Open Source Based Online Forensic Tool for Analysis of Malicious Activity over Email Traffic

Kumar Zinzuvadia[1], Nareshkumar Gardas[2]

[1] *Research Scholar, IT Systems and Network Security, GTU PG School, Gujarat, India*
[2] *Co-ordinator, CDAC-GTU PG School, Gujarat, India*

ABSTRACT

Email is ubiquitous in the contemporary commercial environment, providing a convenient and efficient method for communication. Email brings great convenience to people, but it also had a negative impact. However, involving email-related disputes and crime problems have become increasingly prominent. With the continuous development of network technology, Email has become an important means of information transmission and management in government agencies, enterprises and individuals, however, the subsequent problems of protecting web mail security has become a concern to users. Many cases of inability to determine the authenticity result in a lack of effectiveness of the evidence. In order to determine the authenticity of email evidence, a scientific appraisal is very important. Whenever you send an email message your computer generates additional information relevant to email. This additional information forms a trail of digital artifacts that can be followed when analyzing email messages. Email Forensics is concerned with the extraction of this additional data and using it to recover other information regarding emails. In addition, email has become a potential carrier of criminal evidence. Using email to spam, spread pornography, fraud and other criminal activities have become increasingly rampant. Therefore, technical appraisal of email plays an increasingly important role for analyzing and providing evidence. At present, there exists sizeable amount of support for Open Source based methods for acquiring, identifying and analyzing web-based email, despite its widespread use. There is the need to achieve a systematic process for email forensics which integrates into the normal forensic analysis workflow, and which accommodates the distinct characteristics of email evidence.

Keywords: - *Email Forensics, Email Traffic Analysis, Open source forensics, digital forensics readiness, forensics process.*

1. INTRODUCTION

Nowadays, we can send or receive emails on various kinds of devices. SMTP (Simple Mail Transfer Protocol) was designed to be an easy and cost-effective implementation. This fact, however, makes SMTP a target to be abused. Unsolicited electronic communication, also known as spam, is just one such example of abuse of email. Tracing the origin of spam by using the information contained in SMTP headers is not possible because SMTP is a clear text protocol and can easily be intercepted and modified. [1]

Digital forensic specialists are plagued with sifting through large data sets to find incident information. During the process of analysis it would be much easier if the information to be examined is digital forensic ready, to ensure that the information is valid and usable.

Security monitoring tools have been designed to collect security information in order to detect security breaches within the IT system. [1]

1.1 What is Email Forensics?

Email Forensics is basically the study involving the forensic analysis (identification and analysis) of malicious activity (such as Spam, Phishing, Spoofing, etc.) over Email Traffic and Email Messages.

1.2 Why is it required?

Organizations and persons of all types rely heavily on email communications, making it a crucial factor in every litigation. Deleted emails can often be recovered, even if they are erased intentionally. Metadata, such as email full header information, time stamps, etc., can all be very useful in an investigation if the authenticity of an email is ever brought into question. Email clients and servers are often full database applications, complete with document sources, contact managers, time managers, calendars, and many other features, all of which might be accessed forensically. Erasing or deleting an email does not necessarily mean that it is gone forever. Oftentimes, emails can be forensically extracted even after deletion.

1.3 What is Digital Forensics Readiness?

Rowlingson defines digital forensic readiness as consisting of two objectives. The first objective is to maximize the environment's capability of collecting digital forensic information and the second objective is to minimize the cost of a forensic investigation. Preparing any environment to be digital forensically ready, a mechanism will need to be added to preserve, collect and validate the information contain in the environment. The information gathered from the environment can then be used as part of a digital forensic investigation. [1]

1.4 Types of Email Forensics

Generally Email Forensics include two basic types of techniques:

1.4.1 Email Traffic Analysis:

Email Traffic analysis for Analyzing Email headers and identifying the modification, fabrication, interception, or obfuscation kinds of cyber-attacks; and determining the sender and receiver information, also the geo-location and trace-route path of email packets. All from analyzing raw email traffic packets. Traffic analysis generally focus on capturing raw email packets from live traffic and analyzing that traffic log for traces of anomaly or condition based rules for identifying traces of forgery.

1.4.2 Email Content Analysis:

this includes analyzing the full email message captured and re-compiled for forensic investigation for any abnormality in the message content leading to issue with security of the network for email server, account information, and other forms of compromise privacy information of the account holder and email and domain server.

2. RELATED WORK

Author in paper [1] suggests a need for a unified forensics process methodology that promotes digital forensics readiness for the analyzed content and further refinement of the collected data. Author in paper [12] suggests following a forensics model for efficient evidence identification. Author in paper [11] discusses various types of spam techniques and email viruses. Author in paper [2] suggests construction mechanism of the keywords commonly used in header fields and its application to forensics analysis process. Author in paper [3] discusses detecting the nonobvious artifacts related to email accounts, retrieving the data from the service provider, and representing email in a well-structured format based on existing standards. Author in paper [8] suggests classifying emails based on attributes based on similarity between them.

3. PROPOSED SYSTEM

The system methodology focuses on open-source based client-server application that performs data acquisition from network server and sends imaged data to the web server which follows the identification, analysis and discovery of malicious activity and sends a generated report of evidence to the network server if any abnormality is noticed in email header traffic and/or email content analysis (anti-virus, spam check, phishing, etc.).

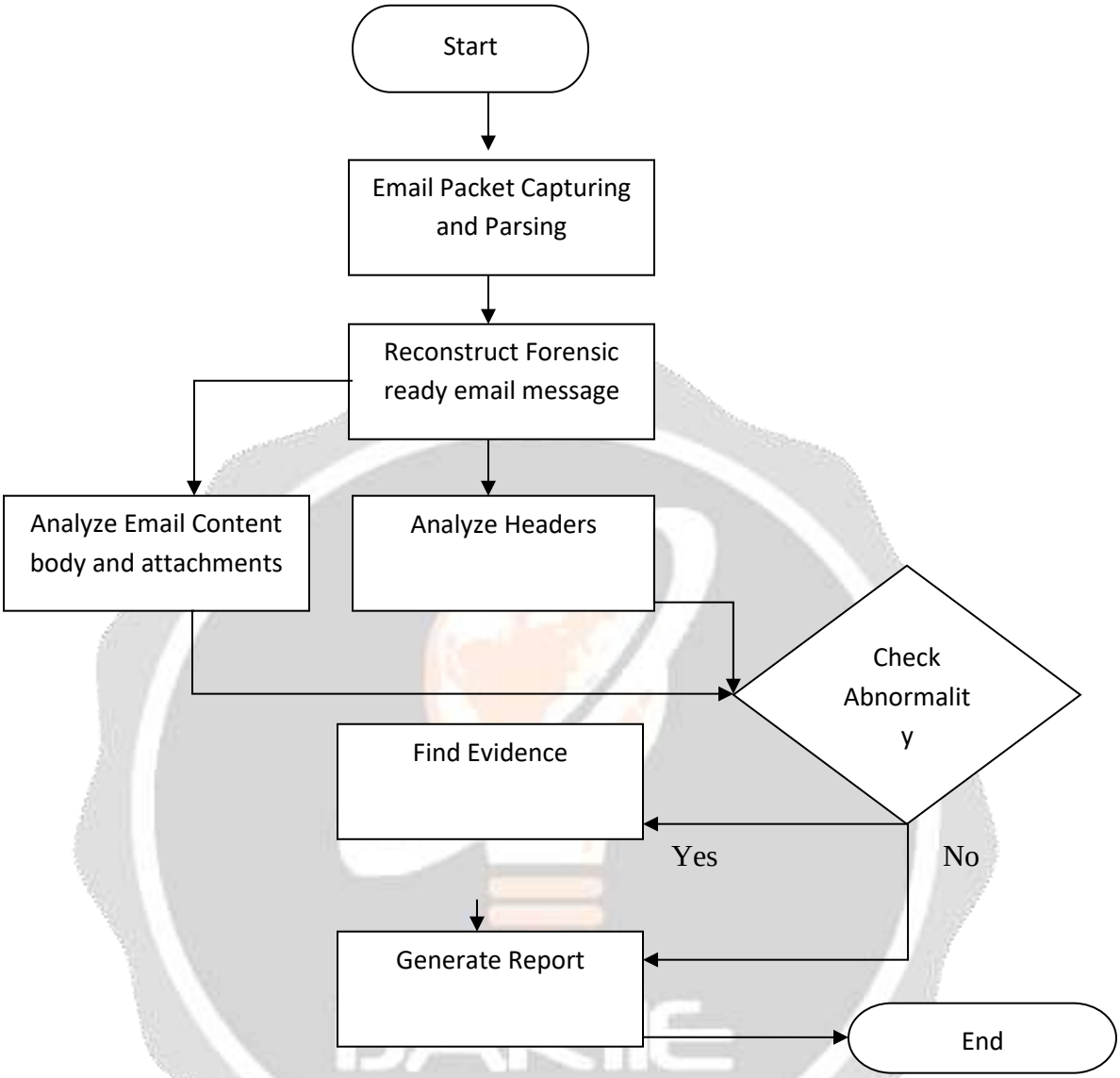

Fig. [1] Proposes System Flow Chart

3.1 Working

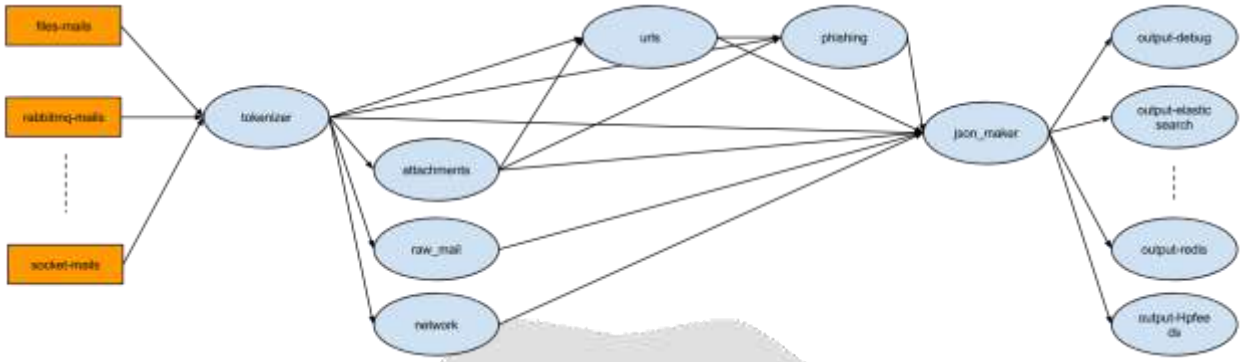

Fig [2]: Proposed System Working

Stage 1: Sniff Email Data Packets, Parsing in accordance with all attributes.
Stage 2: Generate a re-constructed copy of email messages, generate a raw image of email messages and send for forensic analysis.
Stage 3: Receive the raw image, identify all attributes, and analyze all parameters for abnormality in email traffic.
Stage 4: Analyze all parameters for abnormality in email content and analyze if abnormality related email shape and time constraints.
Stage 5: Discovery of evidence if malicious activity found and proof of evidence.
Stag 6: Generate and Send Report containing Evidence Information.

4. EXPERIMENTAL WORK

The implementation would require:
- A python script to sniff data packets
- A python script to parse data packets
- A python script to generate raw image email messages
- A python script to process the forensic investigation

In first step a sniffer is used to sniff, parse data packets and generate raw emails for forensic investigation process. We can see here that the raw data packets are sniffed and parsed for various attributes.


```
{
  "description": "Date: is 6 to 12 hours after Received: date",
  "rule name": "DATE_IN_FUTURE_06_12",
  "pts": 0
},
{
  "description": "ADMINISTRATOR NOTICE: The query to URIBL was block See http://wiki.apach
e.org/spamassassin/DnsBlocklists# for more information. [URIs: villaresidenceandrea.com]",
  "rule name": "URIBL_BLOCKED",
  "pts": 0
},
{
  "description": "RBL: SORBS: sender is an abusable web server [202.29.180.241 listed in dn
sbl.sorbs.net]",
  "rule name": "RCVD_IN_SORBS_WEB",
  "pts": 0.6
},
{
  "description": "RBL: No description available. [202.29.180.241 listed in bb.barracudacent
ral.org]",
  "rule name": "RCVD_IN_BRBL_LASTEXT",
  "pts": 1.6
},
{
  "description": "BODY: HTML included in message",
  "rule name": "HTML_MESSAGE",
```


Fig [3-1]: Parsed Email Sample

```
"received": [
  {
    "from": "host86-187-174-57.range86-187.btcentralplus.com 86.187.174.57 :45321 helo=
    "delay": 0,
    "date_utc": "2016-08-19T14:34:52",
    "hop": 1,
    "date": "Fri, 19 Aug 2016 20:34:52 +0600",
    "with": "esmtpa Exim 4.87 envelope-from <anabelgonzalo@fanox.com> id 1bakrE-000291-
    "by": "localhost.localdomain.com"
  },
  {
    "from": "localhost.localdomain.com mail.revesoft.com 208.74.72.248",
    "delay": 159287.0,
    "date_utc": "2016-08-21T10:49:39",
    "hop": 2,
    "date": "21 Aug 2016 10:49:39 -0000",
    "with": "ESMTP via TCP",
    "by": "mx03.futurequest.net 69.5.6.174"
  },
```

```
  "resourceName": "G64KD.docm",
  "Character-Count-With-Spaces": "6",
  "Last-Author": "1",
  "Character Count": "6",
  "X-TIKA:origResourceName": "G64KD.docm",
  "Page-Count": "2",
  "Application-Version": "16.0000",
  "extended-properties:Template": "Normal.dotm",
  "Author": "1",
  "publisher": "",
  "meta:page-count": "2",
  "cp:revision": "2",
  "meta:word-count": "1",
  "extended-properties:Company": "",
  "dc:creator": "1",
  "dcterms:created": "2017-05-16T12:44:00Z",
  "Last-Modified": "2017-05-16T12:44:00Z",
  "dcterms:modified": "2017-05-16T12:44:00Z",
  "Last-Save-Date": "2017-05-16T12:44:00Z",
  "meta:character-count": "6",
  "Line-Count": "1",
  "meta:save-date": "2017-05-16T12:44:00Z",
  "cp:contentStatus": "Microsoft.XMLHTTPIDEAAdodb.streaMIDEAshell.ApplicationIDEAWscript.shellIDEAProcessIDEAGeTIDEATeMPIDEATypeIDEAopenIDEAwriteI
DEAresponseBodyIDEAsavetofileIDEA\\galaperidol.exe",
  "Application-Name": "Microsoft Office Word",
  "Content-Length": "55473",
  "Content-Type": "application/vnd.ms-word.document.macroenabled.12",
  "Content-Status": "Microsoft.XMLHTTPIDEAAdodb.streaMIDEAshell.ApplicationIDEAWscript.shellIDEAProcessIDEAGeTIDEATeMPIDEATypeIDEAopenIDEAwriteIDE
AresponseBodyIDEAsavetofileIDEA\\galaperidol.exe",
  "X-Parsed-By": [
    "org.apache.tika.parser.DefaultParser",
    "org.apache.tika.parser.microsoft.ooxml.OOXMLParser"
  ],
  "creator": "1",
  "meta:last-author": "1",
  "xmpTPg:NPages": "2",
  "Revision-Number": "2",
  "X-TIKA:parse_time_millis": "1407",
  "X-TIKA:embedded_resource_path": "/G64KD.docm",
  "dc:publisher": ""
}
```


Fig [3-2]: Parsed Email Sample

5. RESULTS AND ANALYSIS

A possible solution for the stated problem constitutes of a SIEM like system that can perform monitoring of email traffic as well as process and generate re-constructed email message that is forensically ready to investigate in order to find any occurrence of malicious activity and also check the email content.

The below graph shows the results about the half an hour of time duration and the packet captured during this time frame.

Graph [1]: Time duration of the packets captured during the data sampling and analysis process.

Spouts (All time)

Search:

Id	Executors	Tasks	Emitted	Transferred	Complete latency (ms)	Acked	Failed	Error Host	Error Port	Last error	Error Time
files-mails	1	1	5920	5920	2015.560	5860	0				

Showing 1 to 1 of 1 entries

Bolts (All time)

Search:

Id	Executors	Tasks	Emitted	Transferred	Capacity (last 10m)	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed	Error Host	Error Port	Last error	Error Time
attachments	2	2	6640	13920	0.000	0.003	5860	6.608	5860	0				
json_maker	1	1	5980	5980	0.005	0.009	35120	3.299	35120	0				
network	2	2	5560	5560	0.000	0.003	5860	1.118	5860	0				
output-redis	1	1	0	0	0.000	0.003	5840	7.488	5880	0				
phishing	1	1	5840	5840	0.045	0.165	17560	44.379	17560	0				
raw_mail	2	2	5740	5740	0.001	0.007	5880	1.103	5860	0				
tokenizer	1	1	28580	34860	0.000	0.000	6080	1904.164	5860	0				
urls	1	1	6000	12000	0.001	0.005	11700	12.443	11700	0				

Showing 1 to 8 of 8 entries

6. CONCLUSIONS

With emergence of new technology there is need to combine two different domains like Email header analysis and Email Content Analysis for a better and more accurate analysis of malicious activity. With the combination of these two domains we can get all the benefits of both the domains which are useful for a systematic collaborative open source approach. Users will be able to detect many aspects of malicious activity like phishing, spamming, spoofing, session hijacking, obfuscation, etc. and also along with necessary legal permissions be able to find out information

about senders and receivers information, the geo-path of the email trace route, etc. This is very useful mechanism for a small scale and medium scale enterprises because it is open source. This work will be extended for development of a web-based open source application that provides the working functionality of the concept methodized here.

7. REFERENCES

[1]. Adding digital forensic readiness to the email trace header Van Staden, F.R.; Venter, H.S. Information Security for South Africa (ISSA), 2010 Year: 2010 Pages: 1 - 4, DOI: 10.1109/ISSA.2010.5588258 IEEE Conference Publications
[2]. Analysis of Email Header for Forensics Purpose Hong Guo; Bo Jin; Wei Qian Communication Systems and Network Technologies (CSNT), 2013 International Conference on Year: 2013 Pages: 340 - 344, DOI: 10.1109/CSNT.2013.78 Cited by: Papers (1)
[3]. Towards comprehensive and collaborative forensics on email evidence Paglierani, J.; Mabey, M.; Ahn, G.-J. Collaborative Computing: Networking, Applications and Worksharing (Collaboratecom), 2013 9th International Conference Conference on 20-23 Oct. 2013
[4]. Condition Factorization: A Technique for Building Fast and Compact Packet Matching Automata, Alok Tongaonkar and R. Sekar, 1556-6013 (c) 2015 IEEE.
[5]. An Email Forensics Analysis Method Based on Social Network Analysis YanHua Liu College of Mathematics and Computer Science Fuzhou University Fuzhou, P. R. China lyhwa@fzu.edu.cn, GuoLong Chen College of Mathematics and Computer Science Fuzhou University Fuzhou, P. R. China cgl@fzu.edu.cn, LiliXie College of Mathematics and Computer Science Fuzhou University Fuzhou, P. R. China 452056919@qq.com 2013 International Conference on Cloud Computing and Big Data 978-1-4799-2829-3/13 $26.00 © 2013 IEEE DOI 10.1109/CLOUDCOM-ASIA.2013.38
[6]. A forensics tool of Foxmail client LiliXie; Guolong Chen Systems and Informatics (ICSAI), 2014 2nd International Conference on Year: 2014 Pages: 400 - 405, DOI: 10.1109/ICSAI.2014.7009322 IEEE Conference Publications
[7]. Software Defined Monitoring of Application Protocols Luk´ aˇs Kekely, Jan Kuˇcera, Viktor Puˇs, Jan Koˇrenek, Athanasios V. Vasilakos DOI 10.1109/TC.2015.2423668, IEEE Transactions on Computers
[8]. Email Forensic Analysis Based on k- means clustering Arya P Nampoothiri Department of Computer Science and Engineering Sree Buddha College of Engineering, Alappuzha, India e-mail: aryapradeep013@gmail.com Minu Lalitha Madhavu Department of Computer Science and Engineering Sree Buddha College of Engineering, Alappuzha, India e-mail: minulalitha@gmail.com 978-1-4799-8792-4/15/$31.00 2015 IEEE
[9]. IP geolocation suspicious email messages Asmir Butković, Saša Mrdović, Samra Mujačić 978-1-4799-1420-3/13/$31.00 ©2013 IEEE
[10]. Cardinality Change-based Early Detection of Large-scale Cyber-Attacks Wenji Chen and Yang Liu and Yong Guan Department of Electrical and Computer Engineering Iowa State University Emails: wenjic, yangl, guan@iastate.edu 978-1-4673-5946-7/13/$31.00 ©2013 IEEE
[11]. Inbound & Outbound Email Traffic Analysis and Its SPAM Impact Seema Khanna National Informatics Centre New Delhi, India seema@gov.in Harish Chaudhry Department of Management Studies Indian Institute of Technology New Delhi, India hciitd@gmail.com Gundeep Singh Bindra Department of Computer Science SRM University New Delhi, India mailbox@gundeepbindra.com 978-0-7695-4821-0/12 $26.00 © 2012 IEEE DOI 10.1109/CICSyN.2012.42
[12]. Research and Implementation of Web Mail Forensics System 978-1-4244-6581-1/11/$26.00 ©2011 IEEE
[13]. A novel method to characterize unwanted email Traffic; Yang Sun, Hongfeng Zhu, 978-0-7695-4202-7/10 $26.00 © 2010 IEEE DOI 10.1109/CASoN.2010.22.
[14]. Identifying and Addressing Rogue Servers in Countering Internet Email Misuse; Wayne W. Liu, 978-0-7695-4052-8/10 $26.00 © 2010 IEEE DOI 10.1109/SADFE.2010.12.
[15]. Generic Network Forensic Data Acquisition from Household and Small Business Wireless Routers; Zhongli Liu, Yinjie Chen, Wei Yu and Xinwen Fu, 978-1-4244-7265-9/10/$26.00©2010 IEEE.
[16]. The Research on Email Forensics Based Network; Wang WenQi, Liu WeiGuang, 978-0-7695-3887-7/09/$26.00 ©2009 IEEE.
[17]. Email Worm Detection by Wavelet Analysis of DNS Query Streams; Nikolaos Chatzis, Radu Popescu-Zeletin and Nevil Brownlee 978-1-4244-2769-7/09/$25.00 ©2009 IEEE.
[18]. Email Shape Analysis for Spam Botnet Detection; Paul Sroufe†, Santi Phithakkitnukoon†, Ram Dantu†, and João Cangussu 978-1-4244-2309-5/09/$25.00 ©2009 IEEE.

[19]. A New Approach for Detecting Abnormal Email Traffic in Backbone Network; Ni Zhang, BinXing Fang, Li Guo, Yu Jiang 1-4244-0605-6/06/$20.00 C2006 IEEE.
[20]. Unifying Computer Forensics Modeling Approaches: A Software Engineering Perspective; A. Chris Bogen David A. Dampier 0-7695-2478-8/05 $20.00 © 2005 IEEE.
[21]. Techniques and tools for forensic investigation of e-mail M. Tariq Banday International Journal of Network Security & Its Applications (IJNSA), Vol.3, No.6, November 2011