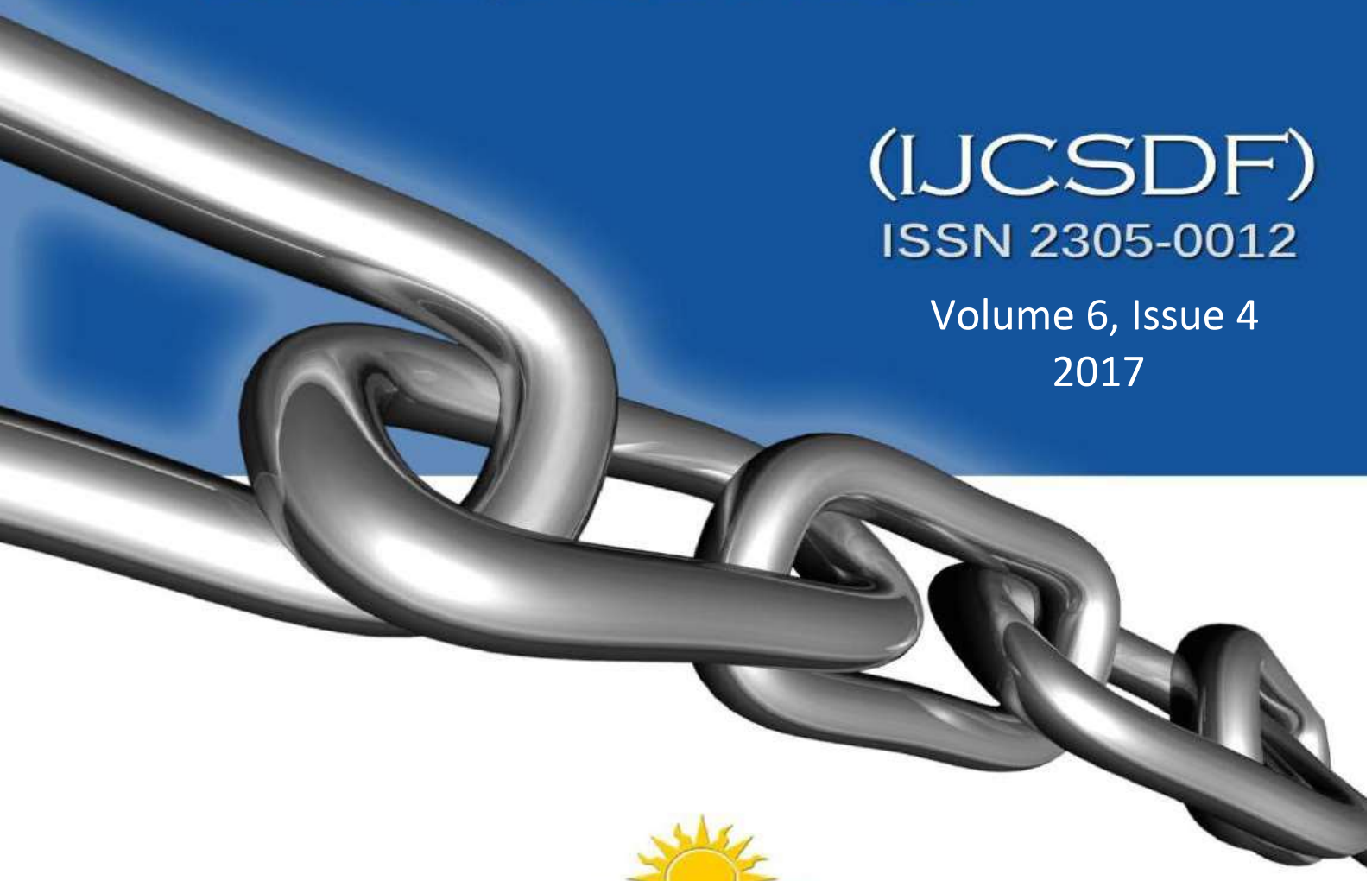
International Journal of Cyber Security and Digital Forensics



(IJCSDF)
ISSN 2305-0012
Volume 6, Issue 4
2017

SDIWC
www.sdiwc.net

ISSN: 2305-0012 (online)

International Journal of
Cyber-Security and Digital Forensics

2017

Overview

The International Journal of Cyber-Security and Digital Forensics (IJCSDF) is a knowledge resource for practitioners, scientists, and researchers among others working in various fields of Cyber Security, Privacy, Trust, Digital Forensics, Hacking, and Cyber Warfare. We welcome original contributions as high quality technical papers (full and short) describing original unpublished results of theoretical, empirical, conceptual or experimental research. All submitted papers will be peer-reviewed by members of the editorial board and selected reviewers and those accepted will be published in the next volume of the journal.

As one of the most important aims of this journal is to increase the usage and impact of knowledge as well as increasing the visibility and ease of use of scientific materials, IJCSDF does NOT CHARGE authors for any publication fee for online publishing of their materials in the journal and does NOT CHARGE readers or their institutions for accessing to the published materials!

Publisher

The Society of Digital Information and Wireless Communications
Miramar Tower, 13 Nathan Road, Tsim Sha Tsui, Kowloon, Hong Kong

Further Information

Website: http://sdiwc.net/ijcsdf, Email:jcs@sdiwc.net,
Tel.: (202)-657-4603 - Inside USA; 001(202)-657-4603 - Outside USA.

The issue date is Dec. 2017.

ISSN: 2305-0012 (online)

International Journal of
Cyber-Security and Digital Forensics

2017

Volume 6, Issue 4

CONTENTS

ORIGINAL ARTICLES

International Journal of Cyber-Security and Digital Forensics (IJCSDF) 6(4): 155-161
The Society of Digital Information and Wireless Communications (SDIWC), 2017 ISSN: 2305-0012

A New Method of Image Steganography: StegBlender

Richard C. Leinecker, PhD
Department of Computer Science, University of Central Florida
Mailing Address: 4000 Central Florida Blvd, Orlando, FL 32816
E-Mail: Richard.Leinecker@ucf.edu

ABSTRACT

This article presents a new method for hiding data within existing image data. The technique falls into the category of steganography, which is an approach of hiding data within other data sets. This new algorithm is named StegBlender, and is a new methodology for hiding data. It is different than traditional steganography methods, and should become part of the standard techniques used.

In this article, a bit shaving technique which has been used extensively is explained in order to give a context. Then, the StegBlender algorithm is explained as a contrast, including advantages over traditional techniques.

KEYWORDS

Steganography, blend, data hiding, obfuscation, embedded data.

1. Introduction

One of the tasks of security professionals is to protect the confidentiality of data. The most common and effective technique of insuring data confidentiality is with data encryption. Encryption converts normal data to what is known as cyphertext. It is a reversible process, and the plain, unencrypted data can be recovered with a process that reverses the encryption. While encryption provides good confidentiality, there is another way to provide confidentiality know as steganography. It is a method of hiding data so that nobody even knows it exists [1]. In this way, the data is not subject to encryption attacks because nobody knows that the data exists.

In order to hide data, you must have a carrier file into which the data will be hidden. The new technique that this article presents uses image files as carrier files. Specifically, BMP files are used. The algorithm presented could easily be extrapolated to use on PNG files. That is not to say that other image formats such as JPG, GIF, and TIF cannot be used, they would each need to take a slightly different approach since their data is laid out differently than BMP and PNG images. But to make the concepts clearer, only application of steganographic techniques for BMP will be undertaken in this article. Applying steganography to JPG would be a valuable addition to the literature, and that is in the planning stage.

2. Bit Shaving – A Common Steganographic Technique

Before delving into a new method, it would be best to discuss one of the more common steganographic techniques in use today which use image files as carriers. It is best described as bit shaving because it sets at least a single bit from the carrier data bytes to a zero value. Most of the time, the least significant bit is set to zero for each byte in the carrier data. For instance, if a carrier byte has the binary value of 10110111 then the least significant bit would be set to zero, and the new value would be 10110110. The only difference between the first byte and the altered byte is that the least significant bit becomes a zero after the alteration. Table 1 illustrates the process of shaving bits from values with several examples. Please note that the resulting decimal value does not always change when the least significant bit is shaved.

Table 1: Decimal and binary Values Before and After Bit Shaving

Decimal Values	Binary Values	Shaved Binary Values	Resulting Decimal Values
187	1 0 1 1 1 0 1 1	1 0 1 1 1 0 1 0 Shaved	186 ←
87	0 1 0 1 0 1 1 1	0 1 0 1 0 1 1 0 Shaved	86 ←
214	1 1 0 1 0 1 1 0	1 1 0 1 0 1 1 0 Shaved	214 ←
119	0 1 1 1 0 1 1 1	0 1 1 1 0 1 1 0 Shaved	118 ←

3. Image Data

First, though, a few words about image data are in order. Most image data, especially for the purposes of this article, are sets of three bytes. These sets of three bytes represent single screen pixels. The three bytes each represent a different color channel. One is for red, one is for green, and one is for blue. Mixing these three color channels provides the range of colors that we see on modern computer displays. Each byte has a value ranging from 0 to 255. The greater the channel value, the greater the effect of that color channel on the resulting color.

We normally refer to the three bytes as RGB triples, standing for Red, Green, and Blue triples. If an RGB triple has the value of 255, 0, 0 then the red value is at maximum, and the green and blues values are at minimum. This results in a pure red pixel on the computer display. In the same manner, an RGB triple of 0, 255, 0 yields a pure green color while 0, 0, 255 yields a pure blue color. Many images contain a fourth byte which indicates an alpha value. This allows for pixels in an image to have varying degrees of transparency. To keep this article clear, only images without alpha values will be considered. Fig 1 shows some common RGB triples and their resulting colors.

Figure 1: Common RGB Colors

4. Hiding a Single Byte of Data

For this discussion, only one bit will be shaved from each carrier byte, although it could be more than a single bit. In order to hide one incoming bit then, we need to shave the least significant bit from one carrier byte. Once the least significant carrier byte has been shaved and is zero, the incoming bit is placed into the position of the least significant bit. Figs 2 through 4 illustrate the process. Each figure uses eight carrier bytes of image data to hide eight bits from the data that is to be hidden.

R1	G1	B1	R2	G2	B2	R3	G3
76	123	205	82	129	211	245	255
Binary							
0	0	1	0	1	1	1	1
1	1	1	1	0	1	1	1
0	1	0	0	0	0	1	1
0	1	0	1	0	1	1	1
1	1	1	0	0	0	0	1
1	0	1	0	0	0	1	1
0	1	0	1	0	1	0	1
0	1	1	0	1	1	1	1 ←Need To Shave Row

Figure 2: Here you see eight carrier bytes: 76, 123, 205, 82, 129, 211, 245, and 255. You also see their binary equivalents. Notice that the least significant bits of each carrier byte are a different color in the figure, and indicate that they must all be shaved, or set to zero.

R1	G1	B1	R2	G2	B2	R3	G3
76	122	204	82	128	210	244	254
Binary							
0	0	1	0	1	1	1	1
1	1	1	1	0	1	1	1
0	1	0	0	0	0	1	1
0	1	0	1	0	1	1	1
1	1	1	0	0	0	0	1
1	0	1	0	0	0	1	1
0	1	0	1	0	1	0	1
0	0	0	0	0	0	0	0 ←Data Set to Zeros

Figure 3: The least significant bits of each carrier byte have been set to zero.

R1	G1	B1	R2	G2	B2	R3	G3
76	123	204	82	128	210	244	255
Binary							
0	0	1	0	1	1	1	1
1	1	1	1	0	1	1	1
0	1	0	0	0	0	1	1
0	1	0	1	0	1	1	1
1	1	1	0	0	0	0	1
1	0	1	0	0	0	1	1
0	1	0	1	0	1	0	1
0	1	0	0	0	0	0	1 ←Data Set to 'A'

Figure 4: Each bit of the character 'A' (decimal 65, binary 010000001) is placed into the least significant bit of the carrier.

Note that each carrier byte holds one message bit. For this reason, we can calculate the amount of carrier bytes needed to hide any given message by multiplying the size in bytes of the message by eight. For instance, a message of 20 bytes is 160 bits. It therefore requires 160 carrier bytes to be hidden, or eight times the message size of 20.

The entire process of bit shaving and replacement can be seen in Fig 5.

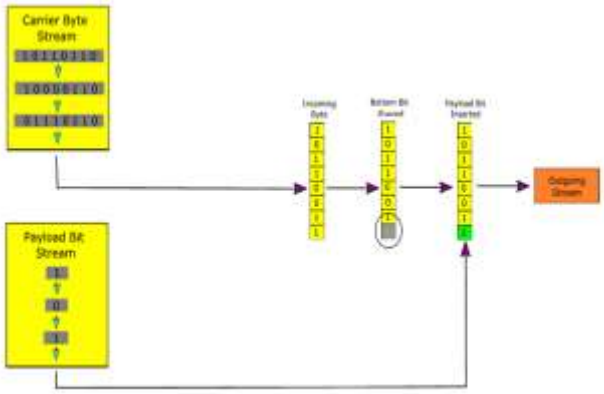


Figure 5: The entire process of bit shaving and replacement.

4. Acceptable Data Degradation

It should be obvious that the carrier data which now contains the message is not exactly the same as the original carrier data. From this, the question arises "can a viewer of the new image detect the difference between the original image and the altered image?" The answer is no for the following reasons.

First, each color channel has a value ranging from 0 to 255. This means that if the least significant bit is set to zero, it only affects the color channel value by .39 percent. And by the law of averages, the least significant bit is only one half of the time, while the other half of the time it is already a zero. So the net result is a change in a color channel of .39 percent half of the time and a result of 0 percent the other half of the time. Averaged over an entire image, this produces a statistical change of .195 percent change, even less than the .39 percent change calculated.

There is a branch of psychology that mentions the amount of change needed in order to be noticed. The concept is known as Just-Noticeable Difference[2]. It says that there is a threshold that must be exceeded in order for a change to be perceived. The general number of 1.25 percent is widely accepted, and is sometimes referred to as Weber's fraction, or Weber's constant. This leads to the conclusion that the .39 percent resulting when shaving the least significant bit is less than the threshold needed for a change in perception.

In order to demonstrate this, a program can be used to compare an original color with a color that has at least one bit shaved. Fig 6 shows an original color and a color with the least significant bit shaved. The change in the viewable color is not perceptible. This program was written especially for this article, and can be found at https://github.com/RickLeinecker/StripColorBits .

Figure 6: This color can be seen in its original hue in the rectangle to the left. With a single bit shaved, the new color can be seen in the second rectangle. There is no perceptible difference between the original color and the color that has had the least significant bit shaved.

5. Bit Shaving and Replacement Detection

There are tools available to detect steganographically hidden messages in images. They are not very effective because the detection tool must know the exact methodology that was used, and then reverse it to retrieve the message.

For the bit shaving and replacement method, there is one giveaway that could tip off a

detection tool. Suppose that before a message was hidden, the entire carrier had each byte's least significant bit shaved. Now the entire image data has the least significant bit set to zero, a statistical improbability. Then suppose that the hidden message only occupies the least significant bits of the first half of the carrier bytes. This means that the entire second half of the carrier still has all least significant bits as zeros, another improbability. A competent detection program would conclude that there was a steganographic message hidden in the carrier bytes.

There are two ways to mitigate this detection hazard. The first is to only shave the least significant bits when there is message data to hide. The other is to go through the unused least significant bits and randomly set some of them to one in order to avoid detection.

6. A New Method of Hiding Data in Image Data

The most important factor in keeping steganographically hidden data hidden is variety[3]. If detection software knows exactly how a message was hidden, it can easily retrieve it. For that reason, a different method was developed. Not only does it have a different methodology to hide data from what this article has discussed, but it is parameterized and can change from use to use. This greatly reduces the ability to detect and retrieve a hidden message. The name of this method is StegBlender.

While the bit shaving technique alters the least significant bit of each carrier byte, StegBlender alters unknown bits in a sequence of carrier bytes. Not knowing which bits have been altered makes it much more difficult to detect hidden messages.

The first concept to define is that of the carrier byte group. This is a grouping of a number of carrier bytes. In bit shaving, there was always a single carrier byte that was acted on for each operation. But with StegBlender it is a group of carrier bytes, and the number of carrier bytes in a group can change from carrier to carrier, thus providing some of the variety.

Once an image has been loaded into a memory buffer, the algorithm moves a pointer from the start of the buffer to the end of the buffer, or as long as necessary to hide the message data. After the StegBlender algorithm is finished with the current group of carrier bytes, the pointer moves up to the next group of carrier bytes. Fig 7 shows the process of moving from group to group.

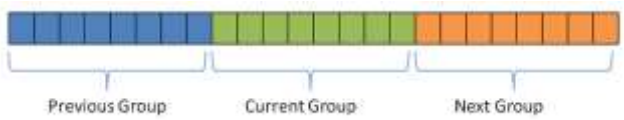


Figure 7: The image data is divided into groups of carrier bytes. After the algorithm finishes with one group, it moves to the next group.

For each group of carrier bytes a total must be calculated. The algorithm relies on having a total of the current set of carrier bytes. This is done by looping through each one of the carrier bytes and adding their values to an accumulator variable. Pseudo code for a function named **countTotal** that does this is shown in Listing 1, while Fig 8 depicts it as a flowchart.

```
countTotal array data, number groupSize
  total = 0
  count i from 0 to groupSize - 1
    total = total + data[i]
  return total
```

Listing 1: The countTotal function which calculates the total of all values within a group of carrier bytes.

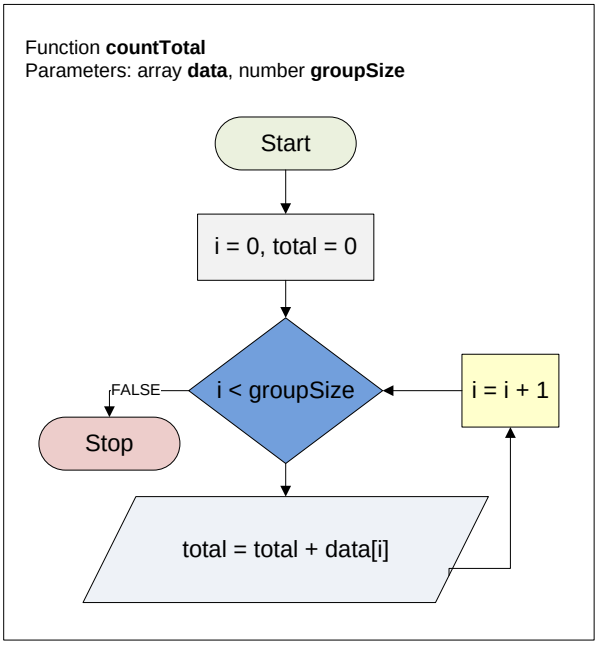


Figure 8: The countTotal function flowchart.

7. Data Group Mod Value

The magic of the StegBlender algorithm isn't the total of the carrier byte groups, but of a derived value from that total and a modulus value. In order to find the hidden value of a group, the modulus of the total of that group and a predetermined modulus value is calculated. For instance, if the predetermined modulus value is 256 and the total of a group is 1234, then the resulting value after performing the modulus operation is 210. The following formula shows this.

hiddenValue = groupTotal modulus predeterminedValue
hiddenValue = 1234 modulus 256 = 210

In other words, the algorithm depends on calculating the modulus value of the total of each group of carrier bytes. The modulus value, when the carrier bytes in the group are adjusted, will equal the next message byte that is to be hidden.

For instance, let us suppose that the message that must be hidden is the set of four characters TEST. Those four characters have ASCII values of 84, 69, 83, and 84. Since there are four bytes to hide, the algorithm requires four carrier byte groups. Ideally the first group will have the mod value of 84, the second 69, the third 83, and the fourth 84. The pseudo code in Listing 2 shows how to determine the hidden values in carrier bytes, the function being **extractMessage**, and Fig 9 shows the flowchart.

```
predeterminedModValue = 256
extractMessage array data,
      number groupSize
  create outBuffer
  while not done
    total = countTotal data, groupSize
    messageValue =
       total mod predeterminedModValue
    append outBuffer, messageValue
    data = data + groupSize
    if exit condition exists
     then done = true
  return outBuffer
```

Listing 2: The extractMessage function determines the hidden message.

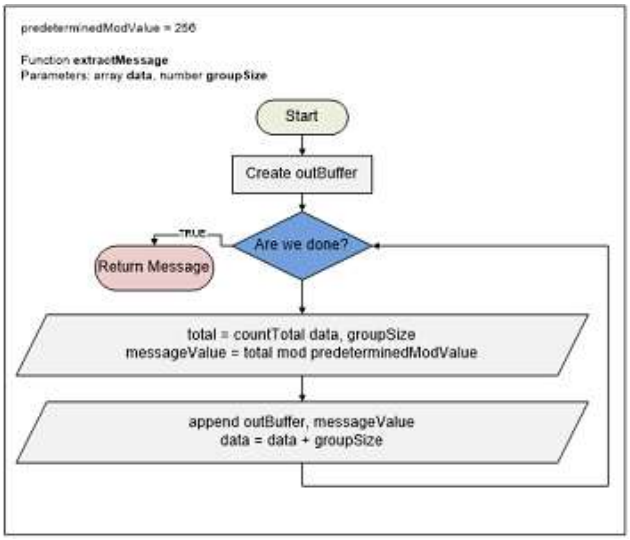


Figure 9: The extractMessage function flowchart.

8. Hiding a Message

It is easier to extract a message than to hide a message. That is because when a message is extracted, no adjustment of the group of bytes is necessary. But for each message byte that must be hidden, the bytes in the current group must be adjusted until the modulus value of the group's total matches the message byte.

Let us suppose that the message byte that must be hidden in this group is 84. Now let us suppose that the modulus operation value does not match that. In order to adjust the modulus operation value, we go through the carrier bytes and either add or subtract one until the modulus operation value matches the message byte.

As a special note, if the group total is above the halfway point, meaning that the average of the group bytes is more than half of the predetermined modulus value, then the algorithm subtracts. On the other hand, if the group total is below the halfway point then the algorithm adds. This ensures that the carrier bytes stay within range without an overflow. The pseudo code in Listing 3 shows the process of adjusting the contents of a group so that its modulus operation value matches the byte that is to be hidden. Fig 10 shows the flowchart for the adjustment code.

```
getModOperationValue array data, number
groupSize
  total = countTotal data, groupSize
  return total mod
predeterminedModValue

calcDelta array data, number groupSize
```

```
  total = countTotal data, groupSize
  halfway = groupSize *
predeterminedModValue / 2
  if total > halfway then return -1
  else return 1

adjustGroup array data, byte
messageValue
  circular = 0
  delta = calcDelta data, groupSize
  while getModOperationValue data,
groupSize <> messageValue
    data[circular] = data[circular] +
delta
    circular = circular + 1
    if circular >= groupSize then
circular = 0
```

Listing 3: The adjustGroup function adjusts the carrier bytes in the group until the modules operation value matches the message byte.

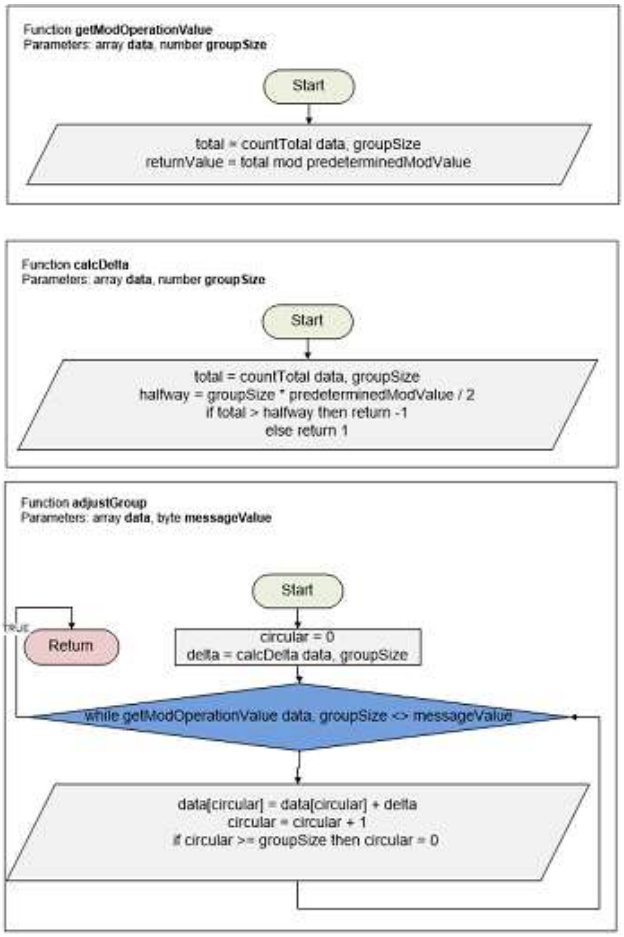


Figure 10: The adjustGroup flowchart.

9. Variations and Minimum Carrier Size

It is easy to determine the size of the carrier data needed to hide a message. The method uses an entire group to hide a single message byte. During the development of the demonstration program, a group size of 12 was chosen. That means that for each message byte, 12 carrier bytes are needed. To find the required minimum carrier size, all that is required is to multiply the message by 12. The following formulas illustrate this.

minimumCarrierSize = messageSize x groupSize

The group size plays a role in the quality degradation of the carrier image. The larger the group size, the less each byte in the group must be altered to arrive at the correct modulus value. On the other hand, larger group sizes may not leave enough room into which the message can be hidden.

The predetermined modulus value can be adjusted in many cases, especially for ASCII values. Since ASCII values range from 0 to 128, the modulus value could be as low as 129.

10. Source Code Project

The source code for the entire project is posted on GitHub at the address https://github.com/RickLeinecker/StegBlender/. The source code should easily compile in any environment.

11. Conclusion

StegBlender represents a new approach to steganography. It is an algorithm that provides a way to hide information while evading detection. Two of its integral values, the group size and the predetermined modulus value, can be changed. With changeable values and with the fact that the addition and subtraction alters unpredictable bits, detection is very difficult.

Future research should include a version that embeds messages in Jpeg images. The reason this is important is because Jpeg is such a pervasive image format. This research is currently underway, and will be completed after the publication of this article.

It is hoped that contributors to the project which has been posted on GitHub will move this research forward, and to provide valuable feedback on the extant research.

REFERENCES

1. Kessler, G. C. (2004, July 2004). An Overview of Steganography for the Computer Forensics Examiner. Forensic Science Communications, 6(3).
2. Link, S. W. (1992). *The Wave Theory of Difference and Similarity*. Hillsdale, NJ: Lawrence Erlbaum Associates.
3. Richer, P. (2003). Steganalysis: Detecting hidden information with computer forensic analysis. Retrieved November 16, 2016, from https://www.sans.org/reading-room/whitepapers/stenganography/steganalysis-detecting-hidden-information-computer-forensic-analysis-1014.

International Journal of Cyber-Security and Digital Forensics (IJCSDF) 6(4): 162-178
The Society of Digital Information and Wireless Communications (SDIWC), 2017 ISSN: 2305-0012

Enhancing AES with Time-Bound and Feedback Artificial Agent Algorithms for Security and Tracking of Multimedia Data on Transition

[1]A.O. Isah, [2]J.K Alhassan, [3]S.S Olanrewaju, [4]Enesi Femi Aminu.

[1]Department of Cyber Security Science, Federal University of Technology, Minna, Nigeria

[2]Department of Cyber Security Science, Federal University of Technology, Minna, Nigeria

[3]Department of Cyber Security Science, Federal University of Technology, Minna, Nigeria

[4]Department of Computer Science, Federal University of Technology, Minna, Nigeria

[1]ao.isah@futminna.edu.ng, [2]jkalhassan@futminna.edu.ng, [3]lanrezubair@yahoo.com, [4]enesifa@futminna.edu.ng

ABSTRACT

Unsecured information travelling over the network from a source to an intended destination are very vulnerable to cyber attacks. Millions of information have been hijacked and stolen across the globe within the last few years, translating to several millions of Dollars in damages. This paper is on the improvement of the security of existing systems that are mainly encryption systems implementing algorithms such as the Advance Encryption Standard (AES), Rivest Shamir and Addleman (RSA) encryption algorithm, without an active alert system for users. This is by designing and implementing an enhanced Advanced Encryption Standard file encryption model to secure and track information on transition. The implementation was achieved by codifying a Timing Circuit Algorithm (TCA) and Feedback Artificial Agent (FAA) into an Advance Encryption Standard algorithm to control decryption and monitor data along the transition path. Information encrypted with this system cannot be decrypted until the due date and time specified, the monitoring agent also sends reports of decryption to administrator's e-mail address and phone number. The model was implemented with Java programming language; the system achieved the desired results when tested.

KEYWORDS

Encryption, Decryption, Multimedia Data, Transition, Tracking, Information Security.

1 INTRODUCTION

Dynamism in the world of Information Technology (IT) is very vital for exchange of knowledge since no individual or group can be an island within the vast oceans of information. However, transportation of sensitive information within an organisation or between organisations is mostly endangered while on transit with an ever increasing array of malicious hackers [1]. The information technology world is in an era of constant change as a result of business ideas turning up from information technology firms. These business ideas could be in the form of text, graphics, audio or videos. Apple is one of those companies that produce new devices on regularly basis to target new market areas [2]. In the launching of iPod, the then Chief Executive; Steve Jobs, presents it to an astonished audience by asking them the use of the tiny pocket in their jeans? He said it is meant for a tiny music player called iPod that can store more than one thousand songs using a small chip of high memory size. This revelation inspires a generous patronage from lots of buyers. This valuable business information could have been stolen by their competitors through data hijacking on transit from Apple's online communication network.

More so, the Chinese economy has been growing tremendously in recent years that it has attracted

United States (US) accusation of individuals in China of cyber attacks and industrial secret information hijack. The US emphasized further that some of the attacks traced down to China are not just about military espionage but are targeted to business information that gives Chinese business owners a competitive edge in the global market [3].

Similarly, State and National Examination bodies are grappling with information leakages in form of examination questions and sensitive materials. This has resulted in increased cases of examination malpractices. The vulnerability of examination questions and materials is inherent in the transition from the examination offices to the examination centres [4].

Last but not the least, is the popular cyber attack against Sony Pictures in the last months of 2014 [3]. The hacking group, Guardian of Peace, had earlier attack and leaked the email accounts of staffs of Sony Pictures, publishing the private emails of some top management personnel that have racial contents in order to instigate people against the leaders in Sony Pictures that contribute to its successes in the Hollywood industry. Thereafter the hackers threaten Sony Pictures not to release their new film titled "The Interview", a comedy film that mimics the assassination of North Korean leader Kim Jong-un, as well as publishing a lot of other upcoming films on online download sites. Sony Pictures lose millions of dollars as they had to heed the threat of hackers by cancelling the release of a film they spent much money advertising to the general public [6]. On the part of top government officials, it was an embarrassment in terms of national security. The US President, Barack Obama has to intervene and ordered the release of the film after many days has passed by and asserts that Sony Pictures made a mistake by heeding to the threats of hackers by not screening the film on the initial date scheduled for it [5].

Consequently, the security and tracking of information on transition is an uphill task for institutions of government and business organisations; hence this study proposes the use of an Enhanced Advanced Encryption Standard (AES) File Encryption System for the Security and tracking of information Transition by embedding Feedback Artificial Agent (FAA) and Time Circuit Algorithm (TCA).

1.1 Motivation

The Authors' motivation is inspired by the high prevalence of data or confidential records breached or stolen across the world. The Risk Base Security (RBS) report of February 2014 for Open Security Foundation revealed that several attacks were successfully carried out that resulted in exposing hundreds of millions of data and confidential information in most advanced countries as the statistics shows in Table 1

Table 1. Statistics of Data exposed and Stolen in 2014 (Risk Based Security report, 2014)

Exposed Records Ranking	Country	Total Exposed Records	Percentage of Exposed Records
1	United States	546,846,693	66.5%
2	South Korea	140,238,121	17.1%
3	Australia	42,672,848	5.2%
4	Sweden	29,000,002	3.5%
5	Japan	22,162,392	2.7%
6	China	12,012,056	1.5%
7	United Kingdom	11,669,949	1.4%
8	Taiwan	6,468,738	0.8%
9	Germany	2,101,718	0.3%
10	Canada	1,564,966	0.2%

1.2 Statement of Problem

Many authors such as [6, 7, 8], all tried to improve the security of data either on transition or at rest, with the use of cryptography and steganography. The problem the existing solutions could not solve is that, the authors were not able to develop a model for monitoring the data on transition or at rest. The concern of the authors of this paper is to add a mechanism for securing and tracking of an encrypted multimedia data of any high security and economic value on transition. The mechanism also alerts the sender via a registered mobile number and email address, of any attempt to either hack sent multimedia data or the successful decryption

of the multimedia data by the genuine consignee. The solution proffered by this study also works effectively on data at rest. However, the focus of the research is on multimedia data on transition as data becomes more vulnerable while on transition between two or more communicating nodes in the internet.

1.3 Aim and Objective

The aim of this study is to provide security and tracking for Information multimedia data on transition using enhanced AES file encryption system. This will be achieved by the following objectives;

i. To design a model for secured multimedia data on transition from a source to destination
ii. To implement enhanced Advanced Encryption Standard (AES) algorithm using java programming language, Feedback Artificial Agent (FAA) and Timing Circuit Algorithm (TCA) to monitor encrypted multimedia data on transition.

To test the performance of the designed model for secured multimedia data on transition.

1.4 Feedback Artificial Agent (FAA)

The Feedback Artificial Agent (FAA) can be defined as the mechanism embedded in the software implementation of this study to effect the sending of alert in the form of short text messages via a registered mobile phone number and email address.

1.5 Time Circuit Algorithm (TCA)

The Time Circuit Algorithm (TCA) can be defined as the mechanism for effecting predetermined time of decryption for any multimedia data encrypted and sent on transition.

2 REVIEW OF RELATED WORK

[6], presents that Steganography allows a user to securely hide messages in a cover media and to extract hidden message from the same media. Singh *et al* agrees that the varieties of steganography techniques; some of which are more complex than others, all have respective strong and weak points. The paper proposes the combined concept of Cryptography and steganography with the use of Elliptic Curve Cryptography (ECC). The main focus of the research is to encrypt the data with Asymmetric Cryptography and apply the steganography technique to hide encrypted data and thus make it a hybrid model. Existing hybrid approach which uses RSA with steganography is suffering from big performance hit. However, the limitation of the work lies in the fact that it lacks a feedback artificial agent that can monitor an encrypted multimedia data on transition.

[7], are concerned about a secured pathway for data transmission, explaining that multicast routing for wireless mesh networks has focused on metrics that estimate link quality to maximize throughput. Nodes must collaborate in order to compute the path metric and forward data. The assumption that all nodes are honest and behave correctly during metric computation, propagation, and aggregation, as well as during data forwarding, leads to unexpected consequences in adversarial networks where compromised nodes act maliciously. The authors identify novel attacks in wireless mesh networks. The attacks exploit the local estimation and global estimation of metric to allow attackers to attract a large amount of traffic. The authors demonstrate both the attacks and defense strategy using On-Demand Multicast Routing Protocol (ODMRP). The limitation of the work is that it is the network pathway of the data being transmitted that is secure, the data becomes vulnerable the moment it enters through a differently configured network and such, it is better to encrypt the data for a more reliably data security irrespective of the network.

[9], explained that security is required for the protection of delivery of multimedia data, thus this security is provided by the use of encryption. The authors used selective encryption for protecting multimedia data that takes less computational workload and provides five levels of security from level 0 to level 4. These five levels of security are; level 0 where there is no encryption, level 1 where the headers from the sequence layer down to the

slice layer is encrypted, the encryption of Intra-frame coded blocks (I-blocks) and the complete encryption of level 1 occurs in level 2, at level 3 the encryption of Intra Frames (I - Frames) and all the remnants of I-blocks and finally at level 4, the complete encryption of multimedia data (video). The study was not able to effects the timing of decryption for improved security assurance of the multimedia data.

[8] argue that Image covers the highest percentage of multimedia data and that its protection is very important. The authors submit that the use of cryptography is an effective tool for assuring the confidentiality of transmitting images over the internet.
The paper presents a review on image encryption techniques of both full encryption and partial encryption schemes in spatial, frequency and hybrid domains. However, the study focused on image only as a multimedia data and very silent on how other multimedia data type can be protected by encryption over the internet.

[10], highlights that Mobile Ad-Hoc Networks (MANETs) established efficiency in the deployment for number of fields, but highly vulnerable to security attacks, a situation that seems to be more challenging for wireless networks. The paper proposes Reliable Data Security Architecture (RDSA) for multi-path multimedia streaming over wireless network to improve multiple path routing efficiency in frequent communication failures due to channel interferences. The proposal framework provides better bandwidth allocation and reduces transmission delay. Simulations are carried out with Dynamic Source Routing (DSR) and Ad-hoc On-demand Distance Vector (AODV) protocols for efficient multi-path multimedia data transmission security scheme. Although, the authors improve the security of multimedia data using RDSA as a methodology, the research could not include a feedback artificial agent system to monitor attempts by hackers to compromise data over the wireless network.

In order to prevent offline and online attacks [11], the authors showed in their analysis of password managers, that the Mobile password manager is the most effective. The limitation of their work lies in the fact that although, passwords are important protective measures in information security but, they only serve as a first line of defense while encryption is last line of defense. This we have utilized with additional elements of defense proposed and implemented in our paper by tracking with time-bound algorithm otherwise herein called Timing Circuit Algorithm (TCA) and the Feedback Artificial Agent (FAA).

[12], studied how to Enhance Multimedia Security in Cloud Computing Environment Using Rivest Shamir Adleman (RSA) algorithm and Advanced Encryption Standard. The authors methodology is system model that merely listed the usually computation of Rivest Shamir and Adleman algorithm as well as the transformation rounds in Advance Encryption Standard. The authors were able to present a concise explanation of the two encryption techniques. The goal of the study was not achieved as the paper could not add to existing knowledge.

[13], tries to impact a change on the RSA algorithm against recent efforts by hackers which the authors called secure RSA for secure document transmission as there are numerous situations where there is the need to secure record transmission, for instance in banking, e-shopping, state security data sharing. The study concentrates on data transfer utilizing Secure RSA, which eliminates some vulnerability of RSA that may keep a hacker from stealing and abuse of information. The authors could not implement this frame work for multimedia data.

[14], noted that most communication channels are no longer one to one, as other devices in the network also receives data generated by a device in the same network through multicast transmission architecture. This is due to the fast improvement of information and network technologies. These multicast systems that enhance rapid delivery of messages in the network also open up loopholes to snooping attacks in the network. The study submits that one to one encryption is no longer effective for the security of data. So, the authors proposed a novel anonymous multi-receiver encryption, in

which receiver's decryption key is fixed. Furthermore, the model provided anonymity of receivers, performance analysis and comparisons with other schemes.

[15], although in their work, emphasizes the complexity and the uncertainty of information security, but they showed that the risk can be minimized by assessing the risk, and then applied Genetic Algorithm (GA) to reducing it by assigning variables to GA and run an iterative tests by the arranged elements. The results were finally compared with the admissible risk volume. However, the authors were able to use the GA to minimize the security risk in information which is one of the concerns of our paper, but securing encrypted information from unauthorized decryption and tracking by the administration could not be solved by the application of the GA proposed by this reviewed work.

2.1 Analysis of the Existing Systems

The need to analyze existing software systems was taken into recognition as it is the bases for justifying this thesis titled Security of Multimedia Data on Transition Using Enhanced AES File Encryption System. This was done by picking and explaining the features and uses of a number of encryption softwares, that are commonly used for online and offline communication security in paragraphs.

The merits of the software are that it has a simple graphical user interface, small memory size as an application and the ability to access text file directly from the computer system. However, the demerits of BCTextEncoder include being text only encrypting software and the need to carefully select the appropriate decryption password from the list of saved keys.

Secretpad in figure 1 is trusty encryption technique based text encryption software that has just three components. The components are the tool bar, where the file, edit, format, view and help features are located. It uses passwords to encrypts texts and only that same password can be use to decrypt it. It makes use of simplified graphical user interface like that of windows notepad. It is useful for the storage of passwords, secret correspondence and other confidential information. The software automatically closes the window in case of long inactivity. Installation is not required as users only need to down load and run it.

Source: listoffreeware.com, 2014

Figure 1. Secretpad Encryption Software.

The drawback inherent in the use of the software is that encryption and decryption icons cannot be easily located, it hidden under the file icon. Also, trusty encryption algorithm is highly vulnerable to hack attacks.

Arcanum Editor in figure 2 is robust encryption software for texts, using a combination of Advanced Encryption Standard (AES), Bytes, Base64, Rot 13 and 1337 Speak encryption techniques. Individual algorithms can be selected at will whenever there is the need to encrypt a classified text based information, as well as choosing a password when using the AES algorithm. The advantages of the encryption software include the present of the different encryption algorithms and a user friendly graphical user interface. The disadvantages include its being text only encryption software.

Source: listoffreeware.com, 2014

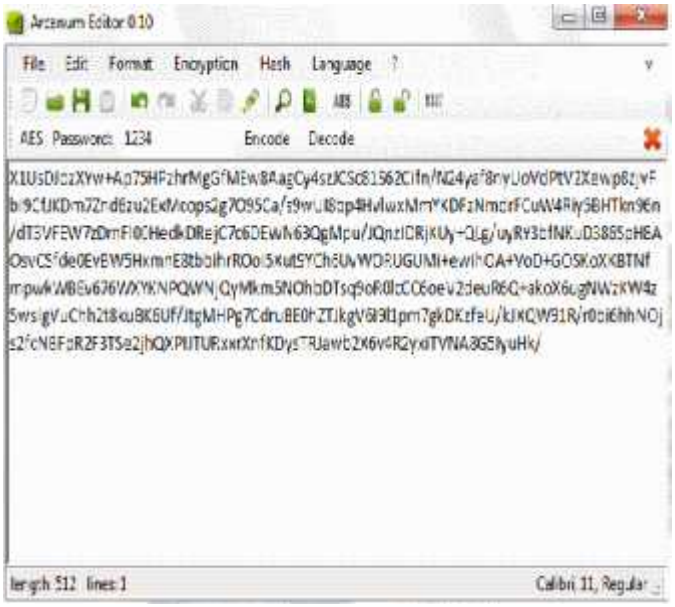


Figure 2. Arcanum Editor and Encryption Software

AES (256-bit) software in figure 3 is Advanced Encryption Standard algorithm technique based encryption software that uses a 256-bit key length password. It can be used to encrypt and decrypt a text while on a removable disk. Encrypted texts are decrypted by entering the same password, failure to do so will pop-up error message. There are other variations of Advanced Encryption Standard, that is, the 128-bits and 192-bit. However, this software uses the 256-bits length only. The merits of the software include a strong encryption and decryption key that is very difficult to crack by hackers and a simple layout of the user interface. The demerits are its inability to encrypt multimedia data and add other variations of AES.

Simple Text Encryptor in figure 4 is a free program, similar to AES encryption software but uses a small encryption key length of 128 bits. It has a simple graphical user interface and has two versions; the one that has to be installed and the other are zipped and stand alone. The zipped only need to be unzipped. Keys are generated randomly to encrypt a text and the same key is entered to decrypt.

Although this is a simply text editor, it has the ability to encrypt texts messages that are intended to be kept confidential. It occupies small memory space and could be installed; the user cannot decide the keys. During execution, the editor allows random generation of keys that can be used to encrypt texts that are saved in .txt file format only and multimedia file cannot be handled by it.

Source: listoffreeware.com, 2014

Figure 3. AES (256bit) Encryption Software.

Source: listoffreeware.com, 2014

Figure 4. Simple Text Encryptor Software

The analysis of existing encryption software in this study includes, Cryptditor shown in figure 5. It adopts Advanced Encryption Standard technique. The software can be used to encrypt any imported text data. Multiple files can be opened in its tabbed windows and encryption and decryption of texts are effected with a password. Cryptditor have basically two components, the tool bar and the text editor. The tool bar is the place holder of file, edit, tab, format and help. Text to be encrypted or decrypted can be typed on the text box or pasted on it. The user then click the file icon and pops out a box showing the Enter password and Repeat password text fields, Passphrase quality. On the box, the user clicks ok to encrypt, decrypt or cancel the process.

Source: listoffreeware.com, 2014

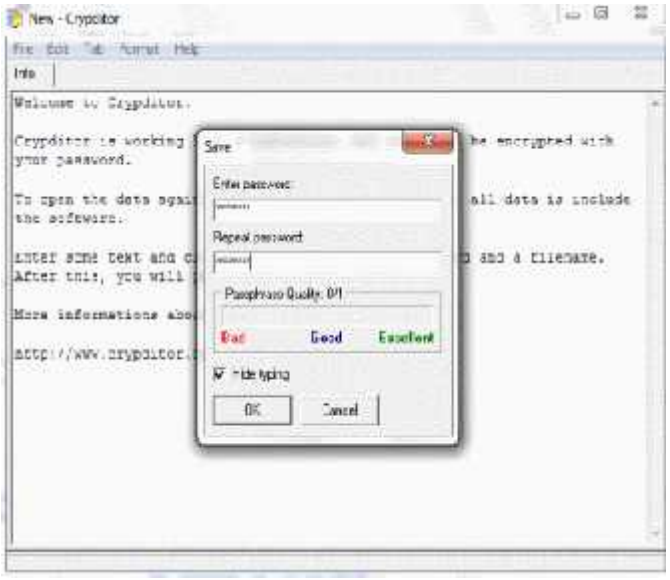


Figure 5. Cryptditor Encryption Software.

Encryption Tool in figure 6 encrypts files with a combination of MD5, SHA1, and the Advanced Encryption Standard. This makes it one of the very strong tools for encryption of confidential texts. It requires the uses of password to encrypt and decrypt files and a mechanism for checking the strength of a chosen password. The Encryption tool has two components, a tool bar and a text box for editing texts. There are file, edit, tools and about icons. The user has to click tools to select the encryption window to open. The encryption window has a Hash algorithm icon to select the desired encryption algorithm, PW iteration icon, and the key size icon.

Source: listoffreeware.com, 2014

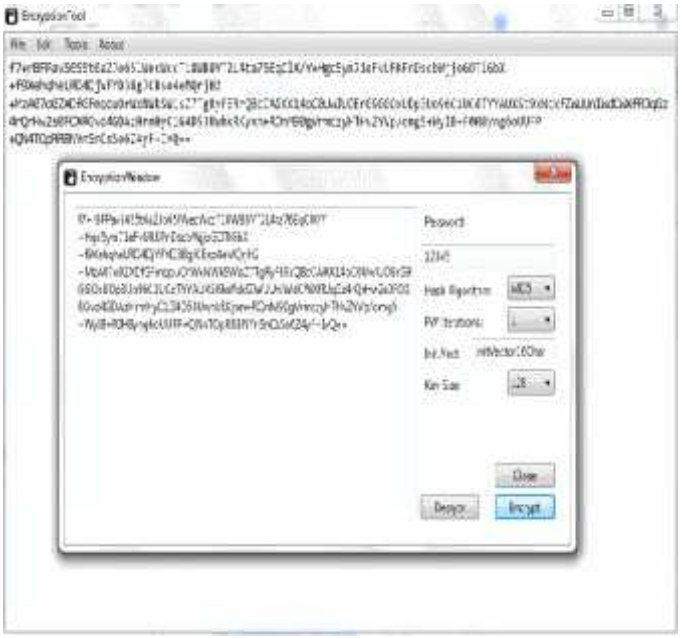


Figure 6. Encryption Tool Software

2.2 Summary

Finally, in this chapter, the authors were able to pin point the methodology used in the related work reviewed, highlight the strengths and weaknesses of the individual studies and then relate it to this study. Available software implementation of different encryption techniques were also analyzed with respect to how it works number of components, the encryption technique adopted, its strengths and weaknesses. All these put together provided a fundamental understanding of the challenges faced with data security on transition and then gives a clue to how it can be improved on in this work; the aim of which is the Security and tracking of Multimedia Data on Transition with an enhanced AES with a Time-Bound and Feedback Artificial Agent Algorithms. The enhancement is based on the addition of Feedback Artificial Agent (FAA) and Time Circuit Algorithm (TCA). In the analysis of existing systems, it can be deduced that most of the encryption softwares are mainly text encryption softwares that cannot encrypt video, graphics and audios. However, some of the few Advanced Encryption Standard based systems that do multimedia encryption does not address the threats to data in the course of transiting from

source node to destination node. The existing research was not able to effect the integration of timing circuit algorithm that can disallow file or folder decryption before the scheduled time of decryption at the destination node. Monitoring agent to keep track of data was also lacking. Hence, this study seeks to bridge these gaps.

3 METHODOLOGY

This section is dedicated to laying out the methodology, the materials and resources used are discussed and the model the authors employed in analyzing the implementation of the proposed work.

In the research, none of the encryption softwares mentioned in chapter two can encrypt video, graphic or audio files with Feedback Artificial Agent (FAA) and a Time Circuit Algorithm (TCA) altogether. Most of the softwares were developed to encrypt text and text files, the few ones that can do the encryption of multimedia data do not have FAA and TCA. Therefore, the authors researched into encryption algorithms that can effectively be used as a scientific basis for implementing the aim and objectives of this study. This is where the Enhanced AES comes into the modeling of the File Encryption System for this work, using the FAA and TCA as the unique models for this work.

3.1 The Advanced Encryption Standard

The security and tracking of multimedia data on transition can be achieved by applying the Advanced Encryption Standard (AES), enhanced and modeled it into a file encryption system, thereby giving it the power to accept any data type such as text, video, graphics and audio. The existing encryption system mostly accepts text. The few that accept multimedia data do not have feedback artificial agent that monitor and notify the source of a data on transition of any attempt to compromise the confidentiality, integrity and the availability of the data and of course the decryption of the data by the authorized receiver. In this model, the real data is digitized into an 8-bit data element in matrix formation and then passed through the Enhanced Advanced Encryption Standard. M represent Multimedia data like; text, video, graphics and audio. Figure 7 shows that data M as a matrix function in which each data element is an 8bits (1byte):

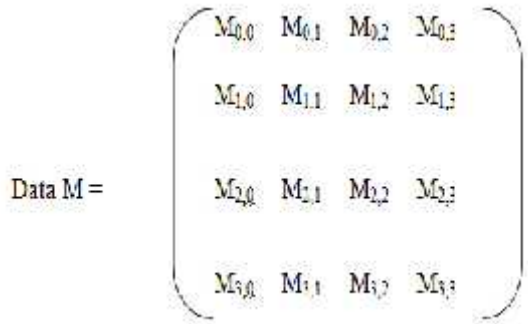


Figure 7. Data M in Matrix Formation.

In the transformation shown in figure 8, the data was converted to an encrypted bit value that is represented in machine language using the Advanced Encryption Standard. This explains the reason why cipher data is presented as a data file since it cannot be read naturally. It is only the deciphered data that has been converted back to normal language that can be read again. The decryption process is the inverse matrix function of the encryption processes.

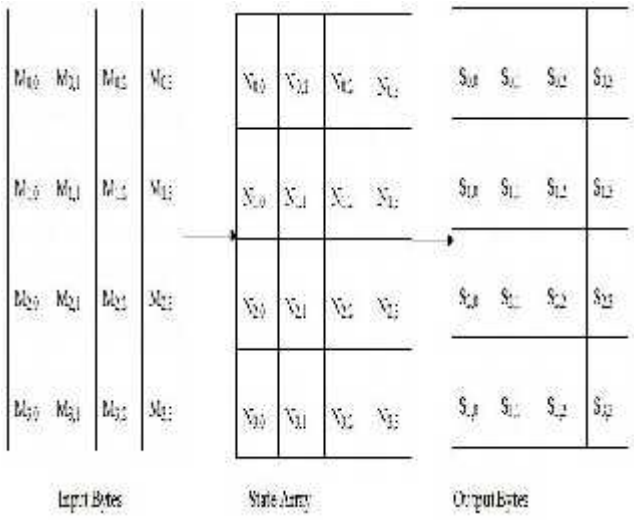


Figure 8. Matrix transformation of plain data in AES.

The internal operation of the AES adopts the concept of repetition called rounds as shown in Figure 8. The plain data M passes through the state array S, where substitutions, transposition and XOR bitwise operations occur to produce the

cipher data N. The inverse matrix of the cipher data N, gives back the original plain data.

3.2 AES Multimedia Encryption Algorithm

The pseudo code for the AES encryption algorithm is shown in capital letters below:

```
START
        SELECT MULTIMEDIA FILE
        ENTER AES ENCRYPTION PASSWORD
        GENERATE CIPHER USING
PASSWORD
        READ IN FILE
        FIGURE OUT FULL SIZE OF FILE
        ENCRYPT FILE DATA
        WRITE ENCRYPTED DATA TO NEW
FILE
STOP
```

3.3 AES Multimedia Decryption Algorithm

The pseudo code for the AES decryption algorithm is shown in capital letters below:

```
START
        SELECT MULTIMEDIA FILE
        ENTER AES DECRYPTION PASSWORD
        GENERATE CIPHER USING
PASSWORD
        READ IN FILE
        DECRYPT FILE DATA
        FIGURE OUT FILE SIZE
        WRITE DECRYPTED DATA TO NEW
FILE
END
```

3.4 Timing Circuit Algorithm (TCA)

The Timing Circuit algorithm is used to determine when a multimedia data encrypted can be decrypted in the file encryption system. It is to make sure that in the event where an attacker was able to hijack a sensitive multimedia data; he would not be able to decrypt it before the time programmed on it for its decryption. The structural model for the Timing Circuit Algorithm is in figure 9.

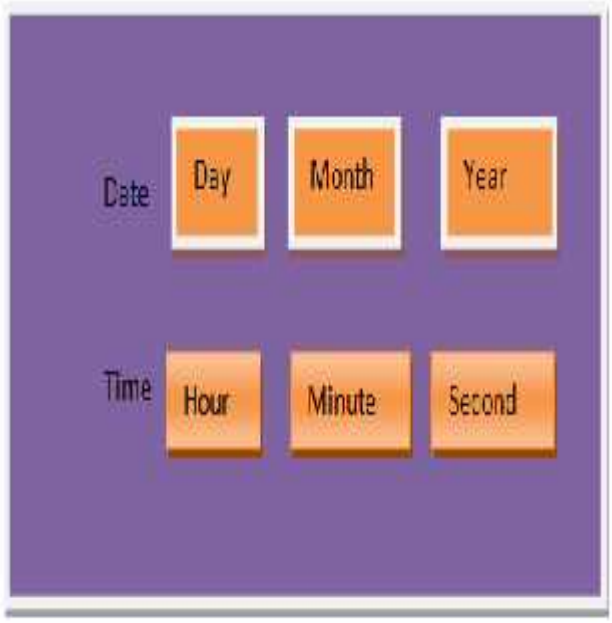


Figure 9. Model for Timing Circuit Algorithm.

3.5 The Timing Circuit Algorithm (TCA) pseudo code

The pseudo code for the Timing algorithm is shown below:

```
START
   GET ENCRYPTED FILE
    GET THE TIMING SECURITY DETAILS OF
         ENCRYPTED FILE
     SET Y=Decrypted due date Year
      H= Decrypted due Time Hour
      M= Decrypted due date Month    MIN=
         Decrypted due Time Minute
       D= Decrypted due date Day
       SEC= Decrypted due Time Second
       y= Current Date Year;  h= Current Time Hour
       m= Current Date Month;  min= Current Time
Minute
       d= Current Date Day;     sec=  Current  Time
Second
       SET                                  Boolean
is_Decryption_Date_Time_Due=x
       x = false
       IF Y > y
          x=false
        ELSE IF Y=y
          IF M>m
             x = false
          ELSE IF M=m
```

```
      IF D>d
        x=false
      ELSE IF D=d
      IF H>h
          x=false
       ELSE IF H=h
       IF MIN>min
          x=false
       ELSE IF MIN=min
       IF SEC>sec
          x=false
       ELSE
          IF SEC=sec
            x=true
          ELSE x=true
       ENDIF
    ENDIF
STOP
```

3.6 The Feedback Artificial Agent (FAA)

The Feedback Artificial Agent (FAA) is the system that monitors the encrypted multimedia data on transition by effecting the sending of short message to a registered phone number and an email address of the sender of data, whenever the holder of the decryption key decrypts the data or any attempt by an attack to hijack and decrypt the multimedia data. The structural model for the feedback algorithm is shown in figure 10. Clicking the process in the figure will trigger the sending of an alert to the phone number and email address registered with the modelled Feedback Artificial Agent for this study.

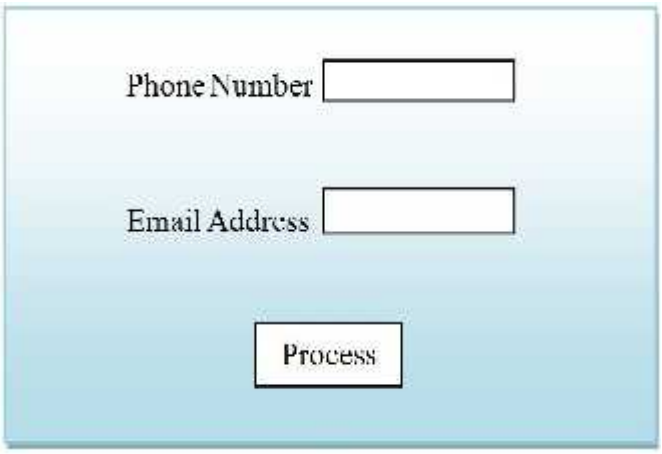


Figure 10. Model for Feedback Artificial Agent Algorithm

3.7 The Feedback Artificial Agent Algorithm

```
START
   IF (DECRYPTION DATE/TIME NOT DUE)
AND
          (PASSWORD IS INCORRECT) THEN
      SEND UNSUCCESSFUL DECRYPTION
      SMS/EMAIL FEEDBACK
   ELSE IF
       DECRYPTION DATE/TIME NOT DUE
AND
        PASSWORD IS CORRECT
        SEND UNSUCCESSFUL DECRYPTION
          SMS/EMAIL FEEDBACK
   ELSE IF
      DECRYPTION DATE/TIME DUE AND
PASSWORD
         IS CORRECT
        SEND SUCCESSFUL DECRYPTION
SMS/EMAIL
           FEEDBACK
    ENDIF
STOP
```

3.8 Algorithm for the Enhanced AES File Encryption System

```
START
   SELECT TARGET FILE
   ENTER ENCRYPTION PASSWORD
   WRITE DECRYPTION DATE/TIME SETTING
TO
         TARGET FILE
   ENTER FEEDBACK RECIPIENT PHONE
         NUMBER/EMAIL
   ENCRYPT FILE
STOP
```

3.9 Algorithm for the Enhanced AES File Decryption

```
START
  SELECT TARGET FILE
  ENTER DECRYPTION PASSWORD
  CHECK DECRYPTION DATE/TIME AND
PASSWORD
  IF DECRYPTION DATE/TIME NOT DUE
AND PASSWORD INCORRECT
```

SEND UNSUCCESSFUL SMS/EMAIL
ELSEIF DECRYPTION DATE/TIME NOT DUE AND PASSWORD CORRECT
SEND UNSUCCESSFUL SMS/EMAIL
ELSE IF DECRYPTION DATE/TIME DUE AND PASSWORD CORRECT
SEND SUCCESSFUL SMS/EMAIL
DECRYPT FILE
ENDIF
STOP

3.10 The Model for the Enhanced File Encryption System

This model developed for the Security of Multimedia Data on Transition using Enhanced AES File Encryption System, is discussed in comparison to the existing AES multimedia encryption system. Figure 11, shows the existing AES file encryption system that has a key generator inputted into the AES encryption of system. The original multimedia data is also inputted into the system resulting in a ciphered multimedia data, the reverse of the process gives back the original multimedia data.

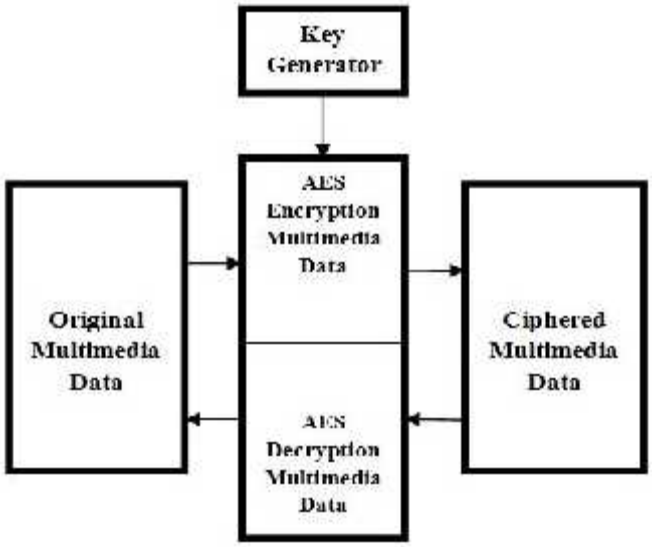


Figure 11. The Model of AES Multimedia Encryption system.

In figure 12, the authors enhanced the AES file encryption system by embedding the Feedback Artificial Agent that serves as a monitoring system for a multimedia data on transition, thereby executing the sending of reports whenever the encrypted data is decrypted and a Timing Circuit algorithm also embedded to the determine when an encrypted multimedia data on transition can be decrypted. The green section is the enhanced section.

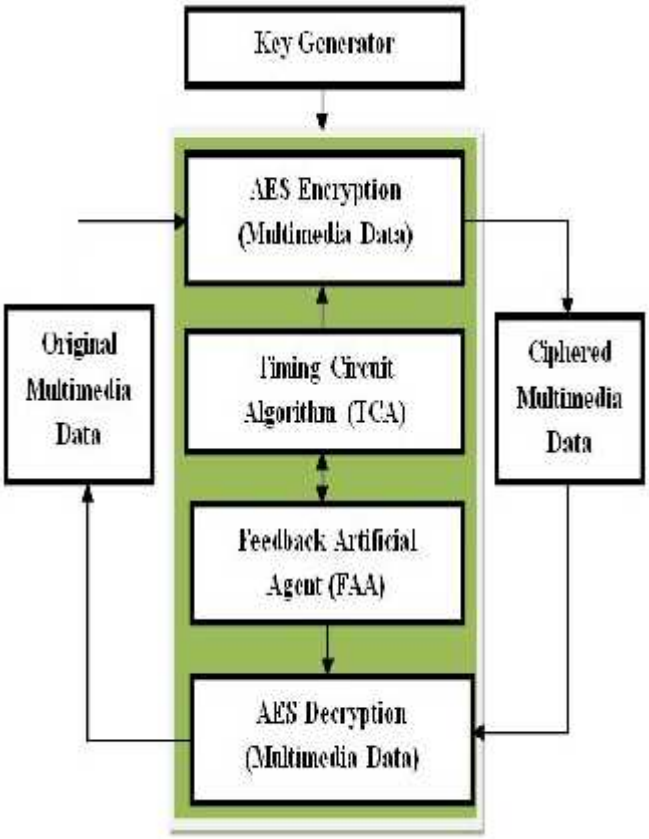


Figure 12. Structural Model for the Enhanced AES Multimedia Data Encryption System.

Figure 13. The Enhanced AES icon

4 MODEL IMPLEMENTATION AND TESTING

The authors discussed the implementation of the software simulation of the Enhanced AES With Time-Bound and Feedback Artificial Agent Algorithm for Security and Tracking of Multimedia Data on Transition.
As explained earlier, the implementation was effected using Java programming Language. This was achieved by some complex coding of the Timing Circuit Algorithm and that of the Feedback Artificial Agent into the existing AES codes to produce a single algorithm (AES Enhanced) for the software implementation of the objectives of this work.

4.1 Launching of the Software Simulation

This system can be accessed directly from the built-up module by double-clicking the Enhanced AES icon to launch or access it from the NetBeans IDE 8.0.2 environment. This is shown in figures 13 and 14 respectively.

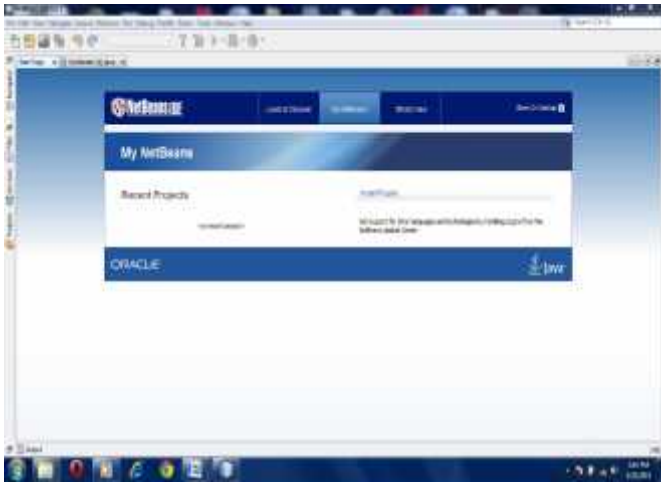


Figure 14. The NetBeans IDE 8.0.2 Environment

4.2 System Application Encryption Interfaces

Figures 15-20 shows the System Application Encryption Interfaces.
It is very important to provide quick step by step process on how to use the system as highlighted in figure 15

Figure 15. System Encryption Interface Displaying the Help Menu.

When the interface is launched, the date and Time security Checkbox can be checked to activate Timing Circuit Algorithm settings. This is shown in figure 16

Figure 16. System Encryption Interface with Date and Time Security Settings.

This interface shown in figure 17 represents the Feedback Artificial Agent (FAA) settings. A default email and phone number can be set here so that reports can always be sent to them by the FAA monitoring the multimedia transition from the source to the destination.

Figue 17. Default Feedback Artificial Agent (FAA) Settings.

Figure 18 represents the interface where a multimedia folder or file to be sent on transit is being access from its location on the computer system, then the expected date and time of decryption were set; also a phone number and an email address were specified where the Feedback Artificial Agent shall be sending its reports to. The ''process'' button is clicked for the encryption process to begin. The Feedback Artificial Agent will automatically be activated to starts its monitoring activities.

Figure 18. Accessing and Encrypting a Multimedia File from Folder

In figure 19, the FAA sends a report displaying the security details and demanding for confirmation before the system finalizes the encryption process.

Figure 19. Inputted Details Confirmation.

Upon the completion of the encrypting process, the FAA sends the report as shown in figure 20. Click the 'OK' to end encryption status.

Figure 20. Multimedia File Encrypted.

4.3 System Application Decryption and tracking Interfaces

Figures 21-29 shows the System Application Decryption Interfaces. The Decryption process involves the simultaneous actions of the Time-bound (Time Circuit Algorithm) and Tracking (Feedback Artificial Agent) mechanisms.

Figure 21 shows the decryption interface of the system. The encrypted multimedia folder has changed to ciphered multimedia folder and the key was entered to decrypt the multimedia folder; the 'process' button clicked to start the decryption process.

Figure 21. The Decryption Interface.

Figure22, represent the interface display the decryption in progress. The FAA is at this moment, monitoring the security details at the encryption stage in order to authenticate them.

Figure 22. Multimedia Folder or File Decryption in Progress.

The interface as shown in figure 23 displays a multimedia file decryption attempt failure because there was no internet connection. The FAA report also fails to deliver as the attempt failed.

Figure 23. Decryption Attempt Offline.

Figure 24 is the decryption interface displaying the message of an invalid password. The encryption and decryption keys used in this system is password based and hence displays 'invalid password' message when an attempt is made to on the multimedia file or folder on transition with wrong passwords.

Figure 24. Decryption Interface displayed attempted with invalid password

This interface shown in figure 25 was an attempt on an encrypted multimedia file on transit when the date and time was not due. Hackers may have succeeded in guessing the password though any method but the Timing Circuit Algorithm that is string to AES Algorithm blocked the decryption.

Figure 25. Decryption Attempted on Transit when time not due

The interface shown in figure 26 represents the successful completion of the multimedia data decryption process at the intended destination when the date and time were also due. The data is decrypted back to the original plain multimedia data

Figure 26. Multimedia Data File Decrypted Successfully.

The interface shown in figure 27 is the action of the Feedback Artificial Agent (FAA) which sends reports to the pre-set email address and phone number on the successful decryption of the multimedia data when the pr-programmed date and time is due.

Figure 27. Feedback Artificial Agent (FAA) Sends Report.

Figure 28 is the interface showing the report that the FAA sent to the administrator's mobile phone as an alert message from the FAA monitoring server.

Figure 28. Report of the Feedback Artificial Agent (FAA) on a Mobile Phone.

Figure 29 is the interface showing the report that the FAA sent to the administrator's e-mail box from the FAA monitoring server.

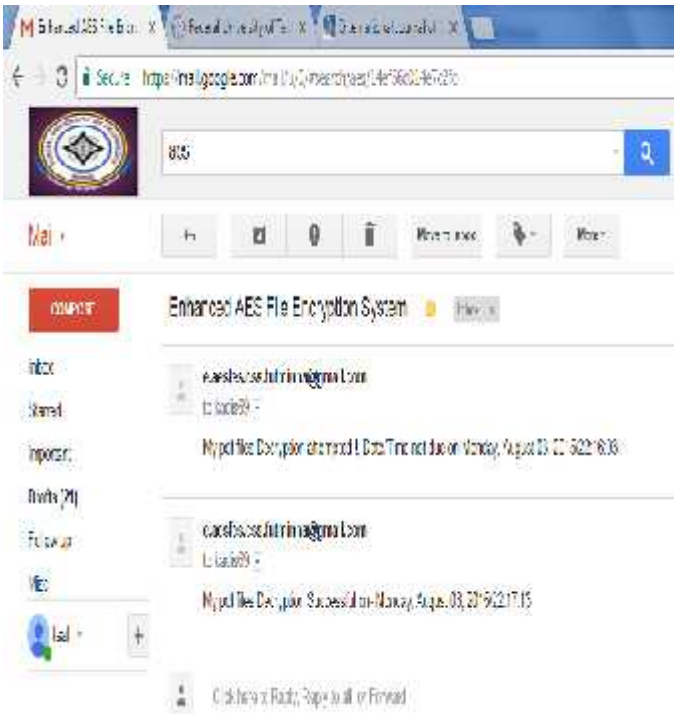


Figure 29. Report of the Feedback Artificial Agent (FAA) Delivered to an e-mail Address.

4.4 Discussion

The system model has been implemented as designed by the authors. Multimedia folder or file can be accessed from any location from the computer system and encrypted as such. This is one of the differences between this very model and most of the encryption system reviewed where the file content has to be copied and pasted or directly written to the text field of the encryption system. The enhancement of the AES with the FAA and TCA has been able to give high reliability to multimedia data transiting via network and also put the data administrator on guard.

5. CONCLUSION AND RECOMMENDATION

The work of this paper, Enhanced AES With Time-Bound and Feedback Artificial Agent Algorithm for Security and Tracking of Multimedia Data on Transition has been able to achieve the aims and objectives which are to design a model for secured multimedia data in transition from a source to destination and also the implementation of the enhanced AES algorithm using java programming language, Timing Circuit Algorithm and Feedback Artificial Agent to monitor encrypted multimedia data on transition. The testing of the performance of the designed model for secured multimedia data on transition was also successful.

The authors recommend further research on this work to enable the monitoring of multimedia data on transition to be tracked, to know the exact location on transit where the data was successfully decrypted or where an attempt was made to decrypt the data. So also, physical security should be taken seriously, as the breach of it can easily thwart all the effort made in securing confidential data.

REFERENCES

1. Avinash Kak "AES: The Advanced Encryption Standard lecture notes on Computer and Network Security" Lecture8: 2017, https://engineering.purdue.edu/kak/compsec/NewLectures/Lecture8.pdf
2. Apple, "Apple Introduces 10GB iPod-2,000 Songs in your pocket', Apple Press Release March 20, 2012. Retrieved from www.apple.com/pr/library/2002
3. Richard Taylor, "Sony cyber attack leaves many questions" BBC Business News, November 26, 2014. Retrieved from www.bbc.com/news/business
4. Olatunde A. Aworanti "Strategies for Managing Examination Malpractice in Public Examinations" 2012, retrieved from http://www.nabtebnigeria.org/wp-content/uploads/2012
5. Alan Yuhas (2014, December 19), Sony CEO insists 'we made no mistake' after US accuses North Korea of hack – as it happened. The Guardian Newspaper. Retrieved from www.theguardian.com/us-news/
6. Pooja Singh, Hardik Upadhyay, Mitesh Thakor, Krunal Suthar, "A Survey: A Hybrid Approach to Secure Transmitted Message by Combining Steganography and Asymmetric Cryptography", International Journal of Innovative Research in Computer and Communication Engineering Vol. 2, Issue 11, November 2014. ISSN (Online): 2320-9801.
7. Adarsh. R, Ganesh Kumar. R, Jitendranath Mungara ''Secure Data Transition over Multicast Routing in Wireless Mesh Network", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-1, Issue-3, August 2012. Retrieved from www.ijitee.com
8. Reena J. Shah, Bhavna K. Pancholi, "Multimedia Security Techniques", International Journal of Innovative Research in Electrical, Electronics, Instrumentation And Control Engineering. Vol. 2, Issue 5, May 2014. Retrieved from www.ijireeice.com
9. Radha. S. Shirbhate, Anushree A.Yerawar, Ankur M. Hingane "Features Preserving Data Encryption Used to Secure Multimedia Data" International Journal of Emerging Technology and Advanced Engineering. ISSN 2250-2459, Volume 2, Issue 1, January 2012. Retrieved from www.ijetae.com
10. Renuka, M. Thangaraj, P "Reliable Data Security Architecture for Multi-Path Multimedia Streaming in MANET", International Journal of Electronics & Communication Technology
IJECT Vol. 3, Issue 1, Jan. - March 2012. Retrieved from www.iject.com
11. Agholor, S. Sodiya, A. S, Akinwale, A. T. Adeniran, O. J. and D. O..: A Preferential Analysis of Existing Password Managers from End-Users' View Point". The Society of Digital Information and Wireless Communications, (ISSN: 2305-0012)
International Journal of Cyber-Security and Digital Forensics, (IJCSDF) 5(4): 187-196 (2016).
12. Nithyabharathi, P. V. Kowsalya, T. Baskar, V. "To Enhance Multimedia Security in Cloud Computing Environment Using RSA and AES" International Journal of Science, Engineering and Technology Research (IJSETR), Volume 3, Issue 2, February 2014. Retrieved from
www.http://ijsetr.org/wpcontent/uploads/2014/02/IJSETR -VOL-3-ISSUE-2-341-345.pdf
13. Jamgekar, R. S. & Joshi, G. S. "File Encryption and Decryption Using Secure RSA" International Journal of

Emerging Science and engineering, 2013 1(4) 2319-6378. Retrieved from www.ijese.org

14. Harn, L., Chang, C. C. & Hsiao, L. W. "An Anonymous Multi-Receiver Encryption Based on RSA" International Journal of Network Security, 2013, 15(4) 307-312. Retrieved from www.ijns.femto.com.
15. Alireza Tamjidyamcholo and Rawaa Dawoud Al-Dabbagh. "Genetic Algorithm Approach for Risk Reduction of Information Security"
The Society of Digital Information and Wireless Communications, (ISSN: 2305-0012)
International Journal of Cyber-Security and Digital Forensics, 1(1): 59-66, (2012)

International Journal of Cyber-Security and Digital Forensics (IJCSDF) 6(4): 179-185
The Society of Digital Information and Wireless Communications (SDIWC), 2017 ISSN: 2305-0012

Deception in Web Application Honeypots: Case of Glastopf

Banyatsang Mphago
Dep of CS/IS
BIUST
Email: mphagob@biust.ac.bw

Dimane Mpoeleng
Dep of CS/IS
BIUST
Email: mpoelengd@biust.ac.bw

Shedden Masupe
Senior Member of IEEE
Botswana Institute of Technology Research and Innovation
smasupe@bitri.co.bw

Abstract

Honeypots are special tools designed to help track and understand attacker?s motives and their attack methods. In web applications, several honeypots have been developed and some have since been abandoned by their developers. But as honeypots are deployed more and more within computer networks, malicious attackers also devise techniques to detect and circumvent these security tools and thereby exposing limitations in most web application honeypots. Dynamic honeypots however, are believed to be the future of honeypots due to their abilities to adjust to the changing environments. Glastpof is one of the more popular if not the most, dynamic web application honeypot currently released to the public. But Glastopf has its limitations too. Once deployed, Glastopf can be easily identified by the attackers due to its performance and appearance, and as such become less useful to the security community. This research describes some of the limitations inherent in Glastopf, and then proposes possible ways to make it more deceptive and more attractive to attackers.

Keywords: *Honeypot, Glastopf, Deception, WSGI, Attack Surface, Web Application, Camouflage*

1 Introduction

Ever since the inception of computers, information security has been a major concern for most organizations. Web application attacks in particular, represent the biggest threats in the organization's security [1–3]. Verizon [2], reported that in 2016 web application attacks are the main source of data breaches, ranked number 1 with an estimate of over 40 percent of all the data breaches. WhiteHack Security [6], also reported that web application attacks represent over 40 percent of all the total data breaches in 2015. Imperva reported that in 2015 content management system (CMS) applications were attacked three (3) times more than non-CMS applications, and Wordpress CMS applications were attacked seven (7) times more than any other CMS applications.

However, web application honeypots are special tools designed to help and understand this kind of attacks. Lance Spitzner [4], argued that dynamic and intelligent honeypots are the future solution to the biggest challenge facing our modern security technologies in configuration, where once the honeypot is deployed it adapts to the changing environment. Budiarto et-al and Abraham et-all [5, 6] also agreed with Spitzner that with a dynamic honeypot, one need not to worry about going back to update the configuration of the honeypot but rather the honeypot adjust itself to the changing environment. Glastopf is one such example of a dynamic low interaction web application honeypot, and the most popular honeypot for its ability to imitate thousands of vulnerabilities. However, the designers of Glastopf honeypot (The Honeynet Project) was not directed much at deception to defeat attackers in the tactical sense, but rather, the dedication was more at learning about the tool, the tactics, motives and sharing lessons learnt [7].

This paper seeks to analyze and propose the methods and techniques that can make dynamic web application honeypots, and in particular Glastopf, more deceptive and still remain attractive to the hacker community. A honeypot that can?t hide its true identity is useless as attackers will simply avoid it, and also a honeypot that is of little interest to the attackers is also of little value to the security community.

2 Honeypot Concepts

The definition of a honeypot is not a straightforward one, mainly because a honeypot is a general technology and not a

solution and do not solve a specific problem. One of the commonly used definition for honeypots is: any security resource that derives its usefulness from being investigated, attacked and compromised by the attackers [8].

There are several ways at which honeypots can be classified, and some of the more popular classifications are based on purpose and level of interaction with the attackers [8–12]. Based on purpose, a honeypot can either be a research honeypot or production honeypot. Research honeypots are those that do not add direct value to the organization but are used by researchers to gain valuable knowledge about the attackers [11, 13]. The primary goal in this type of honeypots is to learn the methods, techniques, and processes the attackers use and pass that knowledge to the security administrators. Production honeypots on the other hand are honeypots deployed in an organization with the primary objective of alerting security administrators of any potential attacks in real time. Based on level of interaction, a honeypot can either be low, medium, or high interaction honeypot. Low interaction honeypot is the one that imitates only services that can be exploited where the honeypot does not provide an operating system for an attacker to interact with, but rather, only simulates services of a particular system [11]. Medium interaction honeypot, just like low interaction honeypot, does not provide an operating system for an attacker to interact with, but rather, the simulated services are more complex technically than the low interaction honeypot. Medium interaction honeypots imitates a collection of systems rather just one in order to present a more convincing interaction with the attacker while still hiding the operating system from the attacker [9, 10]. A high interaction honeypot is the one that gives the attacker to interact with the actual operating system along with real instances of programs rather than their imitations [9, 10].

3 Related Work

This section explores how different honeypot implementations were detected, and how some of these implementations were camouflaged to remain deceptive to the hacker communities.

3.1 Detecting UML Based Honeypots

Detecting honeypots is mostly dependent on the design of the honeypot itself rather than following a predefined criterion out there. Several researchers have used different techniques in detecting honeypots, and all this detections were dependent on how the honeypot is designed and the tools used to design these honeypots.

Innes and Vanapalli [14] discovered that honeypots that use User Mode Linux as their working environment more easily identifiable by the attackers. This is mainly because UML doesn't store data on hard disks but rather on virtual drives that point to already existing portions of file system. Therefore, UML modules were developed to hide where images are mounted (/dev/ubd*) on the UML system, but the major number identifying the /dev/ubd* devices is not the same as the one used for standard IDE or SCSI systems, and this number is 98(0x63).

Holz and Raynal [15], also discovered that honeypots running on UML can also be identified by looking at the /proc tree where there are a number of anomalies that can alert to the attackers that this isn't a real system. First, the cpuinfo of the UML has a model name listed as UML and the mode being reported is Tracing Thread mode. This is because UML by default executes in tracing thread mode, and this is a clear indication to the attackers that this is not a legitimate system. Holz and Raynal [15], also discovered that another way to detect UML based honeypots is to look at the address space of a process in the maps file. The end of a stack in the host operating system is usually indicated by 0xc0000000, but in the UML based it is indicated by 0xbefff000.

3.1.1 Detecting VMware Based Honeypots

Innes and Vanapalli [14], found that prior to 4.5 version of VMware, the hardware was not configurable and remained at defaults where VMware default values can be identified, and this made VMware based honeypots easily detectable. The other discovery by Innes and Vanapalli [14], on VMware based honeypots is the network MAC address that is bound to the network card. The vendor part of the MAC address (the first three octets) on VMware virtual network interface is always one of the following three (3): 00-05-69-xx-xx-xx, 00-0c-29-xx-xx-xx, and or 00-50-56-xx-xx-xx. Innes and Vanapalli also discovered that VMware has an I/O backdoor that was revealed by an analysis of Agobot that it is used to call backdoor functions.

3.1.2 Detecting chroot and Jails Environment

Securing honeypot binaries are often done by the use of chroot and jail environments. Holz and Raynal [15], discovered that the easiest way to tell whether you are inside a chroot() environment is to run an ls-lia command on the root directory, and look at the inodes of '.' and '..' directories. On a standard system, the inodes of this two directories is always two (2), but if you are in the chroot() environment, the inodes of these directories will be something different.

3.1.3 Timing

Another discovery by Holz and Raynal [15], on detecting honeypots is that honeypots running on sebek can be detected by measuring execution time of the read () of the read system where in a system without sebek, minimal time is around 8225 and a scalar product of 0.776282, and in contrast a system with sebek has minimal time of 29999 and scalar product of 0.009930.

3.1.4 Camouflaging Honeypots

Fu et-al [16], discussed ways at which virtual honeypots such as Honeyd can be camouflaged by designing a honey-

pot that supports link latency in the order of one microsecond (μs) instead of the original one millisecond default, to avoid timing signature profiles that attackers might profile against the honeypot hence launching timing attacks against it. They achieved this by changing the Honeyd code to make it support the timing resolution of microseconds, and this involved modifying both Honeyd and the event management library (libevent).

4 The Current Status

Glastopf on its default form comes with some preexisting dynamic web templates. These templates however, have some problems that may hinder the honeypot to achieve its intended goals. First, the templates are too basic for experienced attackers to basically see that it is not a legitimate webserver. The templates are basic in the sense that the text in the template is un-organized and does not make sense to whoever is reading the content. This has the potential to be the first alert to the attackers that the server is not intending to relay information to its visitors but just a tool serving another purpose than that of a web server, which can also affect the page rank of the application. For a search engine to retrieve and display the results, the engine must first understand the text on the page [17]. The second problem on the look and feel of the web templates is that they have no working links on them, and there are no images or any graphics on the template. Although the developers of the honeypot suggest that the default templates can be replaced by one's own templates, the server used does not support use of images and other graphic content on the templates and as such any graphics on custom templates would not be rendered by the server on its default form. This inability to have any working hyperlinks or graphics on the templates has a negative impact on the search engine optimization, page rank, and search engine performance of the application. Search engine optimization is the process of improving the visibility of a website or web page in a search engine results. In addition to textual analysis of web content, search engines examine the hyperlinks between the pages to extract information about the quality of the page, and page rank is one such measure [17].

Another weakness with Glastopf in its current form relates to its performance. Although it is difficult to calculate the response time of a server due to several other metrics factored in such as the size of a file transferred, the medium used, cpu capabilities, static or dynamic content, network input/output, etc, the processing time of a sever however, can be measured, and the best way to quantify this component is to measure it directly, bypassing the internet and other factors limiting the quantification of the response time of a server [18]. In our tests, we measured Glastopf directly on the machine it is running on using Autobench and httperf performance measuring tools, and we found that Glastopf can only process less than ten (10) http requests per second. This finding was also reported by ENISA [7]. Figure 1 shows our test results on Glastopf using Autobench performance testing tool. The red line indicates that indeed the application can only processes less than 10 http requests per second. Figure 2 also shows the results of the same test with httperf measuring tool, and it shows that Glastopf can only process about 8.5 http requests per second. These readings however, can be problematic when compared to that of production servers. Titchkosky et-al [19], measured the performance of Apache web server on both static and dynamic technologies, either connecting or not to the database. In their test, they found that a webserver with dynamic web content of a 2 kB file on Apache 5 to have a response rate of about 600 requests per second or more depending on the dynamic platform used, and about 200 http requests per second before full utilization of the CPU on a server that connects to the database. These overheads are increased by a factor 8 on a static payload. Apache, however, is not the fastest production server out there due to its high memory usage. Other production servers such as lighttpd can reach requests rates of up to 3000 or Nginx can even reach request rates of up to 12000. Having such a low http request processing rate for Glastopf compared to other production severs can have two main problems in achieving its goals. First, the low request processing rate has the capability to alert the attackers that this might not be a legitimate webserver. Secondly, a low performing webserver has the capability to be less interesting to the people it is mostly targeting, the hackers, because of its slowness. "A webserver that fails to deliver its content, either in a timely manner or not at all, causes visitors to quickly loose interest [20]."

The problem with Glastopf performance is largely attributed to server it uses. Glastopf use custom made WSGI webserver, which by default is not design to be a server but an interface to other WSGI compatible production servers. In our tests, we used CProfile to profile our entire Glastopf application in order to indentify where the bottleneck might be in the application. Due to the large size of the diagram produce by our CProfile, for the prupose of this paper we only show a small section of that diagram (Figure 3). In this CProfiling, it was evident that the majority of a application's bottleneck was coming from the WSGI server and the application its imports while running as shown by the red and purple components in the diagram.

5 Proposed Framework

The conceptual diagram (Figure 4) represents the overall proposed Glastopf optimization for its deception, while making it more attractive to attackers. The first 3 levels of the diagram reflect the current status of Glastopf as stated in the current status section above, whereas the last 3 levels of the diagram reflect the propositions that can make Glastopf more deceptive, as described in this section below.

To make Glastopf process requests faster like a standard server, we employed Gunicorn wrapped behind an Nginx proxy server. This required re-programming of WSGI so that the WSGI application callable is in the same script with runner file script for Gunicorn to able to run. In Glastopf,

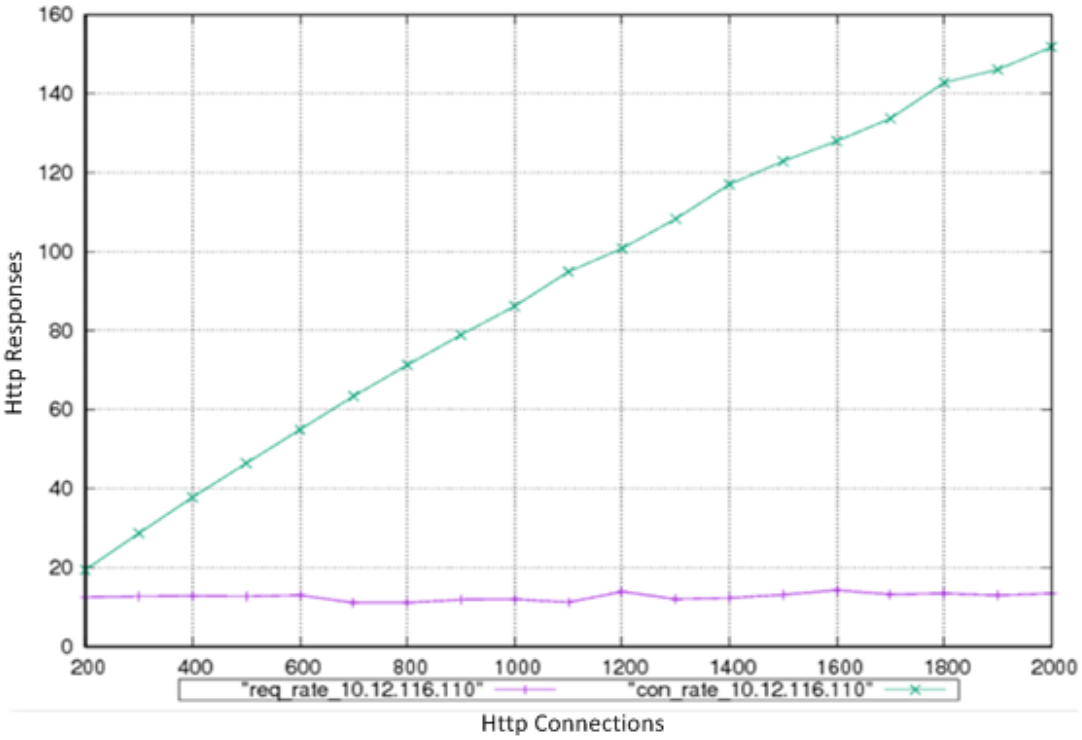


Fig. 1. Glastopf Performance-Autobench

Connection rate: 150.0 conn/s (6.7 ms/conn, <=820 concurrent connections)
Connection time [ms]: min 12.5 avg 3562.9 max 4706.7 median 4545.5 stddev 1864.0
Connection time [ms]: connect 937.1
Connection length [replies/conn]: 1.000

Request rate: 8.5 req/s (118.0 ms/req)
Request size [B]: 74.0

Fig. 2. Glastopf Performance-httperf

the WSGI application callable is the wsgi_wrapper.py file and runner file is the glastopf-runner.py file. By doing this, we have now moved Glastopf from a development server (WSGI) to a production server (Gunicorn) to gain more performance from our application. Gunicorn is based on a pre-fork worker model where one can specify how many workers to deploy for handling requests. In our tests, based on the fact that our test computer had a quad core processor, we then deployed 9 workers, from the general recommended formula of (2*num-of-core) + 1. Then we used httperf performance tool set at 1000 connection rates to measure our application?s perfomance, and we found that Glastopf now has a request processing rate of 999.1 requests/second. Unlike WSGI server, Nginx has the capabilities to serve static files on a web template. By interfacing Gunicorn and Nginx with WSGI, we have optimized our servers so that serving static files is handled by Nginx and any other contents of the request are forwarded to Gunicorn while WSGI provides the interface to the application.

Although running Glastopf on its default WSGI server in a production environment is not ideal due to performance, the problem of serving static files can still be addressed by creating routes/ router for your WSGI application to serve static files. A router is a script that maps URLs to the codes that handles them, and this directly connects content to what is seen on the webpage. To achieve this, we first created a folder in our application directory and then saved all static files in that folder. Then in our WSGI application callable script, we modified the environment path to point to the static folder created so that whenever a static file is needed, the application can fetch the files using the static route in the static folder.

Another proposition reflected in the conceptual diagram is on the attack surface of Glastopf. An attack surface is any system's resource or any application area that is exposed to attack or can be used to attack an application [21]. The bigger the attack surface is, the more attractive the application is to the attackers. To make Glastopf?s attack larger,

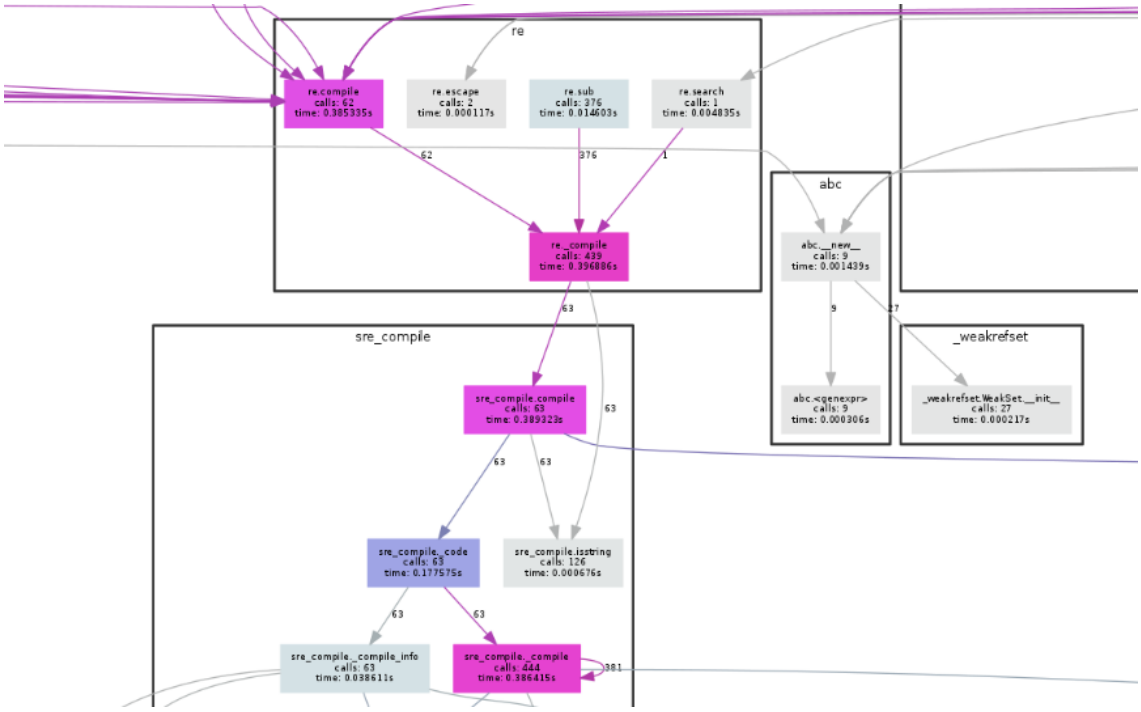


Fig. 3. Glastopf Profile

we determined how attack surface parameters [22, 23] can be deployed in an application in order to make it larger. We achieved this by deploying these attack surface parameters in several formats using different combinations, and then measure the attack surface using Nessus vulnerability scanner. From this test, we found that deploying two same parameters in an application does not increase the size of the attack surface. However, same attack surface parameters with different functionalities would increase the attack surface of an application. Thus, if one attack surface parameter is a subset of another parameter, having both of them in an application would increase the size of the attack surface. Our research revealed that dynamic content in a web application gives the biggest increase in the size of its attack surface than all other attack surface parameters we tested. We also found that if the attack surface parameter does not exist in an application, adding that parameter would automatically increase the attack surface. We created our templates using Wordpress CMS, to bring content based package dependency [**?**] into play, thereby increasing its attack surface. This however, did not show in our test results due to the fact that current vulnerability scanners do not measure content based package dependencies [24].

Lastly, as reflected in our conceptual framework, we propose improving the PageRank of the application in order to make it more visible to the attackers. PageRank is an algorithm, licensed to Google, for performing links analysis. The algorithm works by promoting, among others, pages with high backlinks [25], as they are seen to be more important than pages with less backlinks. Although this is one of our propositions for optimizing Glastopf, for this paper we were unable to test PageRank because out test bed was on a local machine rather than on a registered domain where PageRank could be tested. Therefore, our next paper would focus on how to make Glastopf rank high on Search Engine Optimization by cheating PageRank, Google Panda and Google Penguin algorithms.

With all the above propositions achieved, a perfect honeypot would have been achieved. A perfect honeypot is the one that performs the same manner or as close as possible to the production webserver for easy camouflage, most vulnerable to be more attractive to the attackers, and with a high PageRank for more visibility.

6 Conclusions and Discussions

We analyzed the deceptive qualities of Glastopf honeypot and found out that Glastopf does not handle http request at the same rate as other production webservers. By using a CProfile profiler we found that the bottleneck in Glastopf is caused by its server: a customized WSGI server that can only process 10 http requests per second. We also found out that another weakness in the deceptive qualities of Glastopf is in its attack surface: thus the interfaces the attacker interact with when attacking the honeypot. Glastopf's attack surface is an HTML page consisting of several dorks, where a dork is a vulnerable path to an application and attackers find this dorks through search engines. However, the HTML page used for the attack surface has some problems too that may limit the honeypot from achieving its intended purpose. We found that the HTML page has no working links and the honeypot cannot render any graphics that might be used in the template. A website that doesn't have any working links nor graphics might send alarms to the attackers that

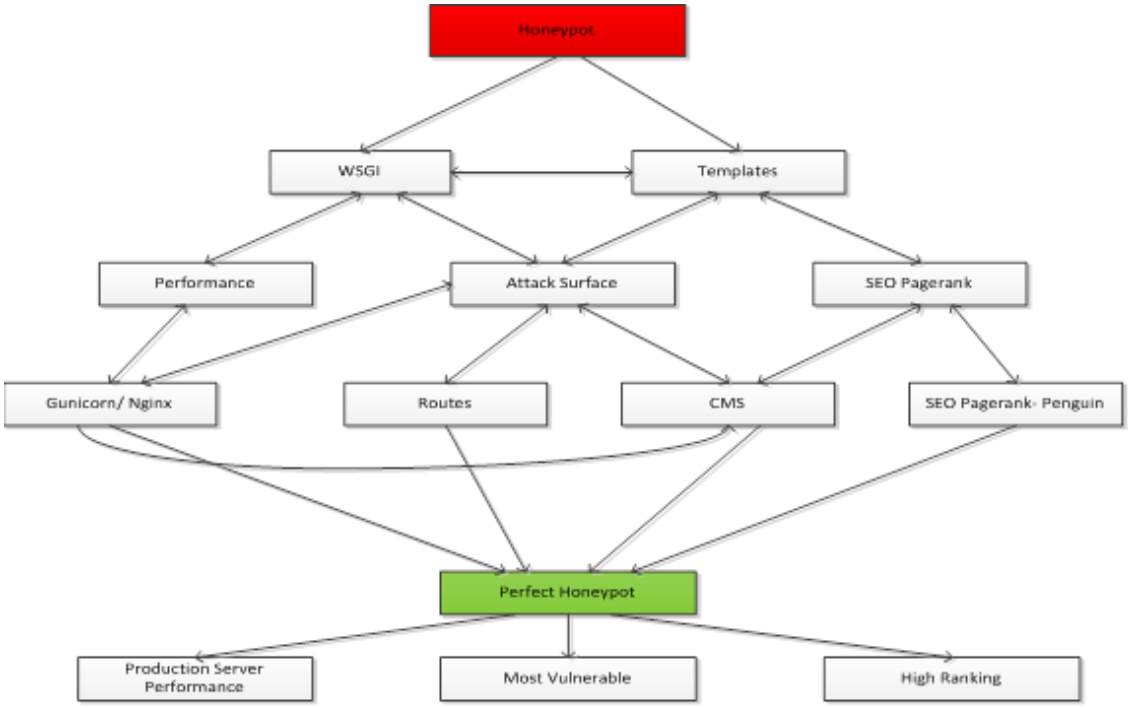


Fig. 4. The Conceptual Diagram

the site might be a honeypot rather than a legitimate website. We found out that the inability of the honeypot to render graphics on web templates was due to its use of WSGI server, which by its designed was built to render dynamic templates and not static files, and therefore cannot render static files such as graphics when not customized. We also found out that the HTML page used for the attack surface has text that doesn't make sense when read. Despite the fact that the honeypot was specifically designed for automated attacks, if it happens that the attacker pulls the interfaces of the site to check what information they have before using the automated tools, he might be able to identify that the website is a honeypot rather than a legitimate website due to the kind of information it has. We therefore proposed a framework that can make the honeypot more deceptive while on the same time making its visibility on the search engines wider. In the conceptual diagram we proposed the use of production web server such as Gunicorn wrapped behind Nginx as a proxy server where the WSGI will act as the interface between the honeypot and the production web server. In this setup, the production server will handle http requests and static files on the web template while the WSGI handles server scripts in the honeypot. From this setup, our honeypot was able to handle about 999.1 http requests per second from the original of about 10 http requests per second, and we can safely say now our honeypot performs like a production server. If production server is used instead of the WSGI, the production server such as Nginx has the capability to render static files, therefore the HTML page can now have graphics and other static files on it so that its attack surface look more like a legitimate webserver. However, if the production server is not used but instead the WSGI is the main server, we proposed proposed creating routes for static files in the WSGI server, and if routes are created, the attack surface will be able to render static files and look more like a legitimate website. We also proposed the use of CMS in creating the attack surface of the honeypot. CMS brings content based package dependency into the honeypot, and content package dependency has the capability to increase the size of the attack surface. We also proposed ways at which attack surface parameters can be used to increase the size of the attack surface of an application. The larger the attack surface of the honeypot, the more visibility it gets when attackers search for vulnerable paths using search engines, and if the honeypot gets more visibility on the search engines, then chances of it being attacked are higher. Lastly on our conceptual diagram we proposed improving the page rank or the honeypot in order to improve its visibility on the search engines. However, due to the requirement of having a public IP to test page rank, in this paper we were unable the page rank of our application.

References

[1] Imperva, "2015 Web Application Attack," Tech. Rep., 2015.
[2] Verizon, "2017 Data Breach Investigations Report Tips on Getting the Most from This Report," Tech. Rep., 2017.
[3] WhiteHat Security, "Web applications security statistics report 2016," 2016.
[4] L. Spitzner, "Dynamic Honeypots," 2003. [Online]. Available: https://www.symantec.com/connect/

articles/dynamic-honeypots

[5] R. Budiarto, A. Samsudin, C. W. Heong, and S. Noori, "Honeypots: why we need a dynamics honeypots?" *Proceedings 2004 International Conference on Information and Communication Technologies From Theory to Applications 2004*, no. May, pp. 565–566, 2004.

[6] R. Prasad, A. Abraham, A. Abhinav, S. V. Gurlahosur, and Y. Srinivasa, "D Esign and Efficient D Eployment of Honeypot and Dynamic Rule Based L Ive," vol. 3, no. 2, pp. 52–65, 2011.

[7] ENISA, "Proactive Detection of Security Incidents - Honeypots," p. 181, 2012.

[8] L. Spitzner, *Honeypots: Tracking Hackers*, 2002.

[9] M. Gibbens, "Honeypots," pp. 1–12, 1999.

[10] R. Baumann and C. Plattner, "White Paper : Honeypots," 2002.

[11] I. Mokube and M. Adams, "Honeypots : Concepts , Approaches , and Challenges," pp. 321–326.

[12] G. Wagener, R. State, A. Dulaunoy, and T. Engel, "Self adaptive high interaction honeypots driven by game theory," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5873 LNCS, pp. 741–755, 2009.

[13] M. Shukla and P. Verma, "Honeypot : Concepts , Types and Working," vol. 3, no. 4, pp. 596–598, 2015.

[14] S. Innes and C. Valli, "Honeypots: How do you know when you are inside one ?" *Australian Digital Forensics Conference*, p. 28, 2006.

[15] T. Holz and F. Raynal, "Detecting honeypots and other suspicious environments," *Proceedings from the 6th Annual IEEE System, Man and Cybernetics Information Assurance Workshop, SMC 2005*, vol. 2005, no. June, pp. 29–36, 2005.

[16] X. Fu, B. Graham, D. Cheng, R. Bettati, and W. Zhao, "Camouflaging Virtual Honeypots," *Work*, pp. 1–17, 2005. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.84.431

[17] D. Gleich, "Models And Algorithms For PageRank Sensitivity," *PhD Thesis*, 2009.

[18] A. Savoia, "Web Page Response Time," *Software Testing and Quality Engineering Magazine*, vol. 2001, no. August, pp. 48–53, 2001. [Online]. Available: https://www.stickyminds.com/better-software-magazine/web-page-response-time-101-primer

[19] L. Titchkosky, M. Arlitt, and C. Williamson, "A performance comparison of dynamic Web technologies," *ACM SIGMETRICS Performance Evaluation Review*, 2003.

[20] D. P. Olshefski, "Measuring and Managing the Remote Client Perceived Response Time for Web Transactions using Server-side Techniques," 2006.

[21] P. K. Manadhata and J. M. Wing, "An Attack Surface Metric," *IEEE Transactions on Software Engineering*, vol. 37, no. 3, pp. 371–386, 2011. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=5482589{\%}0Ahttp://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5482589{\%}0Ahttp://reports-archive.adm.cs.cmu.edu/anon/anon/usr/ftp/2008/CMU-CS-08-152.pdf

[22] S. Goswami, N. R. Krishnan, Mukesh, S. Swarnkar, and P. Mahajan, "Reducing attack surface of a web application by open web application security project compliance," *Defence Science Journal*, vol. 62, no. 5, pp. 324–330, 2012.

[23] T. Heumann, S. Türpe, and J. Keller., "Quantifying the Attack Surface of a Web Application," *Sicherheit*, pp. 305–316, 2010.

[24] S. Zahir, J. Pak, J. Singh, J. Pawlick, and Q. Zhu, "Protection and Deception: Discovering Game Theory and Cyber Literacy through a Novel Board Game Experience," pp. 1–8, 2015. [Online]. Available: http://arxiv.org/abs/1505.05570

[25] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank Citation Ranking: Bringing Order to the Web," *World Wide Web Internet And Web Information Systems*, vol. 54, no. 1999-66, pp. 1–17, 1998.

International Journal of Cyber-Security and Digital Forensics (IJCSDF) 6(4): 186-197
The Society of Digital Information and Wireless Communications (SDIWC), 2017 ISSN: 2305-0012

Preserving Confidentiality and Privacy of Sensitive Data in e-Procurement System

Rajesh Narang[1] Chief Technology Officer, National e Governance Division, New Delhi, India
Tanmay Narang[2], Ex-IT-Analyst, KPMG, DLF Phase 3 Cyber City, Gurgaon, India.
[1] rajesh.narang@digitalindia.gov.in, [2] tanmay.narang92@gmail.com

ABSTRACT

***With the inclusion of direct purchases by Government in e-Procurement System, several security agencies are joining it but they expect their sensitive data to remain hidden. So they expect such system to comply not only to the confidentiality, authenticity and non repudiation guidelines given by World Bank related to Price quotation for e-procurement system, but also to comply with maintaining confidentiality of data related to Government buyer departments. The Security Model proposed here studies threats, vulnerabilities and risks to e-Procurement System, evaluates the suitability of Tokenization, Masking and Encryption techniques by applying them to ensure confidentiality, privacy and integrity of data related to bids, and security agencies. The study finds out that masking need to be applied to ensure confidentiality of data of security agencies, Public Key Infrastructure (PKI) to maintain confidentiality and authorization of price quotation and bid, e-sign to bind both buyers and suppliers to the actions taken by them in life cycle of e-Procurement process. Database security controls need also to be implemented so that data related to security Agencies and keys used to encrypt/ decrypt price quote are put in different tables/vault accessible only to authorized users excluding Data Base Administrator*. Current approaches focus only on PKI.**

KEYWORDS

Tokenization, Masking, Encryption, e-Procurement System, Privacy, Confidentiality

1. INTRODUCTION

Government e-Market (GeM) Place [1], a type of e-Procurement System, is a meeting place of suppliers and purchasers where purchase to pay process is electronically supported and a Government department defines sets of rules for procurement process. E-marketplace is a Business to Business relationship model (B2B) wherein multiple buyers can select products and

Services from pre-sourced catalogues and perform commercial transactions with multiple sellers through a Web platform. The B2B model allowing direct purchases, procurement through bidding and reverse auctioning helps organizations in saving cost and increasing productivity enormously [2], so realizing the benefits of Government e-Market Place, several security agencies are also keen to join it. But they are concerned about the confidentiality and privacy of their sensitive data, the way suppliers were worried about the confidentiality of their price quotation before bid opening time. World Bank [3] issued the guidelines for it and Public Key Infrastructure based technique was used to address it. Similar solution for preserving confidentiality of sensitive data of Security Agencies is studied and being implemented through data masking technique. According to NIST [4] , the growth of business, buyers and suppliers on GeM [1] depends upon its secure and reliable functioning. Studied

by Abdullah et al. [5] found, that security breaches result in loss of reputation, customers and economy of a country. As application design of e-Procurement System solution varies from one environment to another, the design of Indian Government e-Procurement System solution and Ketera used in UK and USA by one of the largest multinational companies in procurement of fuel and energy, studied by Juliette Stephens and Raul Valverde are considered to increase the surface to identify threats, vulnerabilities and associated risks and applicability of the proposed Security Model. This model is designed on the basis of cyber security research methodology, OCTAVE-S [6], ,It spans over four stages: 1) Assets Profiling, 2) Infrastructure Vulnerabilities, 3) Privacy and Confidentiality [7] Techniques and 4) Security Framework. Each stage of the security study is explained section wise in four subsequent sections of the paper and finally findings of the proposed security model are summarized at the end.

2. ASSETS PROFILING

The list of all assets such as servers, firewall, system software, etc. required for hosting e-Procurement System was prepared and assets were demanded from Cloud Service Provider (CSP). The owner Department of Government e-Procurement System checked that CSP produces evidence of implementing ISO27001 [8] Information Security Management System applicable controls in all concerned areas of Infrastructure and operation level. It was ensured that CSP is security certified and CSP has policies in place to prevent spread of viruses through computing infrastructure. It has a documented policy to apply system updates and patches to computing infrastructure provided by it and all the personal edge devices are updated, patched and have up to date antivirus protection.

3. INFRASTRUCTURE AND APPLICATION DESIGN VULNERABILITIES

Here vulnerabilities testing at System and Network, application source code and design, were conducted. Penetration testing tools were used to find the vulnerabilities at System Level, Network Level and application source code. Since tools cannot check vulnerabilities present in Application Design, application threats anticipated, threats identified in other e-Procurement Systems[9], Security incidents of 185 Japanese organizations studied by Abdullah et al. [5], and World Bank guidelines [3] were considered to construct Threat Model which is given as below:

- Actions of Anonymous and Validated Users (especially those performing Reporting and Analytics) were reviewed and checked if it was possible to see:
    - the products,
    - delivery locations and
    - Buyers' details of security agencies in purchase orders.
- Actions of Validated Users in defined roles were examined and checked if it was possible to:
    - access or modify the information and privileges defined for other roles
    - access or modify information of other

Users defined in the same role
    - attach fake budget approval or modify sanctioned amount or perform any
    - other sensitive process of system anonymously
    - attach fake inspection report and release the payment anonymously
- Examined, if it was possible to ignore Informing (through email/ SMS) some of the registered suppliers and service providers about bid notifications to avoid one bid situation and get competitive bid price
- Verified the price quotations received are encrypted and kept in a Vault in a separate table on a separate database server, accessible only to buyer departments before bid opening time
- Verified if there exists any mechanism to bind the suppliers and service providers with their offers including discounts, if any.
- Verified if it was possible to tamper price quotations once opened

Success in any of the action as mentioned above results in compromise in the integrity or confidentiality of the application or back-end data. It will impact the one or the other stakeholders of the e-procurement system as shown in Table 1 and they may lose trust in e-procurement system.

Table I: Threat Assessment Model

Risk	Severity	Actor Impacted
Ability to see data of security agencies (buyers) related to orders, product procured and delivery locations and sharing with adversaries	High	Security Agencies
	Low	Other Agencies
Ability to see price quote of suppliers before opening date and time of bid as it is not encrypted and disclose or modify it	High	Suppliers
Price quotes are in the same table where bid generated and accessible to DBA	High	Suppliers
Common Data Architecture shared between security agencies (e.g. same tables or different tables in same Data Base) and other Government buyers	High	Securely Agencies
	Low	Other Govt. buyers/agencies
Weak Authorization security control with no binding as to who created in following modules : • Order Modules • Billing Modules • Inspection Modules • Payment Modules	 High High High High	Security and other Agencies
No intimation of bids/Reverse Auctions sent to Suppliers/Service Providers when bid/reverse auction initiated	High	Security and Other Agencies

4. PRIVACY AND CONFIDENTIALITY TECHNIQUES

At this stage, the documented processes of e-Procurement System owner, articulating policy and procedure were examined. It was found that policy for following processes were present:

- Registration of Suppliers and Government buyer departments to ensure neither Supplier nor Buyer can be registered without proper e-verification and they cannot register multiple times to avoid submission of three bids by same supplier.
- User accounts and passwords to avoid adversaries to guess user credentials and access system
- Reporting security incidents so that if any event real or suspicious about information compromise comes in notice, it must be reported to Security Group.

It was found that the e-Procurement System owning department needs a policy to handle sensitive data of security agencies to build trust. So a study of three techniques: Tokenization, Masking and Encryption was undertaken to find out the most suitable among them to preserve confidentiality and privacy of the sensitive data.

4.1 Data Hiding Techniques

Security agencies also come on e-Procurement System portal to procure sensitive products such as weapons, uniforms, bullet proof vests, etc. and at the same time they want to maintain confidentiality about such procurements so that adversaries don't know about details of delivery locations and indentors. They want their sensitive data to remain confidential, like:

- Identifier: Attributes that can directly and uniquely identify an individual, such as name, ID and mobile number.
- Quasi-identifier: Attributes that can be linked with external data to re-identify individual records, such as gender, age and area code.
- Sensitive Attribute: Attributes that an individual wants to conceal, such as department name, products ordered and delivery locations.
- Non-sensitive Attribute: Attributes other than Identifier, Quasi-identifier and Sensitive Attribute..

It can be achieved by applying the concept of k-annoymity [10] it is shown in Table 2 where sensitive information is concealed by applying concept of 6-annoymity where information of six attributes is concealed.6-Annoymity With Respect To Department(Dept.), Order No., Product, Delivery Location and Email is shown in Table 2.

Microsoft [11], has given the concept of Differential Privacy where a Privacy Guard sits between user and the database server, it hides the sensitive information. Using techniques of k-annoymity.

The key question is which technique to choose Tokenization or Masking.

a) Tokenization

Tokenization [12] is a way , to replace sensitive data with non sensitive placeholders called tokens. The sensitive data is replaced with tokens in relational databases and files. The tokens are random values which replace original data but have no intrinsic value. For example, in a casino, money is replaced with tokens which is used for making payments and whatever number of tokens are left at the end, they can be converted to money in a controlled environment. A token used in place of cash cannot be used for doing financial transactions.

On the other hand, encryption is a way of protecting data by scrambling it into an unreadable form. It can be converted into readable form using right key. So attackers try to catch hold of encryption keys since with keys original data can be created. Compared to encryption, Tokenization is not scalable and suitable for unstructured data such as large files.

b) Masking

The general principal to preserve privacy is to suppress the sensitive data before it is subjected to disclosure. The sensitive data in the database is hidden. The identifiers which uniquely zero downs someone is suppressed, it is called as

Table 2: 6-Annoymity Privacy for Security Agencies

Dept	Order No.	Product	Delivery Location	Value	e-mail
*	*	*	*	500	*
*	*	*	*	600	*
*	*	*	*	550	*
*	*	*	*	450	*

de-identification. One such technique of de-identification privacy is k-anonymity [10], A personal record is said k-anonymous if every record is distinct from at least k-1 other records over given “quasi-identifier” subsets of attributes. A quasi-identifier attribute such as age, gender and address, is one which gives clue about the identity of a personal record if linked to external dataset available with attacker.

Therefore, quasi-identifier attributes are completely or partially suppressed so that k-anonymity is attained. There are four ways to replace exact attribute value with less informative value, first replace the value such as 27.23 to 25.0; second do top coding, replace age above 60 to 60 and qualify it a senior Citizen, third do generalization, define address though

zip code, fourth use intervals, define age 18 to 15-20, name Sharma to 'S-T', but if the requirement is for deep analysis of sensitive information such techniques cannot be used. Oracle [13], uses four masking techniques: 1) Condition based masking which is applied when two types of formats (e.g. pure numeric in one state and alphanumeric in another state) are used for storing same information, 2) Deterministic masking which assigns same masking value to a Social Security Number across all databases, 3) Compound masking is applied to related columns as a group (e.g. name is stored as first name, last name) to retain same relation and 4) Key based reversible masking where data is masked and recovered in original format using same key.

Masking Performance: According to the benchmarking done by Oracle [13], on an Exadata X2 - 2 full racks with high performance discs, 2 x 6 core Intel Xeon X5675 Processors (3.07 GHz). 1 TB data base containing a table of 6 billion rows whose one column was replaced with random numbers took 12 hours 49 seconds. 100TB data base containing a table of 600 billion rows where one column was replaced with fixed number took 33 minutes.

c) Encryption

Functional encryption is used to encrypt data; its aim is to ensure that an untrusted process learns only the output of a function f(x) about data and nothing else. Earlier engines were not allowing performing SQL operations on encrypted data, but work on CryptDB [14], has shown different SQL operations can be performed with 21-26% reduction in through put. CryptDB [14] , sits between users and data base management system (DBMS) as a security guard. It uses different types of encryption algorithm for different types of columns depending upon the SQL operations performed on them. For example, if addition on two encrypted columns is to be performed, they are encrypted using Paillier [15][, cryptosystem since it allows addition on encrypted data. Similarly if multiplication is to be performed on two columns, multiplicative homomorphic encryption ElGamal [16] cryptosystem is used to encrypt them. For columns on which equality comparison, equi join and GROUP BY are performed such columns are encrypted by applying deterministic (DET) encryption algorithm.

Similarly if on a column ORDER BY or MAX or MIN SQL operations are to be performed then Order-preserving encryption (OPE) of Boldyreva[17] is used to encrypt such column. When a column is encrypted with OPE algorithm of Boldyreva [17] , the server performs range queries giving encrypted constants OPEK(c1) and OPEK(c2) to DBMS corresponding to the range [c1, c2]. Using this encryption method, ORDER BY, MIN, MAX, SORT, etc. can also be performed. The randomized (RAN) encryption does not support any operators, and intuitively, the most secure encryption scheme is chosen for identifiers. It is implemented using Blowfish,[18] to encrypt integer values, taking advantage of its smaller 64-bit block size, and AES [19], encryption algorithm to encrypt everything else with a 128-bit block size.

Throughput of CryptDB [14] compared to MySQL is 21-26% slower and it is impacted maximum by SUM clause, which when performed on encrypted data brings it down 2.0 times less and when used in UPDATE Statement brings it down by 1.6 times. According to

Stephen Tu et al.[20], the disk space requirement of CryptDB [14] is 4.21 times more to the disk space consumed by plain text. If plaintext takes 17.10GB then CryptDB [14], takes 71.98GB Disk space.

5. CONSTRUCTION OF SECURITY MODEL

A security Model is a framework, which refers to a set of well-documented processes defining policies and procedures; and management of security controls of an Information Technology System. A security risk [21] is the loss potential to an asset of organization if at hreat can exploit a vulnerability. The goal of security controls is to contain the security risk and protect information systems, maintain data confidentiality, availability and integrity of business processes. As threats and vulnerabilities vary for each application NIST [4], this Security Model for e-Procurement System is built by taking into consideration the security controls for the threats and vulnerabilities as identified in Table 1 above. These security controls are as follows:

I. Security Control for Risk1 - Data Hiding for Security Agencies

Based on the study of above three techniques, it was found that Tokenization is not scalable and overheads of masking are less than the encryption. Further, masking allows SQL and mathematical operations on masked columns so the data of Security Agencies should be masked. It provides effective safeguard of data when it is at rest and also when used for preparing Reports and data analytics. The Secure Socket Layer(SSL) is used when data is on move.

II. Security Control for Risk 2- Maintaining Confidentiality of Price Quote

The Price Quotations received from bidders are invited in encrypted format using Public Key Infrastructure. Either dongle can be used where Supplier encrypts the price quotation using Buyers Public Key and two employees of Buyer Department decrypt it using their Private Key in the presence of Suppliers at bid opening time and all of them e-sign them to make them tamper proof. Or all the price quotations are encrypted by applying RAN encryption algorithm and master key is stored in a separate Table called as Vault on a separate database server. At the bid opening time, the employee of e-Procurement System logins and the employee of buyer execute the process of decrypting price quote after feeding in One Time Password which the employee receives on his/her registered mobile number in the presence of Suppliers.

III. Security Control for Risk 3 and Risk 4-Maintaining Keys and Data of Security Agencies in separate Tables

All the security controls are at Application Level so a long access control list is maintained. The application design is reengineered to strengthen the security by adding two database level controls, 1) data of Security and other agencies is separated out by storing them in different databases/tables and 2) as said earlier the master key used for encrypting price quote is kept in a separate table on a different database server accessible only

to e-Procurement System owner and buyers.

IV. Security Control for Risk 5- Binding users to critical actions through e-Sign Based Controls

Besides data, for sensitive processes performed by buyers such as entering budgetary approvals, inviting bids, placing orders, accepting consignment and inspecting them, releasing payments; and similarly for sensitive processes such as submitting bids, bills performed by suppliers e-Sign is made mandatory since it brings ***accountability***. Both database level control and e-Sign based control as mentioned for Risks 3 to 5, lower the complexity which otherwise come in application code if they are implemented at application level control level viz., Access Control List .

V. Security Control for Risk 6- Notifications of bids/Reverse Auctions to Suppliers/Service Providers through E-Mail and SMS

Whenever a bid is floated by buyers on e-Procurement System, sending e-mail and SMS are made compulsory to registered suppliers to avoid vendor cartel, biasedness, get competitive bid price and bring ***transparency***. If such intimations are not sent due to one or other failure, it can invite trouble so auditing of actions performed by buyers, suppliers and e-Procurement System owner is made mandatory.

6. IMPLEMENTATION

Based on the information provided in a study conducted for Asian Development Bank [22] for six e-Government Procurement Systems; and information collected for Government e-market Place through interview and code inspection. The key focus of all systems is on keeping confidentiality of bids. The systems can be divided into two groups. There is no information about keeping confidentiality of sensitive information of buyers.

The security framework is applied on systems of two procurement Groups “A” and “B” at four layers: Infrastructure, Application, data, and Process and their security posture is measured.

I. **Infrastructure Layer**: The entire infrastructure needs to be secured at the perimeter level so Firewalls and Intrusion Prevention System must be installed and network segmentation should be done. Also hardening of Operating Systems of Routers, Servers and management Workstations should be done.

Findings: Infrastructure of both Groups is secured.

II. **Application Layer:** The Application design assessed to ensure that it protects the confidentiality, authenticity or integrity ISO 27001 [8] of bid information and other sensitive data of buyers. Here PKI based asymmetric cryptographic Controls and Masking need to be used to encrypt and hide the data. Key management vault shall be

stored separately from encrypted data to support use of cryptographic techniques and DBA should have no access to it. E-sign should be based on two factors authentication [23], the second factor could be one time password or biometric etc. to prevent others to impersonate and sign on a document. The authentication mechanism should ensure credentials are given on a page, which is under SSL.

Findings: The bid is encrypted at Client end in systems of Group "A" and sent through SSL enabled page of client, while in second system "B" bid is encrypted at server side and sent through SSL enabled page of client.

- System "A" has separate vault to store Private Key of Buyer Officers opening the bid, while System "B" stores them on same server where encrypted bid is stored.
- System "A" uses hardware based dongle for Public-Private Key pair whereas System "B" uses e-Sign which uses Social Security Number for signing a document along with OTP.
- E-sign cannot encrypt/decrypt documents as it is not its mandate but dongle can do both bid encryption/decryption at client side to achieve confidentiality and e-sign to achieve integrity and non repudiation.
- So system "B" uses encryption/decryption of bid Price at Server side through RDBMS supported encryption algorithm.

III. **Data Layer**: Sensitive data should be encrypted, masked or hashed in the database.

- Application design should differentiate between data that is sensitive to disclose and must be encrypted using PKI asymmetric key (discussed above).
- Data (bid price after opening the bid) that is sensitive only to tampering must be e-signed and a keyed hash value (HMAC) must be generated.
- The data that is revealing identity of a security agency should be masked to preserve its privacy.
- The data that can be irreversibly transformed (hashed) without loss of functionality (such as passwords) should be treated separately.
- Access control method at Application level should be enforced to provide access to sensitive data and functionality only to permitted users or clients.
- Role-based access controls should be enforced at both database level and application interface. This will protect the database in the event that the client application is exploited.

Findings: It is an important finding so far little attention is paid toward this since it is a post bid opening data.

- Both systems are keeping buyers' information in plain text, which they need to mask.
- Both are using common data architecture for storing information of security agencies and other agencies.
- Both systems are not keeping e-signed bid data.

IV. **Process Layer**: At least following IT security processes and best practices need to be in place for operation and maintenance of Procurement System in line with international standard on Information Security Management System, ISO 27001 [8]

Findings: System "B" disaster recovery site is being designed; it is not available right now. Disaster recovery site must be in place to ensure availability and audit log shipping. As logs are not written for spyware planted at kernel level so logs should not be treated as the sole protection against mala fide acts. Anti spyware, anti spam and antivirus software should be installed. Other findings for both systems are common, which are as follows:

- The application hardening done for Top 10 vulnerabilities defined by OWASP [24]
- Network security assessment done for adequate security through penetration testing and vulnerability assessment as per NIST SP 800-115 [25].
- The software source code evaluated using white box test approach through code review/ inspection process for identifying malicious codes/ Trojan etc.
- End to End transaction workflow checked to verify it is going through the defined path by using dummy test transactions

Keat et. al. [26] conducted a study in Malaysia to know confidence level of people about keeping the data on cloud storage. It revealed that people feel that investigators have limited skill set, limited jurisdiction under their authority and there is shortage of forensic experts so they would not be able to investigate criminal incidents. Further the laws, regulation and guidelines are also inadequate so these will not be able to provide advisory at the time of need. Thus GeM system is hosted on a Private Cloud maintained by a Government agency.

7. CONCLUSION

The proposed Security Model extends existing security foundations of e-Procurement System by adding four security controls at 1) Database layer, 2) Masking and Encryption layer, 3) e-Sign layer and 4) Alert layer. The inclusion of Masking which hides sensitive data but allows SQL and other operations continue to be performed on it, is the hall mark of the proposed Security Model. When masking is compared with Tokenization, it is found Tokenization is not scalable for large number of records. Although encryption as mentioned in CryptDB [14] has come a long way where it is possible to apply different types of algorithms on different columns depending the purpose for which they are used in the application and carry out different types of operations on encrypted columns. Downside is management of encryption keys of different columns, overheads related to extra disk space requirements and degradation of CPU response time. But still, it works better than Tokenization.

For hiding data of buyers, masking technology is better than both Tokenization and Encryption. Masking, similar to encryption scrambles the data without disturbing the format by working on full database or subset of dataset. The aggregation, preservation of sum and average values over masked values can be done. Masking can be performed in such a way that it's extremely difficult to reverse engineer the original values. It helps in guaranteed security of data.

REFERENCE

1. Carsten Block, "A Decision Support System for Market Mechanism Choice in e-Procurement", Dagstuhl Seminar Proceedings 06461 Negotiation and Market Engineering. [ONLINE]. Available: http://drops.dagstuhl.de/opus/volltexte/2007/989

2. Neef, D. :E-procurement from strategy to implementation. FT press, (2001)

3. World Bank : Guidelines Procurement under IDRB Loans and IDA Credits, available on t

4. National Institute of Standards and Technology,: "Framework for Improving Critical Infrastructure Cyber security", Version 1, (2014)

5.Abdullah Almubark , Nobutoshi Hatanaka: . Osamu Uchida & Yukiyo Ikeda,": Identifying the Mechanisms of Information Security Incidents through Corporate Culture Variables and Sampling", International Journal of Cyber-Security and Digital Forensics (IJCSDF) 5(2): pp. 61-74, (2016).

6.Alberts, C., Dorofee, A., Stevens, J., & Woody,: C. Octave-S Implementation Guide. [Online]. Available: http://www.cert.org/octave, (2008)

7. UCI Research (Privacy and Confidentiality): available on http://www.research.uci.edu/compliance

8. ISO / IEC 27001: Information Security Management System Requirements

9.Juliette Stephens & Raul Valverde,: "Security of E-Procurement Transactions in Supply Chain Reengineering", Computer and Information Science; Vol. 6, No. 3, pp 1-20,(2013)

10. Pierangela Samarati and Latanya Sweeney, Protecting Privacy when Disclosing Information: k-Anonymity and Its Enforcement through Generalization and Suppression, DARPA/Rome Laboratory, pp. 1-19, (1998)

11 Microsoft differential privacy for everyone, http://download.microsoft.com , (2015)

12.Adrian Lane: "Tokenization vs. Encryption: Options for Compliance ", pp. 1-10. (2012)

13.Data Masking Best Practice White Paper-Orcleavailablewww.oracle.com/us/products/database, (2013)

14.Raluca Ada Popa, Catherine M. S. Redfield,: Nickolai Zeldovich and Hari Balakrishnan: Protecting Confidentiality with Encrypted Query Processing, in Proceedings of the 23rd ACM Symposium on Operating Systems Principles (SOSP), Cascais, Portugal, October,(2011).

15. Paillier P: Public-Key Cryptosystems Based on Composite Degree Residuosity Classes. In Eurocrypt (1999)

16.ElGamal,:T. A Public-Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms. IEEE Transactions on Information Theory Vol. 31, Issue- 4, (1985).

17. Boldyreva, A., Chentte, N., Lee, Y., and O'Neill, A.: Order-preserving Symmetric Encryption. In Euro crypt (2009).

18. Schneier B.: Description of a new variable-length key, 64- bit block cipher (blowfish). In Fast Software Encryption , Springer-Verlag, pp. 191–204 (1994)

19. Daemen, J., and Rijmen, V.: The Design of Rijndael: AES - The Advanced Encryption Standard. Springer Verlag, Berlin, Heidelberg, New York, (2002).

20. Stephen Tu, M. Frans Kaashoek, Samuel Madden and Nickolai Zeldovich: Processing Analytical Queries over Encrypted Data, In Proceedings of the 39th International Conference on Very Large Data Bases (VLDB), Riva del Garda, Italy, August (2013).

21. Landoll, D. J., The Security Risk Assessment Handbook: a Complete Guide to Performing

Security Risk Assessment. New York: Auerbach Publications, (2006)
22. Ramanathan Somasundaram , INDIA: Case study on e-Government Procurement Development (for Asian Development Bank) 1-85, (2011)
23. Jing-Chiou Liou, "Performance Measures for Evaluating the Dynamic Authentication Techniques", International Journal of Cyber-Security and Digital Forensics (IJCSDF) 5(2): pp 83-93, (2016).
24. OWASP Top 10-2013: The ten most critical web application security risks :available https://www.owasp.org/images/f/f8/OWASP_Top_10_-_(2013).pdf
25. NIST SPTechnical Guide to Information Security Testing and Assessment available (csrc.nist.gov/publications/nistpubs/800-115/SP800-115.pdf)
26. Yee Say Keat, Babak Bashari Rad and Mohammad Ahmadi, "Cloud Computing Security and Forensics Issues and Awareness of Cloud Storage Users in Malaysia", International Journal of Cyber-Security and Digital Forensics (IJCSDF) 6(1): pp.1-13, 2017

International Journal of Cyber-Security and Digital Forensics (IJCSDF) 6(4): 198-205
The Society of Digital Information and Wireless Communications (SDIWC), 2017 ISSN: 2305-0012

Forensic Investigation Technique on Android's Blackberry Messenger using NIST Framework

Imam Riadi[1], Sunardi[2], and Arizona Firdonsyah[3]
[1] Department of Information System, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[2,3] Master of Informatics Engineering, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
(imam.riadi@mti.uad.ac.id, sunardi@mti.uad.ac.id, arizona.f@gmail.com)

ABSTRACT

In the past few years, there has been a rapid increase in the number of smartphone users. this can be seen with various brands and platforms of smartphones that sold almost every week. One of smartphone platform with a huge amount of users is Android. The rapid development of Android smartphone technology has an impact on the growing number of applications developed for Android platform, including instant messaging applications. Blackberry Messenger (BBM) is one of the multi-platform instant messaging applications with the amount of users that increase significantly each year, causing the possibility of digital crimes that occured by digital crime perpetrator is also significantly increased. In the process of investigating digital crime cases, digital evidences are required to solve these cases. To obtain digital evidences, a technique of forensic investigation on physical evidence has been conducted. This paper studies three widely used mobile forensics tools namely, Oxygen Forensic Suite, Andriller, and Belkasoft Evidence Center in extracting data from BBM application that installed on an Android based smartphone using a framework developed by NIST. The results of this research were presented in the form of recorded conversations, BBM Personal Identification Number (BBM PIN), pictures, and conversation timestamp.

KEYWORDS

Android, Digital Evidence, Blackberry Messenger, Digital Forensics, NIST

1 INTRODUCTION

According to Statista [1], in 2016, the number of smartphone users is estimated to reach 2.1 billion. By 2018, more than 36 percent of the world's population is expected to use smartphones, about 10 percent higher than 2011. As shown on Figure 1, the number of mobile phonec users in the world is predicted to pass the five billion mark by 2019.

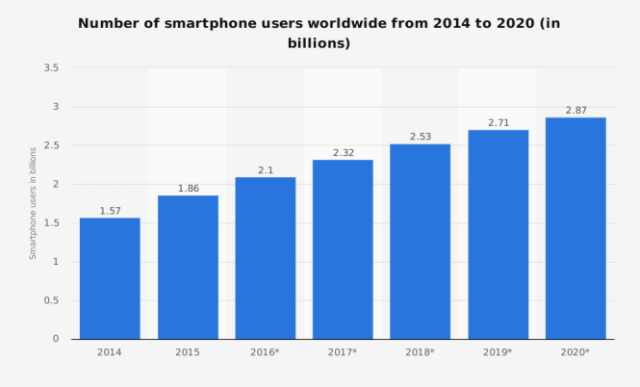


Figure 1. Graphical statistics of smartphone users.

Among those smartphone users, Android and iOS are the two most popular smartphone operating systems in the industry. This rapid development of Android smartphone sales has an impact on the growing number of applications developed for the Android OS, including instant messaging applications. Developers are competing to create instant messaging applications with user friendly features, one of these application is Blackberry Messenger (BBM). A survey conducted by Global Web Index [2], stated that in direct comparison with its competitors, BBM ranked 2nd on worldwide users with 81% registered users, Figure 2 shows the graph.

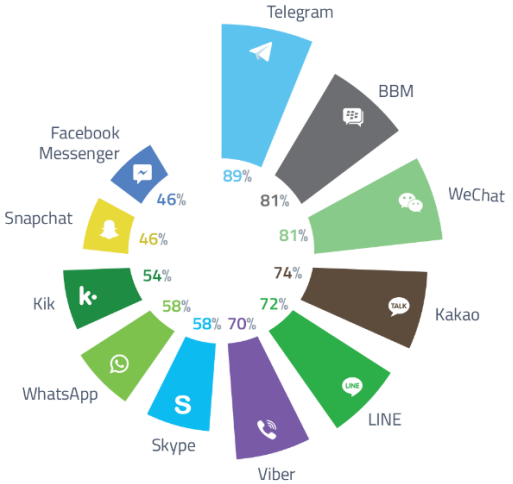


Figure 2. Graphical statistics of BBM users worldwide.

As for Indonesia, a survey conducted by Singaporean online survey organizations We Are Social [3] at 2016, Blackberry Messenger users ranked first with 19% of users, followed by WhatsApp users with with 14% , the graph shown on Figure 3.

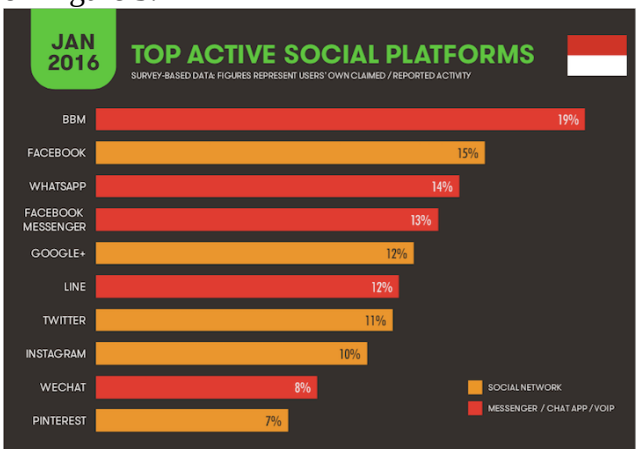


Figure 3. Graphical statistics of BBM users in Indonesia

The increasing of BBM users, in addition to lots of positive impacts that caused, also caused many negative impacts. Many people took advantage of BBM's user friendly features to perform digital crimes such as prostitution, pornography, identity theft, cyberbully, etc. Table 1 shows some examples of digital crimes occured by using BBM at Indonesia [4,5,6,7,8].

Table 1. Digital crime that using BBM (observation)

No	Year	Case
1	2014	Pornography via BBM at Banyuwangi
2	2015	BBM Account hacking at Jakarta
3	2016	Online fraud via BBM at Palopo
4	2016	Identity theft via BBM at Palembang
5	2017	Online prostitution via BBM at Pekanbaru

In the process for solving digital crime cases, necessary supporting evidence is needed. If an android based smartphone became the physical evidence on a case that Blackberry Messenger application was installed, then the application can be analyzed to obtain digital evidence that can be expected to assist law enforcement in solving the cases of digital crimes. The main topic of this paper are to emphasize on the forensic investigation process and to compare mobile forensics tools used in this research by using a framework developed by National Institute of Standard and Technology (NIST) [9].

2 LITERATURE REVIEW

2.1 Mobile Forensics

Mobile Forensics is a science that performs the process of digital evidence recovery from a mobile device using the appropriate way with forensic conditions [10]. The initial research work in this field has focused on acquisition techniques and general forensic analysis of smart devices [11]. This was shown in Burnette's work in 2002 where he discussed the forensic examination of older versions of the BlackBerry and the hardware and software used for acquisition [12].

2.2 Cyber Crime

The website of Interpol in cyber crime section stated "Cyber crime is one of the fastest growing areas of crime". These crime cases include attacks against computer data and systems, identity theft, sexual abuse images, internet auction fraud, the deployment of viruses, and various scams such as phishing [13].

According to the UN (United Nations), cybercrime is: "any illegal behaviour committed by means on relation to, a computer system offering or system or network, including such crime as illegal possession in, offering or distributing information by means of computer system or network". Another definition of Cybercrime is a crime using information technology as instrument or target, and digital forensics, in essence, answer the questions: when, what, who, where, how and why related to digital crime [14]. There are many kinds of cybercrimes: one of the example is cyberbullying, a term that refers to the use of information technology to bully people to send or post text that is intimidating or threatening others [15].

2.3 Digital Evidence

Digital evidence is information stored or transmitted in binary form that may be relied on in court. It can be found on a computer hard drive, a mobile phone, a personal digital assistant (PDA), a CD, and a flash card in a digital camera, among other places. Digital evidence is commonly associated with digital or electronic crime, such as pornography, prostitution, identity theft, phishing, or credit card fraud. However, digital evidence is now used to prosecute all types of crimes, not just digital crimes [16].

2.4 Android

Android is architected in the form of a software stack comprising applications, an operating system, run-time environment, middleware, services and libraries. Each layer of the stack, and the corresponding elements within each layer, are tightly integrated and carefully tuned to provide the optimal application development and execution environment for mobile devices as shown on Figure 4 [17].

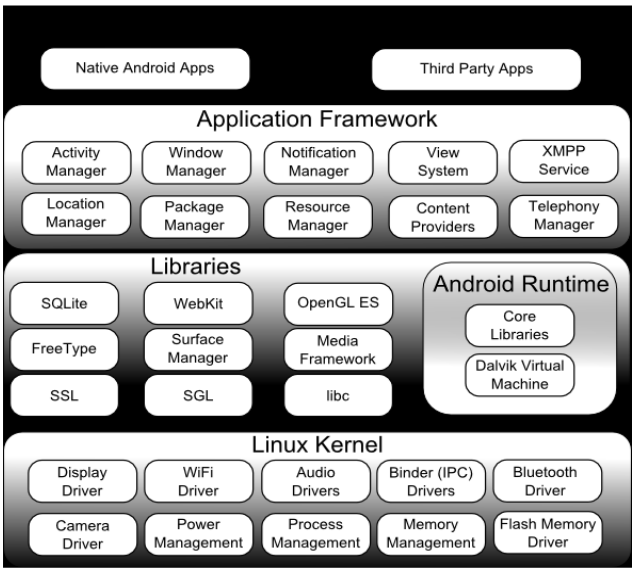


Figure 4. Android Architecture

2.5 Blackberry Messenger

BlackBerry Messenger (BBM) is an application of instant messaging that is developed by Research In Motion (RIM), and originally created for BlackBerry smartphone [18]. As technology advances, BBM is moving beyond Blackberry devices, BBM now become a multi-platfom application that can be installed on Android, iOS, and Windows based smartphone. Since it was created in August 2005, BBM has evolved from a pure messaging application for communication (text and video) to a social eco-system unifying chat, social, commerce and services including games [19].

2.6 Oxygen Forensic Suite

Oxygen Forensic Suite is a forensic software for extraction and analysis of data from cell phones, smartphones and tablets [20]. This tool offers several hash algorithms and one of which can be selected in each investigation case. Oxygen Forensic Suite also has the capability to provide general information about the smartphone and the network that the device was connected to. The other useful capability from this tool is recovered all contacts, SMS, MMS, and user's files [21].

2.7 Andriller

Andriller is a utility which consists of various tools for serving various purposes which includes cracking of screen lock pattern, PIN and passwords, decoding of encrypted databases and files, data extraction automatically and unpacking of android backups. This tool kit solves many of mobile forensics needs for the Android OS [22].

2.8 Belkasoft Evidence Center

Belkasoft Evidence Center is a tool for investigators that can be used to acquire, search, analyze, store and share digital evidence found inside computer and mobile devices. This toolkit will extract digital evidences from multiple sources such as hard drives, drive images, memory dumps, iOS, Blackberry and Android backups. Belkasoft Evidence Center will automatically analyze the data source and lay out the most forensically important artifacts for investigator to review, examine, and report [23].

3 METHODOLOGY

The National Institute of Standard and Technology (NIST) has published a framework for mobile device investigation guide called NIST Mobile Forensics, that contained 4 consecutive steps [24]:

1. Collection: or sometimes called Preservation, this process consist of the steps in gathering and documenting the evidence from the perpetrator, the owner, or at the crime scene. Care has to be taken to preserve other forms of evidence.
2. Examination: or sometimes called Acquisition, in this phase actual data is gathered from the device. In an ideal case the data is forensically copied from the phone as well as from the SIM Card. In some cases technical diffculties can prevent a digital accusation of the device. In a worst case scenario only screen captures of the phone can be gathered.
3. Analysis: Now the gathered data is analyzed for clues regarding the possible crime. The analysis can either be done by hand or with

the help of software tools. There are many software tools available for that purpose. It is important to use different software tools to gather more detailed analysis. Care has to be taken not to miss a crucial piece of evidence solely because one particular tool didn't have the right feature.

4. Reporting: The last step is the most important. Between the gathering of evidence and the presentation in court a signifcant amount of time can pass. An examiner must be able to present his evidence in a conclusive manner and offer the other party information about the tools and methods used.

Figure 5 shows the graphical picture of these 4 consecutive stages.

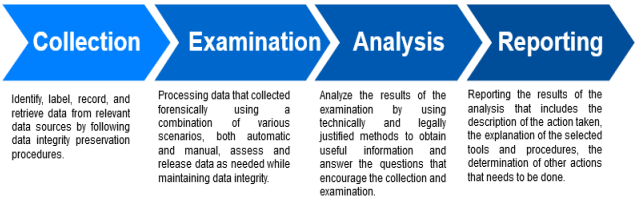


Figure 5. NIST Mobile Forensics Stages

4 RESULTS AND DISCUSSION

4.1 Collection

Collection or Preservation is the earliest stage in mobile forensic methods, and the first thing to do is to search, collect and document the evidence. In this research the evidence is in the form of an android-based smartphone. The result of this stage is as follows:

Table 2. Evidence's Specification

No	Specification Type	Physical Evidence's Specification
1	Brand	Sony
2	Brand Series	Xperia Z
3	Model Number	C6602
4	OS	Android
5	Processor	Quad core 1.5 GHz

Figure 6. Android Smartphone as Physical Evidence

4.2 Examination

Examination is the process of physical evidence backup and retrieval of digital data that contained in it. At this stage the cloning process of physical evidence is conducted. The cloning process can be done using various tools, such as MOBILedit Forensic Express [25]. Figure 7 shows the examination result using Oxygen Forensic Suite.

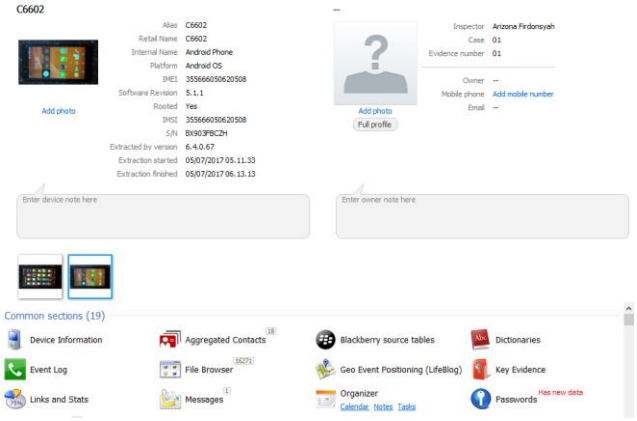


Figure 7. Oxygen Forensic Suite's Examination Result

Oxygen Forensic Suite's examination result provide complete data of physical evidence that contained Device Information, Forensic Examiner's Identity, List of Contact, and Installed Application, BBM included.

For examination process that conducted using Andriller, the result provided is an integrated HTML report that contained all the data extracted from physical evidence. Forensic examiner is able to navigate through the HTML formatted report to find digital evidences. The report shown on figure 8.

Figure 8. Andriller's Examination Result

For Belkasoft Evidence Center, the results given from the Examination stage are relatively similar

to Oxygen Forensic Suite, ie in the form of complete data on physical evidence and applications installed on the physical evidence, including BBM. Figure 9 shows the result of Examination from Belkasoft Evidence Center.

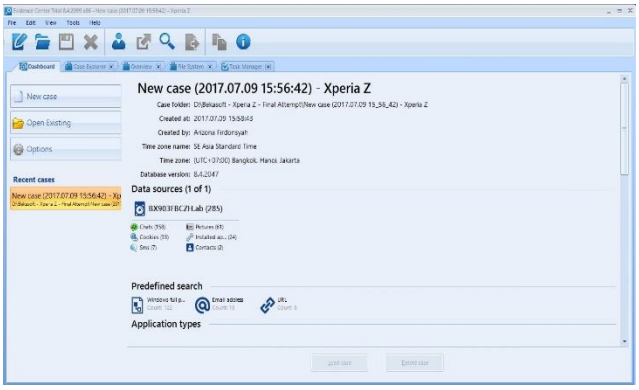


Figure 9. Belkasoft Evidence Center's Examination Result

From the results of this Examination stage, an analysis will be conducted to find digital evidence related to a digital crime case. In this research will be sought digital evidence generated from BBM application.

4.3 Analysis

Analysis is a stage to look Examination result thoroughly to acquire digital evidences. This stage limits the searching process to a certain point that connected to certain data or application. At this research, the search limit is BBM.

Analysis stage conducted by Oxygen Forensic Suite resulted in the form of BBM contact pictures that shows some young girls in an inappropriate pose like shown at Figure 10 (due to inappropriate content, part of the picture's been covered).

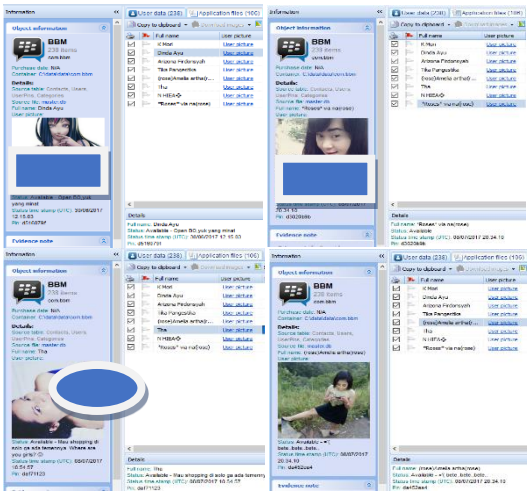


Figure 10. Oxygen Forensic Suite's Contact Analysis

The analysis process also resulted a data conversation with one of the BBM contact that used an unusual and inappropriate language just like shown on Figure 11.

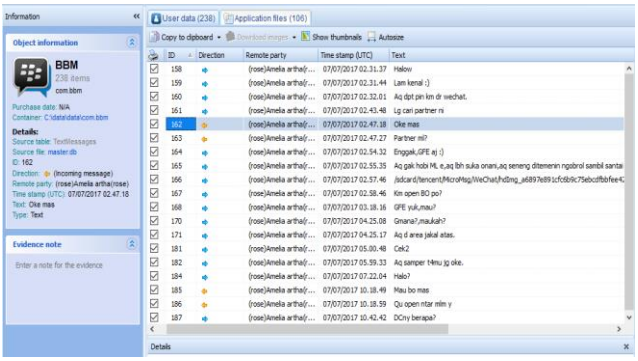


Figure 11. Oxygen Forensic Suite's Chat Analysis

Based on the result above, there is an indication of an online prostitution transaction that happened. To gain more detailed data, analysis of the next tool is conducted. Andriller's examination result did not produce any pictures or photos due to the software limitation but the conversation's data is fully acquired. The analysis shows a conversation that exactly the same like shown on Oxygen Forensic Suite. Figure 12 shows the conversation analysis result.

#	Sender Name	Sender PIN	Recipient PIN	Message	Type	Time
189	Anggara Surya Saputra	dabafdd9	d3020b9b	Salam kenal :-)<	Sent	2017-07-08 10:54:50 UTC+00:00
188	Anggara Surya Saputra	dabafdd9	d3020b9b	Halow,makasih dah accept invite nya :)	Sent	2017-07-08 10:54:39 UTC+00:00
187	Anggara Surya Saputra	dabafdd9	da452ea4	DCny berapa?	Sent	2017-07-07 10:42:42 UTC+00:00
186	(rose)Amelia artha(rose)	da452ea4	dabafdd9	Qu open ntar mlm y	Inbox	2017-07-07 10:18:59 UTC+00:00
185	(rose)Amelia artha(rose)	da452ea4	dabafdd9	Mau bo mas	Inbox	2017-07-07 10:18:49 UTC+00:00
184	Anggara Surya Saputra	dabafdd9	da452ea4	Halo?	Sent	2017-07-07 07:22:04 UTC+00:00
183	Anggara Surya Saputra	dabafdd9	daf71123	Halow	Sent	2017-07-07 06:27:14 UTC+00:00
182	Anggara Surya Saputra	dabafdd9	da452ea4	Aq samper t4mu jg oke.	Sent	2017-07-07 05:59:33 UTC+00:00
181	Anggara Surya Saputra	dabafdd9	da452ea4	Cek2	Sent	2017-07-07 05:00:48 UTC+00:00
174	Anggara Surya Saputra	dabafdd9	d7d4bc3b	Oi	Sent	2017-07-07 04:42:36 UTC+00:00
173	Anggara Surya Saputra	dabafdd9	daf71123	Aq dpt kontakmu dr wechat	Sent	2017-07-07 04:41:51 UTC+00:00
172	Anggara Surya Saputra	dabafdd9	daf71123	Hai	Sent	2017-07-07 04:41:41 UTC+00:00
171	Anggara Surya Saputra	dabafdd9	da452ea4	Aq d area jakal atas.	Sent	2017-07-07 04:25:17 UTC+00:00
170	Anggara Surya Saputra	dabafdd9	da452ea4	Gmana?,maukah?	Sent	2017-07-07 04:25:08 UTC+00:00
169	Anggara Surya Saputra	dabafdd9	da452ea4	PING!	Sent	2017-07-07 04:25:00 UTC+00:00
168	Anggara Surya Saputra	dabafdd9	da452ea4	GFE yuk,mau?	Sent	2017-07-07 03:18:16 UTC+00:00
167	Anggara Surya Saputra	dabafdd9	da452ea4	Km open BO po?	Sent	2017-07-07 02:58:46 UTC+00:00
166	Anggara Surya Saputra	dabafdd9	da452ea4	*Image #10*	Sent	2017-07-07 02:57:46 UTC+00:00
165	Anggara Surya Saputra	dabafdd9	da452ea4	Aq gak hobi ML e,aq lbh suka onani,aq seneng ditemenin ngobrol sambil santai ;)	Sent	2017-07-07 02:55:35 UTC+00:00
164	Anggara Surya Saputra	dabafdd9	da452ea4	Enggak,GFE aj :)	Sent	2017-07-07 02:54:32 UTC+00:00
163	(rose)Amelia artha(rose)	da452ea4	dabafdd9	Partner mi?	Inbox	2017-07-07 02:47:27 UTC+00:00
162	(rose)Amelia artha(rose)	da452ea4	dabafdd9	Oke mas	Inbox	2017-07-07 02:47:18 UTC+00:00
161	Anggara Surya Saputra	dabafdd9	da452ea4	Lg cari partner ni	Sent	2017-07-07 02:43:48 UTC+00:00
160	Anggara Surya Saputra	dabafdd9	da452ea4	Aq dpt pin km dr wechat	Sent	2017-07-07 02:32:01 UTC+00:00

Figure 12. Andriller's Chat Analysis

As for Belkasoft Evidence Center, due to software's trial limitation, this tool is only able to produce a random part of the BBM data as shown on Figure 13.

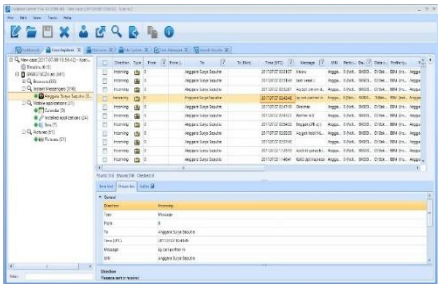


Figure 13. Belkasoft's Analysis Result

4.4 Reporting

The last stage on the mobile forensic investigation method is reporting. At this stage all the analysis's result will be discussed and presented in detail and all artifacts related to the previously obtained from BBM application that showing some indications of a digital crime is documented. The report can be presented based on each tool used on the investigation or by overall report contained all result from all investigation tools.

Oxygen Forensic Suite has the ability to create full reports in various formats, including PDF, the document format that most frequently used. Full Reports from Oxygen Forensic in PDF format is shown in Figure 14

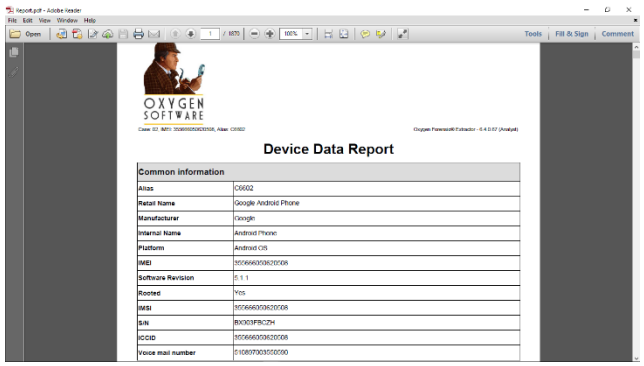


Figure 14. Full Report of Oxygen Forensic Suite in PDF

Oxygen Forensic Suite is also able to create partial reports that refer to one application only, in this research, partial reports are made from the BBM application as shown in Figure 15.

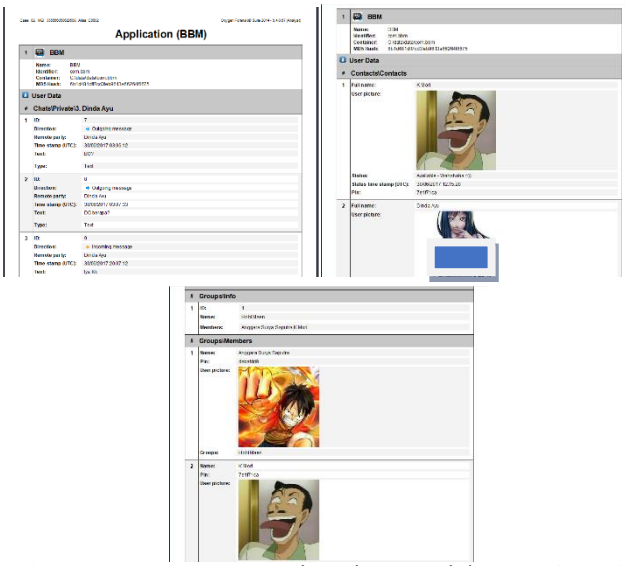


Figure 15. Oxygen Forensic Suite's partial report (BBM)

Andriller's report generation capability is not as good as Oxygen Forensic Suite or Belkasoft Evidence Center. Compared to both forensic tools, Andriller is a lightweight forensic application with simple features. Reports generated from BBM application are just tables containing conversation recordings. Supported formats are also only in HTML and MS-Excel form. Figure 16 shows the report from Andriller.

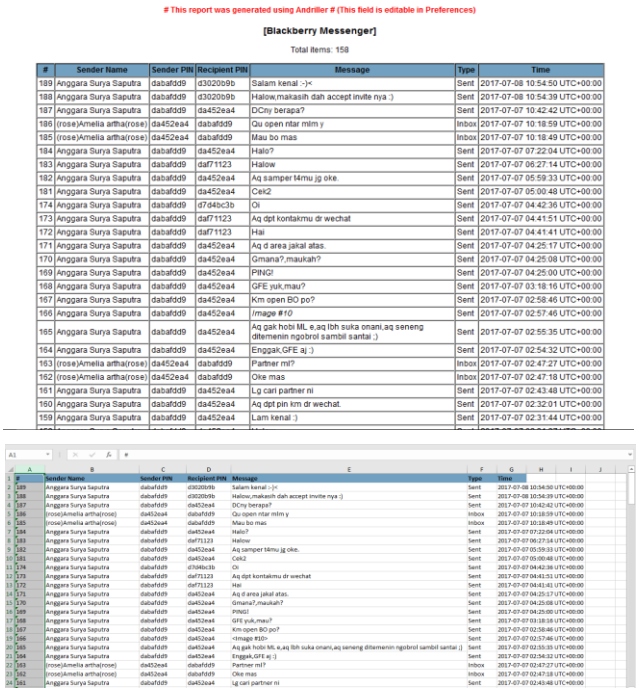


Figure 16. Andriller's partial report (BBM)

As for Belkasoft Evidence Center, the report generation ability is actually quite good, but due to the limitations of the trial version software, the resulting report is only 50%, and taken at random, as shown in Figure 17.

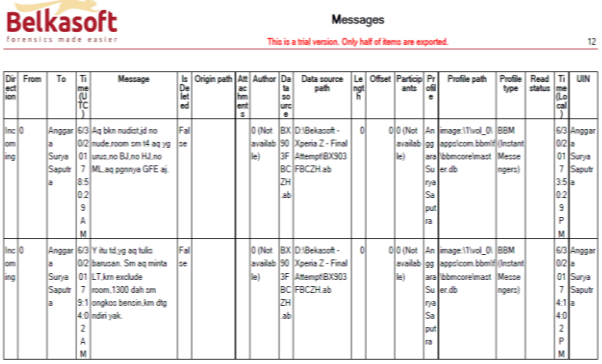


Figure 17. Belkasoft's partial report (BBM)

These generated reports from all forensic tools used are expected to become a supporting evidence that can be used in the investigation and solving process of digital crime cases. The chosen

method and report's presentation format generally based on the policies of local law enforcements, on the other words, local law enforcements might use local presentation format.

5 CONCLUSION

Based on the forensics investigation stages that have been discussed about BBM's digital evidence analysis on Android platform, several things that can be concluded are:
There are various way to get digital datas from BBM that installed on Android-based that depends on various factor also, such as type of vendor, smartphone screen security, transfer protocol used, and also BBM and Android version.
Mobile Forensics is needed because mobile-based services are increasing and getting more users, with the growing popularity of mobile computing and mobile commerce, the need of mobile transactions are also getting higher and the chance for digital crimes occured also increased significantly.
There are various forensic tools that can be used by forensic examiners to acquire digital data from physical evidences, various tools means various capabilities also. Some evaluations on forensic tools can be conducted to get an overview what tool that best for digital forensic investigations.

REFERENCES

1. Statista: Number of smartphone users worldwide from 2014 to 2020 (in billions). Available at https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/. Accessed on August 2, 2017.

2. Jason Mander: Telegram, BBM and WeChat users keenest on transferring money. Available at http://blog.globalwebindex.net/chart-of-the-day/telegram-bbm-and-wechat-users-keenest-on-transferring-money/. Accessed on August 3, 2017

3. Judith Balea: The latest stats in web and mobile in Indonesia (INFOGRAPHIC). Available at https://www.techinasia.com/indonesia-web-mobile-statistics-we-are-social. Accessed on : August 3, 2017.

4. Liputan 6 News: Ditipu Pria Asing Foto Bugil ABG Banyuwangi Menyebar di Internet. Available at http://news.liputan6.com/read/2099669/ditipu-pria-asing-foto-bugil-abg-banyuwangi-menyebar-di-internet?source=search. Accessed on : August 5, 2017.

5. Hernowo Anggie: Blackberry Di-Hack, Ki Kusumo Lapor ke Mabes Polri. Available at http://showbiz.liputan6.com/read/2207179/blackberry-di-hack-ki-kusumo-lapor-ke-mabes-polri?source=search. Accessed on : August 5, 2017.

6. Mei Amelia R: Polda Metro Bekuk Al Gazali Cs di Palopo Atas Kejahatan Hacker BBM dan Penipuan. Available at https://news.detik.com/berita/3120102/polda-metro-bekuk-al-gazali-cs-di-palopo-atas-kejahatan-hacker-bbm-dan-penipuan. Accessed on August 9, 2017

7. Dolly Rosana: Polisi Selidiki Kasus Pembajakan Identitas aplikasi BBM. Available at http://www.antaranews.com/berita/567216/polisi-selidiki-kasus-pembajakan-identitas-aplikasi-bbm. Accessed on August 8, 2017.

8. Safar Sijabat: Aurel Sediakan Jasa Layanan Seks Via Medsos di Kota Pekanbaru. Available at https://www.tribratanews.com/aurel-sediakan-jasa-layanan-seks-via-medsos-di-kota-pekanbaru/, accessed on August 11, 2017.

9. Anggie Khristian, Yessi Novaria Kunang, and Siti Sa'uda: Forensic Analysis of Whatsapp's Artefact On Android Platform. Bina Darma University. Available at http://digilib.binadarma.ac.id/download.php?id=1336, 2016.

10. Ilman Zuhri Yadi and Yessi Novaria Kunang: Forensic Analysis on Android Platform. National Conference of Computer Science (Konferensi Nasional Ilmu Komputer (KONIK)). Available at http://eprints.binadarma.ac.id/2191/1/ilman%20zy-%20analisis%20forensik%20android.2.pdf., 2014

11. Asif Iqbal, Andrew Marrington, and Ibrahim Baggili: Forensic Artifacts of the ChatOn Instant Messaging Application. 8th International Workshop on Systematic Approaches to Digital Forensics Engineering. Available at https://pdfs.semanticscholar.org/1265/425a5b91345656b6b9201b3fe24d0a46d015.pdf., 2013

12. Michael W Burnette: Forensic Examination of a RIM (BlackBerry) Wireless Device. Rogers & Hardin LLP. Available at http://web.archive.org/web/20070718221442/http%3A//www.rh-law.com/ediscovery/Blackberry.pdf., 2002

13. Masoud Nosrati, Mehdi Hariri, and Alireza Shakarbeygi: Computer and Internet: From a Chronological View. Internasional Journal of Economy, Management, and Social Sciences. Available at

http://waprogramming.com/papers/5145ceefb2fa72.01222710.pdf., 2013.

14. Imam Riadi, Jazi Eko Istiyanto, Ahmad Ashari, and Subanar: Log Analysis Techniques using Clustering in Network Forensics. International Journal of Computer Science and Information Security (IJCSIS), Vol. 10, No.7, July 2012, pp. 23-30.

15. Hariani and Imam Riadi: Detection Of Cyberbullying On Social Media Using Data Mining Techniques. International Journal of Computer Science and Information Security (IJCSIS),Vol. 15, No. 3, March 2017, pp. 244-250.

16. National Institute of Justice (NIJ): Digital Evidence and Forensics. Available at https://www.nij.gov/topics/forensics/evidence/digital/Pages/welcome.aspx. Accessed on August 15, 2017.

17. Venkateswara Rao V and ASN Chakravarthy: Survey on Android Forensic Tools and Methodologies. International Journal of Computer Applications (IJCA), Vol. 154, No.8, November 2016, pp. 17-21.

18. Tajudin: The Occurence of Code Switching on Personnal Message of Blackberry Messenger. Journal of English and Education, 2013, pp. 103-112.

19. BBM: About BBM. Available at https://www.bbm.com/en/about.html. Accessed on August 10, 2017.

20. Oxygen Forensic: Smart Forensics for Smartphones. Available at https://www.oxygen-forensic.com/en/. Accessed on August 5, 2017.

21. SeyedHossein Mohtasebi, Ali Dehghantanha, and Hoorang Ghasem Broujerdi: Smartphone Forensic: A Case Study with Nokia E5-00 Mobile Phone. International Journal of Digital Information and Wireless Communications (IJDIWC), 2011, pp. 651-655.

22. Andriller: Andriller – Android Forensic Tools. Available at http://andriller.com/. Accessed on August 7, 2017

23. Belkasoft Evidence Center: Belkasoft Evidence Center 2017. Available at https://belkasoft.com/ec, accessed on August 11, 2017.

24. Lars Woolleschensky: Cell Phone Forensics. Seminararbeit Ruhr-Universit¨at Bochum. Available at https://www.emsec.rub.de/media/crypto/attachments/files/2011/04/cell_phone_forensics.pdf., 2007.

25. Mohammad Junaid, Jai Prakash Tewari, Rajeev Kumar, and Abhishek Vaish: Proposed Methodology

26. on Smart Phone Forensic Tool. Asian Journal of Computer Science and Technology, Vol. 4, No. 2, 2015, pp. 1-5.