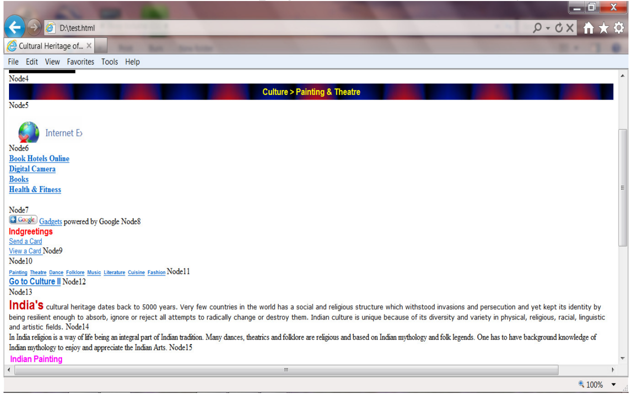
B. Web Page Segmentation

Given a Web Page S, the Web Page Segmentation algorithm partitions the web page based on semantic and visual characteristic to provides Visual Blocks Gi, with the value of i varying from 1 to n, where n is the total number of blocks extracted from the web page, Union of all these blocks combined together must provide you with complete set of information from the web page.

$$\sum_{i=1\ \text{to}\ n}^{n} G_i = S \qquad ..\ (1)$$

The DOM Tree after preprocessing will give a set of valid nodes, which may further be analyzed to eliminate redundant valid nodes or merge sibling nodes based on visual characteristics.

The final output will be valid nodes covering all the visual blocks of the main web page with no redundancy.



Node4

Culture > Painting & Theatre

Node5

Node6

Book Hotels Online

Digital Camera

Books

Health & Fitness

Node7

Google Gadgets powered by Google Node8

Indgreetings

Send a Card

View a Card Node9

Node10

Painting Theatre Dance Folklore Music Literature Cuisine Fashion Node11

Go to Culture II Node12

Node13

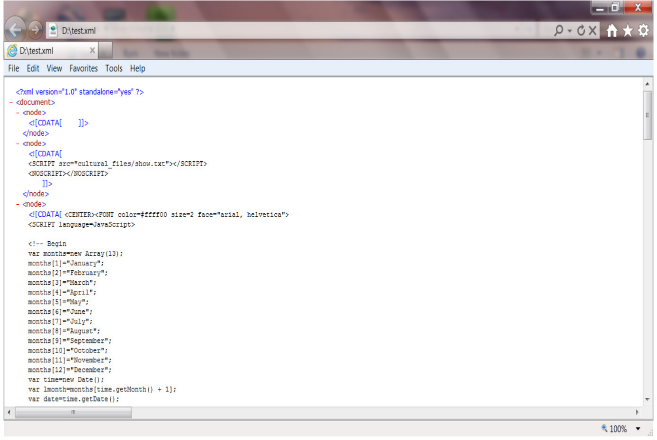
India's cultural heritage dates back to 5000 years. Very few countries in the world has a social and religious structure which withstood invasions and persecution and yet kept its identity by being resilient enough to absorb, ignore or reject all attempts to radically change or destroy them. Indian culture is unique because of its diversity and variety in physical, religious, racial, linguistic and artistic fields. Node14

In India religion is a way of life being an integral part of Indian tradition. Many dances, theatrics and folklore are religious and based on Indian mythology and folk legends. One has to have background knowledge of Indian mythology to enjoy and appreciate the Indian Arts. Node15

Indian Painting

Fig 3: Nodes extracted from the web page

Extracted nodes are stored as objects in the node tag of XML file. These nodes may be utilized further for some other useful purposes.



```
 <?xml version="1.0" standalone="yes" ?>
- <document>
 - <node>
    <![CDATA[     ]]>
   </node>
 - <node>
    <![CDATA[
    <SCRIPT src="cultural_files/show.txt"></SCRIPT>
    <NOSCRIPT></NOSCRIPT>
       ]]>
   </node>
 - <node>
    <![CDATA[ <CENTER><FONT color=#ffff00 size=2 face="arial, helvetica">
    <SCRIPT language=JavaScript>

    <!-- Begin
    var months=new Array(13);
    months[1]="January";
    months[2]="February";
    months[3]="March";
    months[4]="April";
    months[5]="May";
    months[6]="June";
    months[7]="July";
    months[8]="August";
    months[9]="September";
    months[10]="October";
    months[11]="November";
    months[12]="December";
    var time=new Date();
    var lmonth=months[time.getMonth() + 1];
    var date=time.getDate();
```

Fig 4: Extracted nodes stored in the XML file.

IV. Proposed Algorithm

Algorithm : WebPageSegmentation Algorithm(WebPage)

```
Begin
1. Accept web page as input
2. Traverse and Preprocess the DOM tree to remove invalid
   nodes considering size dimensions.
3. Remove the invalid branches of DOM tree.
4. Analyse the remaining valid nodes based on the visual
   characteristics resulting in merging of sibling nodes into a
   single valid node.
5. Traverse the valid node tree to remove duplication of
   nodes incase Parent as well as child node are marked as valid.
6. Output of the module is a set of all valid nodes covering
   the complete web page.
```

Algorithm : ConverttoXML (ValidNode tree)

```
Begin
1. Accept valid node tree as input.
2. Design the Structure of the XML documents and create
   Node tag to store objects.
3. Copy the contents of Valid Node inside the Node tag of
   XML document.
4. Save the XML file.
```

V. Experiment

We have conducted the experiment on the 50 web pages from different web sites related to varied arena i.e, commercial, university, news web sites etc. We evaluated our algorithm based on the following measures, the total number of visual blocks in the web site and the number of extracted blocks.

Based on these values we calculated precision and recall:

Precision = (Correct/Extracted)*100 .. (2)

TABLE I. Result of Selected Web Sites

URL	Precision
www.shoebuy.com/..	98%
www.indtravel.com/..	100%
www.yahoo.co.in	99%
www.indiatimes.co.in	99%
www.planetunreal.com	98%

VI. Conclusion

In this paper, we proposed an algorithm for Automatic Web Information Extraction. The methodology does not need a manual intervention or any kind of training for the Web Page Extraction algorithm. Our experiment shows a good result in terms of precision for extracting information. We finally want to conclude with the fact that our further contribution will be utilizing the information extracted from web page to be used to display web page based on the different display device namely small scale device like mobile phones, smart phones, palmtops and large scale devices like high resolution display, LCD.

REFERENCES

[1] Lan Yi, Bing Liu, Xiaoli Li, Eliminating Noisy Information in Web Pages for Data Mining , ACM 2003.

[2] Nwe Nwe Hlaing, Thi Thi Soe Nyunt, An Approach for Extraction Data Record from Web Page based on Visual Features, International Journal of Advances in Management Sciences, Aug 2011.

[3] Brnhard Krüpl-Sypien, Ruslan R. Fayzrakhmanovy, Wolfgang Holzinger, A Versatile Model for Web Page Representation, Information Extraction and Content Re-authoring, Mathias Panzenböck, Inst. of Information Systems,DBAI Group, TU Wien, Austria,.

[4] Jinbeom Kang, Jaeyoung Yang, Joongmin Choi, Information Extraction, Department of Computer Science &Engineering, Hanyang University, Ansan, Korea

[5] Deng Cai1, Shipeng Yu , Ji-Rong Wen and Wei-Ying Ma, Extracting Content Structure for Web Pages based on Visual Representation, Tsinghua University, Beijing, P.R.China, Microsoft Research Asia

[6] Shian-Hua Lin, Jan-Ming Ho, Discovering Informative Content Blocks from Web Documents, Institute of Information Science, Academia Sinica

[7] G. Hattori, K. Hoashi, K. Matsumoto and F. Sugaya, Robust web page segmentation for mobile terminal using content-distances and page-layout information,

[8] H.Y. Kao, J.M. Ho and M.S. Chen : WISDOM: Web Intrapage Informative Structure Mining Based on Document Object Model.

[9] G. Vineel, Web Page DOM node characterization and its application to page segmentation. In Internet Multimedia Services Architecture and Applications(IMSAA), 2009 IEEE Conference.

[10] X.Li, B. Liu, T. Heng Phang and M.Hu, Unsing Micro Information units for Internet Search. In Proc. Of ACM 11th International Conf. on Information and Knowledge Management.

[11] D. Chakrabarti, R.Kumar and K. Punera. A graph theoretic approach to web page segmentation. In Proc. Of 17th International Conference on Workd Wide web Conf. (ACM Press 2003)

[12] C. Kohlschutter and W. Nejdl. A densitometric approach to web page segmentation. In Proc. Of 17th ACM conference on Information and Knowledge management, 2008.

AUTHORS PROFILE

Neetu Narwal, is Research Scholar, working as Asst. Prof. in the Department of Computer Science Maharaja Surajmal Institute, Affiliate College of GGSIP University, New Delhi. She is MCA and currently pursuing Phd.(Computer Application) from Lingayas University .

Dr. Mayank Singh is Associate Professor in the Department of Computer Application, in Lingayas University. He has done his M.E in Software engineering from Thapar University and PhD from Uttarakhand Technical University. His Research areas are Software Engineering, Software Testing, Wireless Sensor Networks and Data Mining.

On Issues of Relay Nodes Assignment

Asia Samreen
Department of Computer Science
Bahria University, Karachi Campus
Karachi, Pakistan
asia.samreen@bimcs.edu.pk

Gulnaz Ahmed
Department of Computer Science
Bahria University, Karachi Campus
Karachi, Pakistan
gulnaz-ahmed87@yahoo.com

***Abstract*— Wireless sensor Networks are very famous these days due to their coverage and enormous benefits. Cluster making efficiently and dynamically for such networks as to increase the life time of network nodes, is the question in front of researchers. Cooperation among Relay nodes and Edge nodes along with controlled energy consumption requires an efficient way such as LEACH for communication. We have revealed various factors to elaborate the issues relevant to WSNs for instance Cluster-head selection, low data reception rate, relay node placement and number of relay node assignment for optimal power usage.**

Keywords-Cluster-head; Relay node placement; Relay node; Edge node; Power usage; LEACH

I. INTRODUCTION

Wireless Sensor Networks (WSNs) are a novel class of wireless communication network, which combine communication technology, embedded computing and sensor technology [6]. It consists of a large number of spatially distributed autonomous sensor nodes which are connected to each other through wireless medium to monitor different environmental and physical conditions such as sound, vibration, pressure, temperature, etc. and cooperatively pass this information to a main station which is called Base Station (BS). The modern wireless sensor networks are bidirectional and enable to control sensor activity. In the past, military applications such as battlefield surveillance, secure information communication etc. becomes the motivational step to introduce the concept of wireless sensor networks but now a dayȏs Sensor Network is made of from a few to several hundred or may be of thousand " sensor nodesȍ, where each sensor node cooperates with one (or sometimes several) other sensor nodes. The importance of such network will be increased if battery time of these nodes improves employing some efficient techniques. Recent research emphasizes on various issues such as fast transmission of messages, ad-hoc network lifetime and relay nodes placement along with efficient packet transmission across the network. These types of networks are used in different industrial and consumer applications. A Wireless Sensor node normally consists of several parts: a radio transceiver for communication with an external antenna, a microcontroller consisting an embedded electronic circuit for interfacing with the sensors and an energy source, and usually a battery.

Due to limitations of Size and cost sensor nodes result in corresponding constraints on resources such as energy, memory, computational speed and communications bandwidth etc. Sensor nodes are usually powered by battery, so how efficiently and rationally can use energy to extend the network lifetime as much as possible has become one of the core issues of sensor networks [6]. Network routing is generally used to increase the wireless sensor network's lifetime and to make efficient communication. Clustering does help to solve this problem.

The goal of this paper is to discuss new parameters required to increase network lifetime and issues in WSN in detail. Different WSN related issues and approaches to solve those issues addressed by different researchers are also discussed in this paper.

The rest of the paper is divided as follows. In section 2, overview of WSN, Relay and clustering is given in detail. Section 3 is fixed for issues and problems while in section 4 approaches to tackle those issues are discussed. In section 5 conclusion and future work is given.

II. OVERVIEW OF WIRELESS SENSOR NETWORKS

The advantages of *Multiple-input and Multiple-output* (MIMO) systems have been widely seen from the last few decades. Generally cooperation is used to achieve transmit diversity.

It is clearly beneficial for cellular systems; it may not be practical for other scenarios. As we know that size and transmitted power (lifetime) of wireless ad hoc networks are limited in nature which causes a main problem to support multiple transmit antennas. *Cooperative Communication* (CC) makes possible that we can use single antenna mobiles in a multi-user environment by sharing their antennas to generate a virtual multiple-antenna transmitter to achieve transmit diversity. In cooperative wireless communication, we deal with a wireless network of the both cellular and ad-hoc variety, where the wireless agents, which are called users, can increase their quality of service (block error rates or outage probability) with the help of cooperation [1]. Cooperative communication differs in some aspects from the basic relay channel. A basic relay channel is a single-source (S) multiple relay single-destination (d) network while the study of literature shows that in cooperative communication more general cases with multiple sources and multiple relays are under consideration as shown in Figure 1[1].

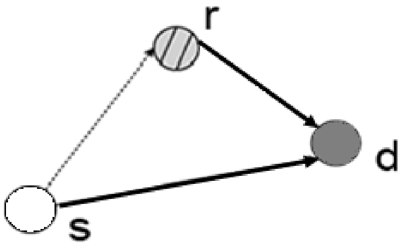


Figure 1: Single source relay channel

Clustering is a very effective method to build a hierarchical architecture in mobile ad-hoc networks. Different clustering schemes are discussed for mobile ad-hoc networks in [13], to meet certain needs of the system such as cost maintenance, to make system energy efficient, for load balancing to distribute overhead of a network. Cluster based organizations of wireless sensor networks have been identified as the best method of sensor organization for reducing the energy consumption of a WSN [11]. Normally there are three types of nodes in cluster networks, Cluster-Head nodes, gateway or sometimes called Relay nodes and normal nodes or member nodes. Cluster-Head nodes are in charge of clusters and receive joint requests from different normal nodes, where normal nodes join a cluster if Cluster-Head node accepts those requests and controlled by a choosing Cluster-Head.

III. Issues and problems

In this section we discuss the issues and problems that affect the performance of wireless networks.

A. Network Lifetime extension Problem:

As we know that every wireless device is limited in size, cost, transmitted power or hardware complexity to one antenna. Due to all these factors mention above the lifetime of a wireless sensor network become limited which is a measure interest. Size and transmitted power (lifetime) of wireless ad hoc networks are limited in nature which cause a main problem due to which working of these networks for a longer duration is not possible [1], [6],[9],[10], [33].

B. Network Bandwidth Extension problem:

One important issue is the use of limited bandwidth in wireless ad-hoc networks [3]. Coordination phase, in the cooperative communication decreases the overall bandwidth efficiency. With the passage of time the increase in traffic has increased the pollution and the accumulation of traffic near junction has caused huge amount of fuel wastage. To overcome this we form vehicle ad hoc networks where security and density estimation is a big problem for the deployment of this network [8].

C. Relay Node placement problem:

Wireless and Sensor Networks have been of great interest from last few years. Researchers are focusing on the problems related to such networks due to limited battery time and structure free architecture of WSNs. Specially, the focused issue nowadays is, how to use such networks efficiently saving the energy with high performance. In this regard an important issue of measure interest is the best place for Relay nodes in the sensing field (RNP).

1) Low data reception rate problem:

We use sensors and transducers called edge nodes to get information. Edge nodes are deployed at some positions which have limited sensing and transmitting power, thus they have short lifetime. A base station is also needed to collect data from these nodes. However, due to geographic problem, sometimes this is not possible to place BS near to ENös which cause a problem of low data reception rate [2].

2) Number of Relay nodes placed problem:

Second problem is that what number of Relay nodes is used in a network to satisfy a specific requirement(s), such as connectivity or survivability and on which candidate locations to gain maximum energy potential to increase network lifetime [10].

D. Relay node Assignment Problem:

Optimal relay node assignment is the focused issue nowadays to use energy resources efficiently with high performance. [4], mainly focuses on the problem of relay assignment in cooperative networks. [5], discusses the relay node assignment Problem in cooperative ad hoc networks.

1) Resource allocation and management problem:

Classical relay channel is discussed previously but more general cases such as multiple sources and multiple relays are still lack of consideration. In above case, resource allocation and management become an important issue [4].

2) Communication problem in cooperative networks:

As there is a limited number of relay nodes therefore multiple sourceï destination pairs compete for the same Pool of relay nodes which cause a communication problem in cooperative networks [4].Here main objective is to assign the available relay nodes to different sourceï destination pairs so as To maximize the minimum data rate among all pairs and to increase network lifetime. RNA for better communication b/w RNôs and ENôs is also discussed in [2].

E. Cluster organization Problem:

Organizing cluster in different size to preserve transmitted power and to achieve higher data rates is also a big problem [6]. Residual energy in a cluster is an important issue in Heterogeneous Energy Wireless Sensor Networks for longer lifetime and the amount of effective messages of the network.

1) Load balancing problem:

Load balancing is required to distribute overhead of a network. A fair distribution of the Cluster-Heads is the big issue in wireless Sensor Networks. By fair distribution of Cluster-Heads, the number of nodes in each cluster can b balanced, which leads to fairy energy consumption of the Cluster-Heads [20].

F. Cluster-Head selection problem:

The Cluster-Head rotation selection is a critical issue in terms of extending the lifetime of the entire WSN. If a Cluster-Head selection phase is triggered with a smaller number of data transmission rounds, it will increase overhead during this phase. On the other hand if the number of data transmission rounds is large before a Cluster-Head selection phase is triggered, the Cluster-Head nodes would not have enough energy to act as ordinary sensor nodes after reestablishment the CH role [11]. In Clustering scheme different clusters are made, within each cluster, a node is elected as a Cluster- Head which is responsible for the resource assignments and cluster maintenances. The clustering algorithm and the selection criteria of the cluster head (CH) are crucial to a clustering ad hoc network [12]. Cluster-Head selection residual and consumed energies are a big issue discussed in [7], [9], [37], [16].

IV. APPROACHES & METHODOLOGIES

There are different methodologies proposed by researchers to tackle the above given problems. Some of them are given below.

A. Techniques for Network Lifetime extension:

In [1] to solve the network extension problem cooperative communication is used, which is used to achieve transmit diversity. This class of methods make possible that we can use single antenna mobiles in a multi-user environment by sharing their antennas to generate a virtual multiple-antenna transmitter to achieve transmits diversity. In [1] three methods of cooperation are used Detect and Forward method, Amplify and forward method and Coded cooperation method. In [6] to make good use of the limited energy, ant Colony optimization (ACO) was applied to inter-cluster routing mechanism. The algorithm utilized the Dynamic adaptability and optimization capabilities of the ant colony to get the optimum route between the cluster-Head**.** Ant colony optimization has a periodic round; each round is divided into cluster formation and cluster route stage. After new ñroundò start, the algorithm firstly divided the cluster, then the data will be transmitted between the CH. In the cluster formation stage, the base station firstly need to use a given transmit power to network to broadcast a signal. After receiving this signal, each sensor according to the received signal strength try to calculate the distance to the base station. [9] focuses on reducing power consumption by considering consumed energy as a factor for cluster head selection of each node to increase network life time of WSN rather than residual energy. Hence in this a new threshold formula of LEACH is proposed.

The invention of consumed energy factor, a new approach to reduce threshold increases Life time better than residual energy. The proposed formula is:

$$t(n)= ? \quad \frac{????}{?? ????\times(?\times ?\ ??\frac{?}{????})} * \tan ?\frac{????}{????}?; \quad ?$$

if n ∈ G

Where Econ=Consumed Energy.

To solve this problem [10] introduce a energy harvesting aware wireless sensor network, because the more energy the placed nodes can harvest the more effective the network can be. For both the connectivity and survivability, *we propose polynomial time constant approximation* algorithms *to* solve the problems. The algorithms aim to deploy a small number of RNs in the candidate locations, such that the overall energy harvesting potential of the RNs is high. A more desirable objective would be to maximize the overall harvesting potential. In [33] a lifetime maximization problem via cooperative nodes is considered and performance analysis for M-ary PSK modulation is provided. With aiming to maximize the minimum device lifetime under a restraint on bit-error-rate performance, the optimization problem determines which nodes should cooperate and how much power should be allocated for cooperation. Moreover, the device lifetime is further improved by a deployment of relay nodes for cooperation in order to help forward information of the general nodes in the network. Optimum location and power allocation for each cooperative relay node are determined with an aim to maximize the device lifetime. A suboptimal algorithm is developed to solve the problem with placing multiple cooperative relay nodes and cooperative nodes. The basic idea behind greedy suboptimal algorithm is to ýnd a node to be helped and a helping node step by step. In each step, the algorithm selects a node to be helped as the one with minimum lifetime span and it has never been helped by other nodes. Then, the algorithm chooses a helping node as the one that maximizes the node lifetime span after the helped node has been served. In this way, the lifetime span of a node can be increased step by step. The algorithm working stops when the node lifetime cannot be remarkable improved or all cooperative nodes have been helped once.

B. Techniques for Network Lifetime extension:

Dual carrier modulation scheme and an adequately designed cooperative space time block code technique is used in [3], for efficient use of network bandwidth. Dual carrier modulation uses two consecutive quadrature phase shift keying (QPSK) symbols to develop two differently mapped 16 quadrature amplitude modulation (16-QAM) symbols. When these two (16-QAM) symbols are carried on far-off subcarriers, which might undergo independent fading paths or might not, it can add certain level of frequency diversity in the symbols. In [8] to solve the problem of density estimation a stable clustering approach for traffic monitoring and routing is proposed. In this approach the Cluster-Head (CH) election is done based on distance and direction information. According to this algorithm each participating vehicle knows its own position using Global Positioning System (GPS). Moreover each vehicle consists digital maps which enable to determine the direction of travel, so that direction information can be computed first of all. Here, the cluster formation and Cluster Head election mainly depends on the TH_DISTANCE which is TH_DISTANCE = (LENGTHMAX+LENGHTMIN) **/ 2**

C. Relay Node placement Solutions:

To solve the low data rate problem in [2], some relay nodes are placed with fixed energy b/w the ENô̂s and the base station to forward data packet with higher data rates. Here only consider two ï hop relay routing from ENôs to the BS. RNs amplify the faded signals coming from ENs and then send these signals towards the Bs. For RNP a weighted Clustering Binary Integer Programming algorithm is made in this paper. WCBIP algorithm makes the data rates high during the network lifetime. For both the connectivity and survivability, [10] propose polynomial time constant approximation algorithms to solve the problems. The algorithms aim to deploy a small number of RNs in the candidate locations, such that the overall energy harvesting potential of the RNs is high. A more desirable objective would be to maximize the overall harvesting potential.

D. Techniques for Relay node Assignment:

In [2] to solve the problem of relay node assignment (RNA) RNA for better communication b/w RNôs and ENôs, the Binary integer programming technique is used. It is a linear programming technique which based on branch-and-bound algorithm. This algorithm first creates a search tree by repeatedly using data called branching. In [4], a flexible vehicle Routing model is used as a solution to the problem of resource allocation and management for relay node assignment in cooperative networks. This model incorporates the problem of clustering and relay assignment into a unified problem and then can be solved more efficiently by using BIP. This model is also useable for other network scenarios.

E. Cluster organization techniques:

[6] Discusses the uneven clustering mechanism. The uneven clustering routing algorithm for WSNs based on ant colony optimization has a periodic round; each round is divided into cluster formation and cluster route stage. After new ñroundò start, the algorithm firstly divided the cluster, then the data will be transmitted between the CH. In the cluster formation stage, the base station firstly need to use a given transmit power to network to broadcast a signal.

After receiving this signal, each sensor according to the received signal strength try to calculate the distance to the base station. [2] Proposes an effective reconfiguration algorithm to solve the load balancing problem by fairly distributing Cluster-Heads (CHs) in wireless sensor networks. This algorithm can be divided into two phases, the first is setup phase and the second is steady-state phase. The steady-state phase is further divided into several frames and all normal sensor nodes transmit the raw data to their Cluster-Heads at each frame. During the setup phase a set of Cluster-Heads is selected randomly. Then each Cluster-Head broadcast an announcement message for declaring itself a Cluster-Head.

F. Techniques for Cluster-Head selection:

A relay based clustering algorithm (RBC) is proposed in [7], to solve the problem of residual energy in a cluster for heterogeneous energy wireless sensor networks. The ñRelayò mechanism is introduced to relay the ñCluster Headò position to its successor by considering the nodesô residual energy. This scheme improves LEACH and is called LEACH-E in this paper. RBC divides the network function in to a number of rounds, and each round have two phases which includes the set-up phase and the steady-state phase, similar to LEACH protocol. It is different from LEACH, the cluster head in RBC, during the current round, assigns the node with the highest residual energy in its cluster the ñCluster Headò position in the next round and a threshold. During the set-up phase of RBC, the non-cluster head node is required to piggyback its residual energy information Together with the Join-REQ to its chosen cluster head. [11],avoid the premature death of a Cluster-Head node by using the an analytical method to identify the optimal point at which a Cluster-Head rotation phase has to be initiate in an energy driven cluster head rotation algorithm. Energy driven Cluster-Head rotation methods use the residual energy of each existing Cluster-Head to determine the point at which to call for a new CH selection phase. The selection phase is triggered once when the residual energy of any of the Cluster-Heads go below a computed threshold value. This threshold value is computed dynamically based on residual energy and a parameter ñCò where C is a predetermined. [12], discussed genetic algorithm for optimal cluster head selection that not only reduces the overhead but also it is stable. This clustering scheme considering the connection duration, concept of critical node, battery power of a node by the Genetic Algorithm (GA). It takes some chromosomes with random number of Cluster-Heads and calculates fitness values for each chromosome. Based on fitness values this Algorithm selects some chromosomes for crossover. Here also it performs the same operations and selects best chromosomes for mutation. After mutation this Algorithm produces the optimal result. [9] Focuses on reducing power consumption by considering consumed energy as a factor for cluster head selection of each node in network. This algorithm adds the concept of consumed energy factor to the LEACH threshold formula of Cluster-Head selection to decrease the threshold value. In [37], an energy-efficient clustering approach name Improved Minimum Separation Distance (IMSD) is proposed. Use of this algorithm improved minimum separation distance between Cluster-Heads, which improves the network lifetime. In [16], a hybrid clustering and routing architecture for wireless sensor networks is discussed. It has three main parts architecture which are a modified subtractive clustering technique, an energy-aware cluster head selection method and a cost-based routing algorithm. In this system each round consists of two phases, a setup phase and a steady-state phase. In the setup phase, the base station divide nodes into clusters, selects most suitable nodes as cluster-Heads for each cluster. During the data transmission phase, the all nodes transmit the sensed data to the base station according to their provided schedule.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we surveyed several network lifetime extension issues rise in wireless Sensor Networks and also discussed the approaches to tackle those issues. The bidirectional nature of the modern wireless sensor networks and capability to control sensor activity leads to a need for a better and an energy efficient hierarchy to increase the network lifetime. There are still many issues are open to research thoroughly like, the assignment of partners and their management in multi-user networks, the development of efficient power control mechanisms for cooperation, designing a better code for coded cooperation method, fixed life time of Relay nodes and Edge nodes, the best place for Relay nodes in the sensing field, the handoff problem between nodes in the network, Synchronization b/w nodes, the design of a cluster and a cluster size decision, A fast and efficient method for collecting and disseminating CSI in moderate and large sized network, To overcome the overhead of ORA algorithm, A more efficient algorithm is needed to save the consumed energy of the network, energy balancing and the study of the ratio of packets being lost due to collision which helps in reducing the end-to-end delay will be addressed.

REFERENCES

[1] Aria Nosratinia, Todd E. Hunter, and Ahmad Reza Hedayat, ñCooperative Communication in Wireless Networks,ò IEEE Communication Magazine, vol. 42, no. 10, pp. 74-80, Oct. 2004.

[2] Wenxuan Guo, Xinming Huang, Wenjing Lou, and Cao Liang, ñOn Relay node placement and Assignment for Two-tiered Wireless Networks,ò Mobile networks and Applications, vol. 13, no. 1-2, pp.186-197, April 2008.

[3] Jee-Hoon Kim, Young Hwan You, and Hyoung Kuo Song, ñEfficient cooperative transmission schemes for Resource-constrained networks,ò Presented at proceedings of the 6th ACM international symposium on Mobility management and wireless access, 2008.

[4] Amir Minayi Jalil, Vahid Meghdadi and Jean-Pierre Cances, "A cross-Layer Approach to Clustering and Relay Assignment based on vehicle Routing Problem,ò Cross layer design (IWLCD), pp. 1-5, Nov. 30 2011-Dec. 1 2011.

[5] Sushant Sharma, Yi Shi, Y. Thomas Hou, and Sastry Kompella, "An Optimal Algorithm for Relay Node Assignment in Cooperative Ad Hoc Networks,ò IEEE/ACM Transaction on networking, Vol. 19, no. 3, June 2011

[6] Jiang Du, Liang Wang, ñUneven Clustering Routing Algorithm for Wireless Sensor Networks Based on Ant Colony Optimization,ò computer research and development (ICCRD), vol. 13, pp. 67-71, Mar. 2011.

[7] Yu Fang, Xiaofu Ma, and Ming Jiang, ñA Relay-Based Clustering Algorithm for Heterogeneous Energy Wireless Sensor Networks,ò Computer science and Automation Engineering, vol. 4, pp. 715-718, June 2011.

[8] Venkata Manoj, M. M. Manohara Pai, Radhika M.Pai, and Joseph Mouzna, ñTraffic Monitoring and Routing in VANETs ï A Cluster Based Approach,ò ITS Telecommunication, pp. 27-32, Aug. 2011.

[9] Desalegn Getachew Melese, Huagang Xiong, and Qiang Gao, ñConsumed Energy as a Factor for Cluster Head Selection in Wireless Sensor Networks, ñWireless Communications Networking and Mobile Computing, pp.1-4, Sept, 2010.

[10] Satyajayant Misra, Nahid Ebrahimi Majd and Hong Huang, "Constrained Relay Node Placement in Energy Harvesting Wireless Sensor Networks,ò Mobile Ad-Hoc and Sensor Systems, pp. 25-34, Oct. 2011.

[11] Sankalpa Gamwarige and Chulantha Kulasekere, ñOptimization of Cluster Head Rotation in Energy Constrained Wireless Sensor Networks,ò wireless and optical communication networks, pp. 1-5, July. 2007.

[12] S.Muthuramalingam1, R.Malarvizhi1, R.Veerayazhini 1 and R.Rajaram2, ñReducing the Cluster Overhead by Selecting Optimal and Stable Cluster Head through Genetic Algorithm,ò computing and processing (software/hardware), pp. 540-545, 2008.

[13] Abolfazle Akbari, Mahdi Soruri, and Seyed Vahid Jalali, ñSurvey of stable clustering for mobile ad-hoc networks,ò Machine Vision, pp. 3-7, Dec. 2009.

[14] P. Tillaport, S. Thammarojsakul, T.Thumthawatworn and P. Santiprabhob, "An Approach toHybrid Clustering and Routing in Wireless Sensor Networks", In Proc. *IEEE Aerospace,* 2005, pp. 1-8

[15] T. Himsoon, W. P. Siriwongpairat, Z. Han, and K. J. R. Liu, ñLifetime maximization via cooperative nodes and relay deployment in wireless networks,ò IEEE Journal on Selected Areas in Communications, vol. 25, no. 2, pp. 306ï 317, February 2007.

[16] Thien, M.C.M and Thien, T., ñAn Efficient Cluster Head Selection Algorithm for Wireless Sensor Networksò, IEEE conference on Intelligent system, modeling and simulation, pp.287-291, Jan. 2010.

AUTHORS PROFILE

Asia Samreen, is an Assistant professor at the Department of Computer Science, Bahria University, Karachi Campus. She is doing her PhD from University of Karachi. Her research interest is in the area of Network Security. She has keen interest in social networks and security issues of ad-hoc networks.

Gulnaz Ahmed, is a student at the Department of Computer Science and Graduate Studies, Bahria University Karachi Campus, Karachi. Her research interest is in the area of Wireless Sensor Networks that maximize innovative patents. She has done MSc. in Applied Physics with specialization in Electronics from Karachi University. She lives in Karachi, Pakistan, with her family, with three siblings and mother. This is her first work. She mostly spends her time for research. On her free time she likes to read books on different topics. She also enjoys hanging out with friends and losing her mind to house music. If you would like to reach her, send her an email to gulnaz-ahmed87@yahoo.com or contact her at this # 0331-7961877.

A New Approach To Monitor Children's Computer Usage Pattern

Neetu Anand,
(Research Scholar, Lingayas University),
Assistant Professor(Deptt. of Computer Sc.),
Maharaja Surajmal Institute,Delhi
neetuanand77@rediffmail.com

Dr. Mayank Singh,
Associate Professor,
Deptt. of Computer Application,
Lingayas University faridabad,Haryana
mayanksingh2005@gmail.com

***ABSTRACT*-Each day, the WWW grows by nearly a million electronic pages, adding to the hundreds of millions previously on-line. WWW is a platform for swapping numerous types of information, ranging from research papers, and learning contents, to audio-visual content and various software's [4]. Mining log data is a very comprehensive research area developed to solve the issues related to usage of computer and internet on single computer or on network of computers. This article provides a analysis of monitoring strategy for computer and internet usage of teenage computer.**

***Keywords:* Usage Mining, Mining Technologies, Monitoring Strategy**

I. INTRODUCTION

Computer and Internet usage has become virtually common among teenager and kids. They access various information and maintain friendships and relationships with their near and dear ones through this never ended social network and involve them to the extent that they forget to give time to their family members and even neglect their physical health. The parents' involvement is needed at this stage to look after the use of technology safely as the young people are, at their dynamic stage of development in which risk-taking behaviours and immature decision making capacities can lead to adverse outcomes. So Parents play a critical role in ensuring their teenage children's responsible and safe use of online and offline services.

Monitoring strategy for the Computer and Internet usage of the teenage children is suggested for parents, and the majority of parents actually want to involve themselves in monitoring their children activities for some of the time. Although the earlier figures show that the use of the computer and Internet by teenage groups is going high and continues to grow, McGrath (2009) advised that young people use technology in a distinctive way to mature people-adult usage trends were more practical or professional purposes, whereas young people were using it for their personal use and they form their group on this social network.

The reasons why parents need to monitor their child's computer and Internet activity are:

- The child creates a lots of friendship groups on network and even build a relationship up to the point where the child is comfortable meeting with them in real life.
- Cyber harassment is the new major threat to underage users.
- They play games and waste there important time of study.
- They swapped too much of their personal information via e-mail, chats, and at social sites
- Download and see banned music and movies videos.
- To check for the time he/she spend on actual work.

II. WEB MINING CATEGORIES

The term Web mining was given by Etzioni **(1996)** to describe how data mining techniques can automatically determine the information from Web resources, and generate patterns on the Web log data. The very first definition of Web mining is to "Discovery and analysis of useful knowledge from the World Wide Web" (Cooley, Mobasher, & Srivastava, **1997,** p. 558). Web mining research overlaps substantially with other areas, including data mining, text mining, information retrieval, and Web retrieval [5].

Web mining is the using of data mining techniques to automatically discover and extract information from Web documents and services. The three main axes of Web mining that have been identified, according to the data used as input in the data mining process, namely Web structure, Web content and Web usage mining.

Web structure mining is categorization of the web pages and generating information such as the similarity and relationship between them.

Web content mining is to retrieve the information (content) available on the Web into more structured forms as well as its indexing for easy tracking information locations. Web content may be unstructured (plain text), semi structured (HTML documents), or structured (extracted from databases into dynamic Web pages).

Web usage mining is the process of identifying browsing patterns by analysing the user's navigational behaviour. This information takes as input the usage data, i.e. the data residing in the Web server logs, recording the visits of the users to a Web site [4].

Web log data is categorised in to three categories based on the source of information. These categories are: *server side*, *client side* and *proxy side* log file. Server side data gives information about the behaviours of all users, whereas the. Proxy side data is somewhere in between the client and server side data [3].

Client Log Files used client side data to give information about a user, using that particular client. *Proxy Log Files* used to capture the user access data i.e. it capture the pages that are being accessed by the users. Proxy server is in many-many cardinality since there are many users accessing many pages. *Server Log Files* are in relationship of many to one since there is only one web server response to many users. Different types of Server Log File include:

a. Referrer Log
b. Error Log
c. Agent Log
d. Access Log

Referrer Log file contains information about the pages that is being referred. *Error Log File* records the errors of web site especially page not found error (404 File not found). *Agent Log File* records the information about the website user's browser, browser version & Operating System. *Access Log file* records all the click, hits and accesses made by the user to the website.

III. METHODS FOR EXTRACTING USAGE PATTERNS

The actual objective of web log mining is to extract interesting and potentially useful patterns that show users correlated preferences in accesses to the web pages being served by a particular web server. There are various methods on web log mining; most of them provide very primitive mechanisms for reporting statistical fact of the accesses, i.e., the number of the accesses to individual files during a period of time, the originality of the users, etc. It is believe that by using data mining techniques and systematically analysing the behaviour of past visitors, more sophisticated knowledge of the users access pattern can be obtained from the web log file. There reasons for using pre-processing techniques on raw log data before being able to apply data mining techniques on it is the incompatibility of data structure and irrelevant information concerning to the specific mining task. Therefore, the basic work is to transform the data into data-mining friendly form and filter out irrelevant information [3].

The most common data mining methods and algorithms applied on log data is association rules, Sequential pattern discovery, clustering, and classification.

Association rule mining is a technique to discover frequent patterns, associations, and correlations or relationship among sets of objects. Association rules are used in order to disclose correlations between pages accessed together during a server session. These types of rules indicate the potential relationship among pages that are habitually saw together even if they are not directly connected, and can expose associations between groups of users with special interests. Apart from being utilized for business applications, these interpretations can also be used as a model for Web site reshuffle, for e.g., by positioning links which connect pages that often watched together, or as a way to improve the system's performance through prefetching Web data.

Sequential pattern is an expansion of association rules mining in that it tells patterns of co-occurrence including the concept of time. In the Web mining such a pattern might be a Web page or a set of pages accessed immediately after another set of pages. By the use of this approach, useful trends users' can be revealed, and predictions concerning visit patterns can be made.

Clustering is used to group items that have similar features. In Web mining, we can classify two cases, user clusters and page clusters. Page clustering group pages that seem to be related according to the users' perception. User clustering results in groups of users that seem to behave similarly when navigating through a Web site. Such knowledge is used to personalize a Web site.

Classification is an approach that maps a data item into one of several predetermined classes. In the Web domain classes usually represent different user profiles and classification is performed using selected features that describe each user's category.

The common classification algorithms are decision trees, naïve Bayesian classifier, neural networks, and so on. There also exist other methods for extracting usage patterns from Web logs.

IV. THE PROPOSED FRAMEWORK OF MONITORING AGENT FOR COMPUTER AND ITS USE

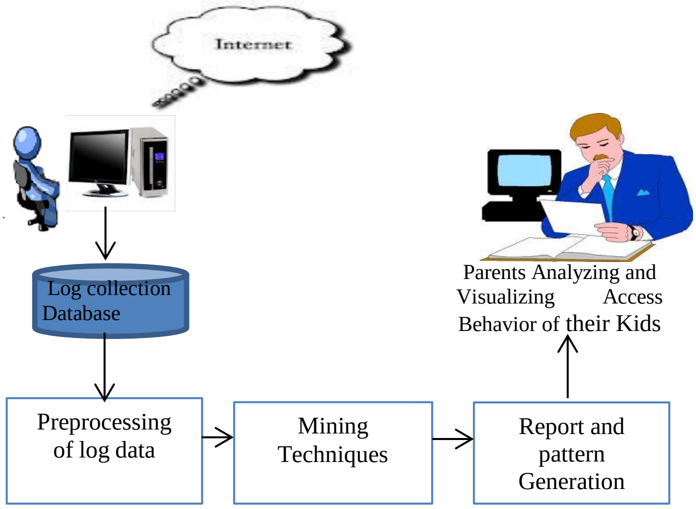

Internet
Log collection Database
Parents Analyzing and Visualizing Access Behavior of their Kids
Preprocessing of log data
Mining Techniques
Report and pattern Generation

Fig. 1. Framework for Monitoring Agent

The research methodology used for the above stated framework is given as:

1. Log data of client computer is collected and stored in a database.
2. Pre-processing activities make a review to the web log data prior to processing. There are several pre-processing tasks that must be performed prior to applying mining algorithm.
3. Report and Pattern Generation, which spent most of all mining activities because these activities do a search to find hidden patterns in the data log.
4. Pattern Analysis is a process to study and conduct an analysis of the results obtained from the search behaviour patterns.

For computer and Internet usage monitoring, software is to be installed on the target computer which allow parents to observe and bound their children's usage of many applications, websites visited and online searches, social networking behavior and other programs. Usually, parents need the four things on their kid's computer:

- Usage controls
- content filters
- monitoring tools
- Computer usage management programs

Usage Controls

Usage controls can be imposed to change a child's behaviour with regards to a computer. With this only for a limited period of time the device will be enabled. Many mobiles, computers, TVs are now coming with usage control options that allow parents to set limits on the times the device may be on.

Content Filtering

It enables the parents to put restriction on the viewing behaviour of their children's mainly for Internet and cable TV. There are many Content filtering software available which allow only age-appropriate viewing of web content, chatting details and social networking interactions. Also there is a way in which parents have to specify some keywords when configuring the software. The software may be configured to ban web pages with the specified keyword, or the keywords are blocked with a series of characters, permitting the child to view the rest of the page. Content may be filtered according to age also. Content filters can be easily installed and configured and the parent can create separate profile for more than one child. Many content monitoring programs allow parents to block certain URLs, keywords and other specifiers. Some content filters may be configured to even alert the parents about the access of blocked contents. Content filtering provides a parent with a level of control over their child's online and offline behaviour

Monitoring Software

It is used to monitor child action and activities on a computer. These software help the parents to observe all aspects of a child's behaviour, including Internet sites visited, Instant Messaging chats, Email, application accessed, as well as other online and offline behaviour.

Monitoring a child's activities is a good way for parents to make sure their child is safe while using a computer.

Computer Usage Management

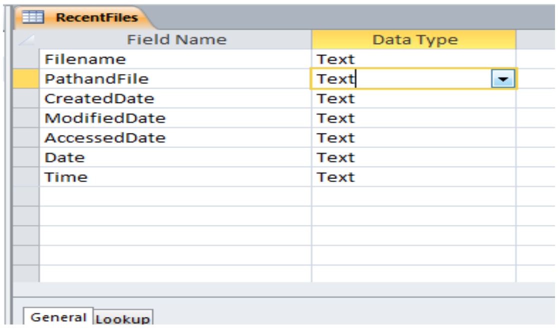
Computer usage management programs are the application software that are used by parents to implement studying behaviour among the kids in order to win computer time for performing activities, such as gaming. These management programs will grant a kid with a quantified amount of computer usage time based on the extent to which they complete their work.

Computer usage management programs are a practical way to give bonus to a kid for completing their schoolwork, or finishing some pre-defined learning task, by permitting them to play innovative computer games once their task is finished. Computer usage management programs may be designed to grant a kid guaranteed time to access the computer for various other activities once the specified task has been completed.

IV. SYSTEM IMPLEMENTATION

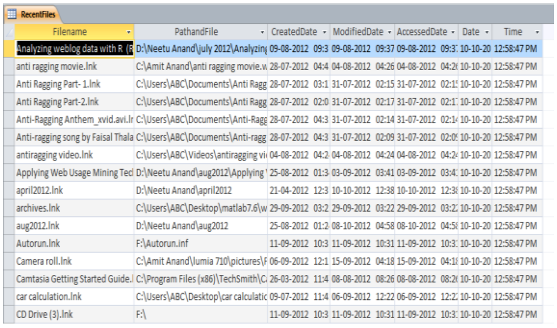
The Proposed intelligent system for users' activity monitoring was implemented with use of VB language and Access database, respectively. Parents want to keep track of how much time their kids spends on watching videos, playing games or engaging in other non-productive activities that may distract them from their schoolwork or sleep.
With the use of intelligent systems, parents can also monitor their children's behavior on a target computer. It can also be used to view all aspects of a child's behavior, including Internet sites visited, Instant Messaging chats, email, application accessed, as well as other online and offline behavior.

The view of the created database to record the content of all the activities on client computer is given below:



Field Name	Data Type
Filename	Text
PathandFile	Text
CreatedDate	Text
ModifiedDate	Text
AccessedDate	Text
Date	Text
Time	Text

Fig. 2. Structure of the Database used to store recent usage files



Filename	PathandFile	CreatedDate	ModifiedDate	AccessedDate	Date	Time
Analyzing weblog data with R (R	D:\Neetu Anand\july 2012\Analyzing	09-08-2012 09:3	09-08-2012 09:37	09-08-2012 09:3	10-10-20	12:58:47 PM
anti ragging movie.lnk	C:\Amit Anand\anti ragging movie.w	28-07-2012 04:4	04-08-2012 04:26	04-08-2012 04:2	10-10-20	12:58:47 PM
Anti Ragging Part- 1.lnk	C:\Users\ABC\Documents\Anti Ragg	28-07-2012 03:1	31-07-2012 02:15	31-07-2012 02:1	10-10-20	12:58:47 PM
Anti Ragging Part-2.lnk	C:\Users\ABC\Documents\Anti Ragg	28-07-2012 02:0	31-07-2012 02:17	31-07-2012 02:1	10-10-20	12:58:47 PM
Anti-Ragging Anthem_xvid.avi.lr	C:\Users\ABC\Documents\Anti-Ragg	28-07-2012 04:3	31-07-2012 02:14	31-07-2012 02:1	10-10-20	12:58:47 PM
Anti-ragging song by Faisal Thala	C:\Users\ABC\Documents\Anti-ragg	28-07-2012 04:3	31-07-2012 02:09	31-07-2012 02:0	10-10-20	12:58:47 PM
antiragging video.lnk	C:\Users\ABC\Videos\antiragging vi	04-08-2012 04:2	04-08-2012 04:24	04-08-2012 04:2	10-10-20	12:58:47 PM
Applying Web Usage Mining Tec	D:\Neetu Anand\aug2012\Applying	25-08-2012 01:3	03-09-2012 03:41	03-09-2012 03:4	10-10-20	12:58:47 PM
april2012.lnk	D:\Neetu Anand\april2012	21-04-2012 12:3	10-10-2012 12:38	10-10-2012 12:3	10-10-20	12:58:47 PM
archives.lnk	C:\Users\ABC\Desktop\matlab7.6\w	29-09-2012 03:2	29-09-2012 03:22	29-09-2012 03:2	10-10-20	12:58:47 PM
aug2012.lnk	D:\Neetu Anand\aug2012	25-08-2012 01:2	08-10-2012 04:58	08-10-2012 04:5	10-10-20	12:58:47 PM
Autorun.lnk	F:\Autorun.inf	11-09-2012 10:3	11-09-2012 10:31	11-09-2012 10:3	10-10-20	12:58:47 PM
Camera roll.lnk	C:\Amit Anand\lumia 710\pictures\F	06-09-2012 12:1	15-09-2012 04:18	15-09-2012 04:1	10-10-20	12:58:47 PM
Camtasia Getting Started Guide.l	C:\Program Files (x86)\TechSmith\C	26-03-2012 11:4	08-08-2012 08:26	08-08-2012 08:2	10-10-20	12:58:47 PM
car calculation.lnk	C:\Users\ABC\Desktop\car calculatic	09-07-2012 11:4	06-09-2012 12:22	06-09-2012 12:2	10-10-20	12:58:47 PM
CD Drive (3).lnk	F:\	11-09-2012 10:3	11-09-2012 10:31	11-09-2012 10:3	10-10-20	12:58:47 PM

Fig. 3. View of the Database

V. CONCLUSION

The Software for monitoring the computer records all activities performed on a computer (launched applications, opened documents etc.), and it will be extended to record all the internet usage data (visited web-sites) and the duration of each activity (when kids actually work with applications/documents and read web-pages).Further the data will be pre-processed and some classification methods will be applied for making prediction system.

REFERENCES

1. Shahnaz Parvin Nina, Md. Mahamudur Rahaman, Md. Khairul Islam Bhuiyan, Khandakar Entenam Unayes Ahmed, Pattern Discovery of Web Usage Mining, International Conference on Computer Technology and Development,(2009).
2. Shivkumar Khosla, Varunakshi Bhojane, Capturing Web Log and Performing Preprocessing of the User's Accessing Distance Education System, International Journal of Modern Engineering Research (IJMER),(2012).
3. F. Tao, P. Contreras, B. Pauer, T. Taskaya and F. Murtagh, Users Interest Correlation through Web log Mining, International Conference in human computer Interface (2001).
4. Magdalini Eirinaki, Web Mining: A Roadmap,(2005).
5. Hsinchun Chen and Michael Chau Web Mining: Machine Learning for Web Applications (2005).
6. Mehrdad Jalali ,Norwati Mustapha, Ali Mamat , Md. Nasir B Sulaiman , A new classification model for online predicting users' future movements,IEEE(2008).

Ms. Neetu Anand is a doctoral student in Lingayas University, Faridabad, Haryana. She is working as an Assistant Professor in Department of Computer Science, Maharaja Surajmal Institute, and Delhi, India. Her research interest lies in Data Mining and Knowledge Discovery. She is having Twelve years of teaching experience.

Dr. Mayank Singh have done his M. E in software engineering from Thapar University and PhD from Uttarakhand Technical University. His Research areas are Software Engineering, Software Testing, Wireless Sensor Networks and Data Mining.

Design and Implementation of Security Framework for Cognitive Radio Networks Resource Management

Obeten O. Ekabua
Department of Computer Science
North-West University, Mafikeng Campus,
Private Bag X2046, Mmabatho 2735, South Africa
(obeten.ekabuao@nwu.ac.za)

Ifeoma U. Ohaeri
Department of Computer Science
North-West University, Mafikeng Campus,
Private Bag X2046, Mmabatho 2735, South Africa
(23989688@nwu.ac.za*)*

Abstract: Designing and implementing a secure communication for any network is an important issue for the optimal control of resource usage in a resource constrain network environment. Therefore, in this paper, we design and implement a joint authentication and authorization framework by transforming the framework requirement analysis. The framework is a security infrastructure capable of monitoring and controlling access to the limited spectrum resources, dynamically managing data and information in CRN, for a secured communication and quality of service (Q_OS). We explained how the various components in the framework interact to ensure a secured communication and effective access control.

Keywords: Network Management, Security, Authentication, Authorization, Access Control.

1. Introduction

Cognitive radio network is a novel technology designed to alleviate the challenges associated with spectrum shortage. Rapid developments in wireless communication have led to development of Dynamic Spectrum Access (DSA) technology involving licensed and unlicensed users. Secure communication is a salient aspect of any network and has remained unexplored in cognitive radio networks (CRN). Consequently, achieving security in cognitive radio network is thus a huge challenge. The dynamic nature of cognitive radios has introduced weaknesses and vulnerabilities which are capable of affecting the quality of service (QoS) of the network [1,2]. Therefore, the main goal of this research paper is to report on the design and implementation of a joint authentication and authorization framework for CRNs, as a fundamental security infrastructure for access control, and dynamic management of data and information. This security framework can use any form of authentication medium based on network security policy (NSP), either, username, password, pin number and so on. This user profile and security data are supplied to the network management database by registration. Moreover, username and password are used often in this framework design for identification. Often times, users make quick conclusions that, the use of passwords for authentication and authorizations are not reliable and capable of providing a secured communication. When this information is transmitted over the network without encryption, they are prone to attacks because all information and data in the device are exposed. Though, this is not within the context of this research project but however, it is necessary to be mentioned it at this juncture [3].

The design aspect of this paper describes the framework layout and its components using designs and other relevant diagrams for explanations. Authentication and authorization are quite interwoven and often misused. However, the major difference between the two is that authentication deals with the identification of the subject (the client) requesting for connection to the (server), the host connection while authorization determines the access right to the resources (services) available in the network. This makes authentication come first before authorization [4].

2. CRN Architecture

Before we introduce the authentication and authorization framework design, it is necessary to first introduce the general design of the CRN

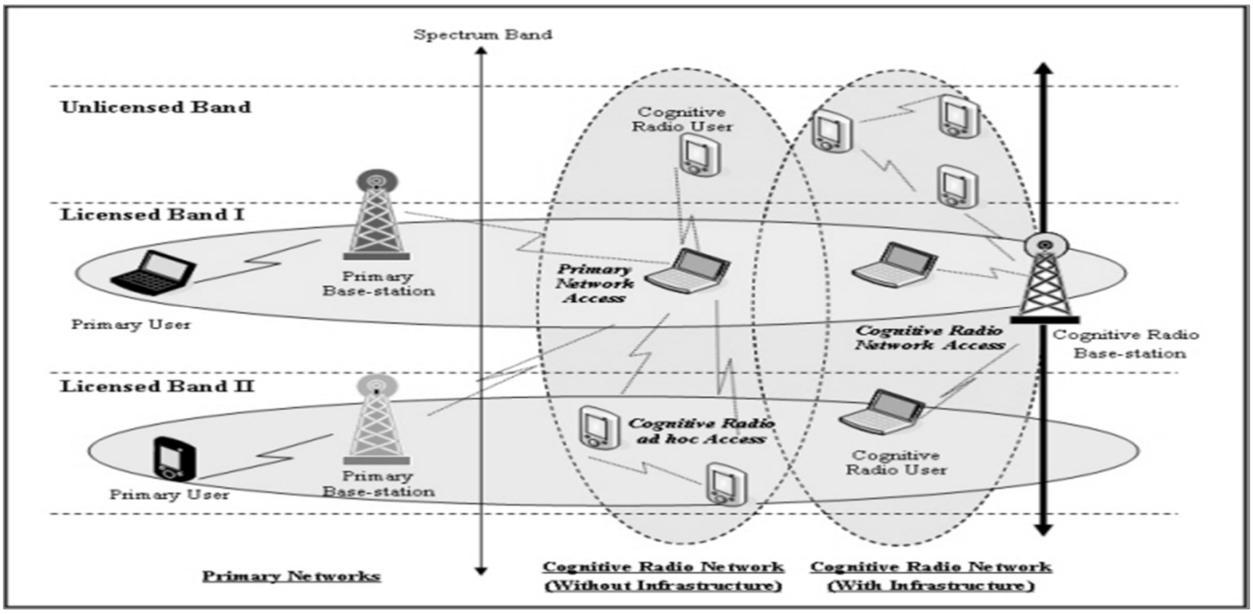


Fig.1: Spectrum CRN Architecture and its Interaction

network for a broad view and understanding of the architecture and other relevant components of the concept since this is the architecture (foundation) upon which the research project work is based on. Cognitive radio Network is dynamic and adaptive in nature. The architecture of CRN below shows the different components of, both functional, operational, and hardware, together with the relationship between them. The spectrum band is infinitely renewable, though limited due to its high demand by the secondary users. The Primary user has the legitimate right to a certain spectrum band, whereas, the secondary user do not have the license to operate in a choice band. The primary and unlicensed networks consist of some basic elements which include; primary user, primary base station, cognitive radio user, cognitive radio base station, cognitive radio network access, cognitive radio ad hoc access and primary network access.

However, the Primary user has the license (right) to operate in a specified spectrum band. This access right can only be controlled and monitored by its base-station and unauthorized users are not allowed interfere or affect its operations. Consequently, the Primary base-station is a fixed wireless infrastructure network component that has a spectrum license but do not have any capability for cognitive radio to share the spectrum with other users of cognitive radio. Therefore, the primary base-station may need to have both the primary and cognitive radio protocols to enable primary network access for the cognitive radio users.

Moreover, the spectrum access is allowed for the cognitive radio users only when not occupied by the authorized users because they do not operate with the spectrum license. Therefore, the cognitive radio user capabilities such as; spectrum sensing, spectrum decision, spectrum handoff and cognitive radio MAC, routing and transport protocols are required to enable communication with the base-station and other cognitive radio users as well.

The cognitive radio base-station in Fig. 3 is a fixed wireless infrastructure component that has cognitive radio capabilities and provides single hop connection to cognitive radio users without the license for spectrum access. The cognitive radio users communicate with each other either in a multi hop manner or through a base-station. Consequently, the cognitive radio network architecture in Figure 1 consists of three different types of network access such as: cognitive radio network access, cognitive radio ad hoc access and primary network access with different implementation requirements.

However, in cognitive radio network access, secondary users have the capability to access the cognitive radio base-station in both the licensed and unlicensed spectrum bands. The entire interactions takes place inside the cognitive radio network, therefore access scheme does not depend on the primary network. In cognitive radio ad hoc access cognitive radio users communicate with each other on both licensed and unlicensed spectrum bands via ad hoc connection. They are also capable of building their own access technology through which they can communicate. In primary network access**,** when the primary network is dormant, the cognitive radio users are able to access the primary base-station via the licensed band.

3. Centralized and Decentralized CRNs

This cognitive radio network architecture consists of both the centralized and decentralized cognitive radio network. It shows the position of the primary network and cognitive network in terms of spectrum usage, and communication that exist within the base station [5]. Fig. 2 and Fig. 3 below show the distinction and variation between the two types of cognitive radio network. It indicates the nature of communication existing in the two networks. In a centralized cognitive radio network as shown in Fig. 2, information is disseminated via a service base station which control and manages transfer of messages within the network.

a) Centralized Network Architecture

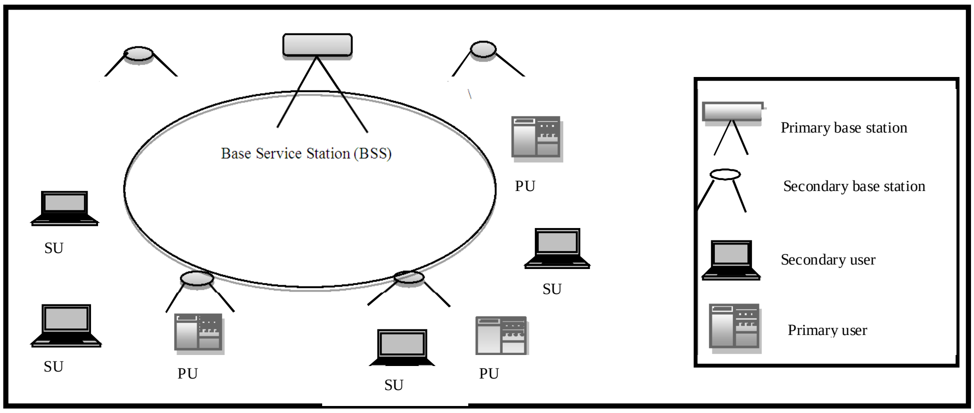


Fig. 2: Centralized Network Architecture

b) Decentralized Network Architecture

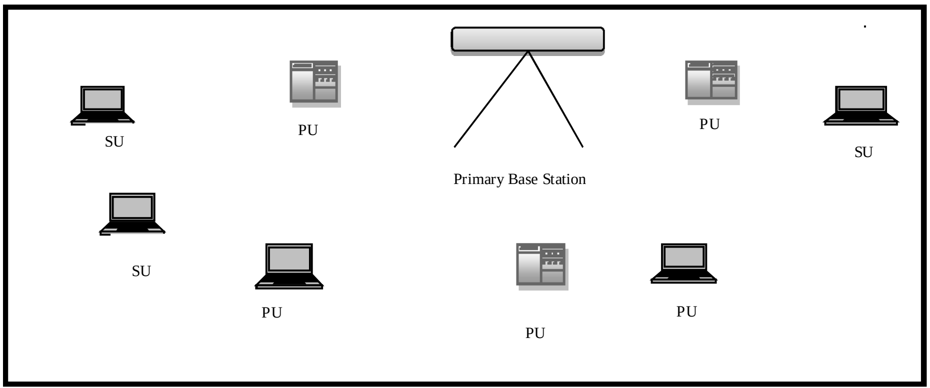


Fig. 3: Decentralized Network Architecture

Data and information are transmitted utilizing radio spectrum frequency bandwidth. Transmission and communication in a decentralized network (fig 3) are transferred directly, but when the devices forming the network are not within a close range, a multi hop is used to enable adequate dissemination of information as used in ad hoc networks.

4. Rationale of Framework

The purpose of the framework in the context of this paper is to ensure a secure communication in cognitive radio network, we use authentication, authorization, as security mechanism, to protect data and information along the line of transmission and also prevent malicious secondary users of the spectrum against network attacks. However, the benefits of are as follows:

a) The framework provides scalability: Typical authentication and authorization configurations depend on a server to or a group of servers to store user name and password. The essence of this is that local databases are not to be built and updated on every router and access server in the network.

b) The framework allows the network administrator configure multiple backup systems. For instance, an access server can be configured to first consult a security server and then the local database before any access is granted.

c) The framework supports standardized security protocols like TACACS +, RADIUS, and Kerberos.

d) The framework provides an architectural capability for configuring two different security measures; authentication, authorization [6].

5. Requirement Analysis

Requirement analysis firstly specifies the underlying requirement for designing and developing the authentication and authorization framework. The host network is the object, while the client host is referred to as the subject. Authentication concentrates on the subject requesting for connection to the network, while authorization concentrate on the subject requesting for a resource.

When the user dials into an access server which is configured using authentication protocol, the access server and spectrum manager prompts the user to make a user name and password available. The security policy decision point (SPDP) which is the request admission control and handoff point, checks to verify if the user is who he claims to be. The security policy enforcement point (SPEP) ensures that the service management policy is enforced by granting or denying access based on network policy.

The access server verifies a user by requesting for user name and password. This verification process is referred to as authentication. At this point the user may either be denied access or granted access. If authentication is successful then the user can be able to execute commands on the network server. The server then determines the commands and resources that should be made available to the user and specifies the privileges and rights the user should have. This process is referred to as authorization.
However, the framework is developed through four operational stages via: “login”, “connection and resource request”, decision and,” grant” or “deny” access stage.

5.1. Authentication

Authentication is a security measure in Cognitive Radio Network (CRN) that ensures that entities (users) are truly who they claim to be. This is verified before access to the network is granted. It actually associates a unique identity to each user in CRN, such as user identification name or password as approved by the service security policy. Using these unique forms identification client (users) can freely request for the spectrum resources. It involves the process of verification and validation of users’ identity (ID).

i) Requirement Name: Login
Description: This feature enables communication with the server.
Justification: This feature allows a new window to open for connection request to the server by the client.
ii) Requirement Name: Server Request
Description: This request will permit the client access into the network for the service he or she wants to access.
Justification: The framework should request the client identity details by requesting for the user identity (user name and password based on the network configuration, authentication, protocols and security policy enforcement point (SPEP).

iii) Requirement Name: Decision
Description: This feature allows the framework to make decision based on the security data and service profile. This stage is handled by the request admission control and handoff which consists of the security policy decision point (SPDP) and SPEP. **Justification**: The framework should ensure that the client is who he claims to be, before permission to access the network is granted based on SPEP and SPDP.

iv) Requirement Name: Grant or Deny Access.
Description: The framework should ensures that all the network services and communications are secured from intrusion and unauthorized access.
Justification: The framework should permit all authenticated client to have access to the services available.

5.2. Authorization

Authorization is a security measure that allows access to only the right entities (users) having the approved privilege to the particular resources requested. Different forms of authorization exist such as; out band authorization, signature authentication and password authentication. Moreover, for any communication (interaction or conversation) involving different parties or entities exchanging information, there should exist, a mutual trust relationship across the multiple domains in CRNS.

i) Requirement Name: Resource Request
Description: This feature will permit the authenticated user, to request for specific services and resources he or she wants to ace ss.
Justification: This framework should validate the users request based on service policies before access is released.

ii) Requirement Name: Decision
Description: This feature allows framework to make decision based on the privileges the client has over the resources available in the in the network. This stage is usually handled by the request admission control and handoff domain which consists of SPD and SPEP.
Justification: The framework makes sure that the user (client) has access to only the resources which he or she has the right or privilege to access.

iii) Requirement Name: Grant or Deny Access
Description: The framework should ensure that all the network resources are protected from unauthorized users.
Justification: The framework should ensure that all users strictly conform to service policies for authorizations based on the privileges given to the user so as to have access to the services and resources provided by the network.

6 Framework Design and Evaluation

This research paper presents a detailed design and implementation of a joint authentication and authorization framework by transforming the information from the framework requirement analysis. The framework is a security infrastructure that is capable of monitoring and controlling access to the limited spectrum resources by dynamically managing data and information in CRN, for a secured communication and quality of service (QoS). It is illustrated using components and interface relationships that describe the operation and functionality of the framework. This chapter also explains how the various components in the framework interact to ensure a secured communication and effective access control.

In a decentralized network, mobile devices exist in different locations and communicate in an ad hoc manner with any fixed infrastructure as shown in Figure 3. Data and information are transmitted utilizing radio spectrum frequency bandwidth. Transmission and communication in a decentralized network are transferred directly, but when the devices forming the network are not within a close range, a multi hop is used to enable adequate dissemination of information as used in ad hoc networks.

6.1 Joint CRN Authentication - Authorization (A-A) Framework

Having designed the authentication and authorization framework separately, it is necessary to also design a joint authentication and authorization (A-A) framework as one security infrastructure or gateway for a CRN. Figure 4 below represents the CRN A-A framework showing the relevant components, and how they interact to form a fundamental security infrastructure for effective dynamic management of data and information in CRN.

Basically, the joint authentication and authorization framework consist of a radio network infrastructure (RNI) and a security policy management center (SPMC). The SPMC In this framework consists of a SPMC agent is installed in each base station to monitor the flow or events within the network. The SPMC agents act like the watch dogs to sense intrusions and malicious attacks. They forwards control messages between the secondary devices and monitor spectrum usage. The SPMC agents are also responsible for service management tasks such as handoff management, secondary user services and all forms of monitoring so that the SPMC is not overloaded.

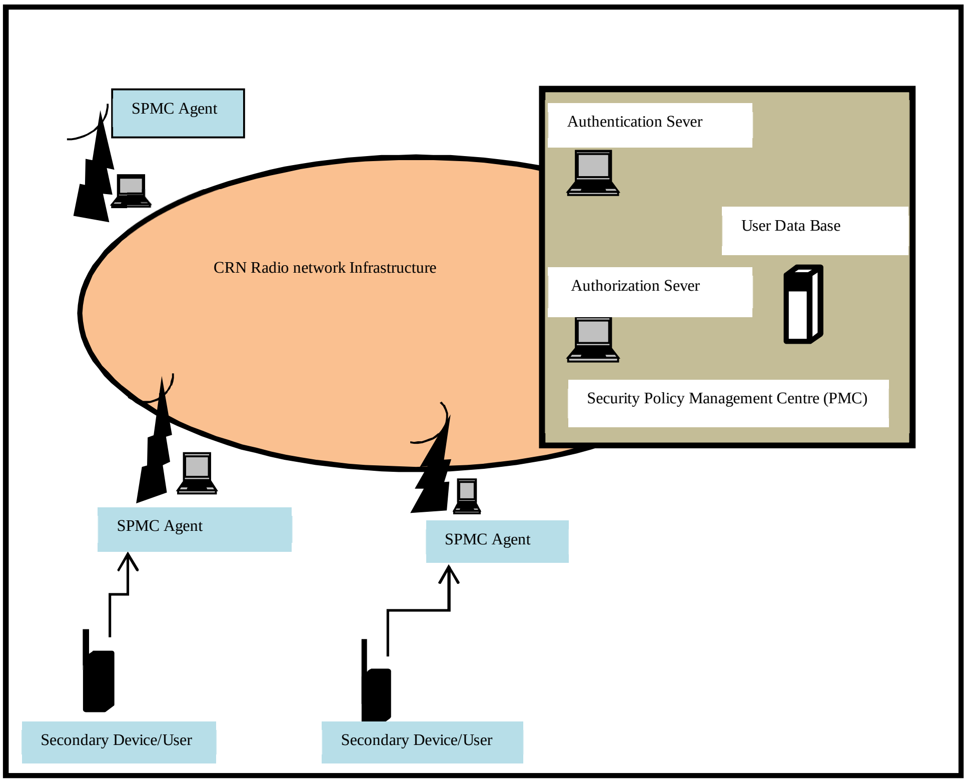


Fig. 4: Joint A-A Framework

the parts which includes; an authentication server, (AS), a user database and an authorization server (AS). The authentication server is responsible for authenticating legitimate users. The authorization server is responsible for the spectrum management. Immediately, a user is authenticated and its service requirement is determined to be acceptable, the authorization server authorizes the user by issuing a registration ticket, with which the user can communicate with other users under a close monitoring by the local SPMC agents.

The wireless infrastructure consists of a base station and the mobile switching centers. Moreover,

6.2. Framework Implication

The reason of this evaluation is to further explain the boarders of the framework. This framework is designed with the assumption that the secondary users or devices adhere to the rules of “inquiring before use” or sensing or listening before use”. This means that before the secondary users or devices listen to the control channel allocation information (CAI), notification of the free spectrum channel to utilize before their messages are transmitted for authentication and authorization request.

Software defined radios (SDR) are the key technology behind the CRNs. Therefore, the framework is designed with the understanding that the secondary devices are able to dynamically adjust the radio wave fronts in accordance to the Federal Communication Commission (FCC) spectrum requirement.

Cryptographic methods and public key infrastructure (PKI) required for encryption and decryption are not within the scope of this research project work. We therefore assume that certificate authority (CA) is available to serve the secondary user services such as; issuing public key certificate to the legitimate users of CRNs. Therefore, the verification of public keys and the actual implementation of this framework are among the future work of this research project.

Consequently, for any effort to evaluate this framework, it is necessary to emphasize that this framework is built on the three pillars of secured communication stated below.

i) Privacy

A secured communication or conversation should be private. Only the sender and the receiver (the parties involved) and the devices involved should be able to understand the communication flow. Privacy in CRN entails confidentiality and trust relationship. Transmission of data and information among the CR devices in the network must be confidential and the parties or entities involved must be in an agreement of trust to ensure privacy. All security credentials and user registration portfolios to enable access to the available spectrum resources are kept private. In CRN authentication and authorization framework embraces privacy as a major responsibility. It restricts access to message and prevents its contents from being exposed to other users who are not involved in the communication (whether legitimate or malicious users). The aim of privacy standard in the authentication and authorization security framework is to protect the transmission, secure communication and dynamically manage data and information in CRN. This enhances access control and can be achieved by the use of automated encryption.

ii) Integrity

A reliable security infrastructure should ensure integrity of the transmitted messages for a secured communication. This ensures that data and information is not altered in an unauthorized manner in transit and that the information received is exactly what is being sent by the transmitter. However, dynamic management of data and information using authentication and authorization security infrastructure ensures that resources are not modified or altered in an unauthorized manner and no third party has unauthorized access to the resources available in the network.

iii) Non repudiation

In CRN non repudiation is a feature that establishes the sender of a message or information to the receiver. It works as an accountability measure but also confirms that data and information is authentic and either parties or entities involved in a communication can deny being a part of it. This monitoring and access control feature ensures denial of (resources) data and information to unauthorized users. This is achieved using encryption of a strong access code for user ID which ensures that data and information in CRN are dynamically managed

6.3 Authentication-Authorization Model

Authentication and Authorization model consists majorly of an engine component called the Authentication and Authorization Engine component. This handles all the decision making activities based on access control policy (authentication and authorization policy).The SPEP for authentication and authorization ensures connection admission control and handoff by enforcing the respective designed policies on the subjects (network users).

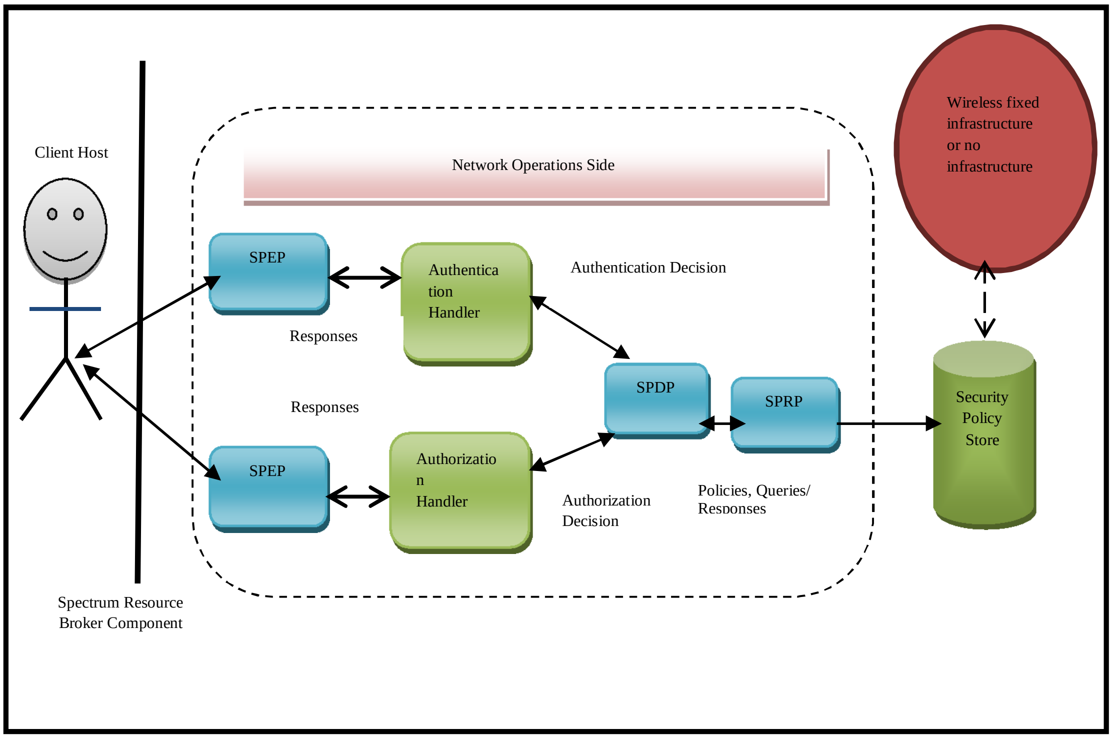


Fig. 5: Authentication - Authorization Engine Component

The authentication handler undertakes the decision making process. It decides on who gets connection, for how long and for what purpose. The result from that component is sent to the SPDP for implementation via SPEP based on the stipulated policy, and send confirmation message to the client. The SPRP fetches the policies from the host store. It grants easy access to the policies and helps in selecting the right policy based on request.

6.4 Spectrum Resource Broker

The Spectrum Resource Broker (SRB) component is the middle man or gateway in the communication line or access path between the client host and the server host and spectrum resources. It manages and controls spectrum resources (data and information). This involves spectrum sharing, spectrum decision and spectrum mobility. All interactions and communications between all cognitive radio networks, both the ones with infrastructure (base stations) and the ones without infrastructure are monitored via the spectrum broker component. It manages all access control mechanisms including; authentication and authorization processes employing the SPEP and SPDP services.

The diagram below indicates what transpires in terms of operations before connections are released from CRNs server and access to spectrum resources is granted.

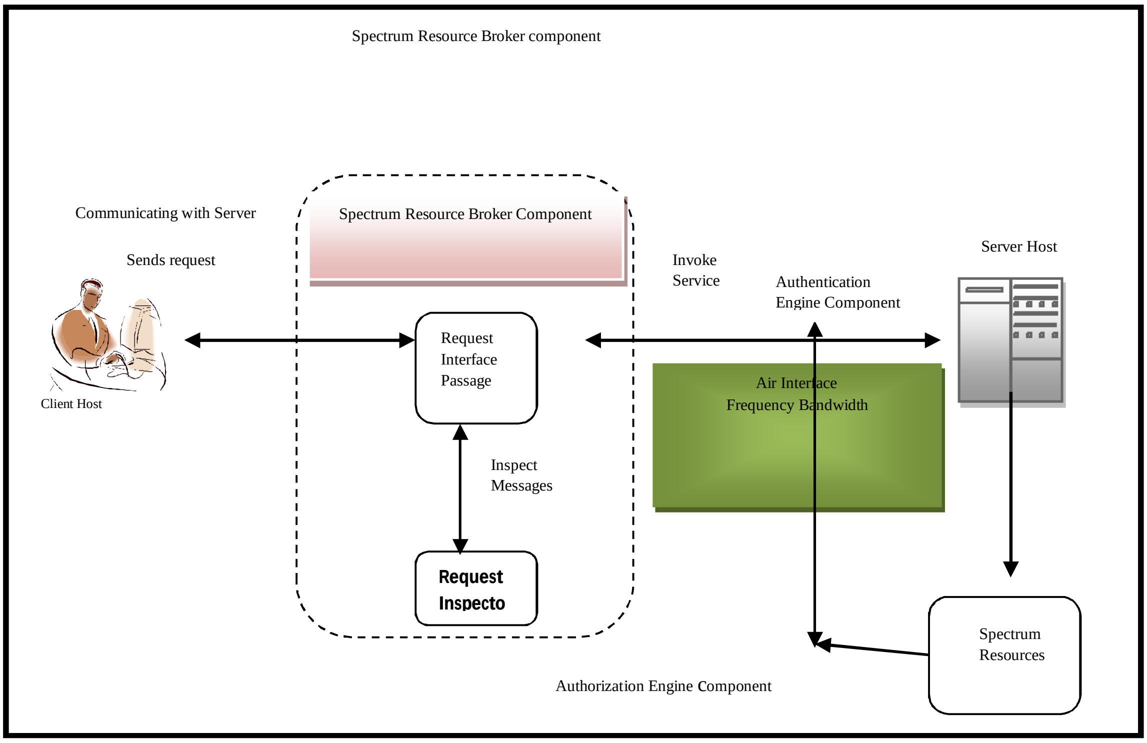

Spectrum Resource Broker component
Communicating with Server
Sends request
Client Host
Spectrum Resource Broker Component
Request Interface Passage
Inspect Messages
Request Inspecto
Invoke Service
Authentication Engine Component
Server Host
Air Interface Frequency Bandwidth
Spectrum Resources
Authorization Engine Component

Fig. 6: Spectrum Resource Broker Component

6.5 SRB UML Sequence

The UML diagram describes the sequence of activities in SRB component of CRNs. It shows the operations of its sub components indicating the request and communication (challenge response) AA protocols.

When the client sends a network or resource request it passes through the air frequency bandwidth because of its wireless nature. The request is delivered to the spectrum resource (SRB) broker that consists of the SPEP and SPDP. The SPEP component of the SRB performs the verification activities based on the security service policy (SSP). The message is then validated in line with the SPDP decision and the network service is invoked. The client is given feedback via the SPEP. The access is either granted or denied depending on the verification outcome.

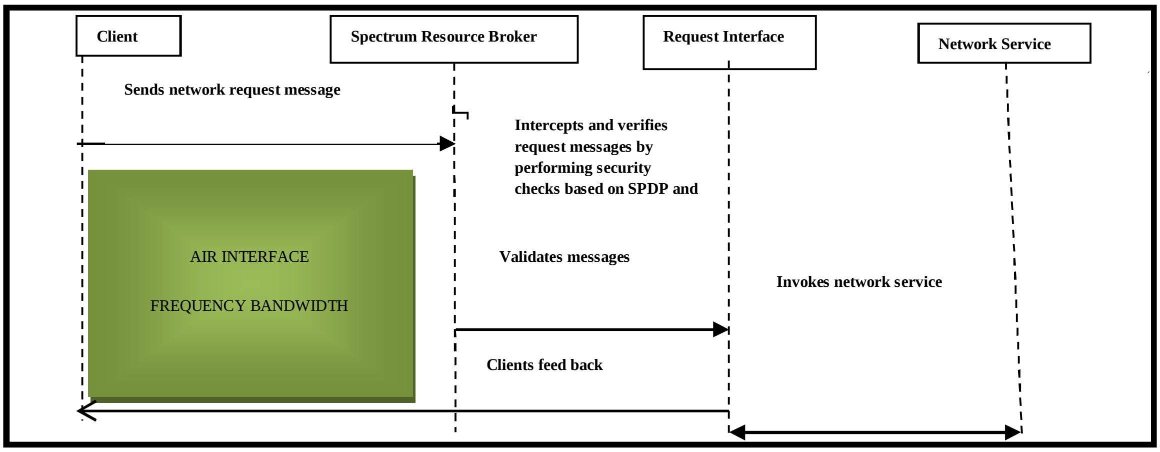


Fig. 7: UML Sequence Diagram - Spectrum Resource Broker

6.6. Security Activity Diagram for A- A Engine Component

The UML Sequence diagram for Authentication-Authorization Engine Component gives a clear description of the relationship and flow of interaction within the A-A engine component and depicts how the service and resource requestor is authenticated and authorized prior to accessing the service and resources and the role each of the components plays in the process of authentication and authorization. Thus, controlling access and dynamically managing data and information in CRN. Authentication takes place before authorization, so it is represented first in the diagram and authorization follows suit.

The Major components of the authentication Engine components such as; the client, the SPEP, the authentication handler (AH), the SPDP, the security policy retrieval point (SPRP) and policy point, are specified in the first column which is the first stage in the sequence.

The arrows pointing downward to the second column specify their corresponding activities and responsibilities respectively, which is the second stage of the diagram. When the client sends the service request message, the SPEP verifies the security details of the client if he is who he claims to be and constructs the authentication decision query and pass over to the SPDP through the authentication handler who certifies the decision query. The SPDP invokes the authentication security policy through the SPRP. The third stage shows the continuous flow of the activities and responsibilities of authentication engine components highlighted in the first column. The arrows pointing to the left hand side in the third column is returning the feedback to the client which is either access granted or denied.

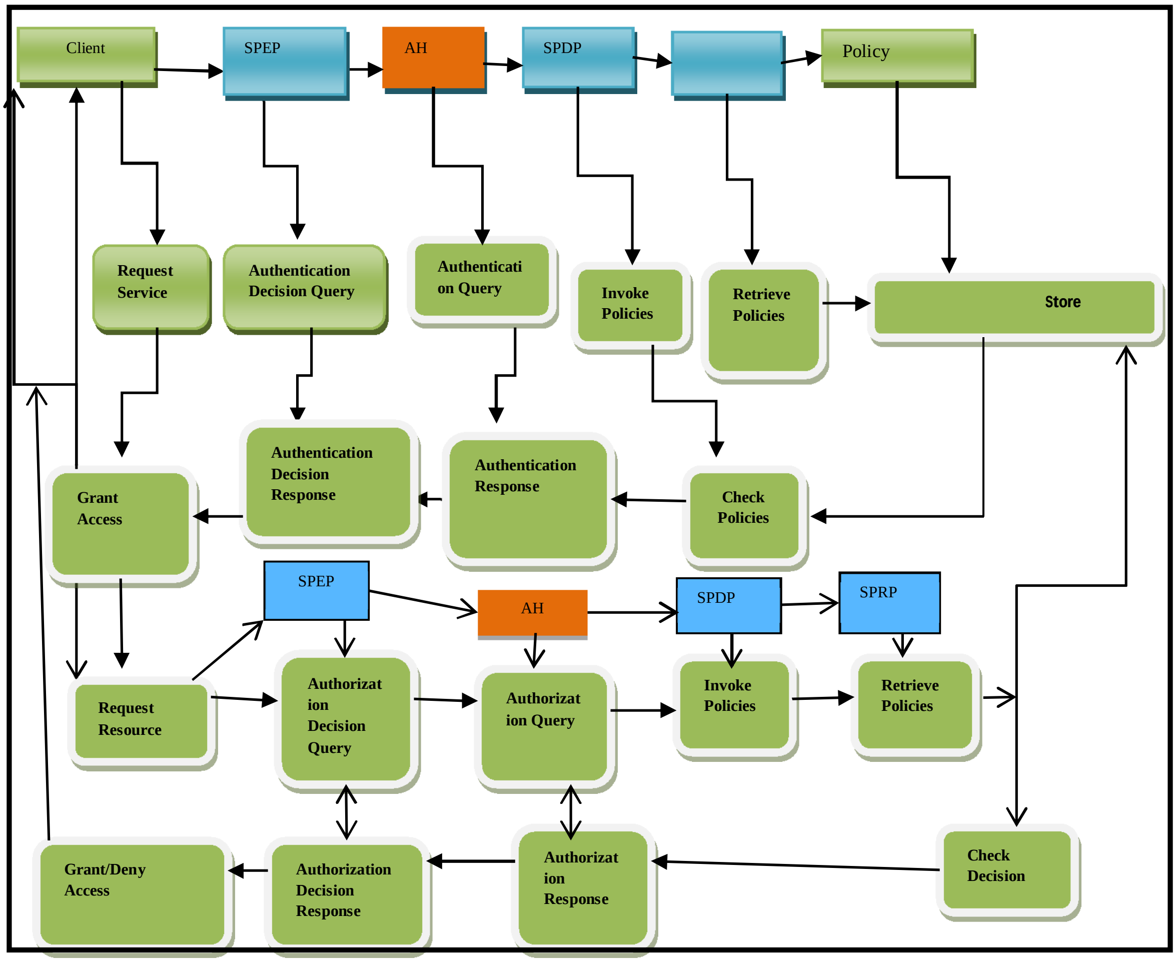


Fig. 8: Security Policy Activity Diagram for A-A Engine Component.

The authorization request follows suit in the fourth stage beginning with the resource request from the client which is usually intercepted at the SPEP to perform authorization decisions and passed over to the authorization handler (AH) for authorization query. It then goes over to the SPDP to invoke the security policies which is in turn retrieved from the security policy store by the SPRP. Before the response is returned to the client, the security policy point checks the authorization decision and returns to the authorization handler for response. The decision response is passed over to the client via the SPEP, which is either access granted or access denied.

6.7 The CRN Usage.

Having understood what CRN and designed its authentication and authorization (A-A) framework, it is also necessary to present the usage diagram for CRN in Fig.12 to show a cross section of the wireless devices in CRN utilizing the spectrum resources. It specifies the several device platforms of CRNS. This means that there is a facility embedded in the devices to enable access to the spectrum resources and enjoy the dividends provided by the network. The service providers are the primary users of the network and they also have end users. The organizations that depend on service providers for the supply and support of the network used to serve their clients constitute the secondary users or end users.

The design clearly explains how the spectrum resources are being utilized and the efficiency of service delivery. Cognitive Radio Network consists of several cognitive radio devices in compatible connection, interacting with each other and the environment to deliver quality services. They interact with the environment in a cognitive cycle which is a core inference mechanism for cognitive devices.

6.8 Spectrum Management Architecture

The spectrum management architecture is a very important aspect of this research project as it shows the different components that are involved in the overall management of the spectrum band. Security as discussed in this research project is an approach for the dynamic management of the spectrum resources (data and information) utilized in CRN. In other words, dynamic management of data and information is majorly about providing a reliable and secured communication of the usage of spectrum resources so as to ensure quality of service (QoS).

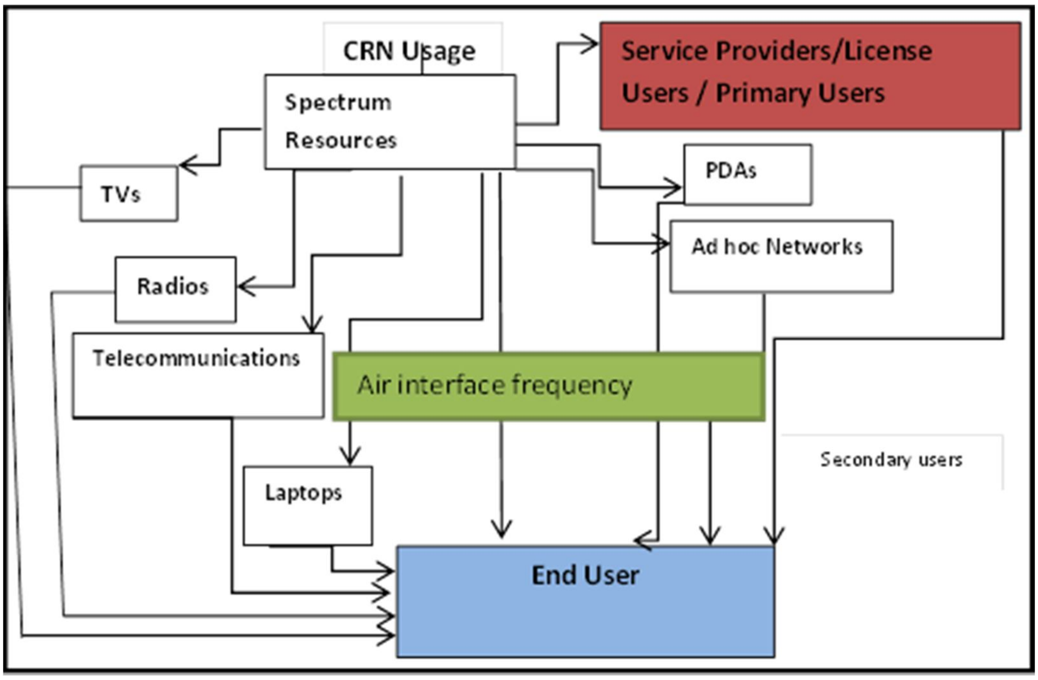


Fig. 9: Cognitive Radio Network Usage Diagram

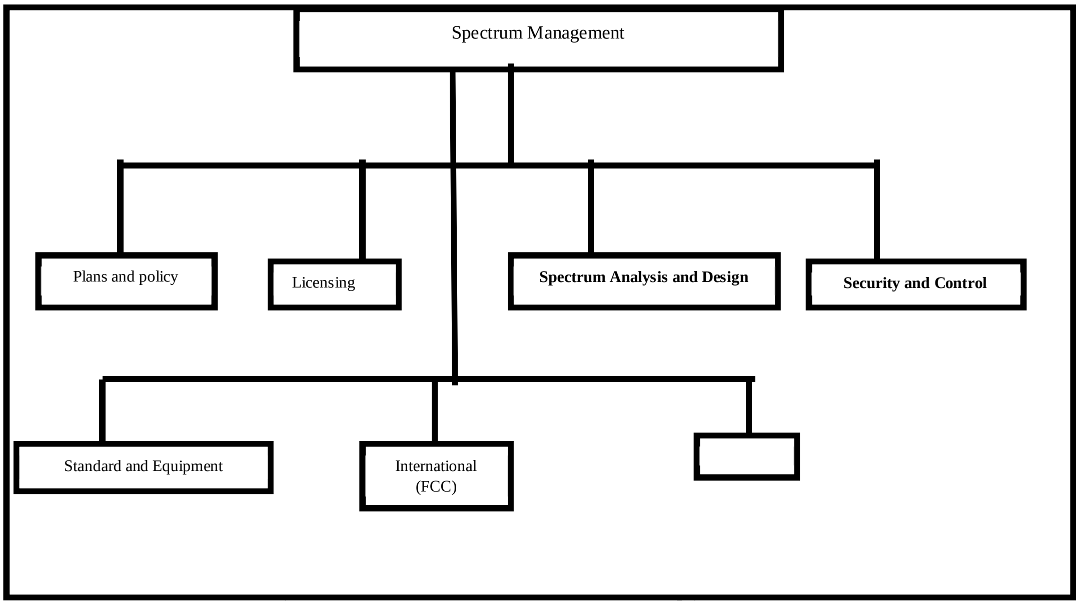


Fig. 10: Spectrum Management Architecture

The spectrum is similar to become a heterogeneous infrastructure, due to its distributed nature and the high rate of usage and deployment of wireless networks. Therefore, management of data, information and communication in such a distributed environment becomes necessary. The wireless devices operating within both the licensed and the unlicensed spectrum band are controlled and monitored to ensure security. However, the diagram above specifies the relationship and flexibility that exist between the spectrum and CR network employing different components of the spectrum management. The plans and policy entity comprises of, the regulatory policy, spectrum allocation and usage. The licensing entity comprises of, the application using the resource, its terms and condition of registration, review and renewal process. The spectrum analysis entity consists of, the design putting into consideration, interference, avoidance and mitigation. The spectrum control consists of, service policy, enforcement, compliance, control, monitoring and inspection. The standard and equipment identity consists of, authentication, authorization and accounting measures. The international entity consists of, the coordinating body, such as federal communication commission (FCC).

7. Framework Implementation Phase

The implementation phase demonstrates how CRN clients interact with the system with the aim of proving the concept of authentication and authorization framework for cognitive radio network.

It also shows how access to the services provided by the CR network is controlled and monitored using authentication and authorization access control mechanism as a protective measure against unauthorized and malicious users.

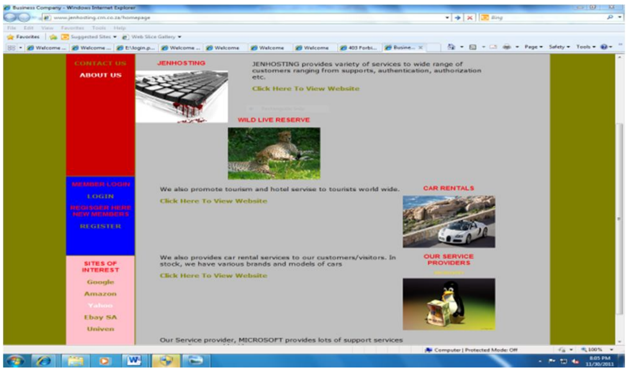
The different interfaces presented in this section indicate the clients' interactions with the system before access is either granted or denied to ensure effective and dynamic management of data and information in cognitive radio network.

7.1 Jenhosting CRN

The framework is implemented using **Jenhosting** Company (JHC). The company provides numerous services among which are mobile telephony, mobile services, mobile internet and fixed telephony as shown in Fig. 15b. It has numerous clients (subscribers) which include Vodacom, MTN, Celtel, Univen and others. The interface of Fig. 12 shows the CRN home page from which you can navigate to other network domain such as services offered by the network as shown in Fig.16, contact information as shown in Fig.15 and other information about the company as shown in Fig.14, including how to register as shown in Fig.16 and the login outcomes as shown in Fig.18a, Fig. 18b and Fig.18c.

a) Jenhosting CRN Company Home Page

The home page of JENHOSTING Company is the main page of the network, which is the entry point to the Cognitive radio infrastructure. It consists of the login button, the register button, including sites of interest shown in Fig.12 and other vital information about the services rendered by the company.



Business Company - Windows Internet Explorer

www.jenhosting.crn.co.za/homepage

JENHOSTING provides variety of services to wide range of customers ranging from supports, authentication, authorization etc.

Click Here To View Website

WILD LIVE RESERVE

We also promote tourism and hotel servise to tourists world wide.

Click Here To View Website

CAR RENTALS

We also provides car rental services to our customers/visitors. In stock, we have various brands and models of cars

Click Here To View Website

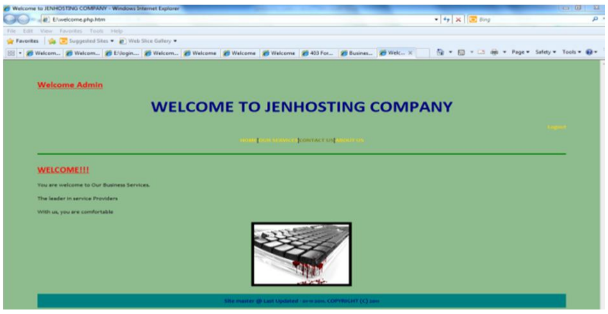
OUR SERVICE PROVIDERS

Our Service provider, MICROSOFT provides lots of support services

Fig.12: Cross Section Jenhosting CRN Home page

b) Jenhosting Welcome page

This shows the page that comes up when the new member button is clicked



Welcome to JENHOSTING COMPANY - Windows Internet Explorer

E:\welcome.php.htm

Welcome Admin

WELCOME TO JENHOSTING COMPANY

WELCOME!!!

You are welcome to Our Business Services.

The leader in service Providers

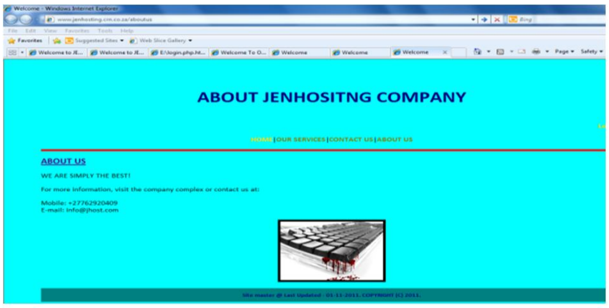
With us, you are comfortable

File Edit View Favorites Tools Help
Favorites
Suggested Sites
Web Slice Gallery
Page
Safety
Tools
Logout
HOME|OUR SERVICES|CONTACT US|ABOUT US

Fig.13: Jenhosting Welcome Page

c) Jenhosting CRN General Information Section

The Fig.14 and Fig.15 interface shows the outcome after the ‘About us’ and ‘Contact us’ button has been clicked from the home page. All necessary information about the network operations, services offered, including the contact information is viewed from these domains.



ABOUT JENHOSITNG COMPANY

HOME|OUR SERVICES|CONTACT US|ABOUT US

ABOUT US

WE ARE SIMPLY THE BEST!

For more information, visit the company complex or contact us at:

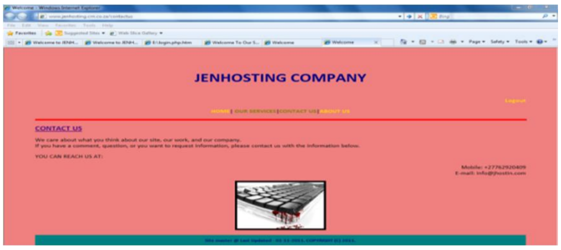
Mobile: +27762920409
E-mail: info@jhost.com

Site master @ Last Updated - 01-11-2011. COPYRIGHT (C) 2011.

Fig.14: Service Inquiries Page

d) Jenhosting CRN Contact Information Section

This page displays the contact information page when the contact button is clicked.



JENHOSTING COMPANY

Logout

HOME| OUR SERVICES|CONTACT US|ABOUT US

CONTACT US

We care about what you think about our site, our work, and our company.
If you have a comment, question, or you want to request information, please contact us with the information below.

YOU CAN REACH US AT:

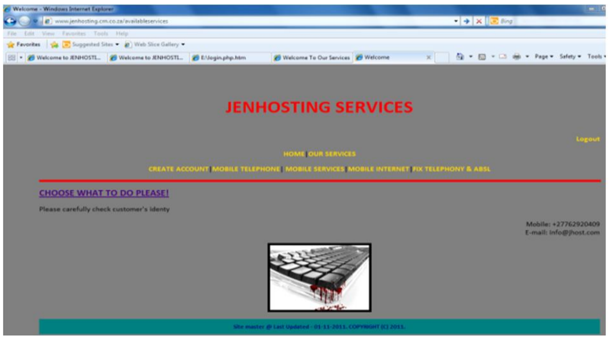
Mobile: +27762920409
E-mail: info@jhostin.com

Site master @ Last Updated - 01-11-2011. COPYRIGHT (C) 2011.

Fig.15a: Contact Page

e) Jenhosting CRN Services

This page displays both the services offered by the cognitive radio network and the available services at the time the service button is clicked.



JENHOSTING SERVICES

Logout

HOME|OUR SERVICES

CREATE ACCOUNT|MOBILE TELEPHONE| MOBILE SERVICES|MOBILE INTERNET|FIX TELEPHONY & ABSL

CHOOSE WHAT TO DO PLEASE!

Please carefully check customer's identy

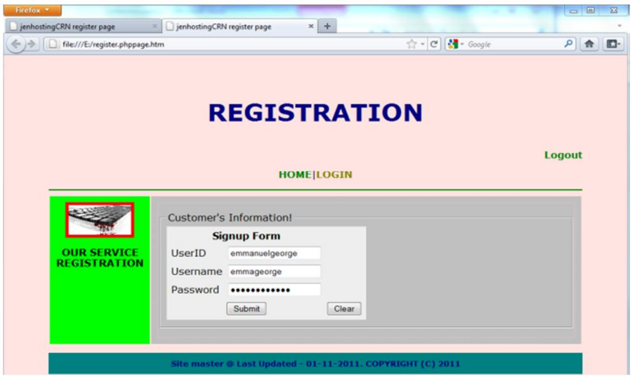
Mobile: +27762920409
E-mail: info@jhost.com

Site master @ Last Updated - 01-11-2011. COPYRIGHT (C) 2011.

Fig.15b: CRN Services Page

f) Clients e-Registration Section

All the basic information required for the registration of the clients based on the network service policy needed for authentication and authorization are captured from this domain and stored in the data base as shown in Fig. 19.



REGISTRATION

Logout

HOME|LOGIN

OUR SERVICE REGISTRATION

Customer's Information!

Signup Form

UserID emmanuelgeorge

Username emmageorge

Password ••••••••••••

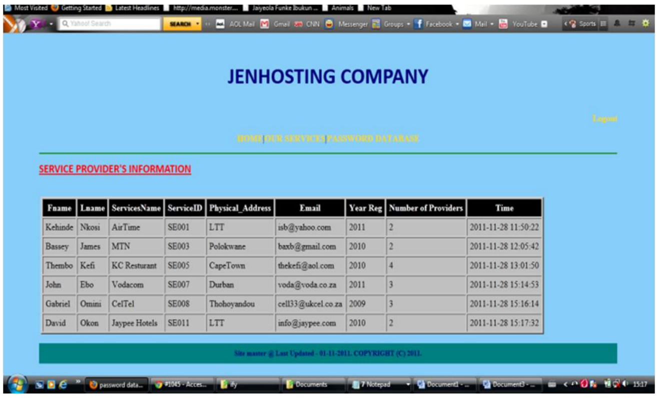
Submit Clear

Site master @ Last Updated - 01-11-2011. COPYRIGHT (C) 2011

Fig. 16: e-Registration Section

g) Jenhosting CRN Database

This represents the authentication and authorization management database and it consists of all the registered clients of the network. The clients name, service name, service ID, password, e-mail and year of registration are clearly specified and stored in this domain for authentication, authorization and security policy services.



JENHOSTING COMPANY

Logout

HOME|OUR SERVICES|PASSWORD DATABASE

SERVICE PROVIDER'S INFORMATION

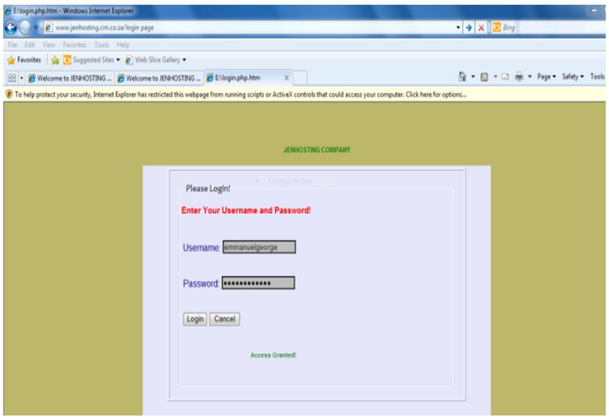
Fname	Lname	ServicesName	ServiceID	Physical_Address	Email	Year Reg	Number of Providers	Time
Kehinde	Nkosi	AirTime	SE001	LTT	isb@yahoo.com	2011	2	2011-11-28 11:50:22
Bassey	James	MTN	SE003	Polokwane	baxb@gmail.com	2010	2	2011-11-28 12:05:42
Thembo	Kefi	KC Resturant	SE005	CapeTown	thekefi@aol.com	2010	4	2011-11-28 13:01:50
John	Ebo	Vodacom	SE007	Durban	voda@voda.co.za	2011	3	2011-11-28 15:14:53
Gabriel	Omini	CelTel	SE008	Thohoyandou	cell33@ukcel.co.za	2009	3	2011-11-28 15:16:14
David	Okon	Jaypee Hotels	SE011	LTT	info@jaypee.com	2010	2	2011-11-28 15:17:32

Site master @ Last Updated - 01-11-2011. COPYRIGHT (C) 2011.

Fig. 17: Jenhosting CRN Database.

h) Successful Login

When a request for services is initiated, the client would need to login to the system by supplying identification details (username and password). The details would then be verified and validated from information already stored in the CRN client membership database. A successful login access is granted only if the user is who he claims to be as verified and validated from the database information. In situation where access is not granted, it therefore implies that the request is invalid and an unsuccessful login message would be displayed.



E:\login.php.htm - Windows Internet Explorer

www.jenhosting.crn.co.za/login page

To help protect your security, Internet Explorer has restricted this webpage from running scripts or ActiveX controls that could access your computer. Click here for options...

JENHOSTING COMPANY

Please Login!

Enter Your Username and Password!

Username: emmanuelgeorge

Password: •••••••••••••

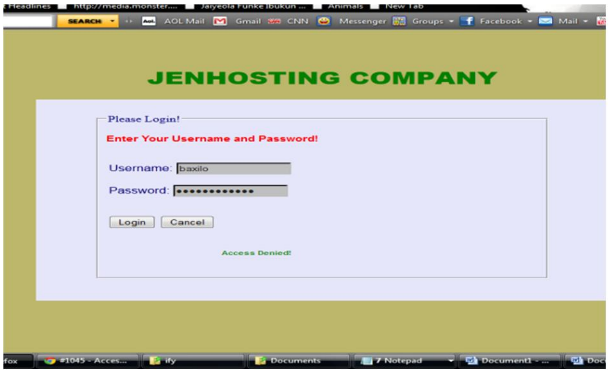
Login Cancel

Access Granted!

Fig.18a: Successful Login

i) Unsuccessful Login

Denial of access to resources during identification of users requesting for services is usually displayed with an unsuccessful login message. This usually happens when a non-registered client is attempting to request for rights of service usage. In such a situation, the system would display unsuccessful login message as a means not to allow malicious intruders into the available services. Unsuccessful login can only be adverted by service requesters registering with the service provider to be allowed access into the CRN resources.



JENHOSTING COMPANY

Please Login!

Enter Your Username and Password!

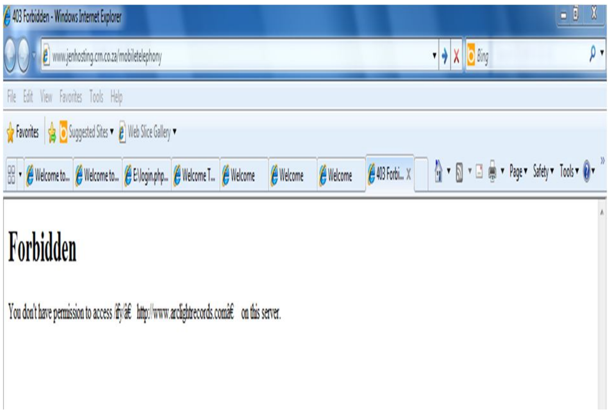
Username: baxilo

Password: •••••••••••••

Login Cancel

Access Denied!

Fig. 18b: Unsuccessful Login



403 Forbidden - Windows Internet Explorer

www.jenhosting.crn.co.za/mobiletelephony

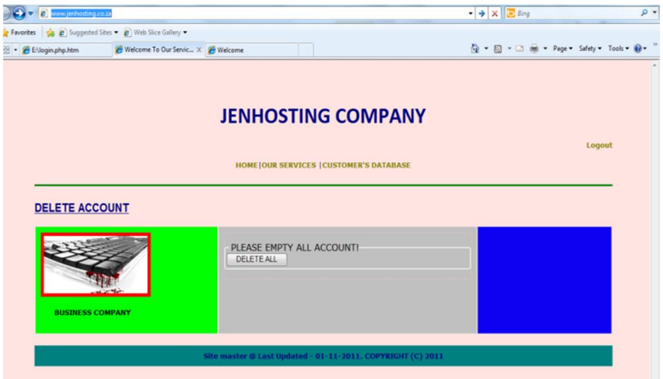
Forbidden

You don't have permission to access /ify/â€ http://www.arclightrecords.comâ€ on this server.

Fig.18c: Unsuccessful Login Section

j) Delete Account Section

This implementation phase ensures that no unauthorized user or malicious user masquerades as a legitimate user to gain access to the network server or the resources available in the network for malicious use. This section of the network has the capability to delete the user account and disable the root connections to such users to ensure efficient access control and effective dynamic management of data and information in the specified CR Network.



www.jenhosting.co.za

JENHOSTING COMPANY

Logout

HOME | OUR SERVICES | CUSTOMER'S DATABASE

DELETE ACCOUNT

BUSINESS COMPANY

PLEASE EMPTY ALL ACCOUNT!

DELETE ALL

Site master @ Last Updated - 01-11-2011. COPYRIGHT (C) 2011

Fig.19: Account Delete Section

7.2. Framework Evaluation

In this paper, we presented an authentication and authorization framework that forms the security infrastructure for access control that can dynamically manage data and information in CRN. It demonstrates how the framework is designed by transforming the artifacts from analysis phase. This paper also has other designs showing the authentication and authorization engine component, the spectrum resource broker component, the UML diagram for authentication and authorization sequence diagram, and the CRN usage diagram. The framework implementation phase consists of various diagrammatic interfaces displaying how the various components of CRN communicate using JENHOSTING cognitive radio network as a model for implementation. Consequently, dynamic management of data and information in CRN provides this reliable security infrastructure as an access control measure to check unauthorized access and all forms of malicious use of the spectrum resources.

Reported in this paper is the design and implementation of authentication and authorization security infrastructure which is able to provide access control and dynamically manage data and information in cognitive radio network to establish control against unauthorized and malicious intruders.

For this controls to be achieved authentication and authorization were introduced. User authentication and authorization is a crucial management component for securing data and information in CRN. Authentication and authorization framework are tightly-coupled mechanisms but also differ in some ways. Authorization process depends on secured authentication mechanism which ensures that a user is who he claims to and thus prevent malicious intruders from gaining access to the secured network resources but also differ in some ways. However, they both offer effective and efficient access control for the dynamic management of data and information in cognitive radio network.

8. Conclusion

The authentication framework designed in this research report is specifically for cognitive radio networks. The A-A server compares a user's authentication details with the user identification details stored in a database. If the details correspond, the user is granted access to the network. If both information differs the authentication process will fail, then access to the network service is denied.

Authorization is a security mechanism which determines the level of access a specific or particular authenticated user should have to the available and secured network resources. It determines whether a user has the authority to issue certain commands. However, the process enforces policies such as determining what types of activities, resources, or services a user is permitted to perform. The features used are compatible to only the cognitive radio network environment. It is designed to provide efficient and effective dynamic management of data and information in cognitive radio networks. It ensures that data and information are protected to enhance secured conversation.

Summarily, reported in this research is the design and implementation of a security framework that enforces access control policies for optimal spectrum resource management.

References

[1]. G. Staple and K. Werbach, “The End of Spectrum Scarcity,” IEEE Spectrum, Vol. 41, No. 3, Mar. 2004, pp.48–52.

[2]. S. Haykin, "Cognitive Radio: Brain-Empowered Wireless Communications," IEEE journal on selected Areas in communications, Vol. 23, No. 2,February 2005.

[3]. Y. Zhou, D. Wu, and S. Nettles. “Architecture of Authentication, Authorization and Accounting for Real Time Secondary Services”, International Jounal of wirwless and Mobile Computing, Vol xx, No x, Jan, 2005.

[4]. O.O. Ekabua, and M.O. Adigun. ”GUISET LogOn: Design and Implementation of GUISET- Driven Authorization Framework,” In Proc. of 1st International Conference on Cloud Computing, GRIDs, and Virtualization, November 21-26, 2010, Lisbon, Portugal pp. 1-6.

[5]. G. Baldini et al. “Security Aspect in Software Defined Radio and Cognitive Radio Networks: A Survey and a Way Ahead,” IEEE Journal, 1553-877x/11/, 2011.

[6]. S. Kumar et al. Ad Hoc Mobile Wireless Networks, www.ubebooks.com-free books and magazines.

CYBER CRIMES ANALYSIS BASED-ON OPEN SOURCE DIGITAL FORESICS TOOLS

[1]Victor O. Waziri PhD,
Department of Cyber Security Science;
School of Information and Communication Technology; Federal University of Technology,
Minna-Nigeria

Okongwu N. O
Economic and Financial Crimes Commission,
Nigeria
.

Audu Isah PhD,
Department of Mathematics/ Statistics
School of Federal University of Technology,
Minna-Nigeria

Olawale S. Adebayo
Department of Cyber Security Science;
School of Information and Communication Technology;
Federal University of Technology, Minna-Nigeria
.

Shafi'í Mohammed Abdulhamid
Department of Cyber Security Science;
School of Information and Communication Technology;
Federal University of Technology,
Minna-Nigeria
.

Abstract:
In this paper, we are present the digital forensic open source tools: Fiwalk, Bulk_Extractor, Foremost, Sleuth Kit, and Autopsy which are all Linux based forensic tools to extract evidences that could be presented in the court of law. Fiwalk reads a disk image and outputs a block of XML containing all the disk image of resident and deleted files. Foremost recovers files by using their headers, footers and data structures. The Sleuth Kit and Autopsy perform various aspects of file system analysis. The Autopsy Forensic Browser is a graphical web interface that presents the results generated by Sleuth Kit. This research project demonstrates the usefulness of the above-mentioned forensic tools for analysis and recovery of obliterated data from hard drives. This paper found that Sleuth Kit, Autopsy Forensic Browser, Fiwalk, Bulk_Extractor, and Foremost all provide effective file system analysis and recovery tool sets. The increasing complexity of storage devices requires that the investigator employs different forensic tool set to complement his arsenal of tools. No single digital forensic tool would be sufficient for an entire digital forensic investigation case. With this consideration, this paper employs various forensic tools. The demonstration of the effectiveness of these digital forensic tools utilized in this paper could serve as an alternative for investigators looking to expand their digital forensic tool set functionality in the court of law. Details of the experiments are fully given at the expense of bulkiness since this works is aim at enhancing the utilities of open source forensics tools applications.

Keywords: *Digital Forensics, Fiwalk, Foremost, Sleut Kits Bulk_Extractor, Autopsy, Linux, Ontologies*

1.

2. Introduction

To start writing on well known discipline such as digital forensics is always ideologically complex to do so. Forensics Computing may be construed as a methodical series of techniques and procedures for accumulating evidence. As in (Anthony et al, 2007) and (Garfinkel, 2009 A), Cybercrimes is on the rise and could be detected using Digital Forensics tools for which the crimes are extracted from various storage devices and digital media. Such systematic analyzed extractions could be presented in the court of Law in a sequential and meaningful format as evidences. Two types of test questions (Brian Carrier, 2002) should be applied by investigators for both computer forensics and traditional forensics to survive in a court of law. These are:

Authenticity: Where does the evidence come from?

Reliability: Is the evidence reliable and free of flaws?

With these basic test questions, computer crime investigations should be predetermined through policy and what is acceptable risk" to every organization. Cybercrime includes the followings as outline in (Marcella et al, 2008), (Anthony et al, 2007) and (David, 2008):

1. **Theft of Intellectual Property.** This pertains to any act that allows access to potent, trade secrets, customer data, sales trends, and any confidential information
2. **Damage of Company Service networks:** This could occur if someone plants a Trojan horse, conducts a denial of service attack, installs an unauthorized modem, or installs a backdoor to allow others to gain access to the network or system
3. **Financial Fraud:** This denotes to anything that may use fraudulent solicitation to prospective victims to conduct fraudulent transactions
4. **Hacker System Penetration:** These occur via the use of sniffers, rootkits and other tools that take advantage of vulnerabilities of the systems or software.
5. **Distribution of Execution Virus and Worms:** These are Some of the most common forms of Cyber crime

Cyber Crime maybe spelt out into three comprises known as the "3Ts":

i. Tools to commit crime;
ii. Targets of the crime (Victim); and
iii. Material that is tangential to the crime

Divergent methods of cyber crimes abound and include amongst others, the following without further explanations due to space constraints: The Computer Facilitated Crimes that involves both insiders and external attacks. All these could be categorized to into various examinational patterns for analyses.

2.2 Rules of Computer Forensics

A good forensics investigator is expected to follow these suggestive rules as outlined in (Anthony et al., ibid):

i. Examine original evidence as little as possible. Instead, should examine the duplicate of the original evidence known as the image
ii. Should follow the rules of evidence and do not temper with the evidence
iii. Always prepare a chain of custody, and handle evidence with care
iv. Never exceeds the knowledge based on the investigative laid down rule by the court
v. Should document any changes of evidence

The rest of this paper runs as follows; Section 2 reviews the related works based on digital forensics devices of open source tools; section 3 outlines the methodologies of the research; in section 4, we perform some experiments based on the open sources tools for the computer forensics. Section 5 discusses the general computing outputs while section 6 gives further hints for future research investigations

3. Related Works

It is admissible that computers have become part of our lives worldwide (Brian, 2003). With the exploitation of web Technologies, so also vast exploits of the technologies have been established by criminals to commit crimes. Due to large and complex involvement of computers in web businesses and other intricate utilities in computing, computer is emerging as legal evidence in both civil and criminal cases.

Computer evidences are admitted in courts of law and these evidences could be anyfile or fragment recovered from the storage devices such as email, browsing history, graphics, photographs, or application documents. These files could be extracted from the hardisks and imaged to recover undeleted and deleted files. Deleted file recovery would require special techniques to retrieve them and this is what we are set out to achieve. These are professionally retrieved in a non-destructive technique. Evidence may be recovered from storage medium installed in digital equipment such as computers, cameras, PDAs, or cell phones [Gialanella et, 2008]. All forensics work should be strategically documented in a clear a system extraction; a principle known as chain of custody; in other to for the evidence to be admissive in the court of law

Computer devices that can establish evidence in the court of law are part of our lives. There have been a lot of some works on digital forensics community to create a common file formats, schemas and ontologies [13]. Despite all these efforts for a common need of affiliation, there has been little concrete standardization. As stated [13], DFRWS started the common Digital Evidence Storage Format (CDESF) Working group in 2006; in which the group created a survey of disk image storage formats in September, 2006. Due to lack of resources, the group disbanded in August, 2007. Hoss and Carver discussed ontologies to support digital forensics (carver and Hoss, 2009), but did not propose any concrete ontologies that can be used. Garfindel introduced an XML, representation for the system metadata (Garfindel et al, 2009), but it has not been universally adopted.

In another development, Richard and Roussev reviewed requirements for the “Next Generation Digital Forensics”. Their works emphasized on system requirements with the argument that that inefficient system design, wasted CPU cycles, and the failure to deploy distributing unified techniques could introduce significant and unnecessary delays that directly translate into ucecessary delays (Richard and Roussev, 2006).

(Politt,2007) reviewed 14 different models for digital forensics investigation but did not attempt to evaluate or catalog them given time constraints. Bradford et al, 2004) argue that it is unwise to depend on upon “audit trails and internal logs” and further postulate that the digital forensics will only be possible on future systems if those systems make proactive efforts at data collection and preservation. Hey proposed a mathematical model for deciding the content and frequency of proactive forensic event recorders. Politt et al., further discussed how virtualization software and techniques could be productively applied to both digital forensics research and education (Polit et al., 2008). They argued that any discussion of virtualization with respect to digital forensics would face an unwelcomed tautology. In effect, the impact of virtualization on forensic examination could virtually be ignored-except when it could not. This is due to the fact that virtualization, and sometimes the subject of virtualization is the subject of the forensic examination, and sometimes the virtualization as a tool is used by forensic examiner.

The literature like other disciplines goes on with different opinions and approaches. For instance, Turnbull et all performed a detailed analysis on the specific-digital media formats being collected by the South Australian Police Elctronics Crime section, theirs appears to be more the first quantitative analysis of its kind (Turnbull et al., 2009)

This paper is concerned with the application of some open sources for digital forensics as evidence in the court of law. We are applying Fiwalk, Bulk_Extractor, Foremost, Sleuth Kit, and Autopsy which are all Linux based forensic tools that could downloaded as open software.

3 Methodology

This section describes the practical experimental methods carried out in this research paper using digital forensic open source tools.

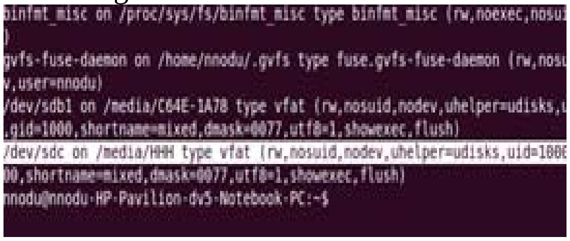
3.1.1 Experimental Analysis

The following tools listed below were used in the experimentations.

1. **Foremost version;**
2. **Fiwalk**;
3. **Bulk_Extractor** were compiled in Ubuntu 10.10; and
4. **Sleuth Kit and Autopsy** was compiled on both Ubuntu 10.10 and Windows 7 (Cygwin).

3.1.2 Steps to recover files from pen drive

1. Using the mount command, the pen drive has been assigned the mount point /dev/sdc, on /media/HHH and file type is fat; as illustrated from Fig 3.0.



```
binfmt_misc on /proc/sys/fs/binfmt_misc type binfmt_misc (rw,noexec,nosui
)
gvfs-fuse-daemon on /home/nnodu/.gvfs type fuse.gvfs-fuse-daemon (rw,nosu
v,user=nnodu)
/dev/sdb1 on /media/C64E-1A78 type vfat (rw,nosuid,nodev,uhelper=udisks,u
,gid=1000,shortname=mixed,dmask=0077,utf8=1,showexec,flush)
/dev/sdc on /media/HHH type vfat (rw,nosuid,nodev,uhelper=udisks,uid=100E
00,shortname=mixed,dmask=0077,utf8=1,showexec,flush)
nnodu@nnodu-HP-Pavilion-dv5-Notebook-PC:~$
```

Fig 3.0: use of the mount command to display current mount points

We then use the 'cd' command to navigate the pen drive and display contents of the drive. The full command is given as follows:

1. Command to navigate to pen-drive: **cd /media/HHH**

Command to list contents of a device/folder and show space it occupies on disk: **ls –s**

2 The command for launching the FIWALK:

fiwalk -X <file> < diskimage> -v
-X<file> = XML output to a <file> (full DTD)

3 The command for launching the bulk_extractor:

bulk_extractor -o output_dir [options] image;

4. Command to launch autopsy: **/autopsy**

Web link for autopsy forensic browser: **http://localhost:9999/autopsy**

Location of the evidence locker: **/cygdrive/j/evidence.**

5. **Command to create an image hard drive**
The command for launching the creation of image hard drive:

dd if=/dev/sdb ibs=512 of=/media/Passport/flashimage.img.dd

At the pen-drive's directory, we used the command above to display file contents and block size in bytes. The files are listed below:

0321680561.pdf, and 0321680561.rar, bluehills.jpg, waterlillies.jpg, sunset.jpg, winter.jpg

Foremost should not be run from the folder/device you wish to recover data from. So navigate to a folder we created on the desktop called recovery. A folder called 'output' will be the result of our recovery.

```
oot@nnodu-HP-Pavilion-dv3-Notebook-PC
otal 0
```

Fig3.1: using the command formost to display the pen drive contents, this is currently empty

3.2 Using Fiwalk to Process a 80-Gigabyte Disk Image in Order to Produce a Digital Forensics XML

We use Fiwalk to produce a digital Forensics XML in this sequential order: The disk image called diskimage.img is stored on an external hard drive. The external hard drive's name is Passport. The syntax to invoke Fiwalk is already given in the last subsection.

3.3 Using Bulk Extractor to Analyze Disk Image for Domain Names, Wordlist, Log-file, and Emails Accessed from Drive

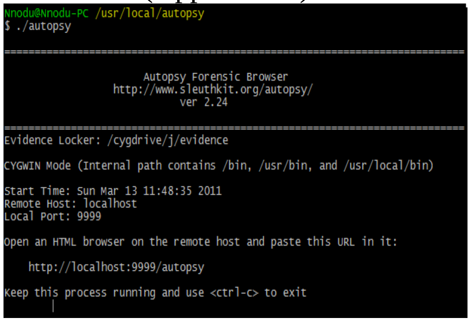
The disk image of 80-gigabyte hard drive is processed below. The Syntax for using bulk extractor is stated in the subsection above.

We navigate to the hard disk drive containing the 80 gigabyte hard drive image called diskimage.img, and invoke bulk extractor to start processing. The output directory is in /media/local/

3.4 Using autopsy and Sleuth Kit to perform Volume and File System Analysis on a 80 gigabyte hard disk

Procedure:

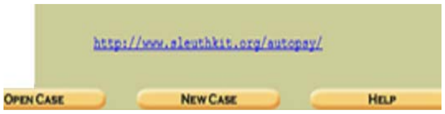
Invoke the autopsy by the use of the command see (Appendix B):



```
Nnodu@Nnodu-PC /usr/local/autopsy
$ ./autopsy

============================================================================

                         Autopsy Forensic Browser
                     http://www.sleuthkit.org/autopsy/
                                ver 2.24

============================================================================
Evidence Locker: /cygdrive/j/evidence

CYGWIN Mode (Internal path contains /bin, /usr/bin, and /usr/local/bin)

Start Time: Sun Mar 13 11:48:35 2011
Remote Host: localhost
Local Port: 9999

Open an HTML browser on the remote host and paste this URL in it:

    http://localhost:9999/autopsy

Keep this process running and use <ctrl-c> to exit
```

Fig3.2: Autopsy Forensic Browser showing that the invocation was successful

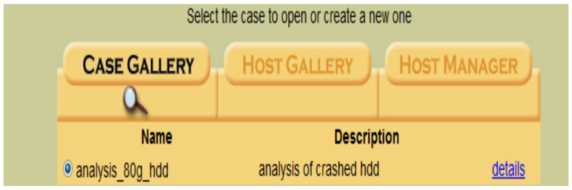
1. From Fig3.2, we will open a HTML browser and paste the address as depicted in the command above



http://www.sleuthkit.org/autopsy/

Fig3.3: The autopsy forensic browser interface opens on a web browser

2. We already have a previous case as shown from Fig 3.4

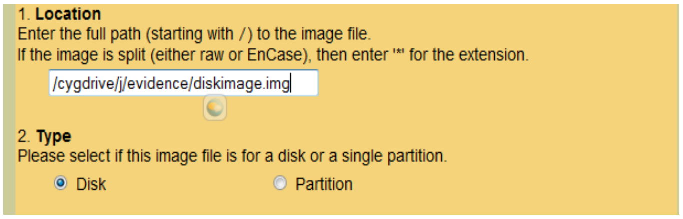


Select the case to open or create a new one

	Name	Description	
●	analysis_80g_hdd	analysis of crashed hdd	details

Fig3.4: Interface showing a previous case file which was called analysis_80g_hdd and was described as having crashed (damaged)

3. A new image is added which must be in the evidence locker and autopsy, accounts for the image file by creating a symlink (symbolic link)



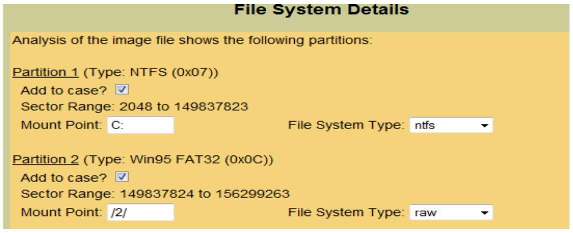
1. **Location**
Enter the full path (starting with /) to the image file.
If the image is split (either raw or EnCase), then enter '*' for the extension.
/cygdrive/j/evidence/diskimage.img

2. **Type**
Please select if this image file is for a disk or a single partition.
● Disk ○ Partition

Fig.3.5: The image file path for the investigation is added and the type '.img' is also stated so autopsy can know what kind of image file is being investigated

From Fig3.5, the image file path for the investigation is added which has an extension of .img From Fig3.6 Analysis of the file system of the 80 gigabyte Hard disk shows two partitions:
Partition 1 of mount point C: is of type NTFS, sector range from 2048 to 149837823. Partition 2 is of type RAW,

sector range from 149837824 to 156299263 With mount point at /2/



File System Details

Analysis of the image file shows the following partitions:

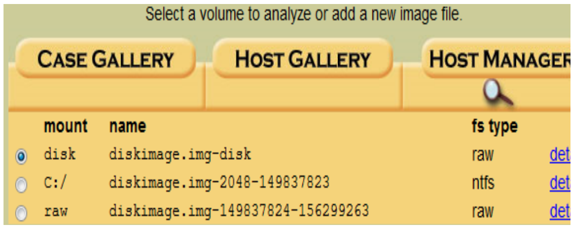
Partition 1 (Type: NTFS (0x07))
Add to case? ☑
Sector Range: 2048 to 149837823
Mount Point: C: File System Type: ntfs

Partition 2 (Type: Win95 FAT32 (0x0C))
Add to case? ☑
Sector Range: 149837824 to 156299263
Mount Point: /2/ File System Type: raw

Fig3.6: After importing the diskimage: Autopsy details a summary of the File System and Image File details

All the units are in 512 byte sectors

The add image button was clicked and the image was successfully added linked to the evidence locker and linked as shown in Fig3.8. The ok button is then clicked again to continue.



Select a volume to analyze or add a new image file.

	mount	name	fs type	
●	disk	diskimage.img-disk	raw	det
○	C:/	diskimage.img-2048-149837823	ntfs	det
○	raw	diskimage.img-149837824-156299263	raw	det

Fig3.7: Autopsy shows the mount points (partitions) and names with file system type

From Fig 3.7, The 80 gigabyte hard disk is displayed on three mount points

3.4 Investigation and Analysis

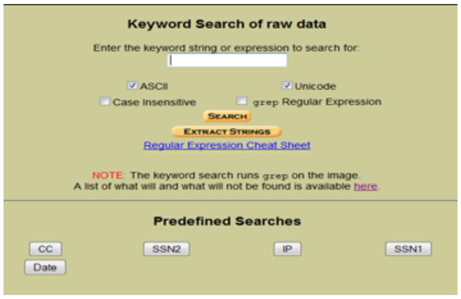
In this section investigation and analysis of the experiments performed are conducted

3.4.1 Analysis of mount disk of name "DISKIMAGE.IMG-DISK"

a. PROCEDURES

1. The tab labeled Analyze when clicked to start the analysis reveals another window opens with three modes of analysis namely:

Keyword Search, image details and Data Unit



Keyword Search of raw data

Enter the keyword string or expression to search for:

☑ ASCII ☑ Unicode
□ Case Insensitive □ grep Regular Expression

NOTE: The keyword search runs grep on the image.
A list of what will and what will not be found is available here.

Predefined Searches

Fig3.7.1: Autopsy shows the keyword option for searches available, predefined searches are also listed to help the investigator

2. A search is performed with the keyword 'bank' but first to make the search faster, the entire strings in the disk image is extracted. This is done by the use of the 'EXTRACT STRINGS' tab.

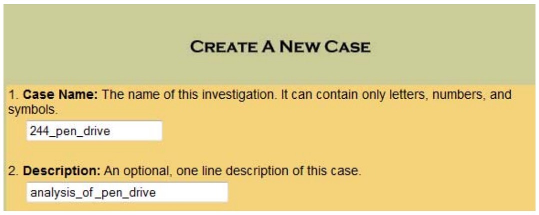
b. Observation

The extraction of string was taking a long time, maybe because of the size of the 80 gigabyte hard drive. Even when the extraction was carried on a quad core windows 7 64 bit system with ram of 8gigabyte, it was still slow. So I decided to try a smaller storage device.

The storage device to be used is the same drive in which I performed a search using Foremost. The state of the pen drive though has changed. An open office document of size 244 megabytes was added to the pen-drive.

3. The **dd** command was utilized in Ubuntu to make an image of the pen-drive the full syntax is shown Appendix B

4. A new case file is created for the 244mb Pen drive as shown it **Fig3.92.** Autopsy recognized the file system type to be fat 16.



CREATE A NEW CASE

1. **Case Name:** The name of this investigation. It can contain only letters, numbers, and symbols.
244_pen_drive

2. **Description:** An optional, one line description of this case.
analysis_of_pen_drive

Fig3.7.2: Opening a new Case in Autopsy

3.5 Materials used in Experimental Analysis

a. Programs Used

Foremost Version 1.1, Fiwalk, Bulk_Extractor and Autopsy

b. Operating Systems Used

Windows 7 Version 6.1.7600 Build 7600, System Type: X64 based PC.

Installed Physical Memory = 3.00 Gigabyte; Ubuntu 10.10 (Linux) and Cygwin

3.5.1 **Choice of Materials**

Foremost was used because the program size is not large and easy to compile in Linux based system. **Fiwalk and Bulk_Extractor:** was used because the programs were easy for use and system resources. The recovery completed without any problems. **Ubuntu, Cygwin:** These operating systems that are Linux based. Most of these Forensic Tools are originally ported from Linux and therefore are easy to compile on Linux based environments.

3.5.2 Foremost

Foremost was used in this research paper to recover deleted data from a 512 megabyte pen drive.

This software recovers files using their headers, footers, and data structures.

The syntax for foremost usage see section 3.1

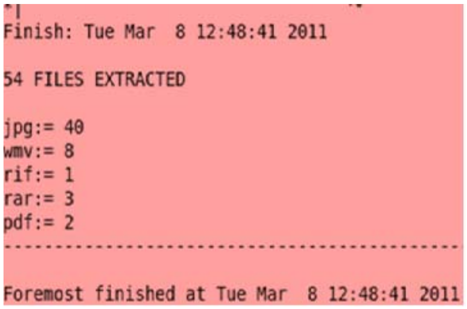
3.5.3 Fiwalk

This was used to analyse a 80-gigabyte Hard drive and an XML report of the entire structure was generated. Bulk_Extractor was used to recover email address, web domain addresses and histogram reports accessed from the Hard drive. Fiwalk makes it easy for non-experts to do significant forensic research and write powerful forensic tools (DEEP.). Based on Sleuth Kit, XML, and the Python programming language, this approach makes it easy for programmers to create tools that perform forensic processing without the need to master domain-specific knowledge (Garfinkel, 2009).

4. The Experimental Results

This section presents the results and findings from the materials and method of section 3. The results are from the experimental analysis performed in the methods sections are hereunder given in this pattern.

4.2 **Results of recovery operation on Pen drive using Foremost**



```
Finish: Tue Mar  8 12:48:41 2011

54 FILES EXTRACTED

jpg:= 40
wmv:= 8
rif:= 1
rar:= 3
pdf:= 2
------------------------------------------

Foremost finished at Tue Mar  8 12:48:41 2011
```

Fig4.0: **This screen shot shows the result of the recovery**

54 files were recovered.

Jpg = 40;   wmv= 8; rif = 1;
rar = 3; pdf = 2

A summary of the audit.txt contains a report of what formost has done using the command in section 3. We will view some contents of the audit.txt file which is displayed below:

Foremost started at Tue Mar 8 12:48:34 2011

Invocation: foremost -T -v -t all -i /dev/sdc

Output directory: /home/nnodu/Desktop/recovery/output_Tue_Mar__8_12_48_34_2011

Configuration file: /etc/foremost.conf

File: /dev/sdc

Start: Tue Mar 8 12:48:34 2011

Length: 244 MB (256900608 bytes)

…..……

……..

Finish: Tue Mar 8 12:48:41 2011

54 Files extracted which are not given here for want of space:

jpg:= 40; wmv:= 8; rif:= 1; rar:= 3; pdf:= 2

Foremost finished at Tue Mar 8 12:48:41 2011

We will then navigate to the output file containing the recovered files with this command as stated in section 3:

The output file contains six files of size 24 bytes: namely audit**.txt, pdf, rar wmv, avi, jpg**

4.2.1 **Examination of JPG Files recovered from Pen drive using Foremost**

Procedures:

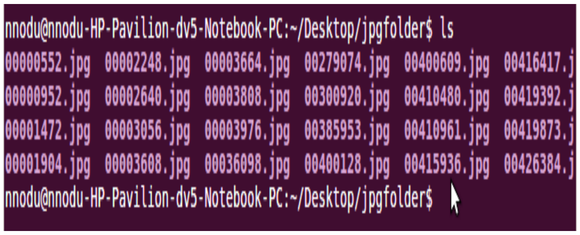
a. **Step1**.

To examine the .jpg folder, (Appendix C3)' , I navigate to the jpg folder (Appendix C 4)

b. **Step2.**

So I then copied the 40 jpg files (Appendix C) from the folder to another folder on my desktop called jpgfolder.

I did this so I would be able to view the jpg files. So from Fig4. I now display the jpg files recovered. The files recovered were tested and found to be in good condition.



```
nnodu@nnodu-HP-Pavilion-dv5-Notebook-PC:~/Desktop/jpgfolder$ ls
00000552.jpg  00002248.jpg  00003664.jpg  00279074.jpg  00400609.jpg  00416417.j
00000952.jpg  00002640.jpg  00003808.jpg  00300920.jpg  00410480.jpg  00419392.j
00001472.jpg  00003056.jpg  00003976.jpg  00385953.jpg  00410961.jpg  00419873.j
00001904.jpg  00003608.jpg  00036098.jpg  00400128.jpg  00415936.jpg  00426384.j
nnodu@nnodu-HP-Pavilion-dv5-Notebook-PC:~/Desktop/jpgfolder$
```

Fig 4.1 showing 40 jpg files recovered

4.3 **Discussion of Results**

The results obtained from the preceding experiments are discussed

a. **Results of Recovery operation performed on Pen Drive**

We discovered that the intial jpg's on the pen drive before the deletion, namely:

Blue hills, water lilies, sunset.jpg and winter jpg were recovered and renamed according to 0003608.jpg, 00003664.jpg, 00003808.jpg and 0003976.jpg by foremost. But because of copyright issues we will not display them.

b. **Examination of Pdf Files recovered**

Just like in the examination of the jpg files I will navigate to the jpg folder by using the 'cd' commnd and repeat the two step process. Using ls I will list the contents of the file .

There are currently two recovered PDFs namely 00030904.pdf and 00273880.pdf in the pdf/ folder. We then repeated steps in examination of jpg files, by using the 'cp' command, we will copy the recovered pdf files from their current folder to a folder on the desktop I created called pdffolder. The command is shown in Appendix C 6:

Fig 4.2 shows the recovered .pdf folders

Fig4.2: A screen shot of recovered pdf files

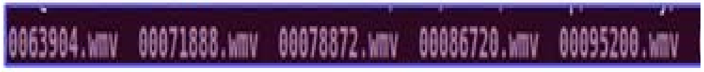
c. **Findings**

1. 00030904.pdf and 00027880.pdf are ' actually one and the same file, which is the same as 0321680561.pdf that was deleted

2. The two files recovered are the same pdf, probably saved at different times on the pen drive. And on examination the pdfs were found to be in good condition.

d. **Examination of WMV Files recovered**

All the procedures above in recovery for the pdf and jpg's are repeated.



```
0063904.wmv  00071888.wmv  00078872.wmv  00086720.wmv  00095200.wmv
```

Fig4.3: showing eight recovered WMV files

From Fig4.3 eight WMV files were recovered. I then copy them to a folder called wmvfolder created on Desktop. This is for easy examination of the files.

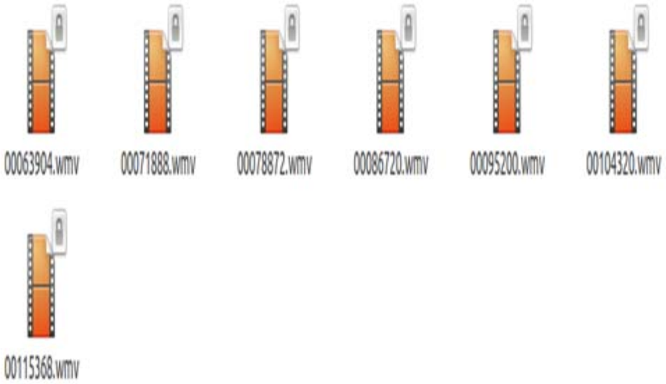
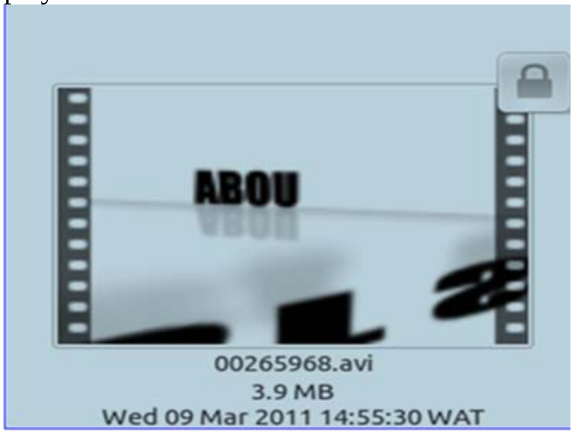


Fig4.4: showing eight recovered wmv files

From Fig4.4, the wmv files recovered were in good condition

e. **Examination of AVI Files recovered**

To examine the avi files recovered, I would simply repeat the two step followed in the recovery of jpeg. An avi file of size 3.9mb was recovered and it is displayed in the screen shot of Fig4.5. The avi file was in good condition when I used VLC to play it.

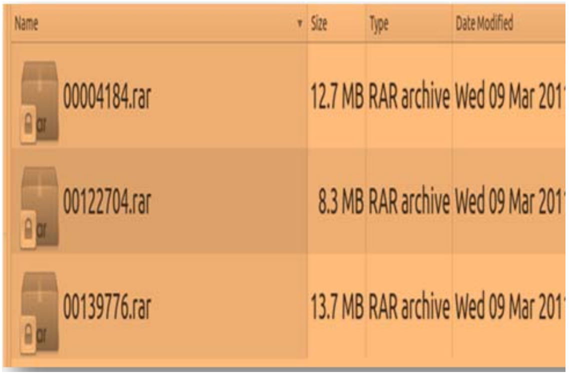


00265968.avi
3.9 MB
Wed 09 Mar 2011 14:55:30 WAT

f Examination of RAR Files Recovered

Repeating the procedures of recovering jpg files, 3 rar files were recovered namely, 00004184.rar, 00122704.rar, and 00139776.rar . This is displayed in Fig4.6

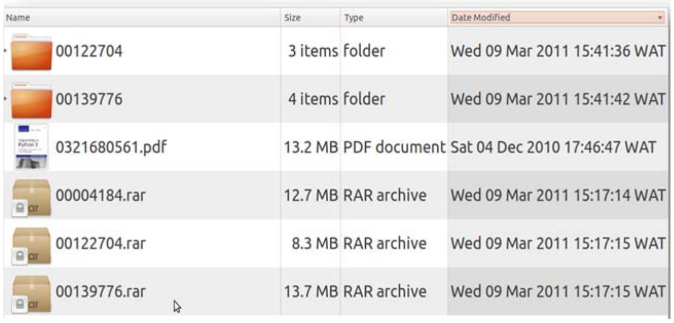
Fig 4.7 is a screen shot of the product of the extraction of the rar files, namely folder 00122704, 00139776 and 0321680561.pdf



Name	Size	Type	Date Modified
00004184.rar	12.7 MB	RAR archive	Wed 09 Mar 201
00122704.rar	8.3 MB	RAR archive	Wed 09 Mar 201
00139776.rar	13.7 MB	RAR archive	Wed 09 Mar 201

Fig4.6: Showing three rar files

recovered with their sizes also displayed



Name	Size	Type	Date Modified
00122704	3 items	folder	Wed 09 Mar 2011 15:41:36 WAT
00139776	4 items	folder	Wed 09 Mar 2011 15:41:42 WAT
0321680561.pdf	13.2 MB	PDF document	Sat 04 Dec 2010 17:46:47 WAT
00004184.rar	12.7 MB	RAR archive	Wed 09 Mar 2011 15:17:14 WAT
00122704.rar	8.3 MB	RAR archive	Wed 09 Mar 2011 15:17:15 WAT
00139776.rar	13.7 MB	RAR archive	Wed 09 Mar 2011 15:17:15 WAT

Fig4.7 shows the products of the

extraction of the rar files

4.3.1 Findings from Fiwalk Recovery Process

1. Archive file 00004184.rar is the same as archive 0321680561.rar I deleted at the start of the experiment.
2. All the archive rar files were tested and found to be in good condition.

The XML report is saved in Passport as diskimage.xml

The result of an extract from the file using fiwalk is displayed below briefly:

```
fiwalk xmloutputversion="0.3">
<metadata>
<dc:type>Disk Image</dc:type>
</metadata> <creator>
          ….
          …..
<command_line>fiwalk -X0
/media/Passport/diskimage.img -
v</command_line>
<uid>0</uid>
<username>root</username>
<start_time>Thu Mar 10 10:29:27
2011</start_time>
```

The result of the bulk_extractor process is briefly shown below:

```
All Threads Finished!
Phase 2. Creating Histograms
ccn ccn_track2 domain email kml rfc822
telephone url zip 0:
make_histogram(://([^/]+),services) ->
/media/local//url_services.txt
          …….
          ……..
# inputs: 154966024  outputs: 209336
# total time: 17288230 msec
```

# elapsed time: 9182.6 seco

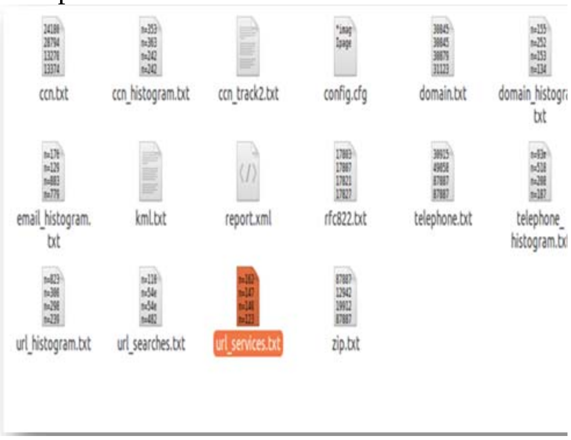


Fig4.8: Screen shot shows bulk_extractor's generated result

From Fig4.8, a summary of the results indicate the following files recovered:

1. **Report.xml** = showing an XML report of the extraction process.
2. **Zip.txt** = shows zip files files described by length, compression method and version.
3. **Url.txt**= a histogram of all URL's by domain
4. **Url_searches.txt** = a histogram of all search items, including Google, Yahoo, Bing
5. **Url_histogram.txt** = shows frequency distribution of the url sites accessed from the drive
6. **Telephone_histogram** = shows frequency distribution of telephone addresses used for a transaction on the system
7. **Telephone.txt** = shows the telephone addresses used for a transaction on the harddisk drive
8. **rfc822.txt**= shows the documents saved on the harddrive, such as letters, memo's etc.
9. **email_histogram.txt** = shows frequency distribution of email addresses accessed from the hard disk
10. **email.txt** =shows frequency distribution of email addresses accessed from the hard disk
11. **domain_histogram** = shows frequency distribution of domain addresses accessed from the hard disk
12. **domain.txt** = shows domain addresses accessed from the hard disk
13. **ccn.txt** = this file reports Federal Express Account numbers

a. **Observation**

In the domain_histogram.txt file, domain frequencies of occurrences were listed from the most occurring to the least occurring.

Then, a summary of the XML report detailing the Bulk_Extractor processes just executed has been summarized in the Fig 4.9. This shows the **url** values values extracted have the highest frequency of 334847 while the lowest are the **zip** files . While the telephone records extracted have the lowest value of 1915 records.

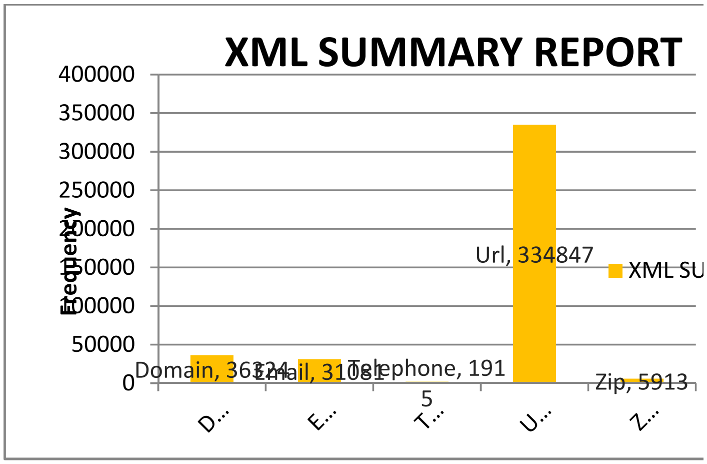
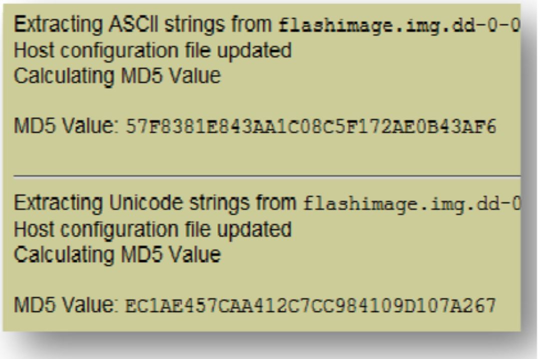


Fig 4.9 The summary of the XML report detailing the Bulk_Extractor processes just executed, the X axis contains the report recovered.

4.3.3 Results of Keyword Search Performed with Sleuth-Kit

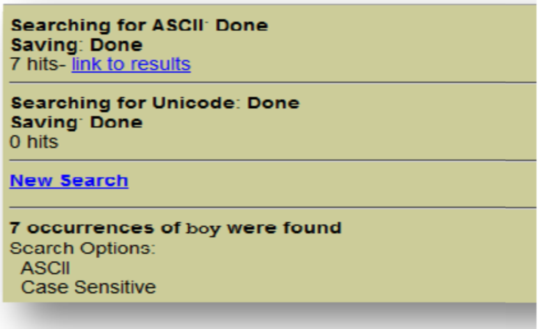
1. An attempt is made to extract the strings for the keyword search to be faster

2. The extraction of strings was successful as seen from Fig4.91



```
Extracting ASCII strings from flashimage.img.dd-0-0
Host configuration file updated
Calculating MD5 Value

MD5 Value: 57F8381E843AA1C08C5F172AE0B43AF6

Extracting Unicode strings from flashimage.img.dd-0
Host configuration file updated
Calculating MD5 Value

MD5 Value: EC1AE457CAA412C7CC984109D107A267
```

Fig4.9.1: The successful operation of string extraction is detailed with the MD5 values stated for ASCII string and Unicode string extraction



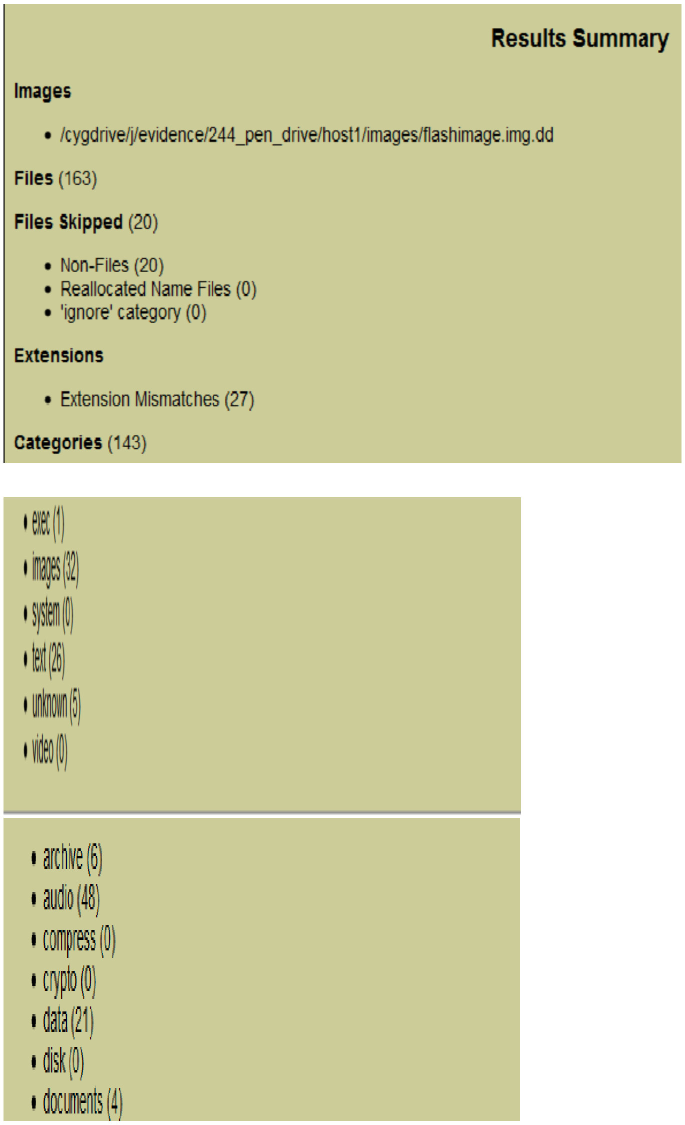
```
Searching for ASCII: Done
Saving: Done
7 hits- link to results

Searching for Unicode: Done
Saving: Done
0 hits

New Search

7 occurrences of boy were found
Search Options:
  ASCII
  Case Sensitive
```

Fig4.9.2: Showing some hits of search for 'boy' under ASCII, but none for the Unicode search

3. A search for boy was executed and the seven hits were obtained from the ASCII section as shown from Fig4.92, but none from the Unicode search. 7 occurrences of boy were found from the search. Sectors 12572, 199482,256193, 305054, 380781, 432376, 452504,

The contents of sector 12572 are of type ASCII. The result can be exported/ saved; this is useful for recovery of documents. In addition, notes can be added to the recovered sector by the investigator to note points of interest. The notes tab is close to the EXPORT CONTENTS TAB.

4.3.4 File Type Sorting By Category



Results Summary

Images

- /cygdrive/j/evidence/244_pen_drive/host1/images/flashimage.img.dd

Files (163)

Files Skipped (20)

- Non-Files (20)
- Reallocated Name Files (0)
- 'ignore' category (0)

Extensions

- Extension Mismatches (27)

Categories (143)

- exec (1)
- images (32)
- system (0)
- text (26)
- unknown (5)
- video (0)

- archive (6)
- audio (48)
- compress (0)
- crypto (0)
- data (21)
- disk (0)
- documents (4)

Fig4.9.3: A result of the file sorting of the pen-drive

In Fig4.9.3, For the Categories, this explains the types of files identified from the pen drive, 163 files were discovered from the file-sorting process.

Archive files found = 6, audio file = 48, compressed files = 0, crypto = 0 Data files = 21, Disk images = 0, Documents = 4

Executable files =1 Images recovered = 32

System = 0 Text = 26 Unknown = 5

Video = 0

S/No	Detail	Number(s)
1	Archive	6
2	Audio	48
3	Compress	0
4	Crypto	0
5	Data	21
6	Disk	0
7	Documents	4
8	Exec	1
9	Images	32
10	System	0
11	Text	26
12	Unknown	5
13	Video	0
	TOTAL	**143**

Table 4.0: Tabular form of report in Fig4.36 summarizing the categories of files found after the File sorting process

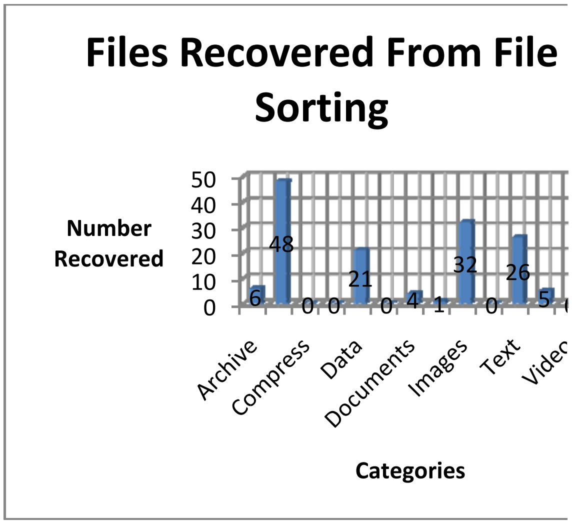


Fig4.94: Bar Chart form of report summarizing the categories of files found after the File Sorting process

4.3.5 **Interpretation of Bar Chart of Recovered Files from Pen Drive**

From Fig 4.9.4 the bar chart shows that the audio files have the highest occurrence rate of 50 files among the recovered items. The cypto, compressed, Disk, executable files, Executable files, and video files were the lowest with zero files recovered.

4.3.6 **Comparing Foremost and Sleuth Kit/Autopsy Recovered Items From Pen-Drive**

the similar search attributes of Foremost and Sleuth Kit named below and detailed in

Table 4.1:

1. Compressed items
2. Images Recovered
3. Video Recovered

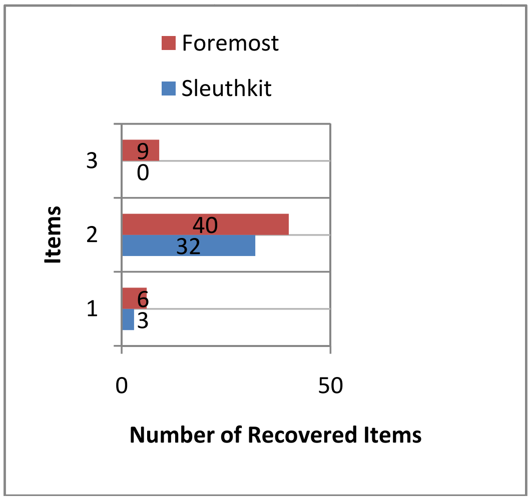
Program	**Number of Compressed Items Recovered**	**Number of Images Recovered**	**Number of Video Recovered**
Sleuthkit	3	32	0
Foremost	6	40	9

Table 4.1: comparison of recovered items by Foremost and Sleuthkit

a. **Analysis of Fig 4.95 Bar Chart**

The Bar Chart shows the frequency distribution of three common recovered items between Sleuth Kit and Foremost. The largest frequency of recovered items is for images, with Foremost having a higher number of 40 items as compared to 32 items of Sleuth Kit. Foremost recovered twice as much compressed items from SleuthKit.

Fig 4.95 A bar graph showing the comparative analysis of the common recovered items between Sleuthkit and Foremost



Legend of Barchart
1 = Video
2= Images
3=Compressed Files

4.4 Summary of Findings

Fifty-four files were extracted by Foremost from the pen drive. Bulk Extractor took 153 minutes to complete its operation, which finished with detailed 18 reports of its activity. Performing searches with Sleuth Kit with large devices seems slow. String extraction is encouraged to be a preliminary activity before performing a search, this make the subsequent searches faster. Foremost recovered more items than Sleuth Kit

5 Experimental Discussion

The case study presented distinct challenges, with different aspects of The Sleuth Kit and Fiwalk toolset. They were utilized to effectively perform a file system analysis. The focus of this research project was based on a case study that is employed to help demonstrate the usefulness of The Sleuth Kit, Autopsy Forensic Browser, and Fiwalk as file system Analysis toolsets.

The scope of case study employed provided a good test of the functionality of the Sleuth Kit, Fiwalk, Bulk_Extractor and Foremost toolkits. Rarely is a single forensic tool ideal for an investigation or recovery process. Therefore, a combination of tools should be applied for flexibility and faster digital forensic investigation process. One major problem affecting all forensic tools is the increasing size of storage media. This often increases the time required for a complete analysis.

The seven objectives mentioned in the chapter titled "introduction" were all achieved and some observations noted briefly discussed below.

The research found that The Sleuth Kit Autopsy Forensic Browser and Fiwalk all provide effective file system analysis toolsets. The flexibility of the tools contained within The Sleuth Kit often lead to complex command line strings, the complexity of which is overcome by the automation provided by the Autopsy Forensic Browser. Not only do The Sleuth Kit and Autopsy Forensic browser provide an effective toolset, they also offer an affordable alternative to expensive commercial or proprietary based toolsets.

Foremost was found to be a program that though compact and small, was a good software for data recovery. Foremost was found not to be complex as compared to Sleuth Kit. Foremost was useful in recovering executable files, video files.

While Sleuth Kit was useful for analyzing a storage device, its recovery feature seems more suited for text files and documents. Observed strengths lies in the "what is contained" and "where is it located" , rather than "how can I extract it?". Also string extraction in preparation for Keyword search with Sleuth Kit and Autopsy was found to be time consuming for large storage media. The web interface is for Autopsy greatly simplified the investigative process.

The demonstration of the effectiveness of The Sleuth Kit and Autopsy Forensic Browser may be used by individuals or Law Enforcement as part of an
Evaluation, when looking to extend their current Digital Forensics toolset, either as an alternative or complement to their current tools
Bulk_Extractor was found to be very efficient in extracting emails, domain addresses, etc. It also seemed fast not minding the size of the storage device. The histogram text files recovered served as detailed and informative statistical tools.
Fiwalk was good for XML generation of a disk image. It was also fast and efficient.

6 Suggestion for Further Research Works

Due to the complexity of new and improving storage devices, an investigator would be well equipped, to be knowledgeable in Hardware and software architecture of various operating system and media devices. Also a working knowledge of programming languages would further his area of expertise in digital forensic science. Since most of the forensic tool-kits are Linux based, a firm grasp of the physiology of multiple platforms would be an added advantage. Further research should be employed in making these tools used in this thesis to be part of a software suite. This means integrating all the tools into one program. In addition, the tools should have a way of communicating with each other, i.e. a symbiotic association between the tools.

References

Anothony Reyes, Jack Wiles (2007): Cybercrime and Digital Forensics Syngress Publishing Inc, Elsevier Inc, 30 Corporate Drive, Burlington, MA 01803; ISBN 13:978-1-59749-228-7; pgs 734

Bradford Phillip G, Brown Marcus, Perdue Josh, Self Bonnie. Towards proactive computer-system forensics. In Proceedings of the international conference on information technology: coding and computing (ITCCa^{TM}04); 2004.

Brian Carrier (2002). Open Source Digital Forensics Tools; the Legal Argument

Brian Carrier (2003).Defining Digital Forensics Examination and Analysis Tools Using AbstractionLayers. International Journal of Digital Evidence. Vol1, Issue 4, Windar, 2003. Available at: http://www.ijde.org/docs/02_winter_art2.pdf

Carv er DL Hoss AM. Weaving ontologies to support digital forensic analysis. 2009

Cohen MI. PyFlag: an advanced network forensic framework. In: Proceedings of the 2008 Digital Forensics Research Workshop. DFRWS, http://www.pyflag.net; August 2008 [accessed 06.03.09].

Corey Vicka, Peterman Charles, Shearin Sybil, Greenberg Michael S, Bokkelen James Van.Network forensics analysis. IEEE Internet Comput 2002;6 (6). ISSN: 1089-7801:60

Garfinkel,S.(2009A)"Automating Disk Forensic Processing with Sleuthkit, XML and Python". (SADFE 2009).

Garfinkel Simson L, Farrell Paul, Roussev Vassil, Dinolt George (2009). Brining Science to Digital Forensics with standardized Forensic corporal. In: Proceedings of the 9th Annual Digital Forensic Re search Workshop (DFRWS); August 2009. Grenier Christophe. Data carving log, http://www.cgsecurity.org/wiki/Data_Carving_Log n;

Gialanella, David, (2008); New Tech, Old Problem. ABA Journal, 94(8), 35

Marcella, Albert J, Menedez, Doug (2008). Cyber Forensics. Boca Raton, FL; Auerbach Publications Taylor & Francis Group

Nance Kara, Hay Brian, Bishop Matt. Digital forensics: defining a research agenda. In: Proceedings of the 42nd Hawaii international conference on system sciences; 2009.

Pollitt Mark, Nance Kara, Hay Brian, Dodge Ronald C, p Craiger Phili, Burke Paul, Marberry Chris, Brubaker Bryan Virtualizati on and digital forensics: a research and education agenda. J Digit Forensic Pract 2008;2 (2). ISSN: 1556-7281:62-73.

Pollitt Mark M. An ad hoc review of digita l forensic models. In: Proceedings of the second inter national workshop on systematic approaches to digital forensic engineering (SADFE'07); 2007

Simon L. Garfinkel (2010): Digital Forensics Research: The Next 10 Years. Journal homepage: www.elsevier.com/located/diin

Turnbull Benjamin , Taylor Robe rt , Blundell Barry. T he anatomy of electronic evidence a quantitative analysis of police e-crime data. In International conference on ava i l ability, reliability and security, (ARES '09); March 16-19-2009. p.143-9. Fukuoka.

IJCSIS REVIEWERS' LIST

Assist Prof (Dr.) M. Emre Celebi, Louisiana State University in Shreveport, USA

Dr. Lam Hong Lee, Universiti Tunku Abdul Rahman, Malaysia

Dr. Shimon K. Modi, Director of Research BSPA Labs, Purdue University, USA

Dr. Jianguo Ding, Norwegian University of Science and Technology (NTNU), Norway

Assoc. Prof. N. Jaisankar, VIT University, Vellore,Tamilnadu, India

Dr. Amogh Kavimandan, The Mathworks Inc., USA

Dr. Ramasamy Mariappan, Vinayaka Missions University, India

Dr. Yong Li, School of Electronic and Information Engineering, Beijing Jiaotong University, P.R. China

Assist. Prof. Sugam Sharma, NIET, India / Iowa State University, USA

Dr. Jorge A. Ruiz-Vanoye, Universidad Autónoma del Estado de Morelos, Mexico

Dr. Neeraj Kumar, SMVD University, Katra (J&K), India

Dr Genge Bela, "Petru Maior" University of Targu Mures, Romania

Dr. Junjie Peng, Shanghai University, P. R. China

Dr. Ilhem LENGLIZ, HANA Group - CRISTAL Laboratory, Tunisia

Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India

Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain

Prof. Dr.C.Suresh Gnana Dhas, Anna University, India

Mrs Li Fang, Nanyang Technological University, Singapore

Prof. Pijush Biswas, RCC Institute of Information Technology, India

Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia

Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India

Mr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand

Mr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.)/ Dimat Raipur, India

Mr. Hayder N. Jasem, University Putra Malaysia, Malaysia

Mr. A.V.Senthil Kumar, C. M. S. College of Science and Commerce, India

Mr. R. S. Karthik, C. M. S. College of Science and Commerce, India

Mr. P. Vasant, University Technology Petronas, Malaysia

Mr. Wong Kok Seng, Soongsil University, Seoul, South Korea

Mr. Praveen Ranjan Srivastava, BITS PILANI, India

Mr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong

Mr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia

Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan

Dr Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria

Dr. Riktesh Srivastava, Skyline University, UAE

Dr. Oras F. Baker, UCSI University - Kuala Lumpur, Malaysia

Dr. Ahmed S. Ghiduk, Faculty of Science, Beni-Suef University, Egypt
and Department of Computer science, Taif University, Saudi Arabia

Mr. Tirthankar Gayen, IIT Kharagpur, India

Ms. Huei-Ru Tseng, National Chiao Tung University, Taiwan

CALL FOR PAPERS
International Journal of Computer Science and Information Security

IJCSIS 2013
ISSN: 1947-5500
http://sites.google.com/site/ijcsis/

International Journal Computer Science and Information Security, IJCSIS, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2011 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

Track A: Security

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc,), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on
its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

Track B: Computer Science

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embeded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail ijcsiseditor@gmail.com. Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at http://sites.google.com/site/ijcsis/authors-notes .

ISSN 1947 5500

http://sites.google.com/site/ijcsis/