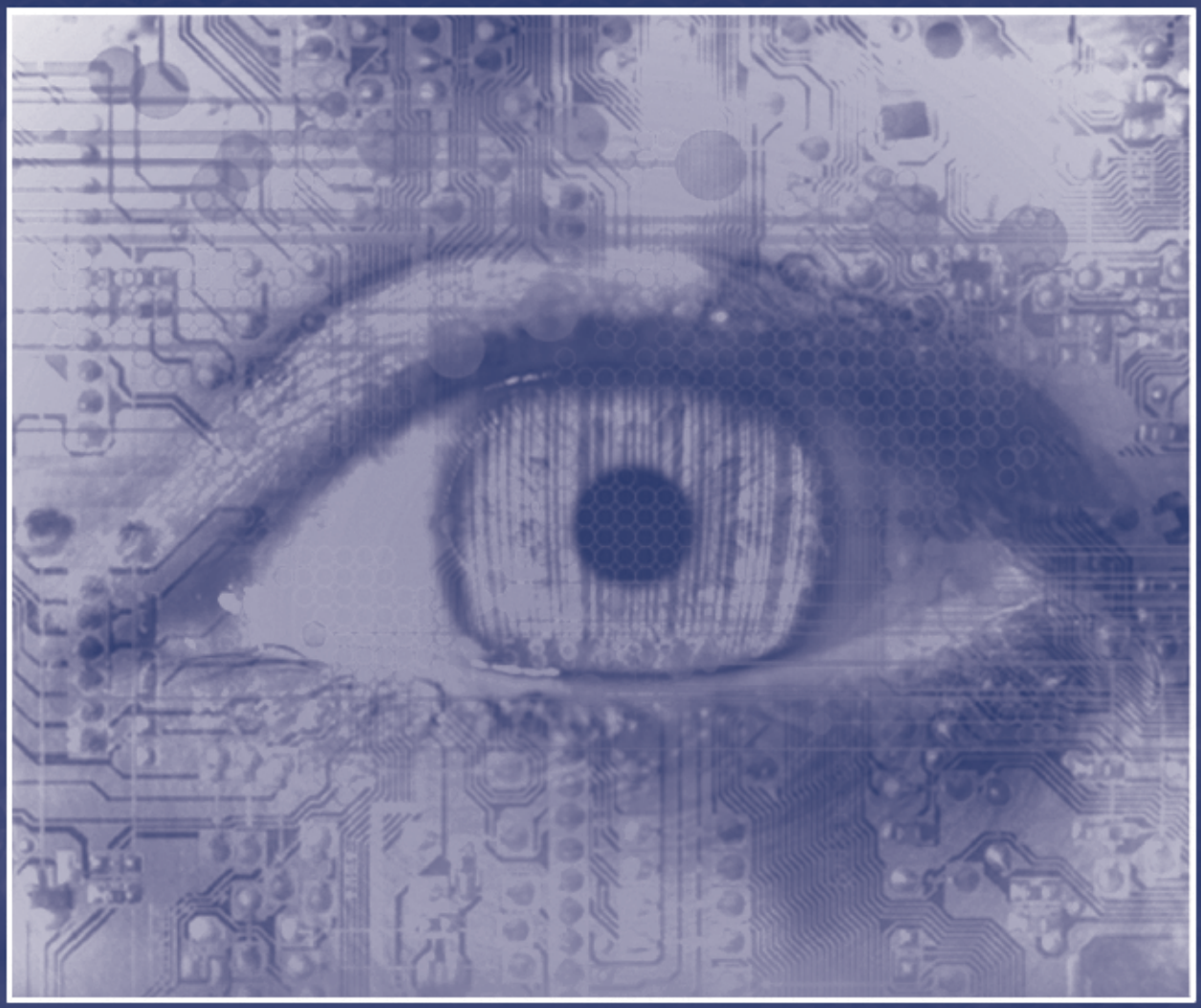
2013

PROCEEDING OF THE EIGHTH INTERNATIONAL CONFERENCE ON FORENSIC COMPUTER SCIENCE

ICoFCS 2013

PROCEEDING OF THE EIGHTH INTERNATIONAL CONFERENCE ON
FORENSIC COMPUTER SCIENCE



www.icofcs.org

Print ISBN 978-85-65069-09-0

By ASSOCIAÇÃO BRASILEIRA DE ESPECIALISTAS EM ALTA TECNOLOGIA - ABEAT

ICoFCS 2013

PROCEEDING OF THE EIGHTH INTERNATIONAL CONFERENCE ON FORENSIC COMPUTER SCIENCE

1st Edition

Brasília, DF
Abeat
2013

Proceeding of the Eighth International Conference on Forensic Computer Science – ICoFCS 2013
ABEAT (ed.) – Brasília, Brazil, 2013, 79 pp. – Print ISBN 978-85-65069-09-0

Address: CLN 309, Bloco D, Sala 103
CEP: 70.755-540 – Brasília/DF - Brazil
Phone: +55 (61) 3202-3006
Email: abeat@abeat.org.br – www.abeat.org.br

Print ISBN 978-85-65069-09-0

COMMITTEE

GENERAL CHAIR

Dr. PAULO QUINTILIANO DA SILVA, Brazilian Federal Police, and University of Brasília, Brazil

SESSION CHAIRS

Dr. ALEXANDRE RICARDO SOARES ROMARIZ, University of Brasilia, Brazil
Dr. ANTÔNIO NUNO DE CASTRO SANTA ROSA, University of Brasilia, Brazil
Dr. FRANCISCO DE OLIVEIRA ASSIS NASCIMENTO, University of Brasilia, Brazil
Dr. JOÃO PAULO CARVALHO LUSTOSA DA COSTA, University of Brasilia, Brazil
Dr. PEDRO DE AZEVEDO BERGER, University of Brasilia, Brazil

REVIEWERS

Dr. Adriano Mauro Cansian, São Paulo State University, Brazil
Dr. Alexandre Ricardo Soares Romariz, University of Brasilia, Brazil
Dr. Anderson Clayton Alves Nascimento, University of Brasilia, Brazil
Dr. Antonio Montes, Renato Archer Research Center, Brazil
Dr. Antonio Nuno de Castro Santa Rosa, University of Brasilia, Brazil
Dr. Ayyaswamy Kathirvel, Anna University, India
Dr. Avinash Pokhriyal, Uttar Pradesh Technical University, Lucknow, India
Dr. Carlos Henrique Quartucci Forster, Air Force Institute of Technology, Brazil
Dr. Célia Ghedini Ralha, University of Brasilia, Brazil
Dr. Clovis Torres Fernandes, Air Force Institute of Technology, Brazil
Dr. Deepak Laxmi Narasimha, University of Malaya, Malaysia
Dr. Dibio Leandro Borges, University of Brasilia, Brazil
Dr. Dinei Florêncio, Microsoft Research, USA
Dr. Francisco de Oliveira Assis Nascimento, University of Brasilia, Brazil
Dr. Ganesan Ramachandrarao, Bharathiar University, India
Dr. Geovany Araujo Borges, University of Brasilia, Brazil
Dr. Hélvio Pereira Peixoto, Brazilian Federal Police, Brazil
Dr. Igor B. Gourevitch, Russian Academy of Science, Russia
Dr. Jaisankar Natarajan, Vit University, India
Dr. Jeimy J. Cano, University of Los Andes, Colombia
Dr. Juliana Fernandes Camapum, University of Brasilia, Brazil
Dr. Luciano Silva, Federal University of Parana, Brazil
Dr. Luiz Pereira Calôba, Federal University of Rio de Janeiro, Brazil
Dr. Marcos Cordeiro d'Ornellas, Federal University of Santa Maria, Brazil
Dr. Mohd Nazri Ismail, University of Kuala Lumpur, Malasya
Dr. Nei Yoshihiro Soma, Air Force Institute of Technology, Brazil
Dr. Nikolay G. Zagoruiko, Novosibirsk State University, Russia
Dr. Nilton Correa da Silva, Evangelic University of Anapolis, Brazil
Dr. Norbert Pohlmann, Fachhochschule Gelsenkirchen, Germany
Dr. Olga Regina Pereira Bellon, Federal University of Parana, Brazil
Dr. Ovidio Salvetti, Italian National Research Council, Italy
Dr. Paulo Licio de Geus, University of Campinas, Brazil
Dr. Paulo Sergio Motta Pires, Federal University of Rio Grande do Norte, Brazil
Dr. Paulo Quintiliano da Silva, Brazilian Federal Police, and University of Brasília, Brazil
Dr. Pedro de Azevedo Berger, University of Brasilia, Brazil
Dr. Pedro Luis Prospero Sanches, University of São Paulo, Brazil
Dr. Renato da Veiga Guadagnin, Catholic University of Brasilia, Brazil
Dr. Ricardo L. de Queiroz, University of Brasilia, Brazil
Dr. Roberto Ventura Santos, University of Brasilia, Brazil
Dr. Sukumar Senthilkumar, University Sains Malaysia, Malaysia
Dr. Vladimir Cobarrubias, University of Chile, Chile
Dr. Volnys Borges Bernal, University of São Paulo, Brazil
Dr. William A. Sauck, Western Michigan University, USA

ABEAT – Associação Brasileira de Especialistas em Alta Tecnologia
SCLN 309, Bloco D, Sala 103
Brassília-DF - CEP: 70.755-540 - Fone: + 55 (61) 3202-3006

CONFERENCE SCOPE

Artificial Intelligence
Artificial Neural Network
Biometrics
Computer Crimes
Computer Forensics
Computer Forensics in Education
Computer Law
Computer Vision
Criminology
Cryptology
Digital Investigation
Image Analysis
Image Processing
Information Security
International Police Cooperation
Intrusion Prevention and Detection
Machine Learning
Management Issues
Network Security
Pattern Recognition
Secure Software Development
Semantic Web
Signal Processing
Simulation
Software Engineering

BEST PAPER AWARD

On this year, the “Best Paper Award” winner is the paper "Greatest Eigenvalue Time Vector Approach for Blind Detection of Malicious Traffic", written by Danilo Fernandes Tenório, João Paulo C. L. da Costa, and Rafael Timóteo de Souza Júnior

The paper "Analyzing Targeted Attacks using Hadoop applied to Forensic Investigation", written by Parth Bhatt and Edgar Toshiro Yano is the runner-up paper of the conference

The choice was made based on the best grades obtained from reviewers of the ICoFCS Editorial Board.

CONTENTS

DOI: 10.5769/C2013001 or http://dx.doi.org/10.5769/C2013001

A Forensic Tool for Signature Authenticity Verification Through Digital Image Processing and Artificial Neural Networks

Deivison Pinheiro Franco[1], Felipe Dantas Barboza[2], and Nágila Magalhães Cardoso[3]

(1) Executive Secretariat of Information Technology, Bank of Amazônia, Belém, Brazil.
Email: deivison.pfranco@gmail.com
(2) Specialization Course in Forensic Sciences, University Center of State of Pará, Belém, Brazil.
Email: barbozafelipe@yahoo.com.br
(3) Specialization Course in Computer Security, Higher Studies Institute of Amazônia, Belém, Brazil.
Email: nagilamagalhaes@gmail.com

Abstract - **This paper aims to propose a computational forensics tool able to verify the authenticity of handwritten signatures in an automated way, to help and optimize this process and act as a tool for decision support. The methodology of this proposal was based on the use of techniques of digital image processing and neural networks through the backpropagation learning algorithm with 500 and 901 approaches. The results showed an average percentage error of 20% in the first and of 5.83% in the second, and the performance of a trained professional to verify the authenticity of signatures has an average error of 6.67%. Thus, we could observe the efficiency of the proposed tool, as well as the difference and evolution of approaches through the relevance of the results.**

Keywords - Authenticity of Signatures, Graphoscopy, Neural Networks, Backpropagation Algorithm, Digital Image Processing.

I. Introduction

According to [1], manuscripts signatures still figures as one of the ways to validate documents authenticity due to its intense individual characteristic coupled with its low cost and practicality, despite of the emergence of various technologies related to this field like, for example, digital certificates and biometrics. Thereat, the fraud by signatures falsification is a crime much practiced in Brazil, reaching generate millionaires losses to people and institutions.

In 2009, KPMG Corporation conducted a research in order to investigate and evaluate the general scenario of organizational frauds in the country and showed that, at the time, 68% of companies interviewed suffered fraud. Of these organizations, 77% had losses of up to R$ 1 million, and 5% of these losses exceeded 10 million. And the type of fraud with higher incidence (29%) was the checks and documents falsification, in which is present the signatures falsification [2]. However, the number of fraud organization cases in Brazil is much greater than the published, because the victims companies are afraid of negative public exposure, which would cause damage to its reputation and image, and even greater financial losses [3].

Graphoscopy is the discipline that certifies a professional to perform the verification of signatures authenticity through concepts and techniques that are the basis for safely conferences with effective results [4]. Thus, the performance of a graphoscopist covers the areas of criminal forensics and litigation, as well as banks, insurance companies, notaries and other financial institutions. And is unquestionable the relevance of its work, once it is directly linked to the security of various institutions where it operates, as well as your users/clients, can perform as decisive evidence in solving crimes and misdemeanors. However, coupled with the intense workload, this professional is subject to many external factors in the exercise of its functions, such as fatigue, stress and personal problems, which could compromise its results.

The physical fatigue might result in misleading observations and mental fatigue favors forgetfulness, unnecessary repetition or omission of any exam. Such failures can bring losses and constraints for both the professional and the organization where he works, and for his customers, or acquit guilty or even incriminate innocent in court [4].

In order to automate the process of analyzing the authenticity of handwritten signatures and assist the professional in graphoscopy with an instrument to support decision making, this paper proposes the creation of a computer forensics tool that can do this verification automated with using techniques of digital image processing and artificial neural networks through backpropagation learning algorithm, which is capable of extracting "signature model" standards for comparison with one or more test signatures and definition of its degree of authenticity. Were also used graphoscopy concepts for signatures classification, analysis and interpretation of the results.

II. Methodology Used to Develop the Tool

Artificial Neural Networks (ANNs) are a branch of Artificial Intelligence (AI) that aims at processing information in a similar way to the human brain [5]. Whereas

DOI: 10.5769/C2012001 or http://dx.doi.org/10.5769/C2012001

backpropagation, according to [6], is a supervised algorithm by error correction for training multilayer artificial neural networks that minimize the error by running the decreasing gradient in the surface errors space weights, where the height for any point in the space corresponds to the measured weights of the error. Thus, the weights begin to be set in units of output, where the error measure is known and continues with the retro propagation of this error between the layers by adjusting the weights to reach the input layer units.

As in the output units the desired and obtained values are known, the adjustment of the synaptic weights is relatively simple. However, for units of hidden layers, the process is not so simple. The weights for a particular neuron in the hidden units, should be adjusted proportionally to the processing unit error which it is connected. Thus, and in accordance with [6], two phases are distinguished in the learning process of the backpropagation: the propagation phase (forward), in which the entries are spread between the layers of the network (from input to output), and the retro propagation phase (backward), whose errors are propagated in the opposite direction to the input stream.

Given the above, and to meet the proposed use of the backpropagation algorithm, the construction of the tool followed sequentially the following steps: Acquisition of Signatures, which were scanned on a common scanner device, Pre-Processing and Digital Processing of Images, whose results are the features extracted for analysis, Creation, Training and Testing of the Artificial Neural Networks, from where the results arose.

Due to the need to review the steps of digital image processing and architecture definition of the artificial neural networks, the study was divided into two approaches, which are described below, whereas initial approach did not meet the expectations of improvement in success rates.

A. Bank of Images

Applying the concepts of digital image processing proposed by [7], the process was developed from the signature collection of three distinct authors: Eric, Felipe and Rodrigo. Each author signed twenty times its own signature, falsified twenty times the signature of the second author, and also falsified twenty times the signature of the third author. This resulted in a total of sixty signatures each, and an overall total of one hundred and eighty samples as shown in Table 1. Therefore, was used a single bank of images, once its acquisition process was the same for both approaches.

The samples digitalization was performed from a common scanner device. Then, each image was resized generating images within the maximum range of 731 pixels wide by 180 high and a minimum of 650 pixels wide by 117 high, all in the ".png" format.

TABLE I. Signatures Collected by Author

Signatures	With the name of author 1 (Eric)	With the name of author 2 (Felipe)	With the name of author 3 (Rodrigo)
Written by author 1 (Eric)	20 (authentic)	20 (fake)	20 (fake)
Written by author 2 (Eric)	20 (fake)	20 (authentic)	20 (fake)
Written by author 3 (Eric)	20 (fake)	20 (fake)	20 (authentic)

B. Approaches

It was used two approaches: the 500 Approach and the 901 Approach - both with features of image processing and different network configurations, where the second came as an evolution of the first.

The two approaches used to prepare the tool have scripts and functions to automate the process and, moreover, are based on the idea of using, from the pixel array, one-dimensional characteristics to the input layer in the recognition algorithms signatures through vertical projection (sum of pixels in each column) and of horizontal projection (sum of pixels in each row of the matrix).

In both approaches it was used the Matlab, version R2008a, to perform the procedures for pre-processing, segmentation and feature extraction of images, and to create, training and testing the ANNs.

The difference between the approaches characteristics becomes more evident from this topic, which differs in the stages of pre-processing and feature extraction, network architectures, and examples of signatures submitted to ANNs in the same training set.

Considering that the architectures are very different and that each one defines different amounts of nodes in the ANNs sensory layers, the initial approach was called 500 Approach, by instituting five hundred entries and the subsequent was called 901 Approach by defining nine hundred and one entries. Such approaches are better described below with their respective processes involved to build the tool.

1) 500 Approach

- Images pre-processing routines:
  - *a) Capture the recognized sample as three-dimensional matrix representing the RGB colors scale;*
  - *b) Transform in grayscale, which matrix format becomes two-dimensional, facilitating manipulation, because it involves fewer variables in the required calculations, both in processing and in training;*
  - *c) Contrast adjust, where the image pixels are highlighted and is highlighted the intensity difference between the darker and the lighter shades;*
  - *d) Histogram equalization adjust, producing an increase in brightness and also in contrast;*
  - *e) Resizing, which reduces the image to the default size of 400 pixels wide by 100 pixels high;*

DOI: 10.5769/C2012001 or http://dx.doi.org/10.5769/C2012001

f) Binarization through logic operation, with the goal of making the background and lighter shades removal, ie, segmentation, such that the signature region becomes black and the remaining regions becames white.

Given the above, Figure 1 below shows images of the same signature in relation to the histogram equalization.

Figure 1. Histogram equalization: (a) Signature captured without equalization, (b) Signature captured with equalization.

• Features extraction:

This step consists in the generation of a concatenated vector of 500 positions for each signature, where the 400 first are related to the sum of the columns (vertical projection) and the others 100 corresponds to the sum of the lines (horizontal projection) of the image, as shown in Figure 2. This vector represents the features extracted from the image of the signature that matches, and acts as a sample entry in the training set or test of the ANN in the 500 Approache.

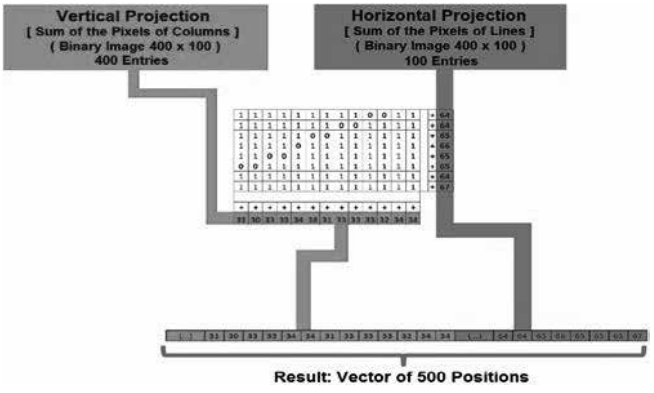


Figure 2. Features extraction of signatures in 500 Approach.

• Artificial Neural Network:

Overall, there were several trainings, however, was highlighted in the 500 Approache only one ANN, called "Eric45", because it was the network that showed better results in this approach. The name assigned to this network is composed by the first analyzed author's name followed by the number of examples presented to this ANN in its training set. Thus, the architecture of the "Eric45" network was composed of:

a) 1 direct network, multilayer, fully connected;
g) 500 entries, corresponding to the vector of 500 positions;
h) 2 intermediate layers with 200 neurons in each;
i) 2 neurons in the output layer, where one is activated in case of authenticity, and the second in case of falsity.

A better view of the architecture of "Eric45" network in 500 Approach can be seen in the Figure 3 below.

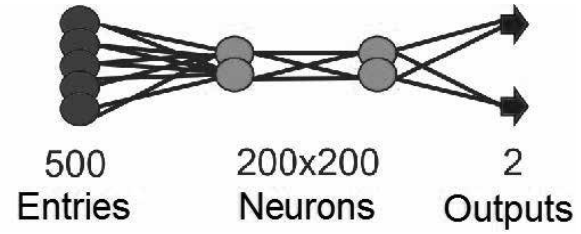


Figure 3. Architecture form of the "Eric45" network in 500 Approach.

Therefore, is possible to observe and assign the following configurations and features of the "Eric45" network:

b) It has up to 1000 times to do the learning;
c) The mean square error to be achieved is 10-3;
d) Uses the backpropagation algorithm cause has great capacity for generalization and enable supervised learning;
e) The training is done with supervised learning, once the classes that the network must be distinguished are known;
f) The learning rate is adaptive;
g) Uses the momentum;
h) Uses the logistic activation function, where the response values are provided in the open interval between 0 and 1.

• Training set:

To the 500 Approach, the training set was created with 45 signatures in order to recognize the author's signature Eric. Each of the three authors were used 15 signatures with the name Eric, being the first 15 authentic, activating the 1st neuron of the output layer, and the other 30 falsifications by slavish imitation (spelled by Felipe and Rodrigo), activating the 2nd neuron in the output layer.

2) 901 Approach

The 901 Approach arose from the need to correct the errors of the 500 Approach, besides trying to improve performance. Many parameters were changed as observed in the description of the following steps.

• Images pre-processing routines:

i) Capture image in RGB format;
j) Transform in grayscale;
k) The contrast adjustment was modified to intensity adjustment. The function used was the same, but the intensity was changed manually, instead of automatically method of the previous approach (500);
l) The adjust in the histogram equalization was not used, cause this practice emphasized the presence of noise in the image, which could interfere in the network learning;
m) Due to the detection of some noise points in the samples submitted to the processing algorithm, it was did a scan in order to turn in white all the pixels in the edges of the image;
n) After the manual removal of unwanted pixels, it has become possible to cut the sample by reducing the image

DOI: 10.5769/C2012001 or http://dx.doi.org/10.5769/C2012001

area of the rectangle that delimits the exact size of the signature;

o) It was created a copy of the image, which has been reduced to a size of 40 pixels wide and 10 pixels high, and then binarized, so that each pixel could be used as input to the ANN;

p) With the original image, it was calculated the ratio between width and height (width/height), so that the result was used as an input of the network;

q) And finally, it was did the procedures of binarization and resizing of the sample to the size of 400 pixels wide by 100 pixels high, withdrawing the sum of each row and columns.

The Figure 4 below shows the main steps of the images pre-processing.

Figure 4. Procedures of the images preprocessing.

• Features extraction:

Unlike previous approach, the resulting vector in the 901 Approach, which is used in the subsequent training and testing process, consists of 901 positions, of which:

r) The 400 first positions corresponding to each pixel of binary value (0 or 1) of the resized image in proportion 40x10, thus representing the positioning feature of the pixels in that image;

s) The next 500 positions of the vector, ie, the 401th to 900th position, corresponding to the sums of rows and columns pixels of the resized image in proportion 400x100, equally to 500 Approach;

t) And the last position corresponding to the result in pixels of the ratio between width and height of the image in 400x100 ratio, calculated in the pre-processing.

The Figure 5 below illustrates the features extraction of the signatures in the 901 Approach.

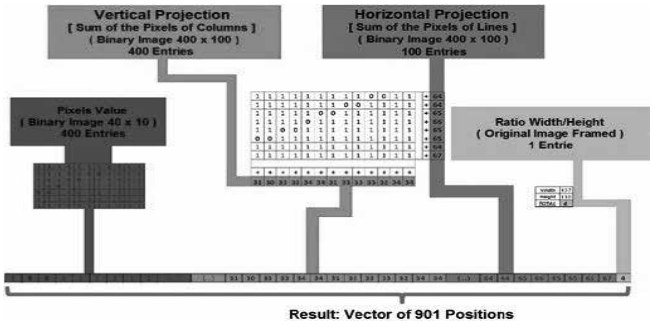


Figure 5. Features extraction of signatures in 901 Approach.

• Artificial Neural Network:

In this approach were highlighted four networks, called "Eric35", "Felipe40", "Rodrigo40" and "Eric40", which have the same characteristics of architecture and configuration, but distinctions among themselves as to its training sets. Each name assigned to the network also consists of the first author's name followed by the number of analyzed samples presented to the ANN in its training set. Thus, the architecture of the "Eric45", "Felipe40", "Rodrigo40" and "Eric40" networks were composed of:

u) 2 intermediate layers with 500 neurons each;

a) 1 neuron in the output layer, which is activated in case of authenticity of the submitted sample;

b) 901 entries, corresponding to the vector of 901 positions, of which: From 1 to 400, the minimum value is 0 and the maximum is 1, since each entry corresponds to a pixel matrix of binary image pixels 40x10; From 401 to 800, the minimum value is 0 and the maximum is 100, because it correspond to the vertical projection of the signature in 400x100 format; from 801 to 900, a minimum of 0 and a maximum of 400, considering the horizontal projection matrix 400x100; At position 901, the minimum value is 0 and the maximum is 10, according to the original size of the acquired images.

A better view of the architecture of "Eric45", "Felipe40", "Rodrigo40" and "Eric40" networks in 901 Approach can be seen in the Figure 6 below.

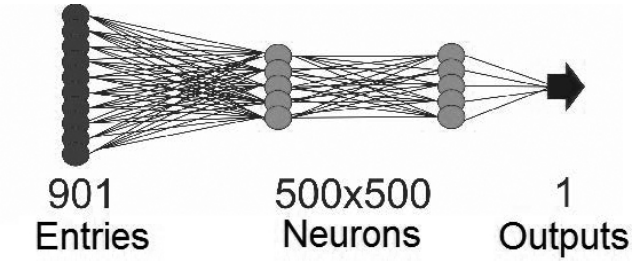


Figure 6. Architecture form of "Eric45", "Felipe40", "Rodrigo40" and "Eric40" networks in 901 Approach.

Therefore, is possible to observe and assign the following configurations and features of "Eric45", "Felipe40", "Rodrigo40" and "Eric40" networks:

v) Comprised 15.000 times for training;

w) Used the backpropagation algorithm with adaptive learning rate and momentum;

x) The outputs in the range between 0 and 1, used logistic activation function;

y) Minimum squared error reached of 10-3.

• Training set:

To the 901 Approach, the training set used had distinctions as to the amount of samples used, as shown in Table 2 below:

DOI: 10.5769/C2012001 or http://dx.doi.org/10.5769/C2012001

TABLE II. Training Sets in the 901 Approach

Network	Analyzed author	Originals samples	Fake samples	Samples with the names of other authors	Total of examples
Eric35	Eric	5	10	20	35
Felipe40	Felipe	10	10	20	40
Rodrigo40	Rodrigo	10	10	20	40
Eric40	Eric	10	10	20	40

Given the above, it was observed that in the 901 Approach the training set of each network is formed by the original samples plus the fakes and plus the names of the others authors. And the first was formed by ten original signatures of the author examined, except "Eric35" network - composed of five signatures.

While the false sample sets are composed of ten falsification by signature slavish imitation of the analyzed author, and being five from each one of the two other authors, the sample sets with the names of other authors consist of: Five authentic signatures of a second author; Five authentic signatures of a third author; Five falsification by slavish imitation of the signature of the second author; and Five falsifications, also by slavish imitation signature, of the third author. So, all behave like falsification without imitation of the analyzed author.

III. Experiments and Results

This section will present and discuss the experiments and results of the tests on the networks that make up the proposed tool, displaying the error rates and the hit rates for each. Also described are the percentages of the two types of error found, such as: acceptance of false signatures (false positive) and rejection of original signatures (false negative).

A. In the 500 Approach

The results of the "Eric45" network were acquired from the creation of a set test with 25 signatures (different from those used in the training set), defined as follows:

z) 5 original signatures of the author Eric;

- *aa) 5 falsification by slavish imitation spelled by the author Felipe;*
- *ab) 5 falsification by slavish imitation produced by the author Rodrigo;*
- *ac) 5 original signatures of the author Felipe, functioning as falsifications with no imitation;*
- *ad) 5 original signatures of the author Rodrigo, also functioning as falsifications with no imitation.*

The learning of the "Eric45" network resulted in 80% hit rate in the tests, i.e., 20 signatures of the sample space presented. The graphs in Figure 7, below, summarizes the information about the results obtained from this network. The first shows the hits and errors for the network in question while the second shows the separation of learning errors types observed.

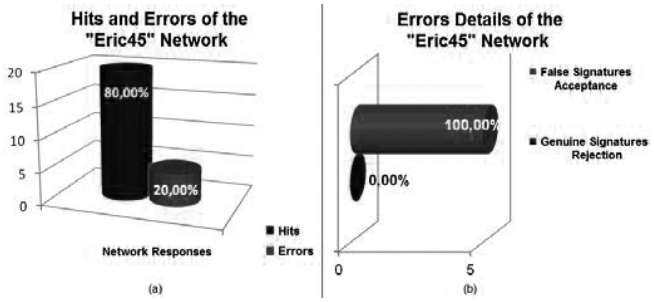


Figure 7. "Eric45" network results: (a) Hits and erros; (b) Erros types.

B. In the 901 Approach

The Table 3, below, summarizes the results of the tests performeds on all networks trained in this approach ("Eric45", "Felipe40", "Rodrigo40" and "Eric40"), with its successes and failures, as well as the values for the acceptance of inauthentic samples and rejection of original signatures.

TABLE III. Results of the Tests Applied to the 901 Approach Networks

Samples in the test sets	Network	Hits	Hits %	Errors	Errors %	Original Signatures Rejection	False Signatures Acceptance
60	Eric35	55	91,67	5	8,33	5	0
60	Felipe40	52	86,67	8	13,33	2	6
60	Rodrigo40	59	98,33	1	1,67	1	0
60	Eric40	60	100	0	0	0	0

C. Test With a Graphoscopist

Tests were conducted with a graphoscopist of the Bank of Brazil, with 15 years of experience as a lecturer of signatures, in order to compare his performance with the network "Eric35" because this is the only network implemented at the time of this professional availability. Therefore, in this procedure was used only samples of the training set (query patterns) and testing (questioned signatures) of this ANN. The results of the tests with the professional are detailed below in Table 4.

TABLE IV. Results of the Tests Applied to the Graphoscopist

Analyzed samples	Hits	Hits %	Errors	Errors %	Original Signatures Rejection	False Signatures Acceptance
60	56	93,33	4	6,67	3	1

D. Results Comparison

Based on the tests results of all networks of the 500 and 901 Approaches, as well as the graphoscopist, it was traced the hit rates comparison, as shown in Figure 8 below.

DOI: 10.5769/C2012001 or http://dx.doi.org/10.5769/C2012001

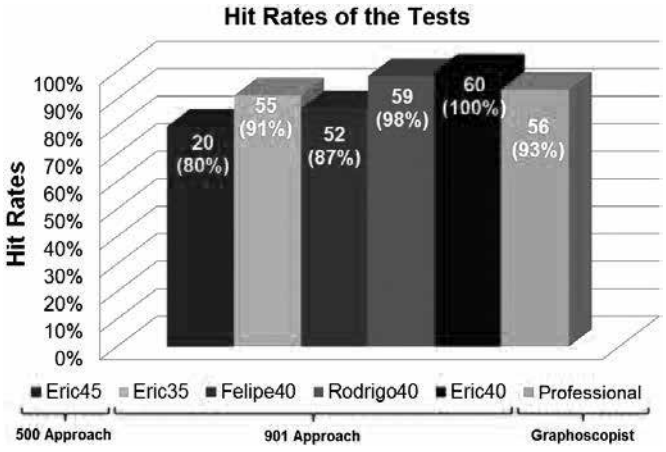


Figure 7. "Eric45" network results: (a) Hits and erros; (b) Erros types.

It was also observed the evolution of the errors types in the tests with the networks and with the graphoscopist, as shown in the graph of the Figure 9, which shows the percentage of false responses acceptance and rejection of true ones.

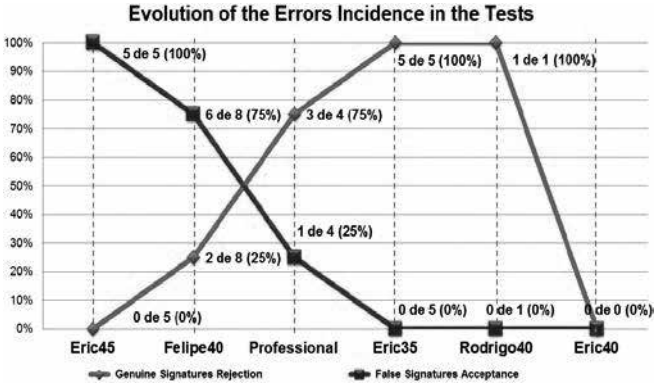


Figure 9. Evolution of the errors types percentage presented by the networks and by the graphoscopist.

Given the results above, it was found that the 500 Approach, despite the considerable success rate, 100% of the network errors have been cases of falsifications acceptance, which consist in serious failure in the signatures authenticity verification, and it causes higher damages compared to the original signatures rejection. These factors lead to the conclusion that the techniques used at that time required modifications in relation to the size, quantity and variety of the examples, as well as changes in the quantities of neurons in the three network layers for better solutions. Thus, the efforts were directed to the reformulation of the resolution method, justifying the existence of the 901 Approach.

In the 901 Approach, the training times have grown exponentially, like the number of times required. The outputs became less diffuse, because the analysis was performed based only on the single neuron in the output layer.

When it draws a parallel between the "Eric45" and "Eric35" networks whose training was to recognize the Eric author's signature, it was noticed considerable improvement, because the initial RNA had an error rate of 20%, among which, 100% were characterized by the false signatures acceptance. And the "Eric35" network presented 8.67% error, which its set was composed only of original signatures rejection.

Comparing "Eric35" network with the graphoscopist, it is clear that the RNA results are under of the professional results relative to the quantity of hits. However, the implementation quality was better, in consequence of the only error committed by the professional included in the set acceptance of false signatures.

The "Felipe40" network, trained to the author's signature Felipe, had the highest amount of errors in the 901 Approach, beyond accept fake samples.

The tests on the "Rodrigo40" network returned only 1,67% hits in the 901 Approach, with much original signatures rejections, allowing the conclusion that there is still difficulty in recognition of signatures standards, so that there are differences inside of the authentic sample set that can be clearly perceived. An example of this is the result of the graphoscopist analysis, which also missed credible examples.

The "Eric40" network, whose results were the best, had 100% hits on the test set. However, due to the find of distinctions between the authentic signatures, there is no guarantee that the network will behave the same way in the case of tests for other examples, even if it is in the standards adopted for the image bank formation.

The fact that the "Felipe40" network have lower hit rate between the three networks of the same training set architecture and configuration ("Felipe40", "Eric40" and "Rodrigo40"), can be attributed to the medium graphical culture of Felipe signatures standards.

The "Rodrigo40" and "Eric40" networks, whose signature standards have high graphical culture, had higher hit rates. However, the "Eric40" network presented higher rate because Eric signatures standards have more facility areas that Rodrigo. Thus, it was assumed that the tests hit rate in the networks is directly proportional to the graphic culture level of the standards signatures by author analyzed.

IV. Conclusions

In the tool development there was little variety of different authors signatures standards, as well as considerable influence on the samples quality, once the device used to scan and the writing strategy of the signatures adapted the environment to facilitate collection and analysis.

The project did not consider external factors such as color, material (pen or pencil), paper form and psychological changes of humor or disposition of the authors. Furthermore, the extracted features combination may still be lower than necessary to enable great generalization without losing the recognition reliability of each signature standard. So that the conclusions inferred on the results may not present the same behavior in all environments and existing signature standards, even if the provided responses are considered perfectly applicable.

DOI: 10.5769/C2012001 or http://dx.doi.org/10.5769/C2012001

Finally, the tool may be considered perfectly valid and viable, once that achieved significant results in the signatures authenticity verification. Fitting to point that there is no consolidated model, but a large increase in technical studies related to the subject. Therefore, this proposal aims to collaborate with the maturing of the theme that revolves around the use of Artificial Intelligence as decision support in situations that present a high degree of complexity, such as the case of signatures verification authenticity.

V. Future Work

For future work, it is proposed: collecting examples of other authors, for both the training set and the test set; add to the networks new extracted features from the images; cross-validation of the image bank samples to extract the best training sets and testing sets to optimize results; consider external factors such as color, material (pen or pencil), paper form and psychological changes of humor or disposition of the authors; and the implementation of a commercial application for the tool.

References

[1] Queiroz, F. e Sousa, A., "Exames Periciais a Documentos Manuscritos", Available from www.queirozportela.com/psicologiadaescrita/pericias.pdf; [Visited November, 2012]; [In portuguese];

[2] KPMG., "Relatório de Pesquisa Sobre Fraudes no Brasil", Available from www.kpmg.com.br/publicacoes/forensic/Fraudes_2009_port.pdf; [Visited May, 2013]; [In portuguese];

[3] Theodoro, A., "Fraude Manuscrita: Um Risco Iminente", Available from www.peritocontador.com.br/artigos/Adriano_Theodoro/fraude.htm; [Visited May, 2013]; [In portuguese];

[4] Gomide, F., "Manual de Grafoscopia", Livraria e Editora Universitária de Direito, 2ª edição 2008; [In portuguese];

[5] Carvalho, A., "Redes Neurais Artificiais – Teoria e Aplicações", LTC, 2ª edição, 2007; [In portuguese];

[6] Rumelhart, D., and Chauvin, Y., "Backpropagation: Theory, Architectures, and Applications", Wiley, 1st edition, 2000;

[7] Gonzalez, R. e Woods, R., "Processamento Digital de Imagens", Pearson, 3ª edição, 2010; [In portuguese];

[8] Cinelli, S., "Grafoscopia - A escrita como Função Eminentemente Cerebral-Central", Available from www.queirozportela.com/psicologiadaescrita/pericias.pdf; [Visited May, 2013]; [In portuguese];

[9] Clark, P. and Niblett, T., "The CN2 Induction Algorithm". Machine Learning 3 261-285, 2009;

[10] Cohen, W., "Fast Effective Rule Induction". In Proceedings of the 12th International Conference on Machine Learning 115- 123. Morgan Kaufmann, Palo Alto, CA, 2005;

[11] DBpedia, "The DBpedia Knowledge Base", Available from http://dbpedia.org; [Visited April, 2013];

[12] Drossu, R. and Obradovic, Z., "Rapid Design of Neural Networks for Time Series Prediction". Washington State University, 2006;

[13] Faraway, J. and Chatfield, C., "Time Series Forecasting with Neural Networks: A Comparative Study Using Airline Data", In: Royal Statistical Society, 2006;

[14] Figueiredo, F., Ricco, J., e Brandão, J., "Norma de Procedimentos de Grafoscopia", Available from www.ibape-sp.org.br/arquivos/norma_de_grafoscopia_logo_novo.pdf; [Visited April, 2013]; [In portuguese];

[15] Frank, R. J.; Davey, N. and Hunt, S.P., "Time Series Prediction and Neural Networks". University of Hertfordshire, 2001;

[16] Hand, D. J., "Discrimination and Classification". Wiley, Chichester, 2001;

[17] Haykin, S., "Redes Neurais – Princípios e Práticas", Prentice Hall, 2ª edição, 2001; [In portuguese];

[18] J. Cannady, "Artificial Neural Networks for Misuse Detection", Nova Southeastern University, 2003;

[19] Jantzen J., "Introduction to Perceptron Networks", Technical University of Denmark, 2008;

[20] K. Fanning, K.O. Cogger, R. Srivastava, "Detection of Management Fraud: A Neural Network Approach", IEEE, 2005;

[21] Patterson D W, Chan K H, Tan C M., "Time Series Forecasting with Neural Nets: A comparative Study". Proc. the international conference on neural network applications to signal processing. NNASP, Singapore pp 269-274., 2003.

[22] Pedrini, H., "Análise de Imagens Digitais", Thomson, 3ª edição, 2007; [In portuguese];

[23] Plummer, E. A., "Time Series Forecasting with Feed-Forward Neural Networks: Quidelines and Limitations". University of Wyoming, 2000.

[24] Quinlan, J. R., "Learning logical Definitions from Relations". Machine Learning 5 239-266, 2000;

[25] R.J. Bolton and D.J. Hand, "Statistical fraud detection: A Review", Statistical Science, vol. 17, No. 3, 2002;

[26] S. Haykin, Neural Networks: "A Comprehensive Foundation". Prentice Hall, Englewood Cli_s, NJ, 2nd edition, 2000;

[27] Santos, C., "Análise de Assinaturas Manuscritas Baseada em Princípios de Grafoscopia". Dissertação (Mestrado em Informática Aplicada) - Pontifícia Universidade Católica do Paraná, Curitiba, 2009; [In portuguese];

[28] T. Koskela et al., "Time Series Prediction with Multilayer Perceptron", Wiley, Chichester, 2001.

DOI: 10.5769/C2013002 OR HTTP://DX.DOI.ORG/10.5769/C2013002

A Remote Biometric Authentication Protocol For On-Line Banking

Anongporn Salaiwarakul

Abstract— ***This paper presents a remote biometric authentication protocol illustrated by on-line banking. The protocol assures three properties which are crucial if the biometric data is involved in authentication process. Even if the biometric data is excellent in authenticating the users as it verifies the users by mean of their personal attributes, the biometric data is sensitive in security prospective because it is hard to keep secret. The biometric authentication works well in supervised situation if the verifier can prove that the biometric data comes from the live presentation of the user at the time of user's verification. Prone to security risk in the unsupervised situation, especially on-line transaction, where a captured biometric data can be presented to the system, a biometric authentication in remote situation that guarantees the security level should be proposed. Even more, the security properties of the protocol should be verified and analysed to promise that the protocol does not manipulate the data with an intruder. The result of the analysis shows that the protocol preserves the three properties: privacy of the biometric data, liveness, and intentional authentication.***

Keywords- ***Biometric Authentication Protocol; Security Propert;, Privacy***

I. Introduction

Biometric data belongs to a particular person. It truly reflects the user's personal characteristics so the data can be used to prove the identity of the user. A password or other methods proof of identity such as smart card or token can be given away or stolen. Therefore, the actual identity of the user could not be confirmed. Even the biometric data has its advantage in term of the authentic user's identity; its security should be considered because of its nature. Biometrics is hard to keep secret and human has a limited number of them. Once they are compromised, they can not be changed, replaced or regenerated as a password or a smart card can.

Biometric security depends on the authenticity of the biometric data. Since the biometric data is in the public domain by its nature, artificial biometric data, e.g. rubber finger that replicates the real user's fingerprint, can be generated so it can often fool a biometric reader. Liveness detection in biometric reader is largely research. Biometric authentication works well in supervised situations but for high assurance situations, the reader should be attended to or at least observed until we get verifiably strong liveness detection. However, this concern mainly relies on research in hardware. Biometric authentication is much harder in the remote or unattended cases.

Authentication in biometric protocol can be compromised in a number of ways: via an attack on the server storing the biometric code, an interception of the biometric data when read by the biometric reader, or an attack during biometric data transmission. This paper presents a secure protocol to enhance security level for a remote biometric authentication. On-line banking is employed as an example for illustration the proposed protocol. The protocol guarantees the live presentation of the user on the time of verification, liveness property. The paper attempts to demonstrate a remote biometric authentication that assures the liveness property; therefore we assume that the related risks of the on-line transaction e.g. key logger or viruses is out of the scope. The protocol reserves the privacy of the biometric data. The intended purpose of the authentication from the user is verified to prevent the forged intentional purpose from an intruder.

II. the protocol

Basically an authentication process is used as a prior step before the user's actual activity can be performed. Therefore, this paper considers that a remote biometric authentication protocol has two major activities: authenticating the user remotely using biometric data, and performing user's activities (requests) e.g. logging into a computer. The former process comprises of two components: user and authentication server whereas the latter process involves the user and the service system.

The proposed remote biometric authentication protocol comprises of three components: the User, the Bank (the service system) and the Biometric Authentication Server. The user provides his biometric data for the user's authentication. He must be ensured that the biometric reader is sincere. In this protocol, *the User* operates his transactions via the local workstation which has Trusted Platform Module (TPM) [1,5] installed. The aim of TPM in this protocol is to verify the correctness and the trustworthiness of the machines the user operates. Upon booting up, the user's workstation and the biometric reader are verified by the TPM. The report of the integrity of the local system is sent to the user. The user decides to continues his activities if he satisfies with the integrity value. This can protect the user from presenting his biometric data to the counterfeit biometric reader which has the potential to steal the user's biometric data or manipulate with an attacker. The *Bank*, the second component, is responsible for the bank transaction which is transferring money from user's account to the other account as requested by the user. *The Biometric Authentication Server* performs the biometric matching process. The server has the biometric template storage which is used when the stored biometric

code is acquired for matching against presented biometric data. The result of the biometric verification is reported to the user. Once the user wishes to perform on-line banking transaction, the biometric verification result along with the requested transaction must be presented to the bank.

The protocol involves two major activities: authenticating the user, and banking transaction (transferring money), each of which has different consideration. A user is required to verify his identity in order to access his account. As the authentication process requests the user's biometric data, the *liveness of the biometric data* must be assured so that the bank is certain that the request comes from the authentic user. For the banking transaction, the message must be verified and guaranteed that the account owner is willing to provide his biometric data for the transaction not other transaction that he is not willing to do so. Hence, the intentional authentication property is verified. Therefore, not only the validity of the biometric verification result is checked but the purpose of the biometric authentication is verified. This assures that the user presents his biometric data for the transferring process not something else.

In this protocol, nonces are used for checking the freshness of messages received and encryption and decryption are also used for the secrecy of message content. Signed messages are used to confirm the origin of the senders. Table 1 shows the communication messages among the three components.

The communication messages commence when the user requests transfer transaction from the bank. As a response, the bank looks for the user's authentication, in this case – the biometric authentication is applied. The bank sends a signed message which includes the user name, the biometric authentication request and the nonce n1. The user forwards this message to the biometric authentication server in order to acquire the user's verification. It triggers the BAS to inquire the user's biometric data; the nonce n2 is included in the replied message. This requested message is signed with the BAS's signature. The user presents his biometric data to the reader. The biometric data is signed by the TPM so that the authenticity of the biometric data can be verified. The encrypted message is composed of user name, the nonce n2 and the signed biometric data with the TPM's signature.

To enhance the security requirement in term of liveness of the biometric data, once the biometric data is submitted to the BAS, the BAS verifies the live presentation of the user by acquiring the user to present verification data. The verification data is a secret data which is known only to him and the BAS. The BAS sends an encrypted message of the request and newly generated nonce n3. As a response, the user presents his biometric data, the verification data and the nonce n3. The message is signed with the TPM's private key. The signed message, together with information the user provided, are enciphered by the BAS's public key. The BAS verifies the message by checking the nonce and the signature. It then validates the authenticity of biometric data. The verification result, the user name and the nonce n1 are signed by the BAS.

Upon receiving the verification result message, the user appends his transfer transaction and sends this message to the bank. The transfer message includes the amount and account he wishes to transfer. The message is encrypted by the public key of the bank. The bank deciphers and verifies the validity of the message. It then checks the matching result. If the result is positive, the bank performs the user's request, transferring the money to his desired account.

TABLE I. the communication messages in the protocol

Communication	Message
User → Bank	uName, amount, acct
Bank → User	$sign_{skBank}$(UName, amount, acct, BioAuthReq, n1)
User → BAS	$sign_{skBank}$(UName, amount, acct, BioAuthReq, n1)
BAS → User	$sign_{skBAS}$(reqBD,n2)
User → BAS	$penc_{pkBAS}$(n2,uName,$sign_{skTPM}$(BD))
BAS → User	$penc_{pkTPM}$(reqVD,n3,$sign_{skBAS}$(reqVD,n3))
User → BAS	$penc_{pkBAS}$(VD,BD,n3),$sign_{skTPM}$(VD,BD,n3))
BAS → User	$sign_{skBAS}$(uName,amount,acct,matchResult,n1)
User → Bank	$penc_{pkTPM}$(uName,amount, $sign_{skBAS}$(uName,amount,acct,matchResult,n1))
Bank → User	Transfered Result

III. Proverif model

This paper verifies the proposed protocol using the Dolev-Yao style adversary [2] in analysing the security goals. This style of adversary is able to read messages over the network and collect in its knowledge set. The attacker can also calculate the attack from its knowledge set. This protocol is verified by Proverif [3] based on applied pi calculus [4]. It can handle unbounded number of sessions of the protocol, generate attacks based on Dolev-Yao style attackers and manipulate messages communicating among components to find any possible attacks to the protocol corresponding to the intended security properties.

The proposed protocol promises the three intended properties. To ensure that the protocol provides the appropriate level of security and the properties it affirmed, ProVerif is used as the verification tool to verify and analyse the protocol. The ProVerif models consist of three major processes: BAS process, Workstation process and the Bank Process. The BAS process represents the duties of the BAS. Its main task is to authenticate the user and ensure the live presentation of the biometric data. The workstation process represents the user's activities: presenting the biometric data for the biometric authentication and request transfer transaction to the bank. The U process is responsible to input the user's verification data and user's biometric data to the workstation. The Bank process represents the bank's business. It verifies the biometric authentication result and performs the transfer transaction as requested by the user if the authentication is successful. The ProVerif models of the protocol are described in the following section.

A. Equation and Signature Theory

This is the method that ProVerif uses to solve the messages. From the equational and signature theory, ProVerif also forms messages from the provided theory in order to solve whether an attacker can acquire the information that the *query* commands ask for. It, in turn, returns the verification results. The *checksign* operation is performed in order to verify the signature of the message and the information in the signed message is presented. The signed message is performed through *sign*. The *enc* is used for message encryption. The messages are encrypted by public key cryptography. To decipher a message, *dec* is performed to decipher an encrypted message from the known key. The model of the protocol in Proverif contains signature and equational theory which generate signed messages and messages encrypted by public key cryptography. This process advises Proverif how to encipher and decipher a message. The *sign* function is used to sign the message whereas the *pk* function is used to generate the public key. The equation *checksign(sign(x,y),pk(y)) = x.* is represented for verifying the signature of a message from the provided public key of the origin. The equation *equation dec(enc(x,pk(y)),y) = x.* represents the deciphement of the public key encryption message.

B. BAS process

When the BAS process receives the authentication request, the BAS responds the request by sending a signed message of newly generated nonce n2 and biometric authentication request. It then waits for the biometric data to be sent to. In order to verify the live presentation of the user, the BAS sends the message that acquires the verification data. This message includes the newly generated nonce n3 and verification data request. The message is signed by the bank to specify the origin of the message. To secure the message and the validation checking purpose, the nonce, verification data and the signed package are encrypted by the public key of the TPM. The BAS expects to receive the biometric data and secret verification data sent from the user's workstation. Upon receipt, it deciphers the message, checks the validity of the data and performs the user's biometric verification. The matching result, user name and the nonce n1 which is received when the BAS has accepted the authentication result are signed by the BAS so that the recipient can validate the origin of the message.

C. Workstation Process

The user's workstation initiates the communication by sending a transfer request to the bank. It then receives a request for biometric authentication and the nonce n1 from the bank. The workstation process forwards this message to the BAS. It then receives the authentication challenge and the nonce n2 from the BAS. The user places his biometric data on the sensor. This results in generation of the biometric data in ProVerif. His biometric data is signed by the TPM to guarantee it origin. The username, the nonce n2 and the signed biometric data are encrypted by the BAS's public key. This message is supplied to the BAS. The workstation process receives the request for the verification data. It then obtains this data from the user and sends it out. It expects to receive the biometric matching result. The workstation process performs the transfer transaction by sending the relevant information to the bank in encrypted format. The transfer transaction is sent out together with the authentication result.

D. Bank Process

The Bank process represents the transfer transaction. It first receives the transfer request from the user or local's workstation. As a response, it then generates a nonce n1 and sends it with the authentication request. The bank waits for the reply message which is expected to be an authentication result. The message the bank receives is encrypted by the public key of the TPM. It then deciphers the message; it checks the validity of the message by checking whether the first nonce is the same as the one it sent to the workstation. If so, the bank checks the signature of the message by checking whether it came from the BAS. The bank checks the matching result. If the result is positive, the bank then performs the transfer transaction.

E. U process

The U process is waiting for the request from the workstation to present his biometric data and verification data.

F. Main process

The BAS process, the Bank process the user's workstation process and the U process are replicated and executed in the main process. The main process generates the keys for each process, these keys including public/private key pairs of the BAS, TPM and bank.

IV. Analysis of the model

The three desirable properties of the biometric authentication protocol are verified: the privacy of biometric data, liveness, and intentional authentication. The privacy of the biometric data is analysed to ensure that the protocol does not have risk of spreading around the user's biometric data without restriction. The Proverif model *query attacker : BD* is used to analyse whether an attacker could intercept the data successfully.

The protocol must be guaranteed that it deny an access from the artificial biometric. The paper models Alice, an intruder, holding an artificial biometric data e.g. rubber finger of legitimate user, Bob, can obtain the positive authentication result as if she is Bob. The Proverif model *query attacker: Alice* is analysed.

Since the protocol is used for a particular purpose, money transfer, the intentional authentication of this protocol can

be illustrated as the money is transfer to the correct account for the correct amount as the user wishes. The analysis of the intentional authentication property for this protocol refers to whether an attacker can capture the authentication result and use it to transfer money by posing as the legitimate user. The Proverif model to check that if the bank process exposes *AliceAccount* to the public channel, the attack is found. The model in ProVerif *query attacker : AliceAccount* is verified. Alice cannot intercept and manipulate the transfer message to transfer the money to her account.

The verification result illustrates that the protocol is secure for use in on-line money transfer in a biometric authentication situation. The protocol accomplishes all the three properties of the protocol.

V. SUMMARIES

This paper proposes a remote biometric authentication protocol. The protocol uses on-line banking to illustrate the protocol. The proposed protocol guarantees three properties: *privacy of biometric data*, *liveness* and *intentional authentication*. In order to secure the biometric data, the TPM is used within the user's workstation. This enhances the security level of the protocol. The user is assured that the workstation and the biometric reader he is using will not manipulate his data and request. The signatures are used to guarantee the origin of the messages. The public key encryptions are applied to secure the messages. The positive results from the verification show that the protocol holds the three security requirements. Therefore, the proposed protocol guarantees the certain level of security when it is used in unsupervised condition.

References

[1] Chen, L., Pearson, S., Vamvakas, A.: Trusted Biometric System. Available at URL http://www.hpl.hp.com/techreports/2002/HPL-2002-185.pdf (2002).

[2] Dolev, D. and Yao, A.C.: On the Security of Public Key Protocols. In Proceedings of 22nd IEEE Symposium on Foundations of Computer Science (1981) 350-357.

[3] Blanchet, B.: ProVerif : Automatic Cryptographic Protocol Verifier User Manual (2005).

[4] Abadi, M., Fournet, C.: Mobile Values, New Names, and Secure Communication. POPL 2001

[5] Tusted Computing Group.: TPM Specification version 1.2 Parts 1-3. Available at URL http://www.trustedcomputinggroup.org/resources/ (2009).

[6] Salaiwarakul, A., Ryan, M.: Verification of Integrity and Secrecy Properties of a Biometric Authentication Protocol. Fourth Information Security Practice and Experience Conference (2008) 1-13.

[7] Salaiwarakul, A., Ryan, M.: Analysis of a Biometric Authentication Protocol for Signature Creation Application. Third International Workshop on Security (2008) 231-245.

[8] Tada, Carlos H. M., Zaghetto, Alexandre and Macchiavello, Bruno: Fingerprint Image Quality Estimation Using a Fuzzy Inference System. In: Seventh International Conference on Forensic Computer Science (ICoFCS 2012), DOI: 10.5769/C2012007, ISBN: 978-85-65069-08-3, pages 41-45.

[9] Fernandes, Mateus and Puttini, Ricardo: Improvements and Adaptations in Fingerprint Processing Techniques for the Creation of a High-Quality Minutiae Database. In: The International Journal of Forensic Computer Science – IJoFCS, Volume 3, Number 1, ISSN: 1809-9807, DOI 10.5769/J200801009, pages: 87-96.

[10] Quintiliano, Paulo and Rosa, Antonio: Face Recognition Applied to Computer Forensics. In: The International Journal of Forensic Computer Science – IJoFCS, Volume 1, Number 1, ISSN: 1809-9807, DOI: 10.5769/J200601002, pages: 51-59..

DOI: 10.5769/C2013003 or http://dx.doi.org/10.5769/C2013003

Cloud Forensics
BEST PRACTICE AND CHALLENGES FOR PROCESS EFFICIENCY OF INVESTIGATIONS AND DIGITAL FORENSICS

José Antonio Maurilio Milagre de Oliveira
Diretor
Legaltech
São Paulo, Brazil
jose.milagre@legaltech.com.br

Marcelo Beltrão Caiado
Chefe da Divisão de Segurança da Informação
Procuradoria Geral da República
Brasília, Brazil
MarceloBC@pgr.mpf.gov.br

Abstract—**Digital forensics is a relative new science that has many challenges to overcome. This has been especially true since the huge adoption of cloud computing, which has its own characteristics, and the fact that many companies and providers are not well prepared to respond an incident in a proper manner. This paper discusses most common assumptions and principles, and proposes a base process for digital forensics in cloud computing.**

Keywords: cloud computing, cloud forensics, digital forensics, procedures, information security.

I. Introduction

There is no doubt that cloud computing is a phenomenon that tends to change the way of delivering services in Information Technology (IT) and Communication. Since 2009, the U.S. Federal Government has announced measures to implement a massive and complex infrastructure with the launch of Apps.gov, an online storefront for cloud services [1].

In Europe, cloud computing is expected to generate 800,000 jobs. In Brazil, it is noted the advance of the Federal Government with public services, which outlines a strategic plan to drive the adoption of cloud services in the country in a program called "TI Maior", presented by the Ministry of Science, Technology and Innovation [2]. The program discusses issues regarding development, regulatory framework and also aspects related to information security as well.

According to a *Kelton Research* survey [3], 74% percent of companies are already using some cloud computing service. Flexibility, IT environment simplification and costs reduction, are just some of the reasons.

On the other hand, there are no doubts that the growth of technology can also carry risks, involving fraud, incidents and electronic crimes. A survey by *CipherCloud* [4] conducted during the cloud-focused Dreamforce event in San Francisco that drew more than 48,000 attendees, shows that among the biggest concerns of companies, when choosing technologies in the cloud, are data security (66%), data privacy (56%), compliance (34%) and data residency (26%).

In this scenario, it is necessary to devise a process of investigation and digital expertise to be effective and that respects the characteristics of business models involving cloud services and especially in accordance with the legislation or applicable international laws. This is the challenge, considering the characteristics of cloud computing that relativize to the extreme the standards and practices adopted in Computer Forensics.

Put together to the challenge a poor doctrine applied to the subject. Among the first papers that keep a relationship with Computer Forensics and problems in cloud environments, there are the ones published by Wolthusen [5], and by Bebee [6] which proves the need to address these issues urgently, considering the astonishing development of technology.

This paper, showing some of the challenges discussed in the international community, has its bedrock on the design of a proposal for the investigation process and digital forensics in cloud environments. It also presents assumptions, principles and practices to be observed in such expertise areas.

II. Planning Information Security in Cloud Forensics

The success of digital forensics in cloud environments is closely linked with information security planning.

Speculations on information security in cloud environments are increasing, from risk analysis to implementations of controls to ensure security metrics are met.

In fact, some of the worrying foreseeable risks in cloud environments that must be included in a risk assessment for a possible implementation or migration are:

1. Improper access to information: Any form of unauthorized access to sensitive or classified information as confidential;
2. Information leakage: The disclosure of communications, data and trade secrets; and
3. Unavailability of services: Attacks targeted to the structure of cloud computing, which somehow disturb or interrupt the service.

A. Which Elements of Tracking Will be Generated

An important aspect to conjecture is related to the systems auditability. In this context, stakeholders should establish metrics, periodicity, scope and format of logs and other records to be created and maintained.

The adoption of an interface to access data records is also critical, mainly in SaaS service facilities, wherein the customer access to records and physical information is more limited. It may also agree upon a forensic API contract, which allows the actual client to initiate the first response.

Finally, it is important that the CSP (Cloud Service Provider) be obliged to inform the customer in cases involving incidents or attempts immediately and with complete documentation about the incident.

B. Human Resources for Forensics Responses

Forensics responses should be predict in agreements between CSPs and customers, especially detailing the procedure, in which case the answer must be forensic imprint and above of all, indicating internal staff as well as contractors or independent third parties that could follow the examinations. There must be a staff of suitable professionals, incident responders and legal body, which must be in the service level agreement (SLA) and in the contract.

1) How to deal with cloud computing

It may be that physical access to the affected device is thousands of miles away from the client, which is why it is important, in the contract, to establish where, physically, the customer wants their data to be, choosing a location with greater forensic maturity and more suitable legislation.

2) Consolidated standards

When detailing the procedure that the human resources will perform, it is essential the adoption of consolidated standards in the community, among which we can mention:

- SAS 70 certification;
- RFC 3227: Guidelines for Collection and evidence Archiving;
- NIST SP 800 86: Guide to Integrate into Techniques Forensic Incident Response;
- ISO / IEC 27037: Guidelines for identification, collection, acquisition, and preservation of digital evidence;
- ISO / IEC 27041: Guidelines for the analysis and interpretation of digital evidence (DRAFT); and
- ISO / IEC 27043: Digital evidence investigation principles and processes (DRAFT).

C. Cooperation in Multi-jurisdiction Cases

The Safety Plan must be designed by knowing how the customers data physical division is performed, considering legal aspects and privacy of each cloud shadow, detailing clearly contacts of response teams and details of the legislation of the countries in cases of incidents.

This preliminary step is critical to the success of any forensic analysis, because in case of any incident, the expert must make the data segregation, which is not an easy task, without having the minimum information. It is important to mention that the cloud provider must present the customer and determine the liability of third parties which are also used to provide the service.

It is therefore confirmed that CSPs and customers need to establish forensic capabilities so that we can reduce the information security related risks in cloud environments.

Best Practices for cloud computing security should be observed when designing, hiring, establishing metrics and service levels across multiple CSPs and customers.

Internationally, the Cloud Security Alliance (CSA), has a good practice guide for information security implementation in cloud environment [7]. Likewise, the European Network and Information Security Agency also has important recommendations on the subject [8].

III. Digital Investigations In Cloud Forensics

A forensic response process to incidents regarding cloud computing should be provided in the Security Management System and agreed with service providers and everyone in the supply chain, considering the maturity of the implemented security as well. There is no doubt that the success of a forensic response process is closely linked with the maturity of information security applied to the cloud structure and especially the willingness of such service providers. Rarely, in an investigation of this nature, there will be the traditional and classic option to seize the equipment.

The digital forensics is an area for identification, preservation, collection and analysis of digital evidence and artifacts (those, when relevant to the case), in the scope of presenting the materiality of an incident (showing whether the event actually happened or not) and mainly by indicating

the source of the incident. This is a science in its infancy, with few more than ten years of groundbreaking research.

Among the fronts of digital forensics, we can identify the *post-mortem,* where analysis have addressed commonly content of discs, recovery, carving, e-discovery, among others, and the live one, which seeks volatile content such as memory, kernel, processes, network states, data that is totally or partially impaired with the shutdown of the equipment.

Nonetheless, cloud computing has elasticity as one of its essential characteristics. The term elasticity refers to the idea of an environment that can be easily extended, according to customer demand.

Cloud forensics, in this context, would be one of the specializations of Digital Forensics, target to the analysis of cloud environments, involving investigations related to incidents, fraud and computer crimes. To Keyun Ruan [9], from the Centre for Cybercrime Investigation, of University College Dublin, cloud forensics would be linked to network forensics, which in turn would be linked to digital forensics. To the authors network forensics techniques could be tailored to cloud computing environments.

Nonetheless, access to data on disk (raw) or snapshot structures will often be essential for understanding what happened to the compromised system, given the elasticity of the cloud service models. We also cannot fail to conjecture the intimate connection with Database Forensics, as sometimes the expert must act in this instance, seeking records from unauthorized modifications of data stored in the cloud.

When we think of cloud, we imagine a model, on demand, in which access is allowed to a shared pool of configurable resources, including but not limited to networks, servers, storage, applications and services. For the forensic expert, initially it will be mandatory in the identification phase, to determinate if it is really a cloud environment or any other form of web service, or even a VPS or VPN. A mistake in the identification of the service, will certainly lead to investigations failure, which may violate standards and best practices.

From the perspective of Computer Forensics, virtualization services on a single physical server brought several points and questions to be addressed by the research community. The ease of deleting data has always been one of the issues pertaining to virtualization. On the other hand, it may be stated that the cloning bitstream (physical) would be facilitated by copying the file that represents the virtual disk.

With cloud computing, we have other issues to be considered. While the cloud become an object of studies by hackers and crackers, in its various instances, from the hypervisor (which manages the resources for virtual machines) to the interface layers, there is also concern about the use of public and private cloud providers as anti-forensic technique. Criminals could be using this technique for improperly accessing virtual spaces, practicing crimes or hosting shells, botnets, access to resources for deep-web, trojans among others. As an example, the Pirate Bay is operating from cloud-hosting providers around the world to escape from authorities [10].

By the other hand, there is the concept of *data abundance* involving artifacts, where screening sample techniques need to be applied to prevent that the forensic never ends.

In this context, the digital investigator must bear in mind that the cloud within the practice of computer incident may be used as:

- Object: When the virtual server in the cloud is the target of cybercriminals, being directly attacked, such as in a denial of service;
- Environment: When the cloud is the environment in which a digital crime is committed, such as unauthorized modification or deletion of data;
- Weapon: When the cloud is one of the tools used to commit crimes or stores digital planning or artifacts that might lead authorship of a possible computer crime. This context is also when cloud is used as anti-forensic technique for stealth connection or attribution of authorship to an innocent person, or even the use of botnets;

In above cases, sometimes customers and cloud providers are at litigation, where an expert will be appointed to evaluate eventual failure in service delivery, which might has generated losses or accountabilities.

Still, it should be noted that forensic investigations in cloud environments will take place in the following interesting prospects:

- Research: Full investigation of violations of law and policy, or even suspicious transactions, rebuilding events and collaborating with authorities and sponsors of expertise in collecting and analyzing evidence; Using from network techniques such as packet capture techniques to disk (dead) capture, and data recovery, encrypted and using stenography;
- Prevention: Through the log monitoring, event correlation and anticipation of supposed incidents; Working in conjunction with the incident response team;
- Compliance: Helping companies and organizations meet the requirements and best practices involving security and response to incidents involving cloud computing;

According to [11], a good research method should always consider different sources of evidence, not only the provider but also the customer terminals, using methods like data

fusions for collection and data correlation.

IV. Cloud Forensics Principles

Although there are much disagreement in regard to assumptions, principles and practices for investigative analysis in the clouds, some assumptions have been consolidated in the international community researchers. Those assumptions are features that need to be considered always in such analyzes. We present some of the most important ones.

A. Consider the Technical, Organizational and Legal Dimensions

Before starting to work on a cloud environment, the expert should divide the initial design of the project in three dimensions: the technical, which will map the entire structure to be analyzed; the organizational, where he will understand the business model, service features, and will map the called dependency chain and human structure for incident response and customer service; the legal, which he will assess the legal issues related to data and to orient the computer examination as evidence acceptance in court, establishing the chain of custody, among others.

B. Consider the Logical and Physical Dimensions

The expert must completely review the structure, in each forensic analysis, understanding the physical dimension that hosts the logical area of the client, and mainly identify which are the physical and logical constraints to access the assets. Seldom, in an examination in such environments, the expert will have full access to the physical dimension, whereas this dimension is considered by many providers their business secret.

C. It Might not Exist Media Control and Access to Physical Infrastructure

The principles, frameworks and best practices are usually based on the assumption that the storage media is always in investigator's control. This changes with cloud computing. Some concepts brought by the principles of the Association of Chief Police Officer (ACPO) of England, and Investigative Process Model (Dip Model) from Digital Forensics Research Conference (DFRS), are put in check when the environment is in the cloud. It must be noted that these frameworks are well regarded by the community in digital investigations.

The non-physical infrastructure must also be characteristics of multi-tenants and multi-ownerships clouds, where information can be stored in different asset owners or where a single physical disk can concentrate data from numerous other clients. In case of access, one could think of privacy violation.

D. Elastic Tools, Elastic Cloud

The community should look for tools that fit the elasticity of the cloud.

E. Provider Cooperation is Essential

Despite some models of services in the cloud facilitate customer access to information and metadata, it is also known that it is virtually impossible to perform an examination in cloud environment without any cooperation.

V. Proposed Process for Cloud Forensics

A process for responding to incidents involving forensic cloud computing should be provided in the Security Management System and agreed with service providers and third parties in the chain of dependence, including the possibility of simulations.

Digital forensics has not been seen as an easy task in cloud computing devices. According to Gartner: "cloud services are especially difficult to investigate, because data access and data from multiple users can be located in several places, spread across a number of servers that change all time" [12].

Starting from the assumption that the company already knows the risks involved in a cloud environment, we have to define a process for the forensic response time, which not only restore services but mainly produces scathing evidence of what occurred in a system, and can be considered in court. Among the steps we propose for an investigation and digital forensics in a cloud environment, there are:

A. Map Technical, Organizational and Legal Dimensions

This is the first step, i.e., before the expert establish an effective plan for forensic analysis, one should divide the assessment into three tabs, and in it, sort and collect all available information, contacts, norms and rules.

At this stage it is important that the expert consider the following, as it will give needed information to advance in the examination:

1. Review the contract, SLA and Security Policy; (cases involving cessation activities, deletion or exclusion, any zero knowledge encryption system, cooperation with authorities), among others; and
2. Assess whether cloud computing characteristics are present (or if we are dealing with other similar services)

As stated by [13], the back-end is generally a three-tier arrangement, comprising: physical machines and storage, virtual machines and a SLA layer (Figure 1). The SLA is responsible for the monitoring of the service contract to ensure

its fulfillment in real-time. All layers should be considered by any expert when evaluating the service contract.

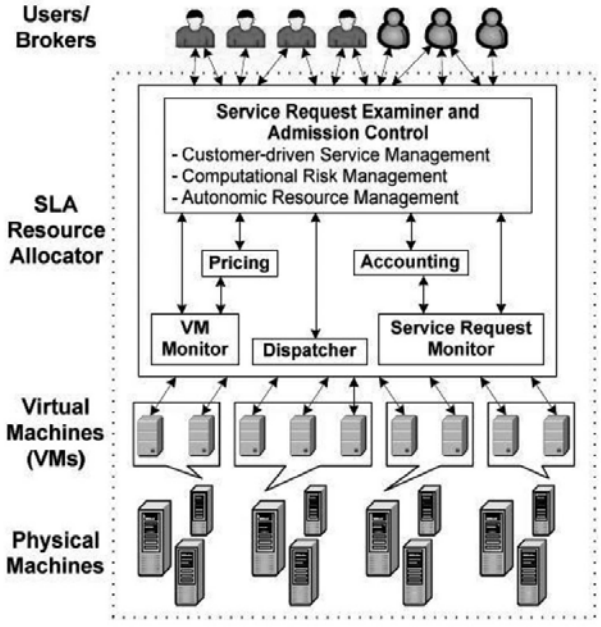


Figure 1: Cloud Computing Layers [13]

The expert must also evaluate organizational configuration or development platform models, which may be:

1. Private Cloud: Infrastructure is operated only by the organization that owns the cloud.;
2. Community Cloud: The cloud is shared by many organizations, because they have a common goal, and is administered by the community;
3. Public Cloud: It contains information from more than one user / customer, maintained by a organization provider; and
4. Hybrid Cloud: Composition involving two or more models, also called *virtual private cloud*. Sometimes used for load balancing in different clouds.

For digital forensics it is important to know the ways of configuring a cloud service, since this will directly impact on the path of data that can be collected as evidence. It should be noted that data stored in the cloud can be stored in one or more distributed physical locations, making the determination as to which law should be enforced or even the procedure and framework applied. This is a complicated issue and that needs to be addressed by the expert.

In addition to the organizational setting, the expert should evaluate the service model supplied by the provider being analyzed. As it is known, there are three basic levels of services involving cloud computing models, namely:

1. SaaS (Software as a Service), where the client can use applications available by the provider cloud, and the interaction is commonly done through web-browsers. As an example, there is the Google Apps suite of applications;
2. PaaS (Platform as a Service), where there is the availability of an application programming interface (API) so that customers can create and host applications. Commonly there is the provision of a development platform; and
3. IaaS (Infrastructure as a Service), which is the assignment of virtualized computing resources such as processing power, memory and storage.

It is critical to identify the service model so that we can prepare the process of digital forensics, considering the variants of each service. The collection is directly influenced by the models of service delivery. In IaaS-based platforms, there is more interaction between the client and platform, which will result on a greater possibility of collecting data for forensic examination, that may not occur in PaaS and SaaS models. Typically, on SaaS and PaaS platforms the expert will not have control of the hypervisor, which would be very important in an investigation.

Another advantage of investigating IaaS environments is related to the fact that in such a model, it is usually possible to make a snapshot analysis, supported by all popular hypervisors like Xen, VMware ESX and Hyper-V. Furthermore, processes need not to be interrupted for forensic analysis, generating no downtime or SLA violations.

On the other hand, the ideal is that SaaS and PaaS interfaces offer or implement an additional interface with the purpose of compliance and forensics. Through the API, clients should receive information about events in their environments [14]. Another alternative may be the compression and encryption of logs that could be sent to third-party servers, preventing the possibility of a shutdown or volatile data destruction.

B. Identify Outlining Stages of Computer Forensics that Will be Overcome and Correlate Them to the Propositions

In this phase the expert will create tabs in his project with all phases of Computer Forensics: a) Identification, b) Preservation, c) Collection, d) Examination, e) Analysis f) Presentation. Within these tabs, he must employ assumptions that are consensus in the research community, as discussed in Part 3 of this work. We must recall that the expert should always be updated with new assumptions, principles and rules.

C. Identify Outlining Stages of Computer Forensics and Propositions with Technical, Organizational and Legal Results

At this point, we propose a data fusion. The expert will merge Computer Forensics phases, given the assumptions related to an examination of this nature, with the result of the mapping of technical, organizational and legal frameworks applied to the case. The result will be a matrix, where the researcher will have assumptions, data and characteristics to be examined at each stage of the forensic examination.

Having this information, the expert can then devise the best strategy for forensic investigation, beginning the execution of

his expert activity. Among the criteria that will emerge and that will guide the work, we list:

1) Identification

The detection of an incident in a cloud environment may differ according to the model adopted for the services. The adoption of cloud in Intrusion detection systems can be implemented by the user in the IaaS or even by the CSP in cases involving SaaS or PaaS. At this time the expert will interact with the professional´s provider for mapping the incident and the extent of damage. It will be identified which access the provider offers to the customer in the event of an investigation. Also, it will be identified if the provider is performing regular snapshots or even object auditing and multiple backups.

2) Preservation

The preservation of evidence in cloud environments is not so peaceful. Implementing preservation techniques may require isolating cloud resources, which can cause performance degradation for other clients. From the best practices, providers should isolate the physical disk connected to an incident. The problem is that data from other customers that share resources could also be copied.

Under the existing frameworks, identifying electronic stored information, commonly sets up procedures considering that the evidence is in the possession of the investigator. In cloud, the providers are in custody of such information. Client control is more difficult. The client can indeed control his data, but do not always have access to the metadata server he uses, and which are fundamental in a computer investigation.

Another issue that needs to be revised in the process of preservation is the chain of custody. In SaaS or PaaS models the customer may not be the first to have contact with the evidence, then the provider shall be responsible for this preservation task, involving the allocation of knowledgeable first responders.

In the conventional model, the chain of custody must start when the researcher has access to physical media. For companies, the challenge remains to implement contracts that allow the investigator access to the evidence, sometimes in a physical way, and not just with network access, or even a chain of custody that begins with the provider and then is transferred to the client.

3) Collection

The challenge of collecting is to have access to data. The investigator may have access to data, or copy over the network, or rely on the CSP team. The evidence collection in cloud environments proposes new challenges to experts, especially due to the lack of tools to assist them with agility. It should initially be pointed out the challenge to the expert who will handle increasing amounts of data, with the storage capacity growth and low cost of such devices. An investigation in a virtualized environment can become extremely costly in the collection phase, due to the existing devices. The elasticity (involving the ability to scale capacity according to the requirements), which is characteristic of the cloud, increases this problem.

One way to minimize this fact is to use screening models, as the model called Screening (CFFTPM) [15] a framework that has been growing among the research community worldwide (Figure 2).

The collection also will deal with the following questions:

1. Multi-jurisdiction: Data can be stored in physical locations with different jurisdictions. One must respect the jurisdiction of where the data resides;
2. Limited access to physical media: For legal or even business strategy, the expert may have limited access to media, needing a further court order;
3. Dead Forensics or Live Forensics: The memory capture and other states might be limited to an interface available to the customer. Similarly, it is virtually impossible to shutdown a machine to remove the disc or boot via live CD, common practices in traditional digital forensics. The expert should establish remote collection strategies.

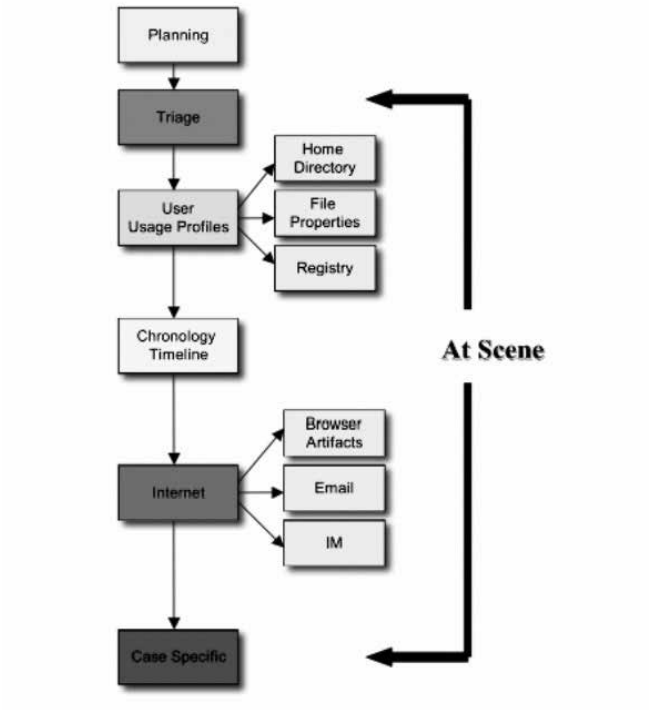


Figure 2: CFFTPM Phases

Commonly, the expert would be performing the imaging of the disc, through duplication bitstream equipment or a command like dd. This changes in cloud, where an interface apparently on a single disk is divided into multiple physical disks. The challenge for the expert is to know each

segment that composes the target and are interesting for the investigation, cloning the devices (defining the start and end clusters) and then have the capability to concatenate them in a investigation environment. Thus, the expert will have to deal with the concept of distributed multi-tenant. A good example of this is Google GFS [16].

It is mandatory for the contractors to predict, therefore, not only the logical access cloud interface, but also situations where physical access is essential. A proposal to collect data that proportionally respects neighbors in the clouds should be among the team's Cloud Service Provider and customers discussions.

If on one hand, cloud providers are striving to provide security to data in the clouds, on the other, such security involving encryption and data traffic can become an enemy in time of investigation involving cyber crimes. The implementation of *zero knowledge system*, a concept that allows all data to be encrypted before being sent to the clouds, may cripple an investigation. There will be the need of a covenant involving the exchange of keys, if the encryption is performed by the client, otherwise the provider should be legally liable. A negative example comes from Google, that to ensure the privacy of users, assures that when a user deletes their data, they are in fact deleted and the pointers of the replicators are also eliminated, which can be a challenge.

On the other hand, in the cloud, we can think about a certain persistence of data, which is an advantage, because unless the customer has administrative access, it becomes difficult to perform a complete deletion of data. Hence, it is important to know the persistence in the form of backups and other data provided by the CSP.

In this sense, the European Union encourages union members to apply the *Data Retention Directive* of 2006 [17], which in Article 5 stipulates communication providers to retain certain data about users, userid, allocated IP, time and date of the communication and time of login and logoff systems. The challenge is whether the legislation includes providers of cloud computing services. We will soon have the first legal signs on the topic.

A proposed solution to the problem of the constant inability of the expert to have physical contact with the evidence to be collected, can be called *organizational cooperation*, where the provider would be responsible for the extraction of forensic image of physical disk or partition, or at least virtual machine created for the client by handling the hypervisor. There should be exceptionally careful when handling the hypervisor, which can be compared to a kernel of the operating system. It will be usual for cloud providers to provide snapshots of the disk and client memory. Good practice recommends that this generation be documented and assisted by an expert for the customer, so it can be used as digital evidence in court. This approach will ensure that the hypervisor was reliable.

Important to remember, in these cases the researcher is not the first to have contact with the evidence and the chain of custody is created by the provider. It is clear and undisputed that cloud providers need to know the procedures of digital forensics, mainly relying on human resources prepared for such tasks.

Despite the cloud provider contract predicting the possibility of going to an expert for the collection of physical evidence, sometimes it would be needed someone from the provider to run the task. This is because we are dealing with different platforms, trade secrets, proprietary technologies, among other logistical issues that make it important that wherever possible the client's expert should be acting in conjunction or can follow the expert's provider in executions of tasks in operation.

It must be reminded that the legal limitation involving the location of each provider can compromise the legitimacy of the collected data.

It is, finally, another good practice to be implemented in providers, the automated generation of hashes of snapshots, dated, the virtual disks, as well as existing files, serving as a basis for comparison after the data is capture by the expert. The investigator then, at the stage of examination, must extract the hashes from snapshots computed by CSP after collection, and compare them by checking the matching.

Other important issues in the collection are:

1. In *live* collection, the expert should consider all endpoints, and the generation of the timeline of events should consider time synchronization, which is difficult and demands specific tools;
2. The expert must pay attention to the segregation of the evidence - collecting information logs from multiple clients can generate legal liability;
3. He should assess whether the system already has a solution to generate hash files from cloud, as well as if provides support for remote binary copy; and
4. The expert should evaluate if the system being examined offers versioning of erased or overwritten files / objects, and how it is possible to access these mirrors.

4) Examination and Analysis

The timestamping should be considered in the collection phase and also in the analysis phase. A knowledge process involving all jurisdictions should be adopted and timestamps applied to services. The community challenge is to design tools to automate this correlation.

Once the collection phase is overcome, by far one of the most problematic stages involving cloud environments, frameworks and practices for analysis can be applied to the analysis of computer artifacts. Many open source tools, data carving, pattern matching, and filtering are recommended, like The Coroners Toolkit, Foremost, Xplico, Autopsy, among others, contained in Linux Forensic distributions, can assist the expert work. Under proprietary software, EnCase and FTK should be considered. Dykstra and Sherman [18], performed one of the first research involving data collection

tests in the cloud with tools like FTK and Encase, in an IaaS environment.

In investigations where network traffic packets were collected, Xplico or Wireshark filters can be used for session reconstruction and even content decoding.

The analysis of evidence in cases involving cloud is similar to analysis of evidence in digital forensics and may involve:

- Processes;
- Memory;
- Files;
- E-mails;
- Logs;
- Network traffic; and
- Web data.

Regarding the logs, it is the expert´s task to be familiar with the most used platforms, knowing the way they are generated, so he can use a parser efficiently, detailing his report in an effective way.

5) Presentation

In this proposal, the presentation phase may consist of legal appraisal or a simple briefing or draft of what happened. It can be used in legal form by a lawyer or even used by the expert for the defense of their findings in court. The forensic reports also work as an input in process improvement and continuous corporate improvement.

The four Daubert Principles [19] (guidelines for acceptance of scientific evidence) should be considered in the presentation of results involving investigations in cloud environment:

1. The key question is know whether the theory can be tested, namely the theory must be tamper-proof. The CSP must maintain evidence for the time agreed;
2. The results should be subject to review by other experts;
3. When applying a determined known technique, the Court must consider the potential rate of error, and the existence and maintenance of standards and controls on their operation; and
4. It should be rated the degree to which the theory and technique is generally accepted by the scientific community. In Computer Forensics this is a difficult task, considering that discussions on international best practices are just starting, which is critical to the advancement in the area.

Nonetheless, it is highly recommended that cloud Provider's technicians sign along the client´s expert report, ensuring uniformity of opinions and avoiding exploitation thesis as self defense on the argument that the provider was unaware or did not recognize what was performed by investigators.

VI. CONCLUSIONS AND FUTURE WORKS

There is no doubt that Computer Forensics in cloud environments is still embryonic and needs to become more mature to be able to equate the efficiency of an investigation with respect to privacy, fundamental rights and guarantees and SLAs between providers and other customers.

It is known that the frameworks, practices and principles, wide discussed and consolidated in the community of Computer Forensics, are not explained in their entirety or accuracy and must promptly be derived, revised and adapted in the design of a minimum standard that meets the concepts, service models and configurations of cloud computing.

Within the present work, a proposal for preliminary e-discovery processes and Computer Forensics was shown, involving cloud environments, not exhaustive and stony, which can be adapted to meet the changing technology and the characteristics of each cloud environment, besides of cases that may be presented.

While many challenges exist in digital research of cloud environments, it is true that the contractual relations are identified as one of the solutions to the problem, and there is urgent need for international regulations. The Computer Forensics must be provided in terms of services, ensuring rights and duties between clients and providers. This is a negotiation that should be made between the parties. Computing is a key element, considering that the elements of compliance in providing cloud services to grow. In Brazil the PL 5344/2013 [20], presented by Mr. Ruy Carneiro, wants to regulate the relationship between users and companies of this type of service.

An example that is worth to mention is the city of Los Angeles, who adopted an e-mail system to 30,000 employees in 2009, hiring Google services [21]. In this contract there are predictions that Google can fix the city in case the system is broken and city data exposed. On the other hand, the Gmail service offered to individual customers, allows Google to processes personal information on servers in the United States and other countries, which can be a deterrent in the face of an investigation involving such servers.

As it can be deduced, considering that bargaining power may be greater on small providers, an expert can find relevant information to an investigation more easily on these cases.

Some issues that can and should be contractually defined are: a) data collection amount and frequency, b) where the information will be stored, c) interface for access to data pertaining to the incident, d) ways that virtual disk images will be provided, e) hash format of the files, f) who handles the evidence on the side of the CSP, g) restrictions on certain datacenter storage locations, which contain no laws on privacy and security, or that do not cooperate in investigations, among other issues that can troubleshoot data spoliation or deterring investigations involving data in the cloud;

From the technical view providers may consider creating automated systems that collect and preserve ESI (Electronic Stored Information) pertaining to customers, for cases involving incidents. Other issues that must be considered in the technical side are: a) Ability to capture specific packages

in relation to client servers b) Potential access to routers and other network components c) Segmented access to Firewall record; d) Access to the service hops, e) Creating an instance for log storage.

In this scenario, although not exhaustive, the first lines to design a model for process efficiency of investigations and digital forensic in the cloud were presented. They can be extended to model specific processes for each one of the different cloud computing technologies.

REFERENCES

[1] V. Kundra, "Streaming at 1:00: In the Cloud". The White House - Office of Social Innovation and Civic Participation. http://www.whitehouse.gov/blog/Streaming-at-100-In-the-Cloud), September 15, 2009.

[2] TI Maior – Programa estratégico de Software e Serviços de Tecnologia da Informação. http://timaior.mcti.gov.br/

[3] Kelton Research, "Global Survey: Has Cloud Computing Matured?". Third Annual Report, Executive Summary, Avanade Research & Insights (http://www.avanade.com/Documents/Research%20and%20Insights/FY11_Cloud_Exec_Summary), June 2011.

[4] CipherCloud, "Data security and privacy stopping cloud implementations", in press.

[5] S.D. Wolthusen, "Overcast: Forensic Discovery in Cloud Environments", In: Proceeding of the 5th International Conference on IT Security Incident Management and IT Forensics (IMF '09), Sttugart, pp. 3-9, 2009

[6] N. Beebe, "Digital Forensic Research: The Good, the Bad and the Unaddressed", In: G. Peterson and S. Shenoi (Eds.), Advances in Digital Forensics V, IFIP AICT 306, Springer, pp. 17-36, 2009.

[7] Cloud Security Alliance (CSA), "Security Guidance for Critical Areas of Focus In Cloud Computing v3.0", 2011.

[8] European Network and Information Security Agency (ENISA), "Cloud Computing - Benefits, risks and recommendations for information security", 2009.

[9] Keyun, R., Carthy, J., Kechadi, T., Crosbie, M. "Cloud Forensics", IN: G. Peterson and S. Shenoi (Eds.), Advances in Digital Forensics VII, 7th IFIP WG 11.9 International Conference on Digital Forensics, Orlando, FL, USA, January 31 – February, 2011, Revised Selected Papers, Spring, pp. 35-46, ISBN 978-642-24211-3, 2011.

[10] S. Anthony, The Pirate Bay moved for the cloud to evade police" (http://www.extremetech.com/computing/ 138037-the-pirate-bay-moves-to-the-cloud-to-evade-the-police), October, 2012.

[11] S. Satpathy, S. Pradhan and B. Ray, "Digital Investigation Tool Based on Data Fusion in Management of Cyber Security Systems", International Journal of Information Technology and Knowledge Management, July-December 2010, Volume 2, No. 2, pp. 561-565.

[12] Gartner Inc., "Cloud Computing – Key Initiative Overview", 2010.

[13] VIVEK School of ERP. Cloud Computing. http://acharyavivek.blog.co.in/2010/04/12/cloud-computing-2/, 2010.

[14] D. Birk, C. Wegener, "Technical Issues of Forensic Investigations in Computing Environments", IEEE/SADFE 2011, 6th International Workshop on Systematic Approaches to Digital Forensic Engineering (in conjunction with IEEE Security and Privacy Symposium), Oakland, CA, USA, 2011

[15] M. Rogers, J. Goldman, R. Mislan, T. Wedge, "Computer Forensics Field Triage Process Model", Conference on Digital Forensics, Security and Law, 2006.

[16] S. Ghemawat, H. Gobioff, S. Leung, "The Google File System", 19th ACM Symposium on Operating Systems Principles, Lake George, NY, October, 2003.

[17] Directive 2006/24/EC of the European Parliament and the Council, 15 March 2006.

[18] J. Dykstra, A. Sherman, "Acquiring Forensic Evidence from Infrastructure-as-a-Service Cloud Computing: Exploring and Evaluating Tools, Trust and Techniques", Digital Investigation 9:S90—S98, 2012.

[19] S. Richard P. In: "FOR 508 Advanced Computer Forensic Analysis and Incident Response, workbook day 5 - Computer Crime US", SANS Institute, May, 2010.

[20] Projeto de Lei 5344/2013. http://www.camara.gov.br/proposicoesWeb/fichadetramitacao?idProposicao=570970

[21] S. David, "Los Angeles Adopts Google Email System for 30,000 City Employees" (http://latimesblogs.latimes.com/technology/2009,10/city-council-votes-to-adopt-google-email-system-for-30000-city-employees), September, 2009.

DOI: 10.5769/C2013004 or http://dx.doi.org/10.5769/C2013004

Analyzing Targeted Attacks using Hadoop applied to Forensic Investigation

Parth Bhatt[1], Edgar Toshiro Yano[2]
Dept. of Electronics and Computer Engineering
Instituto Tecnológico de Aeronáutica
São José dos Campos,SP,Brasil
parthbhatt09@gmail.com[1], yano@ita.br[2]

Abstract—**Conventional intrusion detection and prevention technologies are mostly based to work on traditional methodologies to detect malicious events, while mining on a midsized log data. In recent years, we have seen the evolution of sophisticated targeted attacks performed by well trained adversaries exhibiting multiyear intrusions; therefore existing security toolsets have become insufficient for analysing targeted attacks with necessary speeds and agility.**

Dealing with such sophisticated attacks requires working with huge volume of multiyear security log data. Big Data technologies, such as Hadoop, enable the analysis of large and unstructured data sources, therefore, in this paper we propose our framework based on Hadoop for dealing with Intrusions performed by Targeted threat adversaries, using concept of Intrusion kill chains which will be helpful for forensics analysis.

Keywords-targeted threats ; Hadoop; intrusion kill chain;

I. Introduction

Due to rapid development of internet, all connected and unconnected machines everywhere around the world can be somehow reached and hence can become targeted for malicious purposes. Such activities of selection of targets and launching attacks directed towards a specific target are known as targeted attacks.

Although this term is not new in the area of Computer Security but due to recent advancement and increase in number of such sophisticated attacks, they have become the subject of high alert for every government, organization and industry. The threat of this class are targeted groups or individuals who are willing to spend extra effort to gain their objectives.[1]

A subset of targeted threat that is majorly famed by security industry in these recent years is Advanced Persistent Threats (APT). APT, as defined by a research paper from Lockheed Martin Corporation [2] are "well-resourced and trained adversaries that conduct multi-year intrusion campaigns targeting highly sensitive economic, proprietary, or national security information". These adversaries use sophisticated techniques to elude most of the contemporary security defence mechanisms and once being successful, aims to keep their persistency without getting detected inside their target environments.

Stuxnet was one such most complex APT, which was primarily written to target industrial control systems. The majority of infections were found in Iran specifically targeting programmable logic controllers of gas pipelines and nuclear power plants [5].Another variant of Stuxnet called duqu was recently found.

Operation Aurora aims to steal intellectual property of variety of technological, security and defence companies. A drive-by download attack was used to infect user's machine with a malware exploiting vulnerability [1]. Some other examples of such target attacks are given in case studies in [2], [4].

In this paper, we discuss a framework for dealing with Intrusions performed by Targeted threat adversaries, using kill chains. As described above, targeted attacks are getting advanced and more sophisticated, and for a typical organization dealing with critical information, there may be many adversaries attacking during a common time frame. In such a case, when finding correlation between detected suspicious events is already a challenge, it becomes really important to examine attack patterns to classify them into different groups of adversaries. This problem can be addressed by using intrusion kill chain method [2] which can help in situational awareness, intrusion correlation and intrusion prevention for future attacks.

APT adversaries generally aim to remain persistent inside their targeted environment, and they are likely to continue more intrusions, and these intrusions can be distributed in a big time frame (for example some months or years, as explained in the case study in [2]) . Thus, security log data collected from different sources in any such time period cannot be filtered or discarded, because any intrusion detected in future can be useful to find undetected malicious patterns of similar adversaries in the past. Furthermore, security log data collected from different sources (Network Intrusion detection system or NIDS, Host Intrusion detection system or HIDS , server logs ,mail logs, error logs etc) in a long time period can be classified as unstructured big data[8,9].

In our experiments we successfully tested our framework, which uses Intrusion Kill chain method and Big Data technologies (such as HDFS and map reduce for efficient log management and faster information retrieval of Big Data [3]) can be the future trend for Intrusion detection and prevention systems applied to targeted threats.

The rest of the paper is organized as follows Section II describes related works, Section III is about intrusion kill chains and analysis , Section IV is about technologies used in the framework, Section V explains proposed framework ,Section VI is implementation and results , Finally section VII is about conclusion and future works .

II. Related Work

During the last decade, due to increase in number of sophisticated targeted threats and rapid growth in volume of data traffic, the landscape of analysing log data has drastically changed, as now working with multiyear log data has entered the category of Big Data problem [9].

J. Howes, J. Solderitsch, I. Chen & J. Craighead [8] proposed an analytical security model considering the security analytics using Big Data. Their architecture is directed towards dealing with operational concerns in security organizations that aim to use existing security tools with Big Data analytics. Since their work is aimed towards operational side of security analytics therefore, it does not demonstrate any methodology of practical analysis of security threats as compared to our framework.

J. Therdphapiyanak and K. Piromsopa[3] used Hadoop map reduce model to analyse high volume of log files from server and distributed intrusion detection system and they proved that their frameworks performance was better than a standalone intrusion detection system .They were able to extract important information from the large security logs using their analysis and scalability of Hadoop, but their work was limited to use of K-means clustering algorithm of Mahout for detection of the deviated behaviour Clusters from normal behaviour Clusters. Using the proven capabilities of Hadoop for log analysis as in [3], our proposed framework is directed towards practical analysis of dealing with Targeted threats.

In [16], a blind automatic method for detecting malicious activities in honeypot data is proposed which uses RADOI to successfully identify attacks without any human intervention, while in [14] and [15], blind automatic detection for distributed honeypot data is proposed.

Authors could not find any other academic work that uses practical approaches to deal with detection of targeted threats using Big Data Technologies.

III. Intrusion Kill Chains

To understand the behaviour of targeted threat adversary is a significant problem as it generally deviates some attack vectors for each intrusion [2]. Our event logging systems can capture most of the system events but still correlation of similar group of events to identify behavioural characteristics of an adversary is of great importance, which can be addressed by Intrusion kill chain.

Intrusion kill chain model as given by Eric M. Hutchins, Michael J. Cloppert, and Rohan M. Amin [2] from Lockheed Martin Corporation is a series of phases that an attacker inescapably follows to model and carry out his intrusion. The intrusion kill chain phases are as follows:

The first phase is Information Gathering which involves selection of targets, collecting information about the target, for example searching emails, technologies the target uses, people on which their target trusts. The very next step an attacker will take is Weaponization which is coupling of malicious code with unsuspected deliverable files such as pdfs, docs, ppts and etc. Next in the third phase the attacker delivers the weaponized file to its target environment The most common delivery vectors are email, drive by download through a website link or through USB removable device. Once the malicious weaponized file gets successfully delivered in its target environment, the use of the vulnerability of the target system is taken to execute its malicious code, thus this phase is called as Exploitation

Next, the most important and crucial phase of the Kill chain is the installation of the malicious code inside the target environment .Remote Access Trojan's (RAT) are generally installed which allows adversary to maintain its persistence in the targeted environment.

The second last phase of the intrusion kill chain is Command and control (C2), in this phase the installed trojan or other malicious code generates a communication channel to control its execution and continue its actions to achieve its target

Actions is the last phase of the kill chain in which adversary achieves its objectives by performing activities like data exfiltration.[2] Defenders can be confident that adversary achieves its goal only after passing through all these phases.

Figure 1: Intrusion Kill Chain

After understanding the Intrusion kill chain phases, we need some methods to proceed towards construction or completion of kill chains once an malicious event is found. The following approaches can help us to deal with kill chain Construction.

a) Intrusion reconstruction: when a certain malicious event is detected and its phase is identified, analyst can be sure that the prior phases have been executed successfully [2]. Intrusion reconstruction is done by discovering the previous phases of the kill chain as those phases must have been taken in order to reach the detected phase. This can help defenders to mitigate the future intrusions and to understand the adversary's method of attacking.

b) Intrusion Synthesis: It is important to estimate what might have happened if defenders did not mitigate the intrusion on time. If such measure is not taken then, there is a chance that same type of attack may go undetected in future intrusions [2]. If defenders are able to collect more and more information about the kill chain, they can maintain an advantage over their adversary.

c) Campaign analysis: It consists of analysing multiple correlated intrusion kill chains expected to be from similar adversary over a long period of time (i.e. months or years of intrusion activity). Attacking persistently is an inherent disadvantage for the adversary which can be a great opportunity for defenders to identify the intrusion behaviour and improve their detection for future attacks. Re-using tools and techniques for intrusion is important for adversary to be quick in next intrusion and cost effective. Furthermore, campaign analysis can be very important to identify the adversary's target person or technology [2]

IV. Technologies used in the framework

Although explaining Hadoop and related technologies in details is out of scope of this paper but we provide a brief overview of technology terms that we use in this paper.

a) Hadoop: Apache Hadoop is a framework that allows distributed processing of large collection of data using cluster of computers each having local computation and storage [10]. Hadoop provides high availability, fault tolerance and faster processing speeds of large (structured, semi-structured or un-structured) data sets even with cheap commodity hardware.

Two main modules that Hadoop provides are HDFS and Map Reduce. HDFS is Hadoop distributed file system, which distributes the files across the cluster to provide high-throughput & fault tolerant access. Map Reduce is a programming model for distributed data processing. [10, 11]

b)Hive: It is a data warehouse system for Hadoop, it provides SQL like language HiveQL which becomes comfortable to start working, as SQL knowledge is widespread [12].

c)Pig: "Apache Pig is a platform for analyzing large data sets that consists of a high-level language for expressing data analysis programs, coupled with infrastructure for evaluating these programs" [12].

d)Flume: "Apache Flume is a distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of log data" [7]. It helps to transfer data fault tolerantly from different applications to Apache Hadoop's HDFS.

e) OSSEC: "OSSEC is an Open Source Host-based Intrusion Detection System that performs log analysis, file integrity checking, policy monitoring, root kit detection, real-time alerting and active response [6]."

V. Proposed framework

We are going to address Targeted Attack intrusion management using kill chain approach and test its efficiency using Big Data technologies.

Our proposed framework aims to provide practical implementation to kill chain reconstruction, synthesis and campaign analysis as explained above in this paper using a Hadoop Cluster and Malware Analysis Lab. This framework can be help for management of intrusions for forensics analysis of targeted attacks.

The idea behind using a Hadoop cluster becomes clear when we aim to use large amount of security logs(semi or unstructured logs in text files) from different sources (distributed HIDS, NIDS, server ,mail and etc) collected in huge time frame (1-2 year or more). As explained above, targeted attackers persistently attack on their target environment therefore; using Hadoop cluster gives an advantage for extracting useful information from a huge log data set for campaign analysis.

To simplify the framework we divide it into 5 modules namely, Logging Module, Log Management Module, Malware Analysis Module, Intelligence Module and Control Module.

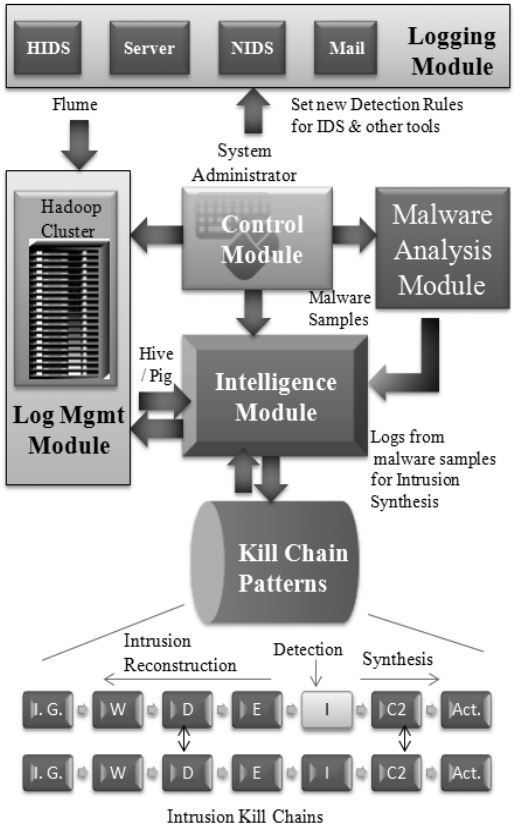


Figure2. Overview of Complete framework

A. Logging Module

This module consists of collecting logs from different sources such as distributed HIDS (Host intrusion detection system) and NIDS (Network intrusion detection system), Web Server Logs, Mail logs and etc. All these logs are generally text heavy semi structured or unstructured data. The rules for IDS s are written by the System administrators who has his hands on control module , these rule set will vary according to the situation and be learned from the Analysis of kill chains.

B. Log Management Module

This module consists of Hadoop distributed file system. The log data sets collected by the logging module are moved into HDFS and can be pre-processed here according to the needs of Intelligence module that would access Hadoop for information extraction. Services like Apache Flume can be used to move such large data into HDFS. Size of Hadoop cluster can vary according to the needs of the organization and size of data to be processed.

C. Intelligence Module

This is the main component of the framework which is responsible for construction and completion of kill chains, correlation between kill chains for campaign analysis and making new rules for Logging module. The suggested rules by Intelligence module can be further checked by administrator and then implemented.

Firstly, Intelligence module will identify the kill chain phase for each trigger event and start with construction of kill chains and as soon as it gets the detected suspicious malware information it will alert the malware analysis module for analysis of suspicious sample.

Trigger events: The trigger events are the events on which the Intelligence module will initiate the kill chain construction for the suspicious events occurred. Trigger events can be rule based or a system administrator input. Generally, a trigger could be a NIDS or HIDS high risk alert and can also be obtained for example: as the deviated behaviours from the Apache server logs after Clustering [3].

As explained earlier that whenever a certain trigger event is detected and found to be a phase of a possible kill chain, then one can be sure that prior events have been already occurred, on this basis Intelligence module can start relating events and identifying the previous phases.

Furthermore, intrusion synthesis can be done after getting log information from the malware analysis module about the malicious delivered payload.

After the intrusions kill chain in completed it can be further analysed and correlated with other kill chains to get more information about the adversary's target and attacking techniques. This process is for the campaign analysis as explained before.

In case, there are some kill chains which could not be completed then, either they are false positives or there is a need of detailed analysis; such cases should be reported to the system Administrator. System administrator can manually decide what type of actions should be taken about the situation. At any stage of kill chain construction and analysis, system administrator can correct or change any information, to improve the automated process of Intelligence module.

Another interesting property of this framework is that it a "Self Feeding framework" as it extracts information from the given resources and from the extracted information it extracts more information to produce highly precise results.

D. Malware Analysis Module

Malware analysis module consist of a malware analysis virtualized Lab Environment (virtual machines configured with different types of event logging tools). One of the most important threat component of targeted attack is malware,[4] therefore Administrator or an Analyst would need detailed knowledge about malware analysis. Although explaining malware analysis in detail is out of scope of this paper, the primary approaches for malware analysis are Code Analysis and Behavioural Analysis [4]. There are varieties of tools that help to perform such analysis of executables. Selection of tools for the analysis depends on the analyst. The primary goal of this module is to examine the suspicious sample. In case found malicious, detailed analysis should be performed and corresponding behavioural log information should be returned to the intelligence module to complete the kill chain synthesis.

E. Control Module

Using control module, the administrator governs the complete framework. The administrator is capable to drive the system in the direction he wants the system to proceed his investigation. This explains how the administrator can use this framework for digital forensic purposes against targeted attacks.

The control module is capable to control all other modules namely, Logging module for setting rules, Log management module for managing the cluster and formatting of logs, intelligence module for maintain the investigation in the right direction and the Malware analysis module for examination of the suspicious samples

VI. IMPLEMENTATION AND RESULTS

In this section we discuss our primary experiments and results about kill chain reconstruction based on search/ correlation algorithm programmed in Java to work with Hive using Hive thrift service on Hadoop.

A Hadoop Cluster was implemented in an academic environment using 5 nodes, which were powered by Intel® Core™ 2 Duo CPU E4500 @2.20Ghz×2 with 2GB of RAM, 80GB Hard Disk, 32bit machines forming a homogeneous cluster. As according to some sources it is not advisable to have heterogeneous cluster therefore, for a better performance analysis of our experiments we preferred to not increase cluster and to let it homogenous as no more same configuration machines were available. A Fast Ethernet Switch was used for the networking within the cluster nodes.

As, persistent targeted attacks are very rare in academic environments therefore, we collected Apache logs, OSSEC logs, Snort logs & mail logs generated at our university for primary experiments and these logs were further simulated according to targeted attack scenarios to perform the tests for Intrusion Reconstruction .

Simulation of a targeted attack using Targeted malicious Email.

A scenario of Targeted phishing Email was created, where attackers sent a phishing Email with attached malicious pdf to two university employees. The Email was well crafted and disguised as an authorized Email from ICoFCS 2013 Conference about invitation to Call for papers. When the pdf is downloaded and opened a benign pdf is extracted and showed to the user while another hidden malicious executable was extracted named as wp8.exe . Corresponding Log entries were simulated in the logs and are put into Hadoop. The intelligence module program was allowed to run over them. In the following we show how our program responded to such events

On June 10, OSSEC detects a malware getting installed into one of the hosts. The log entry about this event is fed into the intelligence module as a trigger event. Upon the reception of the trigger event by Intelligence module, intrusion reconstruction function is invoked that tags it to the Installation phase of kill chain and proceeds as already programmed for this kill chain phase.

Next, it searches in the logs for the location of the malware executable detected by OSSEC, after getting the location "C:\ Users\Master-Infoway\Documents\wp8.exe"; the intelligence module runs another query about timestamp and application that created this executable in the logs of past few months (considering the case that some malwares are intelligent enough to become dormant for some duration of time).

It finds out that the executable "wp8.exe" was created on May 25 upon execution of a pdf. "icofcs.pdf" located at "C:\ Users\Master-Infoway\Desktop" and created on May 25.

This time intrusion reconstruction function starts searching for delivery phase of this kill chain. Delivery is generally made by drive by download, targeted malicious email or USB [Hutchins , Amin & Cloppert 2011]. Finally it searches in mail logs and finds that icofcs.pdf was an email attachment to two employees of the university.

Table 1: Kill Chain Analysis

Info. Gathering	Mailing List ,ICoFCSWebsite
Weaponization	Malware Analysis Lab
Delivery	adm@icofcsconference.com ip : 161.xyz.pq.35 Sub: ICoFCS conference 2013 ICoFCS2013.pdf
Exploitation	0-day PDF
Installation	Malicious File detected " wp8.exe"
C2	Left for intrusion Synthesis phase
Actions	-

This completes the intrusion construction of the Kill chain using Hadoop and Hive queries accessed using our Java Program. Total number of log records fed into Hadoop were 7,049,627 and 5 Hadoop nodes with configuration mentioned above processed it completely in 2 minutes and 12 seconds .Using 5 node Hadoop cluster, we were able to process huge amount of semi-structured logs, Hive queries run the map reduce on Hadoop and the tasks are distributed across the cluster, finally quickly fetching the results.

Comparision with other Log Analysis Technologies

Our framework is completely based on hadoop platform which uses distributed processing of logs but generally all other popular log analysers/intrusion detection systems such as Snort work only on standalone machines. Thus, our framework takes an advantage of utilizing capability of Cluster of machines and bring results faster in comparison to standalone machine based systems. Additionally, our system has no limits to log data size as cluster sizes can be increased dynamically but the capacity of a standalone machine are fixed.

Snort is capable of giving an alert event from log data but our framework gives an alert on finding a pattern of events that form an Intrusion Kill chain.

In general sense our framework is only a log analyser which detects the presence of Intrusion kill chains but technologies such as Snort are intrusion detection and prevention technologies. For example our framework is not capable to detect atomic malicious events.

VII .CONCLUSION AND FUTURE WORK

In this paper, we discussed a framework based on Hadoop for dealing with Intrusions performed by Targeted threat adversaries, using concept of Intrusion kill chains. We simulated a realistic scenario of targeted attack and our framework could detect it using intrusion reconstruction through different sources of semi-structured logs.

The proposed framework has some of the major contributions such as:

- This framework can help in identification of targets, techniques and tactics of the adversary which is useful for forensics analysis of targeted attacks.
- Kill chain construction can help the administrators to build IDS rules to strengthen their posture of defence.
- Other than detecting and analysing targeted attacks, this framework can also help its administrators to identify unknown vulnerabilities (also called 0 day vulnerabilities) in their system that the attacker used.

Although, this framework is greatly promising and well structured for dealing with targeted threats but still it contains some limitation such as following:

- This framework uses Hadoop for managing the log files, while Hadoop is a perfect framework for working with unstructured and semi structured text heavy data sets but, it is not good fit for real time applications and small amount of data set therefore, this deficiency of Hadoop makes the framework slower in response for small data sets in comparison to other relational database systems.
- Classifying of kill chains from common malware to targeted malware, this framework will give some effort for administrator to differentiate a target malware or a common unsophisticated but on the other side, this can be used to analysis of common malwares also.

Using our experiments we successfully tested our framework, which uses Intrusion kill chain method and big data technologies (such as Hadoop HDFS and map reduce for efficient log management and faster information retrieval from semi-structured big data). Finally, according to our analysis, using intrusion kill chain method and Big Data technologies can be the future trend for Intrusion detection and prevention systems applied to targeted threats such as Targeted malicious Emails.

In future, we plan to implement more typical targeted threat scenarios and analyse them with bigger homogenous Hadoop cluster and evaluate its efficiency. We also intend to implement automated correlation of Kill chains for Campaign Analysis.

REFERENCES

[1] A.K. Sood, R.J. Enbody " Targeted cyber attacks: A Superset of advanced persistent threats" Security & Privacy, IEEE Volume 11 , Issue 1 ,pages 54 – 61, Jan.-Feb. 2013

[2] Eric M. Hutchins, Michael J. Cloppert, Rohan M. Amin, "Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains" ,6th International Conference on Information Warfare and Security(ICIW2011) http://www.lockheedmartin.com/content/dam/lockheed/data/corporate/documents/LM-White-Paper-Intel-Driven-Defense.pdf

[3] J. Therdphapiyanak, K. Piromsopa, "Applying Hadoop for log analysis toward distributed IDS", ACM International Conference on ubiquitous Information Management and Communication (ICUIMC) 2013, Article No. 3

[4] Frankie Li, A Atlasis , " A Detailed Analysis of an Advanced Persistent Threat Malware" SANS Institute InfoSec Reading Room,2011 http://www.sans.org/reading_room/whitepapers/malicious/detailed-analysis-advanced-persistent-threat-malware_33814

[5] N. Falliere, L.O. Murchu, and E.Chien "W32.Stuxnet Dossier" Version 1.4 (February 2011)

[6] OSSEC "http://www.ossec.net/"

[7] Flume http://flume.apache.org/

[8] J. Howes, J. Solderitsch, I. Chen & J. Craighead, "Enabling trustworthy spaces via orchestrated analytical security", ACM, CSIIRW 2013, Article No. 13

[9] MacDonald, Neil, 2012, Information Security is Becoming a Big Data Analytic Problem, Gartner, (23 March 2012), DOI= http://www.gartner.com/id=1960615

[10] Apache Hadoop Project http://hadoop.apache.org/

[11] Tom White "Hadoop: The Definitive Guide", 2009, 978-0-596-52197-4

[12] Hive"http://hive.apache.org/", Apache Pig http://pig.apache.org/

[13] "Hadoop Tutorial from Yahoo!", Module 7: Managing a HadoopCluster http://developer.yahoo.com/hadoop/tutorial/module7.html#machines

[14] J. P. C. L. da Costa, E. P. de Freitas, A. M. R. Serrano, and R. T. de Sousa Jr. "Improved Parallel Approach to PCA Based Malicious Activity Detection in Distributed Honeypot Data," International Journal of Forensic Computer Science (IJoFCS), 2012

[15] B. M. David, J. P. C. L. da Costa, A. C. A. Nascimento, M. D. Holtz, D. Amaral, and R. T. de Sousa Jr., "A Parallel Approach to PCA Based Malicious Activity Detection in Distributed Honeypot Data," International Journal of Forensic Computer Science (IJoFCS), 2011

[16] B. M. David, J. P. C. L. da Costa, A. C. A. Nascimento, M. D. Holtz, D. Amaral, and R. T. de Sousa Jr., "Blind Automatic Malicious Activity Detection in Honeypot Data," The International Conference on Forensic Computer Science (ICoFCS) 2011, Florianópolis, Brazil

DOI: 10.5769/C2013005 or http://dx.doi.org/10.5769/C2013005

Cybercrime Investigation Challenges for Gulf Cooperation Council Governments: A Survey

Nasser Alalwan, Ahmed Alzahrani, Mohamed Sarrab

Abstract—***Computer crimes are criminal activities that involve the use of information technology. Modern life is increasingly relying on information technology; however, the more sensitive the information, such as government intelligence, credit card information or personal information the more important it is to improve the systems of computer crime investigation. Crime investigation helps detect illegal access to computer systems with the intent of deleting, modifying, damaging or stealing the data and information. Such activities may cause financial damages and may also result in publishing or destroying secret or confidential information. This survey study focuses mainly on highlighting the main challenges of the Gulf Cooperation Council (GCC) cybercrime investigation system and the need of information security laws in these particular countries. To achieve that personal knowledge, literature review and case study are used to complement and maintain the authors' statement about the challenges of cybercrime Investigation in GCC.***

Keywords- *Cybercrime; Computer Crime; Cybercrime investigation procedure.*

I. Introduction

The revolution of information technology and communication has increased the use of computer systems and networks. The marriage between the information technology and communication has increased the rate of the computer crime all over the world. In addition, with the increase of web applications, the sensitive and critical information become uncontrolled and more vulnerable. Computer crimes or any other digital crimes are criminal activities that involve the use of modern information technology. All aspects of our life have become increasingly relying on modern information technology. The more important the information, such as government intelligence, credit card information and any other private information the more important it is to improve the systems of computer crime investigation to detect any unauthorized access to a computer system with intent of deleting, changing or damaging computer data and information. An illegal access to the data or information in computer systems more potentially causes a financial damage in case of the manipulation or destroy of confidential information and secret or critical information. Computer crimes involve different types of activities such as machine misuse, digital frauds, system interference as well as unauthorized access and they might not necessary include any type of physical damage. Therefore, these different types of computer crimes and the rapid increase on the use of information technology make the work of computer crime investigator became very hard to detect and prevent any type of computer crimes. The issue is that the Arab gulf countries are in different stages of implementing electronic management such as E-government, E-commerce and E-business. For example, 51 cases of cyber-attacks were recorded in 2009 by the Telecommunications Regulatory Authority (TRA) targeting the UAE's information technology infrastructure which warranted the agency to issue the "devastating" effect [1]. The research value is raised from the rapid increase of the cybercrimes all over the world. Some countries have a good progress in the investigation procedures as well as in the law of cybercrime. Their investigators have a good experience in different types of computer crimes and the law in these countries protects the personal and government information. In each of these countries, the law provides criminal penalties for any identifying of the cybercrimes. This may cause other countries as a worth area for the criminal to steal or destroy any type of information. The purpose of this paper is to survey the cybercrime investigation challenges for GCC governments. To achieve that personal knowledge, literature review and case study are used to complement and maintain the authors' statement. The reminder of the paper is structured as follows. Sections II and III provide brief background about cybercrimes and introduce the cybercrime concept. Section IV discusses the state of information security in GCC. Section V and VI provide case study and its discussion respectively. In section VII the paper highlights the cchallenges of GCC cybercrime investigation system.

II. Backgroung

The advent of telecommunication and computer technologies has increased the number of users leading to an increase in cybercrimes such as hacking that causes a lot of damage via the computer. To protect data and information from cybercrimes, it is necessary to create a database to prevent unauthorized use based on confidentiality [2]. Cybercrime investigation is mostly similar to traditional crime investigation. Both have similar investigation procedure including (inspection, collecting evidences, investigations and analysing evidences). Moreover, in both types of crimes the investigators seek to answer the following key questions [3]. *What, When, Where, How, Who and Why?* However, cybercrimes have specific areas to deal with, such as computer machines, network, storage devices and other communication Medias. In addition, in computer crimes a huge record is very necessary to discover any available devices' manuals or any logging files [4, 5]. The most important step in cybercrimes is the strategic plan which is a long term plan that is concerned with national data network infrastructure [6]. Another important factor in cybercrimes investigation is the investigation team.

Creating investigation team with a good experience in computer machines, a data network and software tool is very hard to have one team with all these skills. The investigations team involves the team leader and the team members. The team leader should have a good experience in forensic and cybercrimes investigation [2]. Digital Forensics is the science of identifying, collecting, preserving, documenting, examining, analyzing and presenting evidence from computers, networks, and other electronic devices. It generally classifies and considers the digital evidence in a way that is officially acceptable by courts. Digital evidence contains the collection of several procedures of digital data [6, 7, 8]. Besides the strategic planning, the investigation team and the team leader are the cybercrimes' digital evidence. The digital evidence is an important part of the cybercrime investigation procedure which might include hardware, software, manuals, or phone numbers [9, 10, 11]. Many works have been done on the cybercrimes investigation [12, 13, 14] but to the best of our knowledge no work has been done in this particular area generally Gulf Cooperation Council countries.

III. Conceptualizing Cybercrime

Computer crimes can be defined as any criminal activities that are committed against a computer hardware machine. In computer crimes, the computer machine is used as a target of any type of criminal activities. The types of crimes are not only related to the data, information, software or any other program applications. The criminal activities in computer context is often refers to the computer functions; such as electronic mail and instant messaging services, social media applications, file transfer facilities and audio or visual conferencing programs, … etc. However, cybercrimes are any criminal activities committed using the computer, Internet or other electronic machines as the medium, in violation of existing laws. In other words cybercrimes can be defined as a type of crime that involves the use of computer technology, and for which penalties already exists under existing legislation. Fundamentally, there are no difference between generic individual crimes such as extortion, forgery, fraud, theft, impersonation and their cyber analogues. The cybercrime can also include the use of digital resources to commit any type of traditional crimes such as theft of identifiable card information and other forms of proprietary information or property in both digital and physical form [3, 5, 13, 15].

IV. State of information security In GCC

The state of cybercrime in the GCC is different from other countries all over the world, in which the state of information technology security in GCC and all Middle East regions is affected by many factors such as growth of IT user, IT infrastructure, and poor IT security system, lack of regulation and training of law enforcements.

- **Growth of IT user.**

With decrease cost of the broadband services in the region, the number of new IT users are growing daily and faster than other countries all over the world. According to Internet World Stats, Internet use in the Middle East had reached 2.5% of the total worldwide use by December 2007. Middle East use from 2000 to 2007 increased by 920.2% compared to 259.6% for rest of the world! [17, 18]. Internet World Stats 2013 shows that he growth rate of internet users in Middle East from 2000 to 2012 is 2,639.9%. Table 1 shows the world internet usage and population statistics in June 2012.

Table. 1. World Internet Usage And Population Statistics June 30, 2012

World Regions	Population (2012 Est.)	Internet Users Dec. 2000	Internet Users Latest Data	Growth 2000-2012
Africa	1,073,380,925	4,514,400	167,335,676	3,606.7 %
Asia	3,922,066,987	114,304,000	1,076,681,059	841.9 %
Europe	820,918,446	105,096,093	518,512,109	393.4 %
Middle East	223,608,203	3,284,800	90,000,455	2,639.9 %
North America	348,280,154	108,096,800	273,785,413	153.3 %
Latin America / Caribbean	593,688,638	18,068,919	254,915,745	1,310.8 %
Oceania / Australia	35,903,569	7,620,480	24,287,919	218.7 %
WORLD TOTAL	**7,017,846,922**	**360,985,492**	**2,405,518,376**	**566.4%**

This huge growth of number of users in Middle East has made the Internet more popular, supports the meaning of communication and opens a new online business opportunities. However, this growth in the number of users has increase the potential for IT abuse. Due to the lack of IT security policy enforcement, many Internet users have become victims of cybercrime attacks.

- **Information Technology infrastructure.**

The growth of overall investment in IT infrastructure in the Middle East is extensive, especially in GCC; but this IT infrastructure investment needs to do more in securing IT network infrastructure. In fact, over the past few years banks in the region have invest considerable budget to control the cybercrimes and securing online banking transactions. But, most banks in the region are still vulnerable to phishing attacks and hackers, which indicate that the GCC should invest more in IT security enforcement.

- Poor IT security system.

In all countries in Middle East region, especially in GCC there is a significant lack of security awareness and security policy enforcement among IT and online users. Comparing security awareness and security policy enforcement in the GCC to the other major players such as USA, Europe, China and Russia, there is a big gap between them in the less effort being made to raise awareness and security policy enforcement among IT users in GCC.

- **Lack of regulations and training of law enforcements.**

From the previous mentioned points, it becomes obvious that the GCC lack of regulations and training of law enforcements. These countries need strong security awareness training, targeting native speakers to educate users,

employees and law enforcers to understand the dangers and risks of attacks and hackers [17].

V. Case Study

The case study is about Saudi Arabian Oil Company (Saudi Aramco) where it confirmed the attack of its network occurred due to virus infection. Saudi Aramco is one of the largest energy and petroleum companies all ov ` e r the world. This virus attack could lead information stealing, destroy or any other financial damage. In that time, Symantec announced the discovery of a new malware called "W32. Disttrack" or "Shamoon". The malware infects a PC, steals certain data, send the data to another infected PC inside the compromised network and then overwrites the PC's Master Boot Record, which makes the system useless. The way this malware works might be linked to the Wiper malware which infected Iranian oil terminals in April 2012. The Wiper malware is also considered new variant of Flame as the investigation of the Wiper led to the discovery of Flame, according to Kaspersky Lab.

Kaspersky also provides new analysis of how Shamoon is coded. This type of malware might be used to physically access to a computer device that is connected to Saudi Aramco network then data and information propagation started. The infected device might not be inside Aramco but it can be connected with the company remotely from any other place. In this situation, Aramco needs to conduct thorough investigation to figure out from where this malware accessed their network not only focusing on the recovery from that attack. To identify the identity of the attacker a lot of work and collaboration needs to be done between GCC together and with other countries over the world specially the major players of the dangerous game such as USA, Europe, China and Russia [17, 19]. However, the important question after this Aramco attack is: Are we as GCC ready for the 21st century threats?

VI. Discussion

The paper focusses in this case of Saudi Aramco were the company's machines are infected with the Shamoon virus which requires Saudi Arabia to co-ordinate typical of state-sponsored attacks, and the targeting of critical infrastructure shortens the list of suspects. Another example of hacker attach in the region is the UAE's e-government sites that have been attacked by hackers, which caused financial loss and propagation of secret and confidential information to the public; Moreover, the famous channel news Al-Jazeera website is another example of a big name that has been hacked in the region.

In fact the investments in IT infrastructure have increased the value of e-business and e-governments and have increased and created great opportunities for small and medium businesses in the GCC, which helps with the unemployment problem. This lack of security awareness and security policy enforcement is one of the biggest problems inside IT companies in the GCC. Further, IT users and decision makers in GCC are not aware of the growth of the cybercrime problems. Poor security policy enforcement means that investments and chance to fight in the level of cybercrime are minimal which leaves the business across the GCC vulnerable to cybercrime or online attacks.

VII. Challenges Of GCC Cybercrime Investigation System

Despite the huge potential and many benefits that could be gained from improving cybercrime investigation system. There are still many challenges that face the improvement of cybercrime investigation including: Lack of comprehensive study on the main influencing factors of cybercrime investigation system in the GCC countries: to the best of our knowledge, there is no research study on the main factors/ barriers that influence the work of cybercrime investigators in GCC countries. One of the important factors is the cultural and social considerations such as user behaviors and the lack of knowledge in exchanging cybercrime between the gulf countries. Moreover, these countries are still in the earlier stages of their electronic management implementation, consequently, these countries might be seen by criminal to as a worth area to steal or destroy sensitive information. The following are the main challenges of the GCC cybercrime investigation system.

- Legislation, which may include the criminal offences, requirements to open an investigation, evidences, involvements of and shire knowledge with prosecutor and judge.
- Dedicated Units consist of legal framework, competence offices, field offices, trained and skilled officers, other necessary equipment and software applications.
- Criminal investigative procedures should allow computer access, internet interception, computer search, data preservation and supports procedure complaint and any other cybercrime related information and reports the case to the prosecutor. The procedure should support an investigation, surveillance, identifying IP and phone users and monitoring of phone conversation, internet data.
- Private sector cooperation should assist and exchange of information with the government related to cybercrime victim, evidence, knowledge, training, legislation, protocols, phone companies, banks, ... etc.
- International cooperation, countries all over the world should collaborate in exchanging of information related to cybercrime victims, evidence, public and private, mutual assistance request, contact points, joint investigation.
- Responsibility, In GCC there is big dilemma when discussing the cyberspace related laws. Due to the fact that there is no established cybercrime legislation in the region and there is absence of a government agency or department to be responsible for drafting or dealing with

cyber laws. It can be found that there are many agencies in the government might be involved in cyber related laws such as copyright, E-commerce, E-government, domain name registration and cybercrime. For example more than one government departments can involve in such situation like: Ministry of Interior, Ministry of ICT, Ministry of Justice, Ministry of Interior, Central Bank and even Intelligence and Defense departments. The issue is that any of the above mentioned authorities could claim responsibility of such laws which in fact the big challenge for GCC governments when discussing and drafting cybercrime law. It is very important to establish dedicated department in the government structure to deal with cyber laws. The UAE for example just started and appointed dedicated courts for cybercrime cases.

In addition, the most important challenge in the cybercrime investigation procedure is to understand the criminal activity and prove it [16].

VIII. Conclusion

The GCC are still in the earlier stages of their electronic management implementation. For these reasons, these countries' governments and organizations are concerned about the quality of their investigators and the cybercrimes investigation procedure itself. There is no doubt that the concept of cybercrime is feasible but this paper discussed the main challenges of the GCC cybercrime investigation system and highlighted the need of information security laws in these particular countries. This paper also focuses in the lake of cybercrime investigation experiences exchange between GCC.Therefore, there is a need for a comprehensive study of cybercrime investigation to address all issues and improve the cybercrime investigators procedure. This study should include an exploration of the work of the cybercrime investigators and identifying the types of cybercrimes in GCC. Finally, identifying the skills needed to improve the procedure of cybercrime investigation in GCC and according to their local user behaviour. Finally the study should end up with single information security law in GCC.

References

[1] B. Jay and B. Lubna, Cyber gangs on the prowl in UAE, gulfnews.com Al Nisr Publishing LLC. Online 2011.

[2] E. Timothy, "The field guide for investigating computer crime, part eight: Information discovery," Symantec, 2001.

[3] A. Bahar, Computer crime investigation, Master's thesis, De Montfort University, 2010.

[4] J. Richter, N. Kuntze, and C. Rudolph, "Securing Digital Evidence," in the Fifth International Workshop on Systematic Approaches to Digital Forensic Engineering. 2010.

[5] I. David and S. Karl S, Computer Crime: A Crime fighter's Handbook, OReilly Associates, Inc, Sebastopol, CA, 1995.

[6] V. Loia, M. Mattiucci, S. Senatore, and M. Veniero, "Computer Crime Investigation by means of Fuzzy Semantic Maps," in IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies, IEEE, 2009.

[7] S. Simundic, S. Franjic, and T. Susic, "Databases and Computer Crime," *52nd International Symposium ELMAR-2010*, pp. 195-201, September 2010

[8] H. Yousef, and A. Iqbal, "Digital Forensics Education in UAE," in the 6th International Conference on Internet Technology and Secured Transactions, Abu Dhabi, UAE, 2011.

[9] A. Sammes, and B. Jenkinson, Forensic computing a practitioners guide, Springer Verlag, 2000.

[10] E. Casey, Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet, Academic Press, 1st edition, 2000.

[11] E. Casey, "Error, uncertainty, and loss in digital evidence," International Journal of Digital Evidence, vol. 1, no. 2, 2002.

[12] S. McLean, "Basic considerations in investigating computer crime, executing computer search warrants and seizing high technology equipment," 14th BILETA Conference: CYBERSPACE. Crime, Criminal Justice and the Internet. 1999.

[13] C. Prosise, and K. Mandia, Incident Response: Investigating Computer Crime, Mcgraw-Hill Osborne Media, 2001.

[14] P. Stephenson, Investigating Computer Related Crime, CRC Press, Boca Raton, Florida, 1999.

[15] Warren B. Chik., (2011) "Challenges to Criminal Law Making in the New Global Information Society: A Critical Comparative Study of the Adequacies of Computer-Related Criminal Legislation in the United States, the United Kingdom and Singapore", Available online at www.law.ed.ac.uk/ahrc/complaw/docs/chik.doc visited 28 October, 2011

[16] S. Kabay, Computer Security handbook, Wiley, 2002.

[17] M. El-Guindy, Saudi Aramco cyber attack, are we ready Net Safe. Middle East, 2012. Availbe from http://netsafe.me/2012/08/27/saudi-aramco-cyber-attack-are-we-ready/#more-535

[18] Miniwats Marketing Group, Internet Seven Year Growth, Second Quarter Stats, Internet World Stats News, no. 028 - July, 2007.

[19] C. Bronk and E. Tikk-Ringas, "The Cyber Attack on Saudi Aramco", Survival: Global Politics and Strategy, vol. 55, no. 2, pp. 81–96, April 2013.

DOI: 10.5769/C2013006 or http://dx.doi.org/10.5769/C2013006

Emprego da Engenharia Reversa para caracterização do *modus operandi* das máquinas caça-níqueis quanto à prática de jogo de azar ou outras fraudes

Cleverson Esteves da Silva, Galileu Batista de Souza, and Ricardo Zelenovsky

Abstract—**Despite gambling be forbidden in Brazil since 1941, the growing access to computer equipment brings a new kind of slot machines that, instead of using a specific hardware group, has its environment simulated by a computer program and is built with common and obsolete hardware items. Due to the uncertainty about internal behavior of the management programs that controls these equipments, the forensic exams have not been as conclusive whether they offer properties to allow player influence on final result. Using reverse engineering techniques, this study presents a set of evidences about gambling and frauds in the "Halloween" machines applications. Furthermore, in order to instruct future analyses on machines of different class, here is shown the methodology used to obtaining the necessary information by forensic exam, that could be adapted to other situations involving same kind of applications not covered in this study.**

Palavras-chave: slot machines; gambling; reversing; forense computin; fraudg.

I. Introdução

Assim como a tecnologia tem invadido o cotidiano dos cidadãos de bem, de forma acelerada e irreversível, trazendo comodidade e celeridade para tarefas anteriormente dispendiosas, a criminalidade tem se especializado e utilizado a tecnologia ao seu favor [1].

Algumas modalidades de crime praticadas através de equipamentos eletrônicos encontram uma barreira na análise interna das instruções lógicas responsáveis por gerenciar tais recursos. Um exemplo notório são as máquinas caça-níqueis (videobingo), cujas constantes apreensões têm sido noticiadas nos meios de comunicação em todo o território nacional [2] [3], representando uma grande demanda de trabalho pericial aos órgãos de perícia oficial estaduais e Federal. Porém, um suposto controle por parte do apostador é o principal argumento de defesa para impedir que tais equipamentos sejam enquadrados no dispositivo legal que trata dos jogos de azar, alegando que não se trata de sorte, mas sim de técnica.

As máquinas de videobingo utilizam-se dos principais componentes de microcomputadores de plataforma *personal computer* como placa-mãe, microprocessador, memória e dispositivo de armazenamento (muitas vezes obsoletos e de baixo custo), visto que estes são de fácil aquisição. Tais equipamentos não necessitam de recursos especiais e a aplicação, responsável por simular o ambiente do jogo, é armazenada em um disco magnético ou dispositivo de memória semicondutora. Por usarem componentes padrões da indústria, seus elementos poderem ser facilmente transportados em separadamente sem que desperte qualquer suspeita sobre a sua destinação final.

A simplicidade e padronização do *hardware* contrastam com a incerteza do funcionamento interno dos aplicativos das máquinas de videobingo. Diante da carência de trabalhos que analisem em profundidade o comportamento interno dos programas computacionais que simulam o ambiente de uma máquina caça-níquel, bem como da escassez de documentação técnica sobre os elementos de *software* ou sobre os itens específicos de *hardware*, supostamente atribuídos a suspeita de ilegalidade que paira sobre esses equipamentos. O presente trabalho analisa em profundidade uma família de máquinas caça-níqueis visando caracterizar o comportamento do seu software controlador a fim de caracterizá-lo como jogo de azar e se o mesmo trabalha com parâmetros que permitem a manipulação do resultado do jogo.

Dos equipamentos recebidos a partir do ano de 2009 na Seção de Criminalística de Ji-Paraná (RO) para análise, somente integram o objeto de estudo do presente trabalho as máquinas que apresentam cartão de memória semicondutora conectada ao *slot* IDE da placa-mãe e que se mostraram operantes à época do recebimento. Aquelas máquinas, cuja aplicação está gravada em uma placa de circuito impresso que substitui a placa-mãe ou em placas que são inseridas em *slots* PCI ou ISA, foram descartadas da amostra por corresponderem a uma geração de equipamentos que está cada dia mais obsoleta devido à especificidade do *hardware*.

O universo de estudo do presente trabalho consiste de 08 (oito) exemplares de máquinas caça-níquel, todas pertencentes à categoria denominada "Halloween", que tem predominância absoluta no Estado de Rondônia e apresenta a mesma forma externa de interação e a mesma forma interna de exibição de resultados de máquinas analisadas por órgãos periciais de outras regiões brasileiras. Apesar da pequena quantidade de máquinas, o universo de estudo oferece ao todo 18 (dezoito) opções de jogos aparentemente distintos, mas com jogabilidade semelhante.

A realização de um experimento estatístico, traçando um perfil do seu funcionamento após exaustivas execuções das máquinas de videobingo [4], seria inviável, pois nas máquinas caça-níquel convencionais existe somente um elemento de interação com o equipamento, representado pela alavanca, ao passo que as máquinas que se pretende analisar no estudo proposto possuem diversos elementos de interação, representados pelos botões do painel, todos com finalidades específicas, o que aumenta de forma exponencial a quantia de combinações necessárias para se traçar um perfil de funcionamento confiável.

Mais precisamente, o estudo apresentado neste artigo tem por objetivos:

- Traçar um perfil de comportamento de tais máquinas com base no modo de operação inferido a partir dos elementos materiais coletados com técnicas de Engenharia Reversa. Assim, equipamentos similares podem ser examinados mais rapidamente por Peritos Criminais de todos os Estados da Federação. Ademais, poderão eles ser conclusivos em seus laudos no que se refere à prática de jogo de azar e outras fraudes, encerrando definitivamente a discussão que vem se prolongando no âmbito forense desde o surgimento de tais máquinas;
- Elencar pontos de similaridade de jogos aparentemente distintos, possibilitando que versões do aplicativo principal de máquinas tipo "Halloween" não analisadas neste trabalho sejam inferidas quanto à prática da modalidade de jogo de azar ou fraudes diversas, tanto através da comparação direta do arquivo executável, como através da identificação das características operacionais e comportamentais apresentadas pelo presente estudo;
- Apresentar uma metodologia de análise que auxilie o exame pericial de equipamentos similares apreendidos em todo o território Nacional, fornecendo diretrizes que visem à obtenção de informações relevantes à atividade pericial, inclusive em aplicações da mesma natureza que não pertençam à família "Halloween".

A organização do trabalho é como segue. Na seção II são apresentadas considerações acerca da definição dos jogos de azar a luz da legislação. A seção III compreende a análise interna dos jogos através da Engenharia Reversa, nas suas modalidades de análise *online* e *off-line* e tem como foco a comprovação do enquadramento como jogo de azar. Por fim, a seção IV exemplifica a metodologia empregada na análise, permitindo que máquinas similares sejam prontamente analisadas e que metodologia similar à descrita no artigo possa ser utilizada para os demais casos.

II. Jogos de Azar

O jogo de azar é aquele onde a única ação disponível ao jogador é aguardar o resultado processado pelo mecanismo do jogo, quer ele seja manual ou automatizado, na esperança de que seja compatível com o valor escolhido, sem que haja qualquer possibilidade de interferência direta ou indireta depois de iniciado o processo, estando o praticante da ação "a mercê da sorte" [5].

Assim, para se caracterizar um determinado jogo como sendo de azar, deverá haver aleatoriedade suficientemente capaz de impedir a previsão de resultado, mesmo que haja circunstâncias matemáticas que contribuam para o aumento ou redução da probabilidade de determinado resultado ocorrer. Apesar de os jogos de azar variarem em relação à probabilidade de ganho ou perda, seus resultados, ainda assim, dependem exclusivamente do fator sorte.

Promulgado em 03 de outubro de 1941, pelo então Presidente da República Getúlio Vargas, o Decreto Lei nº 3688/41 institui a Lei de Contravenções Penais e define em seu art. 50 §3º o jogo de azar como sendo o jogo em que o ganho e a perda dependem exclusiva ou principalmente da sorte do apostador, acrescentando ainda ao rol de tal prática delituosa, as apostas em qualquer modalidade esportiva.

No caso das máquinas caça-níqueis, devido ao fato de os equipamentos possuírem um conjunto de botões que passa ao apostador uma falsa impressão de controle, tem sido constatada uma divergência de entendimento entre as sentenças prolatadas pelo Poder Judiciário em todo o território nacional [6] quanto ao enquadramento das máquinas caça-níquel atuais, na legislação especial sobre jogos de azar.

III. Análise das Máquinas Caça-níquel

Ao examinar preliminarmente as máquinas de videobingo como um todo, observa-se que a caracterização do seu funcionamento em um aspecto amplo e geral está diretamente ligada a análises de itens específicos e de menor abrangência como as características físicas inerentes às formas de interação do usuário com o equipamento; o hardware que a compõe; os softwares que dão suporte ao hardware e ao ambiente do jogo; e o ambiente de execução da aplicação e o software aplicativo, propriamente dito, responsável por implementar toda a regra de negócio que irá determinar a resposta dada a cada interação sofrida.

Nesta ótica, a Engenharia Reversa de Software apresenta-se como importante ferramenta na análise forense de tais equipamentos, pois "(…) é executada com o objetivo de obter uma melhor compreensão de um sistema existente" e "(…) é composta de uma série de técnicas utilizadas para a descoberta de informações a respeito de um sistema de software" [7], sendo responsável pela obtenção de um modelo abstrato do comportamento do sistema [8].

A. Usabilidade

A máquina apresenta um conjunto de jogos; uma vez escolhido um, ele é aberto em tela cheia exibindo em seu pano de fundo o motivo que dá nome ao jogo (Figura 1). No terço superior da tela são exibidas as informações referentes ao "crédito", que se refere ao saldo do apostador no equipamento

e ao "prêmio", que é calculado com base na aposta que foi realizada, caso haja um resultado favorável ao apostador.

O valor do prêmio tende ao crescimento a cada resultado desfavorável. É possível multiplicar o valor da "aposta" (consequentemente do prêmio), através do produto da quantia de combinações (linhas) escolhidas, por um fator multiplicador entre 1 e 10, escolhido através do botão "aposta" e cujo valor será deduzido do crédito do apostador.

No terço inferior da tela principal, o jogo exibe o valor "acumulado" (*jackpot*) pela máquina e que corresponde à premiação especial paga em caso de resultado compatível com uma sequência pré-estabelecida. Quanto maior o valor acumulado por uma máquina, maior é o apelo para que os apostadores a escolham. Um determinado jogo, conforme a sua versão, pode apresentar um ou mais acumulados.

Figura 1. Tela principal do jogo com a opção "linha 9" ativada.

O terço médio da tela, por sua vez, apresenta a área de exibição das figuras sorteadas, a qual é composta por 05 (cinco) colunas, contendo 03 (três) figuras cada, representando uma matriz de 3 x 5, que totaliza uma quantia de 15 (quinze) quadrantes, onde são distribuídas as 10 (dez) figuras distintas que compõem o jogo, conforme o sorteio realizado.

À direita e à esquerda da matriz são exibidos valores numéricos que variam conforme a quantidade máxima de linhas (apostas) suportadas pelo equipamento. As linhas são pré-estabelecidas pelo desenvolvedor da aplicação e são sempre formadas por 05 (cinco) imagens dispostas sequencialmente na horizontal ou diagonal.

Quanto mais linhas o apostador escolher, maior será o valor deduzido do seu crédito e disponibilizado na forma de aposta na rodada corrente, já que os créditos são deduzidos na proporção de 1 (um) para cada linha escolhida, na modalidade de aposta simples ou 10 (dez) para um, na modalidade de aposta máxima.

$$Aposta = n(Linhas\ Escolhidas) \times Fator\ de\ Aposta \qquad (1)$$

Até que sejam inseridos créditos, o equipamento permanece no modo de demonstração. Depois de inseridos créditos o apostador poderá:

- escolher em quantas linhas deseja apostar, pressionando qualquer um dos botões referentes a linhas;
- exibir a tabela de premiação através do botão "tabela";
- realizar uma jogada automática, com escolha de linhas e valores de aposta por parte da própria máquina, através do botão "auto";
- optar pela aposta máxima permitida pelo equipamento, configurando automaticamente o valor máximo de linhas e a modalidade máxima de aposta, através do botão "aposta máxima";
- escolher um valor entre 1 e 10 para a modalidade de aposta, que será multiplicado pela quantia de linhas escolhidas, através do botão "aposta";
- realizar o sorteio depois de configuradas as suas opções de aposta, através do botão "jogar";
- encerrar sua sessão na máquina, resgatando os créditos existentes, através de botão afixado na lateral do gabinete, caso esta opção seja oferecida pelo equipamento.

B. Aspectos Físicos

Apesar de variar em sua forma externa, que geralmente consiste de uma caixa em madeira compensada pintada na cor preta, as máquinas de videobingo apresentam basicamente uma quantidade que varia entre oito e dez opções diretas de interação, representadas pelos botões do seu teclado, mas que se presume ter a capacidade de proporcionar diversas outras opções indiretas quando combinadas entre si.

Tais botões apresentam rótulos que fazem alusão modalidades de apostas como "cartelas" e "linhas", que apresentam opções numeradas, ou ações do apostador tais como "pagar", "aposta" e "jogar".

As máquinas de videobingo são compostas por uma placa-mãe da plataforma *personal computer,* com um cartão de memória representando o armazenamento persistente da máquina e com capacidade não superior a 512MB. Armazena o sistema operacional, a aplicação principal quanto os programas auxiliares, destinados à preparação do ambiente..

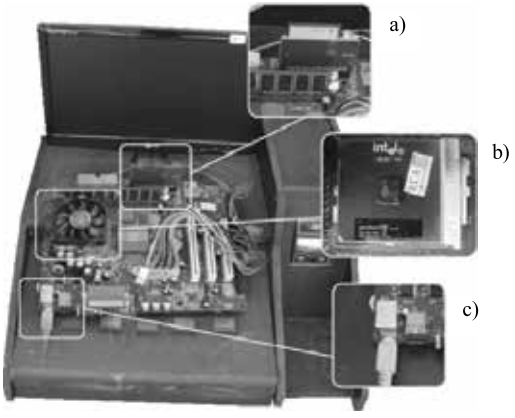


Figura 2. Composição gráfica ilustrando o interior dos equipamentos.

Durante a análise, observou-se que todos os conjuntos microprocessador pertencem à plataforma PC (*personal computer*), com predominância dos modelos Pentium III do fabricante Intel e K6 II do fabricante AMD, conforme destacado pela alínea "b" da figura 2, embora possa ser executado em qualquer outro processador da família 8086, superior ou inferior, conforme exames realizados.

A conexão do equipamento com o seu teclado customizado dá-se indiretamente através de conectores DIN 5 ou PS/2 com o auxílio de uma placa de circuito impresso, conforme destacado pela alínea "c", ou diretamente mediante a conexão de uma placa de circuito impresso, dotada de porta paralela, que recebe os valores produzidos pelos botões do teclado com o auxílio de um conector DB25.

A alimentação de créditos na máquina é realizada através de cédulas de moeda corrente com o auxílio de um dispositivo noteiro, diretamente ligado ao teclado customizado. Ele envia uma quantia *n* de pulsos, previamente estipulados para cada padrão calibrado [9] como se a tecla "P" houvesse sido pressionada *n* vezes no teclado convencional. Em suma, para o resto do equipamento, o noteiro e o teclado customizado comportam-se como um teclado convencional, com a correspondência de teclas apresentada na Tabela I.

TABELA I. Correspondência dos pinos da porta paralela

Funcionalidade	Tecla	ASCII	Hexa	Pino 1	Pino 2
Linha 1	Y	89	59	5	13
Linha 5	J	74	4A	5	12
Linha 9	H	72	48	5	11
Linha 15	M	77	4D	5	10
Linha 20	N	78	4E	5	9
Tabela	T	84	54	4	13
Auto	V	86	56	4	12
Aposta Máxima	F	70	46	4	11
Aposta	G	71	47	4	10
Jogar	B	66	42	4	8
Encerrar jogo	R	82	52	5	6
Noteiro	P	80	50	24	25

C. Ambiente Operacional do Jogo

A preparação do ambiente operacional do jogo foi caracterizada através da identificação e classificação dos arquivos presentes no dispositivo principal de armazenamento, separando aqueles que são de uso geral pelo sistema daqueles de uso exclusivo e que possuem influência direta na aplicação. Além disso, dentre os arquivos listados, foram identificados aqueles protegidos por senha ou dissimulados com o objetivo de dificultar o seu reconhecimento pelo sistema operacional e ferramentas de análise forense.

1) Identificação dos arquivos

A análise do interior dos dispositivos de armazenamento revelou que todos os equipamentos possuem em comum um conjunto de arquivos armazenados na raiz da única partição (FAT) presente no dispositivo de armazenamento. Ainda na raiz da partição há um diretório denominado "dos" e outro denominado "data", destinados ao armazenamento de arquivos da aplicação e da base de dados, respectivamente. O sistema operacional MS-DOS dificultou bastante qualquer análise, haja vista a carência de ferramentas de monitoramento e Engenharia Reversa para essa plataforma.

Ao examinar o arquivo "autoexec.bat", verificou-se que durante a inicialização da máquina, além de configurar a variável de ambiente "path" para o diretório "c:\dos", levanta-se a hipótese de criação de uma unidade virtual representada pela letra "d", onde é chamada posteriormente a execução do aplicativo "programa.exe".

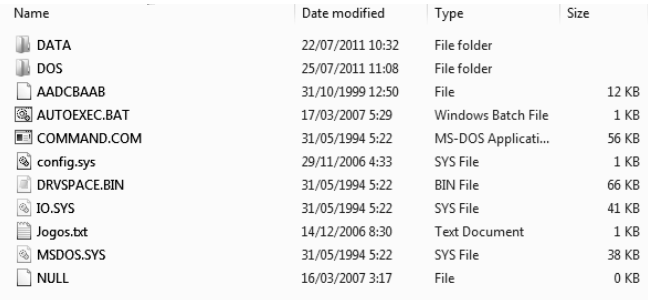


Figura 3. Arquivos presentes na raiz do dispositivo de armazenamento

Dentre os arquivos armazenados pelo diretório "data", somente o arquivo "base.dbf", apresenta-se relevante. Além de apresentar nome que sugere uma base de dados, possui assinatura compatível com o formato dBase. Ao ser editado, tal arquivo revelou uma listagem de campos aparentemente destinados à configuração das máquinas. Porém, parte dos campos não possui valor armazenado, sugerindo serem utilizados somente quando da execução da aplicação, enquanto que outros apresentavam conteúdo codificado. Apenas os campos que traziam valores em ponto flutuante, referentes aos índices de acumulação da máquina, apresentaram-se legíveis durante a análise *off-line*.

Da mesma forma, dos arquivos armazenados no diretório "dos", somente alguns apresentaram relevância para o presente estudo, como o arquivo "menu.exe", responsável pela exibição da tela de seleção de jogos e consequentemente por carregar o jogo escolhido e o arquivo "jogos.bin", que apresenta assinatura compatível com a base de dados em arquivo dBase.

Além desses dois arquivos que as várias máquinas possuem em comum, foram identificados arquivos com a extensão "sys", onde os nomes e a quantidade de arquivos apresentados por cada equipamento são compatíveis com os itens exibidos na tela de seleção de jogos. A convergência entre as opções apresentadas na tela inicial de seleção e os arquivos armazenados, indica que cada arquivo corresponde a um dos jogos oferecidos pelo equipamento.

Ao todo, nos equipamentos analisados, foram encontrados 23 (vinte e três) arquivos com extensão "sys", embora a análise preliminar dos equipamentos tenha apresentado uma lista de 18 (dezoito) jogos distintos.

Verificou-se que, com base na assinatura do cabeçalho, tratam-se de arquivos compactados no formato ZIP, com a

ferramenta de compressão de código "PKLite, com senha de proteção. A existência do arquivo "pkunzip.exe" no diretório "dos", que apresenta apenas a capacidade de descompressão do formato ZIP, sugere que tais arquivos são descompactados durante a execução da máquina.

Após descompactar os arquivos conforme descrito na próxima subseção, verificou-se que apesar da divergência de tamanho do arquivo final, cada um dos jogos consiste de uma quantidade variável de arquivos na extensão "dat", com nomes compatíveis entre si, e um único arquivo executável denominado "programa.exe". Nos vinte e três jogos presentes foram identificadas três variações de tamanho para o arquivo "programa.exe", sugerindo que ao invés de cada jogo possuir uma estrutura interna distinta, vários deles compartilham do mesmo arquivo principal.

TABELA II. Comparativo de conteúdo dos arquivos "sys".

Nome	Executáveis	Tamanho	Total
bigb1ac.sys	programa.exe	273 KB	339
buca2.sys	programa.exe	281 KB	335
buca2ac.sys	programa.exe	281 KB	341
camp2ac.sys	programa.exe	281 KB	314
camp2007.sys	programa.exe	281 KB	321
fruta.sys	programa.exe	281 KB	315
fuga1ac.sys	programa.exe	273 KB	372
fuga2acm.sys	programa.exe	281 KB	376
gcard1.sys	programa.exe	273 KB	333
goldcard.sys	programa.exe	273 KB	333
hallo1.sys	programa.exe	273 KB	335
hallo1ac.sys	programa.exe	273 KB	335
hallo2.sys	programa.exe	281 KB	335
hallo2ac.sys	programa.exe	281 KB	335
hallo2tk.sys	programa.exe	281 KB	331
hallofor.sys	programa.exe	677 KB	335
oro2ac.sys	programa.exe	677 KB	236
pantanal.sys	programa.exe	273 KB	373
sexy2.sys	programa.exe	281 KB	333
sexy2ac.sys	programa.exe	281 KB	373
trago1ac.sys	programa.exe	273 KB	331
vacalo2.sys	programa.exe	281 KB	374
vakalo1.sys	programa.exe	273 KB	331

2) Senha de proteção do programa

Devido ao fato de os jogos serem carregados na memória somente quando escolhidos na tela de seleção de jogos, a descompressão é realizada automaticamente pela aplicação, sugerindo que a senha ou era armazenada em uma variável local da própria aplicação responsável pela descompressão, ou em algum arquivo que pudesse ser lido quando necessário.

A segunda hipótese mostrou-se verdadeira quando os arquivos de nome "jogos.bin" de cada máquina, que foram reconhecidos como sendo tabelas dBase. Eles revelaram uma tabela com as colunas "nomejogo", "mascara" e "senha", todas do tipo texto, porém com conteúdo codificado. A quantidade de registros da tabela é compatível com os itens apresentados pela sua tela de seleção de jogos e com a quantidade de arquivos com extensão "sys" disponíveis em cada equipamento.

Apesar de as informações armazenadas apresentarem-se ininteligíveis em um primeiro momento, indicando a adoção de alguma espécie de codificação, foi possível constatar que as senhas cadastradas para todos os jogos possui a mesma representação codificada, com comprimento de seis dígitos.

TABELA IV. Exemplo de conteúdo do arquivo "jogos.bin"

NOMEJOGO	MASCARA	SENHA
§³´¸Á°±»Ž¸¹	§³´¸œ™¿ÆÁ	²§¹«ª¼
§³´¸Á°±»Ž¸	§³´¸›™¿ÆÁ	²§¹«ª¼
¬µ³¬®¸‹¯®À³	¬©¨º›™¿ÆÁ	²§¹«ª¼
¸»·»Š¾±ÅÇ	¸«¿Á›˜¾ÅÀ	²§¹«ª¼
§¸¨»²¶°µ¿¯¾£¢£«	¨§´¸›š›£›ÁÈÃ	²§¹«ª¼
«¸¼¼²¸³	«¸¼¼ª˜¾ÅÀ	²§¹«ª¼

Na tentativa de se encontrar um padrão entre os símbolos exibidos na coluna "mascara" e os nomes dos arquivos com extensão "sys", percebeu-se que o algoritmo de codificação consiste na substituição do caractere original por outro cuja representação numérica seja compatível com o resultado da soma de uma centena com o valor numérico do caractere que se deseja codificar, somado ainda ao valor referente à sua posição na cadeia de caracteres.

$$texto\_cifrado[i] = texto\_claro[i] + 100 + i; \quad (2)$$

Tendo elucidado o algoritmo de codificação, realizou-se a tradução das informações apresentadas pela tabela "jogos", que além de fornecer a senha de descompactação (MARCAR), demonstrou que os valores presentes no campo "mascara" referiam-se aos nomes dos arquivos que contém o jogo compactado, conforme ilustrado pela tabela a seguir.

TABELA VI. Tradução do conteúdo do arquivo "jogos.bin"

NOMEJOGO	MASCARA	SENHA
HALLOWEEN II	HALLO2.SYS	MARCAR
HALLOWEEN I	HALLO1.SYS	MARCAR
GOLDEN CARD	GCARD1.SYS	MARCAR
SUPER SEXY	SEXY2.SYS	MARCAR
BRASILEIRAO 2007	CAMP2007.SYS	MARCAR
FRUTINHA	FRUTA.SYS	MARCAR

D. Estrutura Interna

Uma vez coletadas as informações disponíveis através da análise *off-line*, referentes ao ambiente operacional das máquinas caça-níquel, procedeu-se com a análise *online* do arquivo "programa.exe", extraído dos arquivos de extensão "sys", cuja assinatura de cabeçalho indica tratar-se de binário compilado em plataforma de 16bits.

Durante a análise binária do aplicativo principal do jogo, foram identificados três níveis de segurança para acesso dissimulado a parâmetros de leitura ou de ajuste de comportamento das máquinas caça-níquel.

1) Ambiente de Configuração

Ao depurar o arquivo "programa.exe" através da ferramenta CodeView, verificou-se que durante a exibição da tela principal do jogo, além de aguardar pelo código das teclas listadas na coluna "Hexa" da Tabela I, é realizada uma comparação do conteúdo do registrador AL com o valor hexadecimal 0x34, que corresponde à tecla "4" do teclado convencional.

Esta *backdoor* leva a uma tela de administração do sistema com opções de leitura estatística de entradas e saídas e configuração de parâmetros do jogo (Figura 4). Esta funcionalidade é exibida somente caso o apostador ainda não tenha iniciado uma sessão com a inserção de créditos.

Dentre as opções apresentadas pela tela de configuração, a primeira delas é a de "leitura parcial", acessível através da tecla "Y" (Tabela I), que exibe as estatísticas da máquina quanto aos valores totais de "entradas", "pagamentos", "apostas" e "prêmios". É possível ainda obter o percentual de prêmios pagos e um balanço que se refere ao lucro da máquina, que corresponde ao resultado da subtração dos "pagamentos" pelo total de "entradas". O percentual de prêmios pagos, por sua vez, é obtido através da proporção do número de prêmios pagos em relação ao número de apostas.

```
        MENU PRINCIPAL - HALLOWEEN VS 2.01 - 09/12/2006 - (FS)


                LINHA 1  >> LEITURA PARCIAL

                LINHA 5  >> LEITURA OFICIAL

                LINHA 9  >> ULTIMOS PAGAMENTOS

                LINHA 15 >> ULTIMAS ENTRADAS

                LINHA 20 >> ULTIMOS 10 JOGOS

                TABELA   >> PARAMETROS

                JOGAR    >> VOLTAR AO JOGO


NUMERO DE SERIE 02-08/12-000
```

Figura 4. Tela de administração do sistema.

2) Primeiro nível de segurança

Ao pressionar a tecla "J", referente à opção denominada "leitura oficial" (Figura 5), tem-se a impressão de que a aplicação não executa qualquer ação. Porém, ao analisar o seu código binário, verificou-se que ela aguarda o envio de um valor através do teclado. Este valor, por sua vez, é confrontado com uma cadeia de caracteres armazenada na seção de dados do arquivo executável, que corresponde à senha correta. Essa prática de armazenamento de senha é conhecida como *hardcoded* [9]. A senha para o primeiro nível de segurança, que é composta por 05 (cinco) dígitos numéricos, é o primeiro valor que antecede a cadeia de caracteres "Leitura Oficial" no arquivo executável do programa controlador.

A divergência entre as telas de "leitura parcial" e "leitura oficial" consiste tão somente na possibilidade de reiniciar os contadores da primeira leitura, sem que os valores totais da máquina sejam perdidos. Com isso, durante a análise de uma determinada máquina, a "leitura parcial" pode não refletir os valores totais arrecadados, sendo necessário confrontá-los com os valores da "leitura oficial".

```
        HALLOWEEN - LEITURA OFICIAL - 02-08/12-000 VS 1.35 (FS)

TOTAL DE ENTRADAS                          512600          5.126,00
TOTAL DE PAGAMENTOS                        445754          4.457,54
TOTAL DE APOSTAS                           761500          7.615,00
TOTAL DE PREMIOS                           701888          7.018,88



% PREMIOS                                                     92.17
TOTAL DE RODADAS                                              14313
TOTAL DE JOGOS GANHOS                                         11522
BALANCO                                     66846           668,46


<JOGAR> VOLTAR
```

Figura 5. Tela de leitura oficial acessível através de senha.

3) Segundo nível de segurança

Ao contrário das demais opções que realizam mera consulta aos valores armazenados na base principal de dados, ao informar o código correspondente ao pressionamento da tecla "T", a aplicação permite a alteração de alguns parâmetros de comportamento do equipamento. Porém, o acesso a essa funcionalidade é controlado por um mecanismo de checagem de senha a ser informada pelo usuário.

```
DIGITE A SENHA CORRETAMENTE...
```

Figura 6. Tela de controle de acesso por senha de segurança.

```
DELAY CTR VELOCIDADE (0->1)        0.10

TERMINAL COM DISPLAY (S/N)   N         PERC ACUMULADO              0,30
TRAVA (S/N)                  S         PGTO. BOTAO (S/N)           S
VS COLOMBIA (S/N)            N
CONFIG IMPRESSORA            0
ACUMULADO(1) INICIAL       200
ACUMULADO(1) ATUAL         511         ACUMULADO(2) ATUAL          260
ACUMULADO MAXIMO          1000
PRECO $ CREDITO (* 100)              1
CREDITOS POR PULSO                 100



                                                    AP   -> + 1
                                                    MX   -> - 1
                                                    (S)AU-> + 100
INTERFACE (0, 1, 2)                  2 TIPO-2       (N)TB-> - 100
ANO      (00-99)                     8              PG   -> + 1000
MES      (01-12)                    12              20L  -> + 10000
SERIE    (0-999)           020812000                15L  -> +,01
< PAGTO > CONFIRMA VALORES                          09L  -> ZERA
OUTRA TECLA: DESCARTA                               05L  -> SAI
```

Figura 7. Tela de configuração de parâmetros.

Observando o comportamento da aplicação principal em relação ao valor de senha informado pelo usuário, verificou-se que este é confrontado com uma cadeia numérica com cinco dígitos de comprimento, armazenada também no próprio arquivo executável, porém diferente da senha utilizada para acessar a tela de leitura oficial. A senha para o segundo nível de segurança pode ser encontrada após a única ocorrência da cadeia de caracteres "**informe a senha corretamente**", depois de extraídas as cadeias de caracteres do arquivo binário.

Uma vez que a senha tenha sido informada corretamente, os parâmetros de configuração são exibidos um a um para que sejam ajustados conforme as instruções exibidas no quadrante inferior direito da tela (Figura 7). Todos os parâmetros são livremente configuráveis, exceto aquele intitulado "Config Impressora".

4) Terceiro nível de segurança

Ao selecionar este parâmetro ("Config Impressora"), ao invés de o valor informado ser armazenado no registrador correspondente, é realizada uma comparação com um terceiro valor armazenado no segmento de dados do arquivo, indicando tratar-se novamente de uma senha, diferente das duas senhas anteriores. A senha para o terceiro nível de segurança é o primeiro valor numérico contendo cinco dígitos que sucede a cadeia de caracteres "**Config Impressora**" no arquivo executável.

Nos vinte e três jogos analisados, foram encontradas ao menos duas senhas distintas para cada nível de segurança. Para o primeiro nível, parte dos jogos utilizava a senha "30114", enquanto que o restante utilizava a senha "43210". O segundo nível, por sua vez, apresentou as senhas "30113", "30118" e "01234" e o terceiro nível, as senhas "30116" e "36546" em predominância.

5) Decodificação da base de dados principal

Diante do fato de a base de dados principal, representada pelo arquivo "base.dbf" presente no diretório "data", apresentar conteúdo ininteligível, procedeu-se com a modalidade de análise *online* com o intuito de verificar o comportamento da aplicação principal em relação aos valores armazenados.

Durante a análise, constatou a existência de codificação das colunas originalmente definidas como do tipo "string" que armazenam valores numéricos. A codificação é trivial: cada byte que forma o número é registrado na forma da sua representação ASCII. Tendo em mente que a arquitetura x86 é *little endian*, a ordem é do menos significante para o mais significante. Como exemplo, considere o valor originalmente informado como "100", que era gravado como "d" (a representação ASCII do número).

Uma vez analisados os valores armazenados nos respectivos arquivos dos equipamentos sob investigação, foi possível constatar a função das principais colunas da tabela abrigada pelo arquivo "base.dbf". Assim, percebe-se que as colunas que trazem o termo "jack" referem-se sempre à ação de pagamento do "prêmio especial" ao apostador. "Jack" e "jackpar" armazenam os montantes referentes aos prêmios pagos pela máquina, sendo que o segundo é o valor parcial, pois pode ser zerado a qualquer momento pelo administrador do sistema. "Seqjack", por sua vez, guarda a quantidade numérica de prêmios pagos desde a instalação do equipamento.

Com isso, ao examinar o comportamento da aplicação em relação ao valor armazenado em "ctrjack", verificou-se que este obriga que a máquina limite-se a acumular o valor definido em "acummax".

6) Parâmetros fraudulentos de configuração

As colunas "perce" e "trava" são aquelas que evidenciam o maior poder lesivo da aplicação principal. A primeira delas, definida na tela de configuração como "Config. Impressora", cujo acesso é realizado através de dupla verificação de senha, refere-se ao percentual de retenção do equipamento, aceitando valores numéricos entre 0 e 4. Quanto menor o valor informado, maior a retenção praticada pela máquina. Em todas as máquinas analisadas, o valor de "perce" estava configurado como "0".

Tal ajuste influencia negativamente na aleatoriedade do resultado, uma vez que a aplicação apresenta um laço de repetição que obriga a realização de novo sorteio caso o nível de retenção do equipamento não tenha sido atingido.

A coluna "trava", também configurada através da tela ilustrada pela Figura 7, quando marcada positivamente, impede a exibição das figuras compatíveis com as sequências de bônus, sem prestar qualquer esclarecimento ao usuário da máquina, levando-o a erro.

Outro indício de fraude foi constatado na tela de bonificação denominada internamente como "maçã". Quando da montagem da interface em tempo de execução, os valores da matriz não são previamente sorteados. Ao contrário disso, o sorteio acontece tão somente quando um dos quadrantes é escolhido pelo apostador. Entretanto, assim como ocorre com o sorteio na tela principal, o resultado é confrontado com o nível de retenção do equipamento antes que este seja exibido ao usuário. Caso seja necessário, outro valor é sorteado sem que o usuário tenha ciência disso.

IV. Metodologia de Análise Proposta

Assim, com base na adaptação do método utilizado por Vênere [11] na análise de código malicioso para atender ao modelo proposto pelo NIST [12], infere-se a seguinte metodologia de análise de máquinas caça-níquel que pertençam ou não à família Halloween:

- Devido à existência de parâmetros não identificados de configuração do ambiente operacional, que normalmente impedem a execução de uma imagem do dispositivo de armazenamento em um ambiente

controlado, deve-se realizar uma cópia bit a bit do dispositivo de armazenamento, a fim de preservar a integridade do conteúdo armazenado no dispositivo original;

- O equipamento deve ser iniciado com o dispositivo que contém a cópia bit a bit e deve-se realizar uma análise comportamental inicial com base na usabilidade da interface, onde serão identificadas as principais funcionalidades da aplicação;
- De posse da análise comportamental, deve-se realizar um inventário dos itens de hardware do equipamento, identificando os componentes de interação com a aplicação, tais como dispositivos de entrada e saída. Durante a confecção do inventário, as funcionalidades identificadas na etapa anterior são mapeadas aos códigos produzidos pelos dispositivos de interação;

Tendo uma ideia formal inicial do comportamento da aplicação principal, havendo similaridade gráfica ou elemento que indique tratar de jogo pertence à família Halloween, deve-se determinar a relação seguindo os seguintes critérios:

- Localizar os arquivos de jogos com extensão "sys" gravados no diretório "DOS";
- Localizar o arquivo "jogos.bin" e, com base na fórmula apresentada, traduzir as informações da base de dados que contém as senhas de descompressão dos arquivos de jogos;
- Examinar os arquivos de imagens dissimulados e o arquivo binário "programa.exe", que representa o núcleo da aplicação;
- Extrair as cadeias de caracteres do arquivo binário e localizar as senhas dos três níveis de segurança.

Uma vez que atendam os quatro passos descritos acima, três abordagens distintas para identificação de similaridade entre as diferentes versões disponíveis [17] comprovou que os jogos da família Halloween possuem essencialmente o mesmo núcleo operacional, adotando comportamento semelhante apesar das aparentes diferenças que possam existir quanto à temática do jogo ou à exibição de um ou mais prêmios acumulados (*jackpot*).

Porém, caso a máquina não se enquadre como da família Halloween por não apresentar similaridade aparente ou por não atender aos quatro requisitos descritos anteriormente, deve-se:

- Realizar uma análise do ambiente operacional da aplicação, identificando arquivos responsáveis por subsidiar a execução do programa que gerencia os dispositivos de interação. Nesta etapa devem ser catalogados e separados os arquivos referentes à preparação do ambiente de suporte dos arquivos diretamente vinculados à aplicação;
- Dentre os arquivos vinculados à aplicação, devem-se identificar aqueles que correspondem ao jogo propriamente dito e a possíveis bases de dados utilizadas pela aplicação principal ou por aplicativos que deem suporte a ela, pois a existência de bases de dados sugere a utilização de parâmetros de configuração do ambiente e registro estatístico de jogadas;
- Abrir as bases de dados encontradas e caso o conteúdo esteja codificado, tentar a decodificação delas através das fórmulas propostas. Caso as fórmulas não sejam aplicáveis, buscar a decodificação usando padrões recorrentes em nomes de jogos similares ou pela análise *online* do código binário com o auxílio de um *debugger*;
- Uma vez decodificadas as bases de dados, realizar um mapeamento inicial entre os valores armazenados, referentes a parâmetros de configuração do ambiente e registro estatístico de jogadas, e os elementos apresentados pela interface;
- Caso os arquivos referentes a jogos fornecidos pela máquina estejam na sua forma binária, abri-lo através de uma ferramenta *debugger* e realizar a análise on-line retificando ou ratificando o mapeamento inicial realizado entre os valores armazenados na base de dados e os elementos de entrada e saída apresentados pela interface da aplicação principal;
- Caso os jogos tenham sido comprimidos com a utilização de senha, verificar se alguma das bases de dados identificadas possui as senhas armazenadas e realizar a descompactação. Por sua vez, caso os jogos tenham sido submetidos a uma ação de empacotamento, utilizar uma ferramenta de *unpacking* para ter acesso ao código-binário original ao invés do código-binário do desempacotador. Em alguns casos, o arquivo principal pode ter sido submetido primeiramente a uma função de empacotamento e posteriormente a uma função de compressão. Neste caso, deve-se realizar primeiramente a descompressão e em seguida o desempacotamento;
- Durante a análise *online* deve-se procurar por funções de captura de texto e geração de números randômicos em arquiteturas de 32-bits ou superiores, ou interrupções de relógio interno e teclado em arquiteturas de 16-bits, para facilitar a identificação de áreas ocultas de configuração de parâmetros ou das evidências da prática de jogo de azar, respectivamente;
- Ainda durante a análise on-line, quando da identificação de regiões obscuras de acesso a configurações de parâmetros com verificação de senha, buscar pela existência de senhas *hardcoded* gravadas na área de dados do código binário ou simplesmente realizar a alteração do *flag* responsável por decidir o desvio adotado por determinado salto condicional presente no código diante da verificação da senha de acesso;
- Realizar testes de operação, ajustando os parâmetros e observando a mudança de comportamento da aplicação, a fim de produzir um modelo comportamental mais apurado. Esse modelo visa constatar as fraudes provenientes da dissimulação de funcionalidades identificadas na etapa de análise inicial do comportamento e que induzem o apostador a acreditar que o equipamento comportar-se-á de maneira distinta daquela que foi programada.

V. Considerações Finais

Diante da incerteza que paira sobre o comportamento das máquinas caça-níquel simuladas por programas de computador, em todo o território brasileiro tem havido divergência no entendimento dos magistrados quanto ao seu enquadramento ou não no rol dos jogos de azar. Por simularem certa jogabilidade, tais máquinas despertam a falsa impressão de que a habilidade do apostar pode influenciar na probabilidade de ganho ou perda.

Com o objetivo de pacificar a discussão, o presente trabalho aplicou técnicas de Engenharia Reversa para realizar a análise do comportamento geral da aplicação das máquinas tipo "Halloween" desde o momento em que o apostador aciona o botão que dispara o mecanismo de sorteio, até o momento em que o resultado é exibido na tela.

Durante a análise foram encontrados parâmetros de configuração ocultos ao apostador com até três níveis de segurança por verificação de senha de acesso. Ao relacionar esses parâmetros com o comportamento da aplicação, percebeu-se que alguns deles influenciam diretamente no percentual de retenção (lucro) da máquina e no pagamento de premiação. Foram evidenciadas ainda dissimulações de arquivos e de opções de configuração.

Além de o sorteio seja realizado de forma aleatória, sem qualquer hipótese de o apostador influenciar no resultado depois de disparada a ação correspondente, foram constatados parâmetros de configuração que impedem que uma determinada sequência de premiação seja exibida na tela. Com isso, é evidenciado não só a prática de jogo de azar (aleatoriedade do sorteio) como também a possibilidade de fraude (bloqueio não transparente de opção de premiação).

As contribuições do presente trabalho têm seu foco na caracterização da prática de jogo de azar ou fraudes. São elas:

- Caracterização do comportamento interno dos programas de computador que controlam as máquinas tipo "Halloween", para subsidiar a instrução de laudos periciais confeccionados por Peritos Criminais ao analisar jogos idênticos aos abordados pelo presente trabalho;
- Apresentação de uma metodologia de análise que permite aos Peritos Criminais realizar, por analogia, o exame de programas que controlam máquinas incompatíveis com o tipo "Halloween".

Como trabalho futuro, sugere-se o aprofundamento da pesquisa referente ao comportamento interno da aplicação para instruir o desenvolvimento de uma ferramenta que realize a aferição automática dos percentuais de ganho ou perda do equipamento. De forma completamente automatizada, a aplicação realizará *n* interações com o jogo e verificará os resultados gerados, mantendo um histórico independente da relação aposta x ganho, otimizando a abordagem proposta por Nogueira [4] em seu estudo.

Referências

[1] Costa, Marcelo A. S. Lemos. Computação Forense. 2.ed. Campinas: Millennium, 2003, p.5-8.

[2] G1. Videobingo. Disponível em: g1.globo.com/brasil/noticia/2011/12/operacao-tio-patinhas-apreende-108-maquinas-caca-niqueis-no-para.html. Acesso em: 31 de janeiro de 2012.

[3] G1. Videobingo. Disponível em: g1.globo.com/parana/noticia/2012/01/policiais-civis-fecham-cassino-que-funcionava-em-mansao-em-curitiba. html. Acesso em: 31 de janeiro de 2011.

[4] Nogueira, José Helano Matos. Máquinas Caça-Níqueis. Perícia Federal, Março de 2002, a.4, n.12, p.18.

[5] Bello, Leo. Aprendendo a Jogar Poker: Princípios, Técnicas e Prática. 2.ed. Rio de Janeiro: Nova Fronteira, 2008.

[6] Silva Júnior, A. Lopes. A contravenção de exploração de jogo de azar. Disponível em: jus.com.br/revista/texto/10110/a-contravencao-de-explo racao-de-jogo-de-azar. Acesso em: 5 de janeiro de 2012.

[7] Eilam, Eldad. Reversing: Secrets of Reverse Engineering. Indianapolis: Wiley, 2005.

[8] Pressman, Roger S., Engenharia de Software. 6.ed. São Paulo: McGraw-Hill, 2006.

[9] ICT. Products/Bill Acceptor. Disponível em: www.ict-america.com/product/bill_ acceptor.asp. Acesso em: 17 de agosto de 2011.

[10] Narvaja, Ricardo. Introducción al cracking con OllyDBG. Disponível em: Ricardo narvaja.info. Acesso em: 5 de setembro de 2011.

[11] Vênere, Guilherme. Engenharia Reversa de Código Malicioso. São Paulo: Escola Superior de Redes/Rede Nacional de Pesquisa, 2009.

[12] Peters, James F. & Pedrycz, Witold. Engenharia de Software: Teoria e Prática. Rio de Janeiro: Campus, 2001. p.561-2.

[13] Schneier, Bruce. & Ferguson, Niels. Practical Cryptography. Indianapolis: Wiley, 2003.

[14] Flake, Halvar. Structural Comparison of Executable Objects. DIMVA. Disponível em: citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.83.6632. Acesso em: 21 de novembro de 2011.

[15] Dullien, Thomas & Rolles, Rolf. Graph-based Comparison of Executable Objects. Disponível em: citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.96.5076. Acesso em: 21 de novembro de 2011

[16] Eagle, Chris. The IDA Pro book: The unofficial guide to the world's most popular disassemble. San Francisco: No Starch Press, 2008.

[17] SILVA, C. Esteves, ZELENOVSKY, Ricardo, SOUSA, G. Batista. Emprego da Engenharia Reversa para caracterização do modus operandi das máquinas caça-níquel quanto à prática de jogo de azar ou outras fraudes. Dissertação de Mestrado, Publicação PPGENE.DM - 103 A/2012, UnB: Brasília, 112p.

DOI: 10.5769/C2013007 or http://dx.doi.org/10.5769/C2013007

Greatest Eigenvalue Time Vector Approach for Blind Detection of Malicious Traffic

Danilo Fernandes Tenório[2], João Paulo C. L. da Costa[1,2], and Rafael Timóteo de Souza Júnior[1,2]

(1) Laboratory of Array Signal Processing
(2) Laboratory for Decision-Making Technologies (LATITUDE)
Department of Electrical Engineering
University of Brasilia (UnB)
URL: www.pgea.unb.br/~lasp

***Abstract*—Recently, blind techniques have been applied to detect malicious traffic and attacks in honeypots. The honeypot traffic can be divided into legitimate and malicious traffic, where the legitimate traffic corresponds to DHCP, broadcasting, and synchronization. In practice, other servers connected to the network may be also targets for attacks and malicious traffic. Therefore, it is crucial to develop detection techniques for malicious traffic for such computers. In this paper, we propose a solution that blindly detects malicious traffic for any computer connected to the network. We validate our proposed solution considering two types of malicious traffic: synflood and portscan.**

***Keywords*—Eigenvalue Decomposition; Model Order Selection; Detection.**

I. Introduction

Nowadays one of the greatest challenges in Internet is security assurance, obtained by integrity, availability and confidentiality of data. There are several ways to provide security, taking into account both technical aspects, through the use of safety equipments or systems, as administrative and personal, related to establishment of a security policy and awareness campaigns. Regarding safety equipments or systems, we can use for instance firewall, Intrusion Detection Systems (IDS) and Intrusion Prevention Systems (IPS).

Several methods have been proposed for identifying and characterizing malicious activities. Classical methods typically employ data mining [1] [2] and regular file parsing [3] for detecting patterns which indicate the presence of specific attacks in the analyzed traffic. Recently, automatic blind malicious traffic detection techniques have been developed for honeypots [4] [5]. However, the honeypot traffic is simpler since there are no legitimate applications running.

The use of Model Order Selection (MOS) Schemes to detect highly correlated components as significant network activities; and identifying malicious activities in honeypot network flow datasets without any previous information or attack signatures by applying model order selection schemes has been proposed in [4].

In this work, we propose an automatic blind malicious traffic scheme to be used in any server of a network. Inspired by [4] [5], we model a real network traffic data into three components: the legitimate signals, the malicious signals and the noise.

Our proposed scheme is based on the eigenvalue decomposition, however, in contrast to [4] [5], we consider the time variation of the greatest eigenvalue. We show that based on this variation, attacks such as portscan and synflood can be detected.

This paper is organized as follows. In Section II, we define the notation used in this paper. In Section III, we discuss about data log, and how we model it as signals and noise. In Section IV, we characterize the portscan and synflood attacks. In Section V, we propose our scheme for the normalized and nonnormalized case. In Section VI, we explain in details the experiments with real data, and evaluate several MOS scheme presenting experimental results which attest the validity of our approach. In Section VII, we make our concluding remarks.

II. notation

In this paper the scalars are denoted by italic letters (*a, b, A, B, α, β*), vectors by lowercase bold letters (**a**, **b**), matrices by uppercase bold letters (**A**, **B**), and $a_{i,j}$ denotes the (*i, j*) elements of the matrix **A**. The superscripts ^{T} and $^{-1}$ are used for matrix transposition and matrix inversion, respectively.

III. Data Collection

The log information of a computer connected to the network is formed by timestamp, protocol, source IP address, source port, destination IP address, destination port and additional information, depending on the type of transport protocol used.

In order to exemplify these collected data, we consider the following TCP traffic log:

21:00:34.099289 IP 192.168.1.102.34712 > 200.221.2.45.80: Flags [S], seq 2424058224, win 14600, options [mss 1460, sackOK,TS val 244136 ecr 0,nop,wscale 7], length 0

and the UDP traffic log:

21:24:42.484858 IP 192.168.1.102.68 > 192.168.1.1.67: BOOTP/DHCP, Request from 00:26:9e:b7:82:be, length 300

In this paper, we consider only the following information from the log data timestamp, port type and port number.

A. Data Model

The reduced log data is divided into q time slots of N samples, where each sample is collected in a certain time period. Each element $x_{m,n}^{(q)}$ represents the number of times that the port m appears at the n-th time period, at the q-th time slot.

The collected data at the q-th time slot is represented by $\mathbf{X}^{(q)} \in \mathbb{R}^{M \times N}$, where M represents the amount of ports, and N represents the amount of time samples. The matrix $\mathbf{X}^{(q)}$contains all traffic (signal, noise and attack), and we can model it as:

$$\mathbf{X}^{(q)} = \mathbf{S}^{(q)} + \mathbf{N}^{(q)} + \mathbf{A}^{(q)} \quad (1)$$

where $\mathbf{S}^{(q)}$ is the matrix that represents the legitimate traffic, $\mathbf{N}^{(q)}$ represents the noise, and $\mathbf{A}^{(q)}$ the malicious traffic.

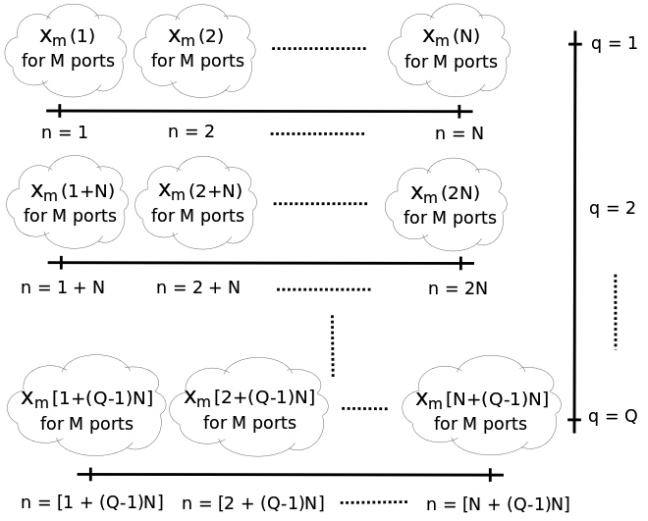


Figure 1. Traffic matrix $\mathbf{X}^{(q)}$, for q = 1, 2, …, Q.

Our goal in this paper is to detect the rank of the matrix $\mathbf{A}^{(q)}$, given only the matrix $\mathbf{X}^{(q)}$. Thereby, if the rank $\{\mathbf{A}^{(q)}\} \neq 0$, we have a malicious traffic; otherwise, if rank$\{\mathbf{A}^{(q)}\} = 0$, there is no malicious traffic.

IV. Characterization of Portscan and Synflood Attacks

In this section, we show important properties of the portscan and synflood. These properties are important to explain the validity of the proposed solution.

In the Fig. 2, the portscan transmits only two packets for each TCP port and one packet for each UDP port. Note that there is a high correlation since the traffic is equal.

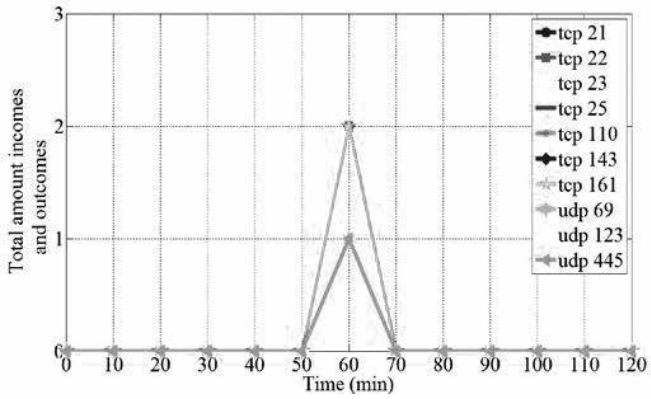


Figure 2. Malicious traffic over M ports vs n time slots (M = 10 and n = 120). This traffic profile represents the traffic characterized by portscan, consisting of TCP and UDP portscanning.

In the Fig. 3, the synflood attack consists of sending hundreds of packets with the SYN flag active in a short period of time. In our case, considering this attack, if a server with port 80 open, the server is overloaded and may cause the unavailability of the service rendered by it. In a time interval of ten minutes, there were more than two hundred ten thousand packages related to the attack, an unusual traffic in a data network, especially by the fact of being concentrated in a short period of time.

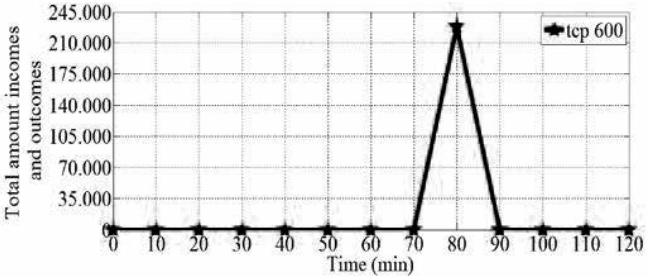


Figure 3. Malicious traffic over M ports vs n time slots (M = 1 and n = 120). This traffic profile represents the traffic characterized by synflood.

II. Proposed Solution

Basically, the model order of a dataset is estimated as the number of main correlated components with energy significantly higher than the rest of uncorrelated components. In other words, the model order can be characterized by a power gap between main components and the noise components. In the context of network traffic, the principal components are represented by outstanding network activities, such as highly correlated network connections which have, for example, the same destination port [4]. The efficacy and efficiency of methods based on Principal Component Analysis (PCA) depend on the MOS scheme adopted, since each scheme has different probabilities of detection for different kinds of data [6].

We consider two cases, one case normalizing $\mathbf{X}^{(q)}$, and the other one nonnormalizing it. The purpose of this was to adapt the solutions to the characteristics of portscan and synflood

attacks. Thus, we built two correlation matrices. One from the normalized case $\mathbf{R}_{xx}^{'(q)} \in \mathbb{R}^{M\times M}$, and other from the nonnormalized case $\mathbf{R}_{xx}^{(q)} \in \mathbb{R}^{M\times M}$.

A. Normalized case

In detecting portscan we have low values associated with this attack (Fig. 2), but repeatedly, since we are scanning in multiple ports (correlated traffic). Then, when we normalize the $\mathbf{X}^{(q)}$matrices, the portscan can be collected.

The normalization of the vectors can be obtained by the following equation:

$$\mathbf{x}_m^{'(q)} = \frac{\mathbf{x}_m^{(q)} - \bar{\mathbf{x}}_m^{(q)}}{\sigma_m^{(q)}} \tag{2}$$

for $q = 1, ..., Q$ where $\bar{\mathbf{x}}_m^{(q)}$ is the mean of $\mathbf{x}_m^{(q)}$ and $\sigma_m^{(q)}$ is the deviation standard of $\mathbf{x}_m^{(q)}$.

Once obtained the vectors $\mathbf{x}_m^{'(q)}$, we can construct the matrix $\mathbf{X}^{'(q)}$, and then determine the correlation matrix in order to find the eigenvalues.

$$\mathbf{R}_{xx}^{'(q)} = \frac{1}{N}\ \mathbf{X}'^{(q)}\mathbf{X}'^{(q)^{\mathrm{T}}} \tag{3}$$

where N is the number of the sample time.

The eigenvalue decomposition of $\mathbf{R}_{xx}^{'(q)}$ is given by:

$$\mathbf{R}_{xx}^{'(q)} = \mathbf{E}'^{(q)}\mathbf{\Lambda}'^{(q)}\mathbf{E}'^{(q)\mathrm{T}} \tag{4}$$

where $\mathbf{\Lambda}^{'(q)}$ is a diagonal matrix with the eigenvalues, and the matrix $\mathbf{E}^{'(q)}$ has the eigenvectors corresponding to each eigenvalue. However, for our model order selection schemes, only the eigenvalues are necessary.

By selecting only the main diagonal of the matrix $\mathbf{\Lambda}^{'(q)}$, via diag{$\mathbf{\Lambda}^{'(q)}$}, and by ranging $q = 1, ..., Q$, we can build a matrix $\mathbf{C}' \in \mathbb{R}^{M\times Q}$.

Assuming that the eigenvalues $\lambda_m^{'(q)}$ are in the descending order, i. e., $\lambda_1^{'(q)} > \lambda_2^{'(q)} > \cdots > \lambda_{m-1}^{'(q)} > \lambda_m^{'(q)}$, the first column of the matrix $\mathbf{C}'$ has the Greatest Eigenvalue Time Vector (GETV).

As shown in the Section VI, by using the GETV in the model order selection schemes, it's possible to detect the presence of malicious traffic even if applications are running.

B. Nonnormalized case

In detecting synflood we have a huge uncorrelation traffic. Thus, differently of portscan we have in this case only one traffic, but with a high value. So, we cannot normalizing this traffic since the normalization would cause an abrupt reduction of the high value associated with this attack, causing it to disappear.

Thus, in order to detect the synflood, we cannot normalize the matrix $\mathbf{X}^{(q)}$. However, except by normalization, the whole procedure to find the GETV is equal to the one shown is Subsection V.A.

C. GETV combined with Model Order Selection Schemes

Each model order selection scheme has different characteristics. We used the following method in our simulations: AIC [7] [8], MDL [7] [8], EDC [8] [9], RADOI [10], EFT [11] [13] and SURE [12].

The EFT and EDC models showed the satisfactory results for our scheme. In case of EDC, the information criterion is a function of the geometric mean, *g(i)*, and arithmetic mean, *a(i)*, of the *i* smallest eigenvalues. Note that $c_{1,q}$ and $c_{1,q}^{'}$, $q = 1, ..., Q$, should be in descending order.

The estimate of the model order d can be represented by $\hat{d}$, through the following expressions:

$$\hat{d} = \operatorname{argmin}(J(i)) \tag{5}$$

$$J(i) = -2N(Q - i\ +1\ )\log\left(\frac{g(i)}{a(i)}\right) + (i-1)p(Q,i,N) \tag{6}$$

where $p(Q,i,N) = [2Q\ -(i-1)]\sqrt{N\log(\log(N))}$.

To use the (6) we have firstly to put the vector of eigenvalues in ascend order.

For the EFT based schemes, i.e. R-D EFT II, R-D EFT, M-EFT, and EFT, we have has to compute the threshold coefficients, as shown in [13]. Without the threshold coefficients, the EFT based schemes cannot be applied. By computing these coefficients and applying the EFT it's possible to find the model order of our scheme.

VI. Simulations

In this section, we describe the performed experiments in order to validate our proposed scheme for detecting portscan and synflood attack in a computer.

A. Data Analysis

For this simulation we used a computer (based on Linux operational system) performing common tasks (web access mainly) during an interval of three hours. The application tcpdump was used to capture the network traffic, as shown in Fig 4.

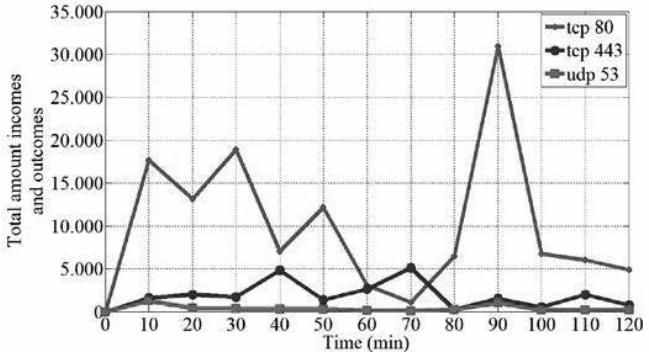


Figure 4. Signal traffic over M ports vs n time slots (M = 3 and n = 120). This traffic profile represents only the signal, without any kind of attack or noise.

At 21:54h we conducted with the portscan, in order to simulate an attacker who wants to know the status of the following ports: TCP 21, 22, 23, 25, 110, 143 and 161; and UDP 69, 123, and 445. At the range time from 22:10h to 22:19h we conducted with the synflood attack (log below), in order to simulate an attacker who wants to cause a Denial of Service (DoS), causing unavailability of services.

22:10:04.986927 IP 192.168.1.104.64263 > 192.168.1.102.600: Flags

[S], seq 3652238756, win 1365, length 0

22:10:04.986961 IP 192.168.1.102.600 > 192.168.1.104.64263: Flags [R.], seq 0, ack 3652238757, win 0, length 0

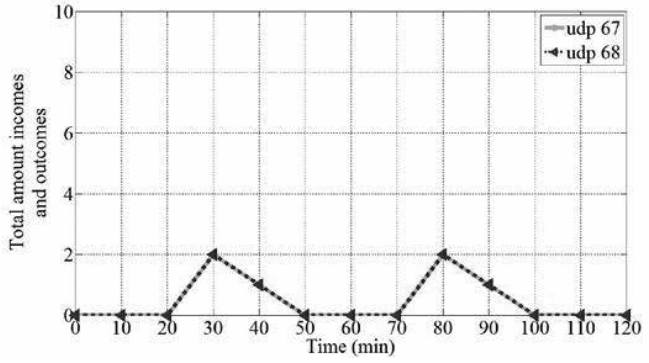


Figure 5. Noise traffic over M ports vs n time slots (M = 2 and n = 120). This traffic profile represents only the noise, consisting of udp 67 and udp 68 traffic.

The Fig. 4 shows the signal traffic consisting of requests and responses on TCP port 80, TCP port 443 and UDP port 53. The TCP port 80 is associated with unencrypted web access, the TCP port 443 to encrypted web access, and UDP port 53 is associated with name resolution, DNS.

The noise traffic (log below) is formed by UDP port 67 and UDP port 68, associated with the Dynamic Host Configuration Protocol (DHCP). It can be seen in Fig. 5.

21:24:42.484858 IP 192.168.1.102.68 > 192.168.1.1.67: BOOTP/DHCP, Request from 00:26:9e:b7:82:be, length 300

21:24:42.487652 IP 192.168.1.1.67 > 192.168.1.102.68: BOOTP/DHCP, Reply, length 548

B. Eigenvalue Decomposition (EVD)

As described in Section III, the total simulation time of 120 minutes was fragmented into Q = 6 periods of N = 20 minutes each, where each period we use the sampling time of 1 minute. As the simulation began at 21:00h, the first period goes from 21:00h until 21:20h (T_1), the second from 21:20h until 21:40h (T_2), the third from 21:40h to 22:00h (T_3), the fourth from 22:00h until 22:20h (T_4), the fifth from 22:20h until 22:40h (T_5), and finally the sixth from 22:40h to 23:00h (T_6). Thus, it was possible to build Q = 6 matrices $\mathbf{X}^{(q)}$ of the total traffic (signal + noise + attack). Obviously not every period there is attack, only at T_4 occurred the synflood attack (Fig. 3), and at T_3 the portscan (Fig. 2).

Once we have the $\mathbf{X}^{(q)}$ matrices for each period, it is now possible to obtain the correlation $\mathbf{R}_{xx}^{'(q)}$ and $\mathbf{R}_{xx}^{(q)}$ matrices, related to each matrix $\mathbf{X}^{(q)}$. With that it was possible to obtain the set of eigenvalues for that correlation matrices, generating a total of 2Q vectors of eigenvalues: 6 vectors related to $\mathbf{R}_{xx}^{'(q)}$, built from the normalization of $\mathbf{X}^{(q)}$ (Fig. 6), and 6 vectors related to $\mathbf{R}_{xx}^{(q)}$, built from the nonnormalizing of $\mathbf{X}^{(q)}$ (Fig. 7).

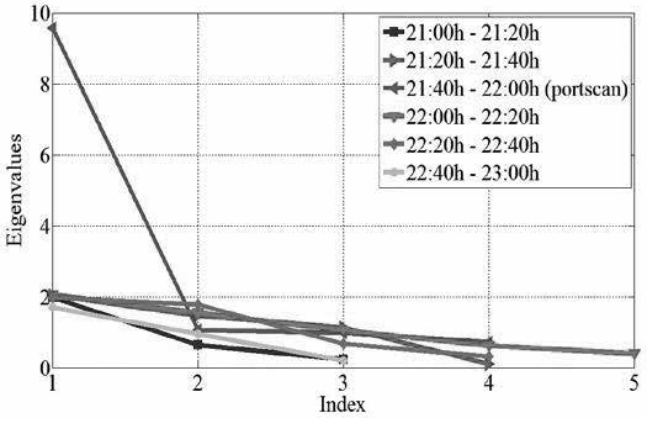


Figure 6. Eigenvalues of the normalized case over each time slot. In this figure is the greatest eigenvalue related to the portscan is much greater than the others.

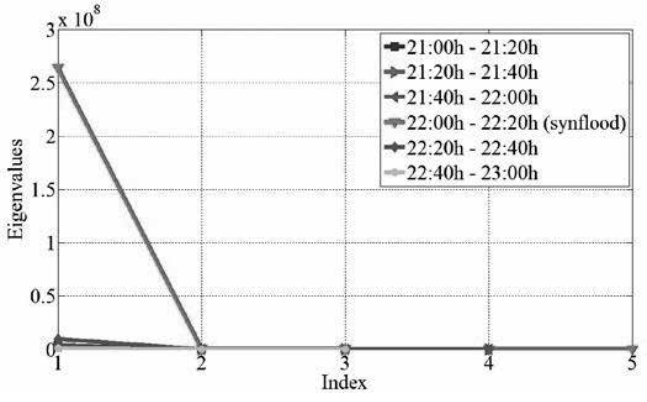


Figure 7. Eigenvalues of the nonnormalized case over each time slot. In this figure is possible to see the eigenvalue related to the synflood (purple line index 1).

Calculating the eigenvalues of each $\mathbf{R}_{xx}^{'(q)}$ and $\mathbf{R}_{xx}^{(q)}$ matrices, we can reduce the size of our matrix $\mathbf{X}^{(q)}$, and get some interesting conclusions, derived from the eigenvalue decomposition properties of the correlation matrices, such as: the eigenvectors associated with each eigenvalue are orthogonal to each other, and also linearly independent; and the eigenvalues are real and nonnegative.

C. Applying Model Order Selection to Data Analysis

Although the variation of the eigenvalues related to the attacks, the job is not complete until we find a model that applies to this scheme. The estimation of the model order by visual inspection is performed by following subjective criteria such as considering only the eigenvalues greater than one and visually identifying a large gap between two consecutive eigenvalues. Then, to let this work the most complete and objective as possible we tested several MOS approach, like AIC, MDL, EDC, RADOI, EFT and SURE.

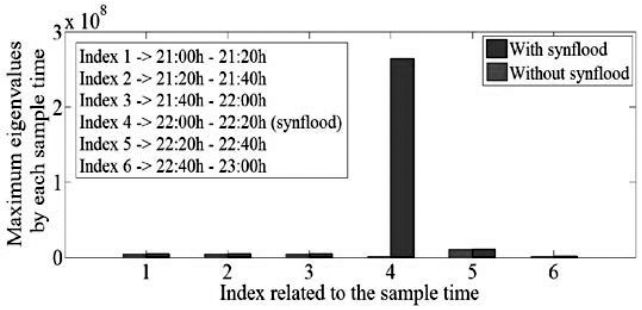


Figure 8. Greatest eigenvalue time vector approach related to the nonnormalized case. It is possible to see the eigenvalue related to the synflood, much greater than the other ones (brown bar index 4).

Before we show the results obtained with the application of the models, we will discuss the input values for each MOS approach. The Fig. 8 and Fig. 9 show the greatest eigenvalues obtained in each period. Thus, we applied the Greatest Eigenvalue Time Vector (GETV) approach.

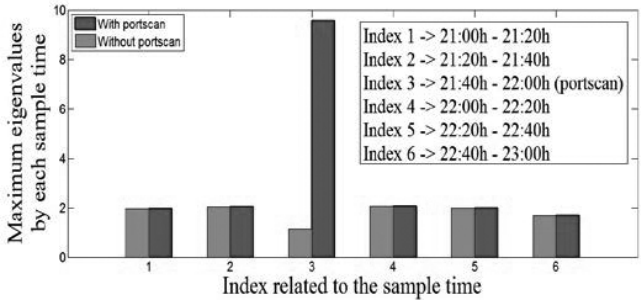


Figure 9. Greatest eigenvalue time vector approach related to the normalized case. It is possible to see the eigenvalue related to the portscan, greater than the other ones (brown bar index 3).

The method consists of selecting the largest eigenvalues of the q = 6 time slots, and apply them to the existing model order selection schemes in order to detect malicious traffic.

In Figs. 8 and 9, we show the greatest eigenvalues of the q = 6 time slots. In Fig. 8 we have the nonnormalized case, used to detect synflood attack. In this figure it is possible to compare the values of the eigenvalues with and without the attack. We can see clearly how the component related to the attack stands out from the rest. In Fig. 9 we have the normalized case, used to detect the portscan. In this figure it is possible to compare the values of the eigenvalues with and without the attack.

TABLE I. Nonnormalized Case

Case	Method					
	AIC	*MDL*	*EDC*	*RADOI*	*EFT*	*SURE*
With synflood	2	1	1	4	1	14
Without synflood	4	1	0	1	0	13

The tables I and II were obtained after we apply the methods cited in our scheme. The table I give us the results in the nonnormalized case, and the table II in the normalized case.

TABLE II. Normalized Case

Case	Method					
	AIC	*MDL*	*EDC*	*RADOI*	*EFT*	*SURE*
With portscan	1	1	1	1	1	6
Without portscan	0	0	0	1	0	1

According to the tables, we see the various model orders and found two that stood out. The Efficient Determination Criterion (EDC) and Exponential Fitting Test (EFT) showed us the correct order models, which is equal to one, indicating that there was an attack. The methods are efficient in detecting both attacks. The efficiency of the model as it behaves when there is no attack, in this case showing that the model order is zero, indicating that there is no attack, neither portscan nor synflood.

VII. Conclusion

In this paper we propose The Greatest Eigenvalue Time Vector (GETV) approach for detecting portscan and synflood in a network traffic flow data collected at a computer. First we showed the data log used, and the propose of the model for network flow data, in order to verify the validity of our approach through simulation results with real log files collected at a computer. Several model order selection methods were experimented with the simulation data, showing that EDC and EFT yields the best results for this type of data.

Since our proposed scheme is blind, it does not require previous collection of data and learning periods.

References

[1] Y. H. Hu "Parallel eigenvalue decomposition for toeplitz and related matrices," International Conference Acoustics, Speech, and Signal Processing (ICASSP'89), 1989, pp. 1107 -1110, vol. 2, 1989.

[2] H. T. Wu, J. F. Yang and F. K. Chen. "Source number estimators using transformed gerschgorin radii," IEEE Transactions on Signal Processing, 43(6):1325–1333, 1995.

[3] R. R. Nadakuditi and A. Edelman. "Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples," IEEE Transactions of Signal Processing, 56:2625–2638, 2008.

[4] B. M. David, J. P. C. L. da Costa, A. C. A. Nascimento, M. D. Holtz, D. Amaral, and R. T. Sousa Júnior. "Blind automatic malicious activity detection in honeypot data," pp. 02-04, ICoFCS 2011.

[5] J. P. C. L. da Costa, E. P. de Freitas, B. M. David, A. M. R. Serrano, D. Amaral, and R. T. Sousa Júnior. "Improved blind automatic malicious activity detection in honeypot data," ICoFCS 2012.

[6] J. P. C. L. da Costa, "Parameter estimation techniques for multidimensional array signal processing," Shaker, First edition, 2010.

[7] J. P. C. L. da Costa, A. Thakre, F. Roemer, and M. Haardt, "Comparison of model order selection techniques for high-resolution parameter estimation algorithms," in Proc. 54th International Scientific Colloquium (IWK'09), Ilmenau, Germany, Oct. 2009.

[8] M. Wax and T. Kailath, "Detection of signals by information theoreticcriteria," IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. ASSP-33, pp. 387–392, 1985.

[9] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai. "On detection of the number of signals in presence of white noise," J. Multivar. Anal., 20:1- 25, 1986.

[10] E. Radoi and A. Quinquis, "A new method for estimating the number of harmonic components in noise with application in high resolution radar," EURASIP Journal on Applied Signal Processing, pp. 1177–1188, 2004.

[11] A. Quinlan, J.-P. Barbot, P. Larzabal, and M. Haardt, "Model order selection for short data: An exponential fitting test (EFT)," EURASIP Journal on Applied Signal Processing, 2007, Special Issue on Advances in Subspace-based Techniques for Signal Processing and Communications.

[12] M. O. Ulfarsson and V. Solo, "Rank selection in noisy PCA with SURE and random matrix theory," in Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008), Las Vegas, USA, Apr. 2008.

[13] J. P. C. L. da Costa, F. Roemer, F. A. de Castro Junior, R. F. Ramos, S. Schwarz, and L. Sabirova, ``Ilmenau Package for Model Order Selection and Evaluation of Model Order Estimation Scheme of Users of MIMO Channel Sounders,'' XXIX Simpósio Brasileiro de Telecomunicações (SBrT'11), Curitiba, Brazil.

DOI: 10.5769/C2013008 or http://dx.doi.org/10.5769/C2013008

Investigação de Crimes Relacionados à Pedofilia Utilizando Metadados de Imagens

João M. Ceron, Paulo César Herrmann Wanner, Lisandro Z. Granville, Bruno Werneck

Resumo — A investigação de crimes relacionados à pedofilia é um processo complexo que compreende a análise de imagens. Nota-se, no entanto, que cada vez mais dispositivos móveis são utilizados para a captura de imagens. Boa parte dos celulares e *tablets* possuem sistema de GPS interno o qual permite que fotos sejam capturadas com informações de localização. Valendo-se disso, este trabalho busca apresentar uma arquitetura para coleta, análise e correlação dos dados analisando informações embutidas nas fotos. Para isso, os autores do trabalho descrevem uma arquitetura e implementam um protótipo para auxiliar peritos na resolução de crimes relacionados com pedofilia.

Keywords—exif; metadada; investigação; pedofilia;

I. Introducão

A violações de leis criminais que envolvem o conhecimento de tecnologia para a sua perpetração está em franco crescimento. Dispositivos informatizados são frequentemente utilizados em alguma etapa das diferentes tipificações criminais [1].

Em crimes relacionados a pedofilia, em especial, a análise de evidências digitais pode ressaltar características fundamentais para a resolução de um crime. A análise de fotos existente num telefone celular podem, por exemplo, revelar dados do autor e até mesmo informações geográficas.

Analisar um dispositivo informático com imagens relacionadas à pedofilia é uma tarefa árdua e psicologicamente desgastante. Nesse processo, muitas vezes o investigador passa por inúmeros diretórios de fotos com conteúdo sexual envolvendo menores. A utilização de ferramentas para auxiliar nesse processo de investigação, de fato, pode facilitar a perícia criminal.

Essa demanda pode ser preenchida por ferramentas que classificam imagens segundo características intrínsecas. Existem ferramentas especializadas em reconhecer padrões de imagens, e outras em mapear características inseridas no próprio arquivo da imagem. No entanto, as atuais ferramentas apresentam lacunas que podem ser endereçadas para aprimorar investigações digital.

Sendo assim, este trabalho descreve uma arquitetura especializada na análise de metadados de imagens com o objetivo de auxiliar na investigação de crimes relacionados a pedofilia. Para isso, os autores descrevem uma arquitetura baseada no modelo cliente servidor que possibilita a coleta de informações e o armazenamento dos dados numa base centralizada. Deseja-se, deste modo, possibilitar que informações de diferentes investigações possam ser armazenadas e correlacionadas entre si.

Toda a arquitetura baseia-se apenas nos dados presentes no metadados das imagens periciadas. As informações do metadados são analisadas e exportadas para uma base de dados, como por exemplo: data da captura da imagem; configurações da câmera; e coordenadas geográficas. Tais informações das imagens agrupadas com demais dados da investigação são armazenadas numa base de dados. Dessa forma, por meio de uma interface (API), um investigador pode consultar os dados presentes na base de dados e correlacionar as informações. Adicionalmente, diferentes meios de visualização podem auxiliar na resolução da investigação:

a) *Mapas:* localização onde as fotos foram tiradas ilustrando a área de atuação do autor (ex. *Google Maps*, *Bing Maps*);
b) *Estabelecimento*: locais próximos que podem auxiliar na investigação, por exemplo, solicitar imagens de câmeras de vigilância de estabelecimentos comerciais (ex. *Google Places*, *Foursquare*);
c) *Redes Sociais:* comentários próximos ao local das fotos podem revelar observações de transeuntes (*Twitter*, comentários georreferenciados);
d) *Dispositivos:* classificação de fotos por dispositivo o que permite identificar as câmeras utilizadas.

Como resultado, os autores deste trabalho apresentam um protótipo da solução e descrevem o seu uso numa investigação de crime relacionado a pedofilia. Este trabalho está organizado da seguinte maneira: na seção II são descritos os trabalhos relacionados; na seção III é apresentado a solução proposta; na seção IV é descrito o protótipo implementado e um estudo de caso; por fim, as conclusões são apresentadas na seção V.

II. Ferramentas de Investigação

Durante uma investigação, o perito criminal precisa trabalhar com uma grande quantidade de dados: arquivos de texto, imagens, vídeos, programas executáveis, base de dados e outros. Torna-se um grande desafio extrair dessa massa de dados informações que podem comprovar uma atividade criminosa.

A fim de identificar informações relevantes para uma investigação são utilizadas ferramentas capazes de analisar diferentes tipos de arquivos em diferentes sistemas de armazenamento. Existem ferramentas que são amplamente

utilizadas para processar uma grande quantidade de dados categorizando e indexando informações encontradas a fim de facilitar o trabalho pericial. A diversidade de ferramentas para análise forense é muito alta. Boa parte dessas ferramentas são comerciais, mas também existem soluções gratuitas de boa qualidade [2].

A comunidade acadêmica constantemente busca identificar novas tendências e avalia a aplicabilidade de certas ferramentas no contexto de uma investigação. Por exemplo, em [3] os autores avaliam a ferramenta *Forensic ToolKit* (FTK) muito popular entre os peritos. A FTK é uma suíte de ferramentas para analisar um sistema de arquivos, incluindo a identificação de arquivos removidos e diversos relatórios com característica dos dispositivos. A ferramenta *Encase*, talvez a mais conhecida entre os peritos, é estuda por Garber em [4]. Assim como o FTK, a ferramenta Encase também é uma suíte de ferramentas para investigação forense fortemente customizável apresentando, até mesmo, uma linguagem de programação *script* para automatização de tarefas.

Outra gama de ferramentas forenses são as especializadas em buscam extrair informações de arquivos específicos. Tais ferramentas visam interpretar dados de arquivos de registro do Windows; informações de programas de comunicação instantânea; registro de eventos do sistema operacional; e ainda recuperar arquivos do espaço não alocado no disco utilizando técnicas de *data carving* [5]. Num contexto mais específico, algumas ferramentas são capazes de identificar imagens com pornografia infantojuvenil. O *NuDetective Forensic Tool* [6], por exemplo, é uma ferramenta especializada na busca de imagens com nudez infantil. O *NuDetective* utiliza técnicas de busca por arquivos conhecidos com base em listas de resumos criptográficos (*Hash List*), técnica também utilizada por ferramentas comerciais de amplo uso.

Observa-se, também, ferramentas com nicho específico de atuação. A ferramenta *Cellebrite Universal Forensic Extraction Device (UFED)* [7] é um exemplo. Essa ferramenta é especializada na análise de dispositivos móveis como celulares e *tablets*. Tal ferramenta faz uma varredura no aparelho e disponibiliza diferentes tipos de relatórios. Além de identificar fotos capturadas, a ferramenta disponibiliza uma mapa com o histórico de localização do dispositivo (baseado nas coordenadas GPS e torres de celular).

Apesar da flexibilidade das ferramentas forenses as mesmas apresentam certas lacunas que, em escopos específicos, ficam evidenciadas. Numa investigação cujo objetivo é analisar metadados das imagens de uma maneira mais detalhada as ferramentas generalistas de investigação forense apresentam limitações. Por exemplo, é possível obter metadados de imagens na grande maioria de ferramentas e diferentes sistemas operacionais (vide Fig. 1) no entanto as mesmas não permitem, de forma direta, correlacionar dados ou, até mesmo, atuar nos dados embutidos nas imagens.

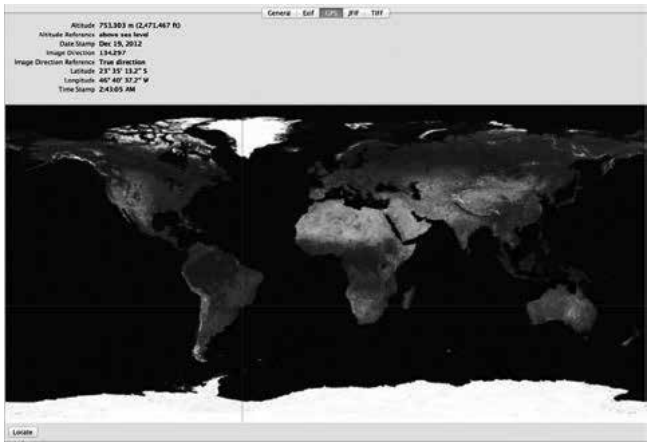


Fig. 1. Metadados de uma foto visualizada na interface nativa do sistema operacional MAC OS X v10.8 Mountain Lion.

Diante do exposto, cabem aprimoramentos nas atuais soluções para as necessidades reais definidas no escopo deste trabalho. Desenvolver uma arquitetura que possa ser utilizada como base de dados para diferentes investigações cujo objetivo seja investigar crimes relacionados a pedofilia continua sendo uma lacuna a ser explorada. Na sequência é descrita a arquitetura proposta pelos autores a fim de endereçar as atuais limitações das ferramentas existentes.

III. Arquitetura da Ferramenta de Investigação

Uma investigação forense pode atuar em diferentes tipos de mídias e formatos de arquivos de imagens. Da mesma forma, sabe-se que o universo periciável é bastante dinâmico e sofre constantes mudanças com a inserção de novas tecnologias. Sendo assim, a arquitetura proposta foi definida utilizando o conceito modular a fim de ser facilmente extensiva para adequar-se as diferentes necessidades de investigação.

Além da modularidade, é importante que a arquitetura seja concebida para adequar-se a diferentes domínios administrativos. Por exemplo, na Fig. 2 é ilustrado um cenário com diferentes domínios administrativos representados por DEICC -- (Delegacia Especializada de Investigações de Crimes Cibernéticos) -- de 3 regionais RS, DF, SP. Os diferentes domínios administrativos podem atuar segundo o conceito de federação, e também podem atuar de forma independente. Isso significa que, no exemplo, uma delegacia (DEICC) pode atuar de forma isolada, dependendo apenas do modo de operação da arquitetura configurado.

As federações representadas na Fig. 2 comunicam-se entre si via comunicação segura (VPN). Além Além disso, é descrito um módulo opcional de gerenciamento da arquitetura. Esse módulo gerencial só faz sentido quando são utilizadas diferentes federações onde deseja-se configurar a interação entre as mesmas.

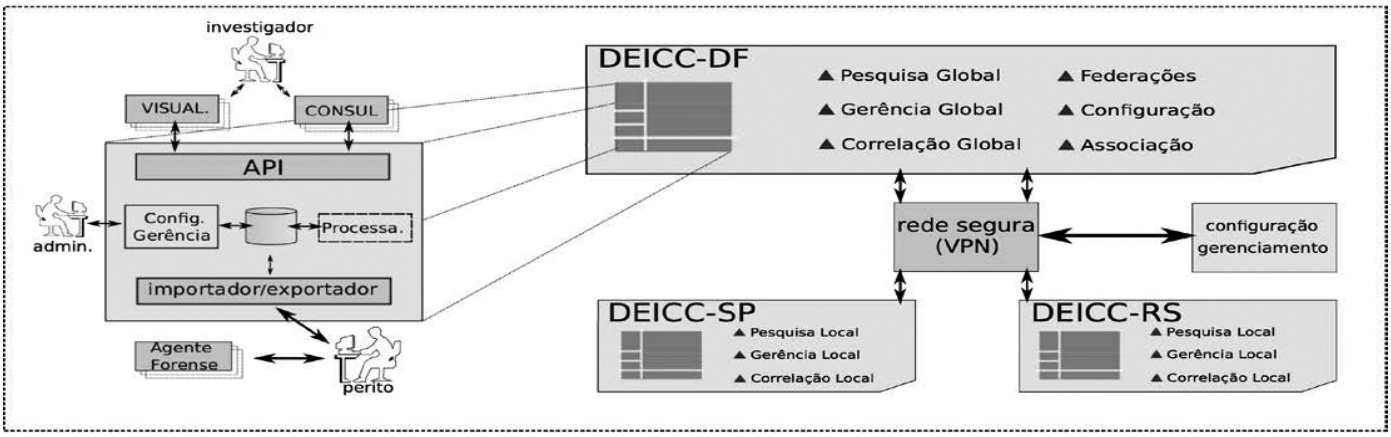


Fig. 2. Arquitetura proposta utilizando um cenário com diferentes docmínios administrativos.

Com isso, o usuário Administrador da arquitetura pode gerenciar acessos, monitorar, e configurar os diferentes módulos. Assim como o usuário Administrador, são definidos outros perfis de usuários conforme descrito abaixo:

a) *Administrador:* responsável por definir controles de acesso a base de dados e configuração. Dessa forma, torna-se possível restringir o acesso aos dados tendo em vista as permissões de um perito.

b) *Investigador*: neste perfil, o usuário pode consultar a base de dados e correlacionar informações distribuídas pela arquitetura. Por exemplo, pode-se verificar a existência de um *hash* criptográfico em diferentes investigações.

c) *Perito:* este usuário é responsável pela coleta de evidência no dispositivo periciado. O perito executa o software e faz a submissão dos dados para a base de dados.

Os diferentes módulos que compõe a arquitetura podem ser observados em destaque na Fig. 2. Na parte superior são especificados módulos para consulta e visualização dos dados armazenados na base de dados; na parte central os elementos responsáveis pelo processamento das informações e configurações da base de informações; e, por fim, na parte inferior os agentes forenses responsáveis pela coleta de dados nos dispositivos periciados.

A visualização de dados (VISUALI.) e consulta de informações (CONSUL.) permitem que o usuário com perfil Investigador consulte os dados presentes na arquitetura e faça correlações. Para que diferentes módulos de visualização e consulta pudessem ser implementados, foi desenvolvida uma interface de programação (API).

Utilizando a API desenvolvida é possível realizar consultas via linha de comando (CLI) e também via interface Web dando maior flexibilidade a arquitetura. Um exemplo de uso da API é apresentado na seção IV, no entanto a descrição da API foge do escopo desse trabalho e será endereçada em trabalhos futuros.

Os módulos responsáveis pelo controle e processamento das informações estão alinhados na parte central da arquitetura sendo respectivamente: API de consulta; importação/exportação; configuração e controle. Utilizada como interface entre os módulos de visualização e consulta, a API de consulta tem acesso a base de dados observando as políticas de controle definidos no módulo *configuração e controle*. O controle de acesso descreve políticas de acessos ao banco de dados. Essas políticas podem ser definidas segundo usuários ou valendo-se do conceito de *visões*. As visões devem considerar quesitos como permissão do usuário, operações de investigações, domínios administrativos (federações) e outros. Além dos controles, demais configurações são definidas por esse módulo, tais como: usuários e configuração da arquitetura.

O módulo importação/exportação é responsável obter as informações dos agentes forenses e armazená-las na base de dados. De forma complementar, o módulo possibilita que a base de dados seja exportada para uma outra base de dados remota, quando a arquitetura for configurada de forma hierárquica -- federações.

O Agente Forense, por sua vez, é o software responsável analisar os metadados de imagens e extraí-las para uma base de dados. Para isso, o software examina o sistema de arquivos de um dispositivo e identifica metadados no formato *Exif* (*Exchangeable Image File Format*) [8].

O *Exif* é uma especificação que define um conjunto de dados que podem ser embutidos em arquivos de imagem e áudio. Essa especificação é bastante abrangente e atualmente é implementada por diversos dispositivos, como celulares, câmeras digitais, *tablets* e outros.

Além de informações técnicas de uma imagem, o *Exif* pode opcionalmente conter informações georreferenciadas (utilizando o GPS do dispositivo). A Listagem 1 ilustra parcialmente o metadados de uma fotografia capturada utilizando uma telefone celular.

```
ExifTool Version Number : 8.90
File Name : iphone5.png
MIME Type : image/jpeg
Make : Apple
Camera Model Name : iPhone 5
Exposure Time : 1/15
F Number : 2.4
ISO : 800
Exif Version : 0221
Date/Time Original : 2013:01:14 23:41:48
Create Date : 2013:01:14 23:41:48
GPS Altitude : 80.1 m Above Sea Level
GPS Latitude : 41 deg 1' 52.80" N
GPS Longitude : 73 deg 46' 27.60" W
Image Size : 3264x2448
```

Listagem 1: Conjunto parcial de informações embutidas numa imagem descrita pela especificação *Exif*.

As informações contidas nas imagens, tais como, dispositivo utilizado, data da foto e localização podem ser fundamentais numa investigação. Compondo esses dados, um perito pode identificar os dispositivos utilizados bem como a região geográfica onde o possível crime ocorreu. Sabe-se que informações do metadados podem ser adulteradas, no entanto, essa análise está fora do escopo do nosso trabalho [9].

Numa primeira análise, o *Agente Forense* identifica imagens com informações georreferenciadas. Na sequência, o metadados dessas imagens é inspecionado bem como outras informações do arquivo são extraídas (nome do arquivo, assinatura *hash* do arquivo, e outras). Como resultado, o *Agente Forense* é responsável por compilar as informações encontradas e armazená-las num arquivo de texto num formato aberto e padronizado para exportação (XML, JSON). Da mesma forma, são incorporadas ao arquivo de exportação informações sobre a investigação e identificação do perito. Por fim, o perito tem um conjunto descritivo das informações encontradas no dispositivo e pode exportar diretamente para a base de dados centralizada.

É importante destacar que toda a arquitetura atua apenas com a análise de metadados dos arquivos. Os arquivos originais não são alterados tampouco transferidos para arquitetura. Na sequência os demais elementos da arquitetura são descritos com maior detalhamento.

IV. Implementação e Estudo de Caso

Baseando-se na arquitetura apresentada na seção anterior foi implementado um protótipo do sistema. Esta seção descreve detalhes de implementação bem como um estudo de caso com a aplicação do protótipo. Por questões de organização esta seção está organizada da seguinte forma: na primeira etapa é descrita a implementação do protótipo que foi dividida em *Módulos de Coleta* e *Módulos de Processamento*; e na segunda etapa um estudo de caso é apresentado.

A. Módulos de Coleta

Os módulos de coleta são aqui representados pelos *Agentes Forense*. O *Agente Forense* é o software responsável por varrer um sistema de arquivo em busca de metadados de imagens no formato *Exif*.

No estágio inicial da varredura o software solicita de forma interativa algumas informações para o perito, tais como: identificação do investigador, identificação da investigação, e identificação da prova a ser periciada. Com base a isso, o software produz um arquivo texto no formato aberto -- JSON ou XML -- com todas as informações inseridas pelo perito e também com os metadados encontrados na varredura. Um fragmento do arquivo gerado é ilustrado na Listagem 2.

É importante que o *Agente Forense* seja portável e com baixo nível de complexidade para ser executado nos mais diferentes cenários de investigação. Para isso, o mesmo foi implementando utilizando a linguagem *Perl* com bibliotecas específicas para lidar com informações de metadados de imagem [10].

```
<item>
  <created:time>2013-06-02 01:00:09 +0000</created:time>
  <operation>Chakal_2013</operation>
  <researcher>SP_BR_0324A</researcher>
  <md5>d07625f10c9c85c70d97880dfe81c713</md5>
  <filename>iPhone5.png</filename>
  <exif:captured>2013:05:04 18:58:13</exif:captured>
  <exif:geo:lat>-23.62754</exif:geo:lat>
  <exif:geo:long>-46.66082</exif:geo:long>
  <exif:make>Apple</exif:make>
  <exif:model>iPhone 5</exif:model>
  ....
  ...
<item>
```

Listagem 2: Fragmento do arquivo XML gerado pelo *Agente Forense*.

Na Listagem 2 é possível identificar alguns campos armazenados do arquivo resultante de uma varredura do *Agente Forense*. Observa-se dados que identificam a investigação (*operation*) e o perito (*researcher*). Adicionalmente, observa-se informações sobre o metadados da imagem, como data da captura, fabricante do dispositivo, informações geor-referenciadas, entre outras.

De posse do arquivo gerado pelo *Agente Forense*, cabe ao perito submeter o arquivo para os demais módulos de processamento de dados.

B. Módulos de Processamento

Os módulos de processamento correspondem a parte mais complexa do sistema, onde os dados coletados pelos agentes são armazenados e posteriormente processados para investigação e correlação.

Toda a implementação do protótipo foi realizada num único servidor. O servidor consiste num máquina Linux com recursos modestos, utilizando uma base de dados Sqlite versão 2.8.17. Os *Módulos de Processamento* foram implementados na linguagem *Perl* acessando a base de dados via DBI (*The Perl Database Interface Module*). Diferentemente dos demais módulos, o módulo de visualização implementado (mapa)valeu-se de *JavaScript* executado localmente no servidor Web Nginx 14.0.

De forma inicial, os *Módulos de Processamento* analisam o arquivo submetido pelo perito e armazena as informações no banco de dados da arquitetura. As informações do arquivo JSON/XML são convertidas em diferentes campos do banco de dado, dando maior flexibilidade para acesso dos dados. A interface de consulta via linha de comando, por exemplo, dá a possibilidade para um investigador consultar informações utilizando filtros. A Listagem 3 exibe uma consulta na base de dados solicitando a todos os dispositivos presentes.

```
# list_devices
 id | device                | insert_timestamp     | operation_tag   |
----+-----------------------+----------------------+-------------------- +
001 | Apple, iPhone 5       | 2012:01:04 18:56:57  | Chakal_2013     |
002 | Asus, Nexus 7         | 2012-10-24 22:46:55  | Chakal_2013     |
003 | Motorola, MB525       | 2012-10-24 22:46:55  | Chakal_2013     |
004 | Apple, iPhone 5       | 2013-04-14 08:16:11  | Chakal_2013     |
```

Listagem 3: Listagem de dispositivos presentes na base de dados.

É possível observar na Listagem 3 os diferentes dispositivos presentes na base de dados e o horário de inserção no sistema, bem como, a identificação do operação. De forma direta, essas informações representam que foram analisados diferentes dispositivos na investigação *Chakal_2013* e foram encontradas fotos georreferenciadas em 4 diferentes dispositivos.

Diferentes filtros podem ser aplicados diretamente no metadados das imagens. Assim como consultar a base de dados por um dispositivo específico, pode-se buscar por *hash* criptográfico de fotos, data de captura da foto, e região geográfica. Na Listagem 4, por exemplo, são representadas duas diferentes consultas: a primeira procura todos os dispositivos do tipo ``Apple, iPhone 5'' inseridos na base de dados no último ano. E, a segunda consulta, lista todas as informações das fotos extraídas no dispositivo especificado. O comando *list_photos* exibe informações das fotos cuja localidade seja o estado de SP (segundo API de localização utilizando Google Places [11]) e que tenham sido capturadas no ultimo mês.

```
# list_photos -device="Apple,iPhone 5" --last-year
 id | device | photos | operation_tag |
-----+--------------------+---------+--------------------+
001 | Apple,iPhone 5 | 15 | Chakal_2013 |
004 | Apple,iPhone 5 | 34 | Vortice_2013 |
# list_photos -device-id="001" --local="SP" --last-month
MD5 | capture_timestamp | lat | long |
----------------+-------------------------+-------------+------------+
d076....c713 | 2013-05-04 18:58:13 | -23.62754 | -46.66082 |
a3d1....0528 | 2013-05-19 16:17:04 | -23.61047 | -46.66833 |
8a8f....c212 | 2013-05-07 10:23:50 | -23.60339 | -46.68374 |
bcac....d107 | 2013-05-29 10:23:51 | -23.60839 | -46.69112 |
```

Listagem 4: Uso da interface via linha de comando para consultar informações armazenadas na base de dados.

De forma complementar a interface de linha de comando, foi desenvolvida a interface via navegador Web. Essa interface é importante, pois permite que os dados presentes na base de dados possa ser correlacionados com outros serviços Web descrevendo o conceito de *mashups* [12].

A Fig. 3 representa uma configuração da interface de visualização implementada. No mapa da figura são ilustrados pontos representando a localização de fotos capturadas por um dispositivos periciado. É possível observar que a localização das fotos concentram-se no estado do Rio Grande do Sul e São Paulo. Logo, as fotos presente no dispositivo podem identificar características de um crime que aconteceu nos dois estados.

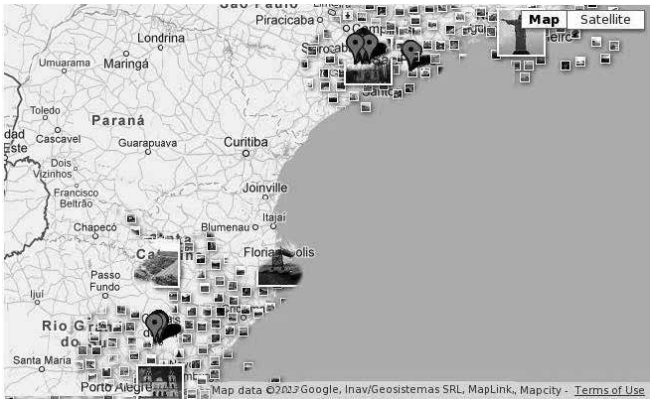


Fig. 3. Interface ilustrando os pontos onde as fotos foram capturadas e composta com uma camada de fotos do Google Panoramio.

Na sequência o protótipo implementado, descrito nesta seção, é utilizado numa investigação simulada demonstrando possíveis usos da ferramenta.

C. Estudo de Caso

Esta subseção visa ilustrar as principais funcionalidades da arquitetura. Para isso foram utilizados dados sintéticos de 3 dispositivos móveis distintos. Dessa forma, nenhuma imagem com pornografia infantil foi analisada, mas sim imagens arbitrárias georreferenciadas armazenadas nos dispositivos. Os seguintes dispositivos foram analisadas:

- Motorola Defy: celular com 40 fotos georreferenciadas dispersas no estado de São Paulo e Rio Grande do Sul;
- Asus Nexus 7: *tablet* com 10 fotos georreferenciadas no estado de São Paulo;
- iPhone 5: celular com 15 fotos dispersas no estado do Rio Grande do Sul.

Em nosso estudo, os diferentes dispositivos foram montados numa máquina e o *Agente Forense* implementado foi executado.

Inicialmente, o *Agente Forense* fez uma varredura completa no dispositivo e salvou os metadados no formato XML, conforme exemplificado na Listagem 2. Sendo assim, foi criado um arquivo com informações dos metadados para cada dispositivo: 2013-02-28_12-24-01_503_apple_dev.xml, 2013-02-28_12-30-34_138_motorola_dev.xml, 2013-02-28_12-42-11_107_asus_dev.xml.

Os arquivos XML com informações dos metadados encontrados foram submetidos para o servidor onde os módulos de processamento foram implementados. A submissão dos arquivos foi feita utilizando o SSH2, no entanto

outros meios poderiam ser utilizados, como uma interface Web de submissão. Com isso, o módulo de importação dos dados faz uma análise no arquivo XML e o armazena na base de dados. A partir desse momento, as informações coletadas pelo perito podem ser consultadas via diferentes interfaces implementadas.

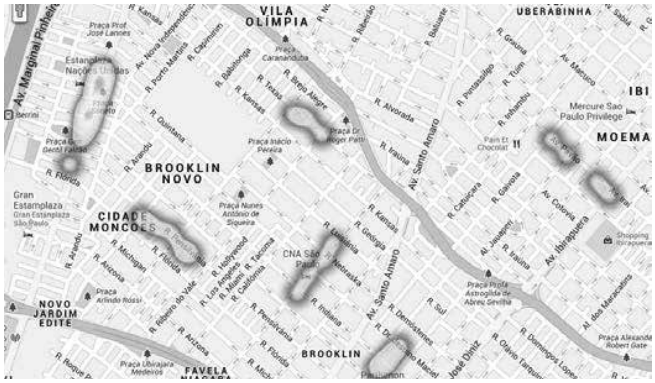


Fig. 4. Mapa utilizando gradiente de cores (*heatmap*) representando a concentração geográfica onde as fotos analisadas foram capturadas.

Na arquitetura proposta, as consultas são realizadas utilizando uma API de consulta (vide seção III). Com a mesma API, diferentes módulos de visualização podem acessar a base de dados da arquitetura. Na Listagem 3, por exemplo, foi exemplificado uma consulta utilizando a interface de comando (CLI). No entanto, outras formas de visualização podem ser implementadas. Em nosso protótipo foi desenvolvida uma interface de mapas para filtrar e ilustrar a localização das fotos. Para isso, as coordenadas geográficas das fotos -- armazenadas na base de dados – são processadas e convertidas num arquivo GeoRSS [13] e posteriormente exportadas para a interface do aplicativo Google Maps [11].

Na interface de mapas implementada é possível exprimir diferentes representações. Uma representação delas ilustra a visão geral de onde as fotos foram capturadas (vide Fig. 3). Já em outra configuração, um mapa com outras características dos dados pode ser detalhado. Na Fig. 4, por exemplo, é ilustrado um mapa no formato *heatmap*. Neste formato, é utilizado um gradiente de cores para representar a concentração geográfica onde as fotos foram capturadas. Essa informação pode identificar local por onde o suspeito passou e, até mesmo, auxiliar na identificação de um possível logadouro onde o crime foi efetivado.

Assim como a interface de linha de comando, a interface de mapas implementada permite o uso de filtros. Logo, torna-se possível plotar informação de um dispositivo específico, uma operação em particular, ou até mesmo, filtrar pela data de captura das fotos.

Uma outra representação da interface de mapas é ilustrada na Fig. 5. Nesta figura é possível definir pontos de interesse próximo ao local de captura de fotos. Através de uma pré-configuração na interface Web, o investigador pode procurar por estabelecimentos específicos para adicionar a plotagem. No exemplo, são ilustrados pontos que correspondem a escolas.

A representação de estabelecimentos próximos ao local das fotos capturadas pode auxiliar na resolução de crimes. Tem-se o conhecimento onde investigações foram solucionadas com ajuda de estabelecimentos próximos. Em uma dada investigação um perito encontrou fotos de pedofilia num dos dispositivos analisados. Nas fotos eram expostas cenas com crianças utilizando um uniforme escolar específico. Utilizando a arquitetura proposta, seria possível identificar a área de atuação de um criminoso e também escolas próximas que poderiam auxiliar na investigação.

Fig. 5. Mapa utilizando gradiente de cores (*heatmap*) representando a concentração geográfica onde as fotos analisadas foram capturadas.

O protótipo implementado possibilitou demonstrar as principais características da arquitetura proposta. São ressaltados alguns recursos que podem contribuir numa investigação. Em especial, destaca-se como fontes externas de informações georreferenciadas podem ser compostas aos metadados das fotos numa investigação criminal.

V. Considerações Finais

Nos últimos anos, no entanto, o campo de fotografia digital foi impulsionado com o uso de dispositivos móveis. Como diferencial, os dispositivo móveis podem embutir informações geográficas nas próprias fotos. Logo, analisar os metadados de fotos capturadas com dispositivos móveis tornou-se atrativo.

Embora algumas ferramentas permitam a visualização dos dados de localização, não se tem conhecimento de uma ferramenta específica para tal propósito. Sendo assim, este trabalho apresentou uma arquitetura baseada no metadados de imagens georreferenciadas para a investigação de crimes relacionados à pedofilia.

Como resultado, os autores implementaram um protótipo da arquitetura e demonstraram como informações de fotos georreferenciadas podem auxilar na resolução de casos. A interface implementada permitiu que as informações geográficas de fotos periciadas pudessem ser compostas com outros serviços *online* (*mashup*). Com isso, um investigador pode visualmente identificar características do possível crime.

Adicionalmente, acredita-se com um maior número de investigações armazenadas na base de dados seja possível

usar a própria arquitetura como base de conhecimento. Em investigações futuras, crimes semelhantes podem ser correlacionados e características reincidentes podem ser mapeadas de forma direta.

Por questões de *design*, a arquitetura proposta atua somente em imagens com informações geográficas embutidas. Como limitações, as funcionalidades descritas não são efetivas em imagens não georreferenciadas.

Em trabalhos futuros, os autores desejam implementar outras interfaces de visualização e permitir a correlação de dados com outras fontes de informação.

Referências

[1] E. Wendt e H. V. N. Jorge, Crimes Cibernéticos Ameacas e procedimentos de investigacao. Brasport, 2012.

[2] B. Carrier, "The sleuth kit.". Disponível em: http://www.sleuthkit.org/sleuthkit/desc.php, 2013.

[3] X. Ding e H. Zou, "Time based data forensic and cross-reference analysis," in Proceedings of the 2011 ACM Symposium on AppliedComputing. ACM, 2011, pp. 185–190.

[4] L. Garber, "Encase: A case study in computer-forensic technology," IEEE Computer Magazine January, 2001.

[5] M. I. Cohen, "Advanced jpeg carving," in Proceedings of the 1st international conference on Forensic applications and techniques in telecommunications, information, and multimedia and workshop. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008, p. 16.

[6] M. de Castro Polastro e P. da Silva Eleuterio, "Nudetective: A forensic tool to help combat child pornography through automatic nudity detection," in Database and Expert Systems Applications (DEXA), 2010 Workshop on, 2010, pp. 349–353.

[7] R. Mislan, "Cellphone crime solvers," Spectrum, IEEE, vol. 47, no. 7, pp. 34–39, 2010.

[8] Technical Standardization Committee on AV & IT Storage Systems and Equipment, "Exchangeable image file format for digital still cameras: Exif Version 2.2," Tech. Rep. JEITA CP-3451, April 2002.

[9] T. Gloe, "Forensic analysis of ordered data structures on the example of jpeg files," in Information Forensics and Security (WIFS), 2012 IEEE International Workshop on. IEEE, 2012, pp. 139–144.

[10] Phil Harvey, "Image::ExifTool: Módulo para processamento de metadados de arquivos de mídia". Disponível em: http://www.sno.phy.queensu.ca/~phil/exiftool/, 2013.

[11] Google, "Google Maps API". Disponível em: http://code.google.com/apis/maps/.

[12] C. Santos, R. Bezerra, J. Ceron, L. Granville, e L. Rockenbach Tarouco, "On using mashups for composing network management applications," Communications Magazine, IEEE, vol. 48, no. 12, pp. 112–122, 2010.

[13] Y.-F. R. Chen, G. Di Fabbrizio, D. Gibbon, S. Jora, B. Renger, e B. Wei, "Geotracker: geospatial and temporal rss navigation," in Proceedings of the 16th international conference on World Wide Web, ser. WWW '07. New York, NY, USA: ACM, 2007, pp. 41–50. Disponível em: http://doi.acm.org/10.1145/1242572.1242579.

DOI: 10.5769/C2013009 or http://dx.doi.org/10.5769/C2013009

Making Sense of E-Government development in Saudi Arabia: A Qualitative Investigation

Osama Abdulaziz Alfarraj[1], and Thamer Alhussain[2]

(1) Computer Science Department, Community College, King Saudi University, Riyadh - Saudi Arabia,
Email: oalfarraj@ksu.edu.sa

(2) E-Commerce Department, College of Administration & Financial sciences, Saudi Electronic University,
Email: talhussain@seu.edu.sa

Abstract—***The implementation of eGovernment across countries is rapidly increasing. However, with this increase in the development of eGovernment projects especially in developing countries such as Saudi Arabia, there are still some difficulties facing the proper development of eGovernment. This paper aims to explore how eGovernment implementation and development can be understood in the context of Saudi Arabia based on the developers' perspectives. An attempt is made to identify the factors influencing the development of eGovernment and contribute to cause the delay of its initiatives at government organisations in Saudi Arabia. To achieve the aim, an unstructured interview within a qualitative approach was adopted in this study. Grounded theory techniques based on Strauss and Corbin approach (1990) were employed in this study in order to analyze the collected data.***

Keywords-eGovernment; Saudi Arabia; Development; Grounded Theory Techniques

I. Introduction

The focus of this study is the eGovernment concept in Saudi Arabia as one of these countries that still facing difficulties in implementation of its eGovernment. In actual fact, the Saudi Arabian government has already commenced implementation of its eGovernment concept named "Yesser" in 2005 [1]. Yesser is an umbrella for all eGovernment activities, procedures, legislations and other related issues and acts as the government's controller. The program has been launched and regulated in cooperation with three entities, which are the Ministry of Communication and IT, the Ministry of Finance and Communication, and the IT Commission [1]. Therefore, some eGovernment facilities are already in place. However, the duration of the eGovernment program, which has been set by the Saudi government, was not seem to be enough to achieve the expected outcomes according to what has been done so far and published in the literature. In particular, the Saudi government's clear statement regarding eGovernment, mentioned by several researchers such as [2, 3] as well as in several websites such as the Yesser eGovernment website (the official Saudi eGovernment website launched for the purpose of eGovernment implementation), asserted that, "*By the end of 2010, everyone in the kingdom will be able to enjoy from anywhere and at anytime – world class government services offered in a seamless user friendly and secure way by utilizing a variety of electronic means*" Yesser Vision.

It is now 2012; Yesser eGovernment program has changed its vision from offering electronic services to be supporting the infrastructure projects especially at the government organisations due to the noticed weakness in the infrastructure at public sectors [1].

II. Aim and significance of this study

The eGovernment phenomenon has become a wide area for research and study [7]. Yet, despite this emphasis on the concept of eGovernment in the literature, there is still a lack of research, especially on the factors that impede its applications and the reasons for this, specifically in Saudi Arabia [4, 5, 6]. Much of the published research regarding eGovernment in Saudi Arabia was considering the adoption side to the concept of eGovernment. However, most of the reviewed literature in relation to the eGovernment implementation at government organisations in Saudi Arabia was very few and their outcomes were as an expectation for the factors that might affect eGovernment during implementation process because the program of eGovernment has not accomplished during conducting previous research. Furthermore, most of the previous research about eGovernment implementation in Saudi Arabia used different research approaches which sometimes play role in reaching the results and clarifying the phenomenon being studied. In this study, the factors that influencing the implementation and development of eGovernment will be explored from the view of point the people who involved in the implementation of eGovernment and we call them here as developers.

III. Research Methodology

This section provides information about the methodological stance that will adopt it in this study. This study adopts the unstructured interviews method within a qualitative approach. Moreover, the techniques of grounded theory based on the approach of Strauss and Corbin (1990) [8] were employed to analyze the collected data.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

A. Interview method

Qualitative interviewing is a type of interview method that is often associated with qualitative research and the one adopted in this study. It is not just a normal interview that stressed interviewing skills as it has generic characteristics, which include the flexibility in style of interview, focusing on people's actual experiences more than general beliefs, and stressing the relationship between the interviewer and interviewee that are considered as crucial to the method [9].

Twenty one in-depth interviews were conducted with different groups of participants involved in the implementation and development of eGovernment. These groups include IT managers, IT experts, members from eGovernment program, and IT academics engaged in the development of eGovernment. The current study adopts purposive or purposeful of sampling as it is considered to be the best for this study within a qualitative approach. Sampling in grounded theory is called 'theoretical' by most of researchers rather than 'purposeful' however, the two terms are interchangeable [10].

B. Grounded Theory Techniques

As mentioned, this study adopted the techniques of grounded theory derived from the approach of Strauss and Corbin (1990) [8]. As identified in the literature there are four main approaches/types of grounded theory used within IS research (as illustrated in the Table 1) and analytic which is the use of grounded theory technique is one of them.

Using of grounded theory techniques here as an Analytical method, means using only the techniques and procedures of grounded theory to analyze the collected data and generate meaning for the area under study. The usage of grounded theory techniques for coding can be employed any or all of the three phases of coding (open, axial, and selective) and it does not required for multiple rounds of interviews as well as it does not require to stick with any particular formulation of grounded theory [11, 12]. Researchers using this approach usually come up with diagrams that explain the situations, events, people, and activities being researched through defining the relationships between categories and concepts that formed by codes and then create understandable meaning of this.

TABLE I. Four grounded theory approaches used in IS research

Approach	Principles	Coding	A priori Theory	Paradigm model	Typical Refs
Glaserian	Required	Open, Selective	No	Viewed as family of codes	Glaser & Strauss (1967); Glaser (1992)
Straussian	Required (Glaser disputed adherence)	Open, Axial, Selective	No	Greater emphasis	Strauss & Corbin (1990, 1998)
Analytical	Not necessarily	Any or all used	Maybe used	Some times used	Variety
Mixed	Not necessarily	Any or all used	Maybe used	Some times used	Mingers (2001)

Source: [38]

IV. The use of grounded theory procedures to analyze the Data

Next sections will briefly explain the used of grounded theory techniques and procedures.

A. Open Coding

It is called initial coding which is the first phase/step in coding collected data. It is defined by [8] (p. 61) as "the process of breaking down, examining, comparing, conceptualizing, categorizing data". Data in this phase of coding is broken down into small pieces in order to manage it and conceptualize it through assigning a label to it that represent its meaning [8].

In this study, open coding is considered as an initial step in the analysis process. A total of 320 codes were emerged and created based on 21 interviews. Two methods of coding were employed which are (i) In Vivo as referring to using the codes and terms that participants assign to their ideas and concepts during the interviews in order to preserve participants meaning regarding their views [13] and (ii) Simultaneous Coding as referring to "the application of two or more different codes to a single qualitative datum, or the overlapped occurrence of two or more codes applied to sequential units of qualitative data" (p. 55) [14].

B. Axial Coding

It is the next procedure in grounded theory that comes immediately after the open coding step where the process of putting data back together takes place in this step in order to make connection and links (relationships) between categories [8]. It is also called theoretical coding where the process of referring sub-categories to their categories and

making relationships among them is taking place in order to start creating meaning [16, 11]. This meaning should reflect what the empirical data is about regarding the reasons caused the delay in eGovernment implementation.

In this analysis phase, codes were refined to find out core codes in order to compare these codes to others for the purpose of finding similarities and differences in terms of concepts that can be placed together within sub categories. The total major categories created in this phase of coding and after refining the categories are twelve major categories and given the names of cooperation and collaboration, organisations and needs at organisations, IT professionals and IT skills, eGovernment implementation and challenges, awareness and training, provision of electronic services, education about the concept of eGovernment, financial allocations and incentives for IT staff, regulations & procedures and plans, e-readiness, ICT infrastructure, motivators. These main categories presented in Figure 1.

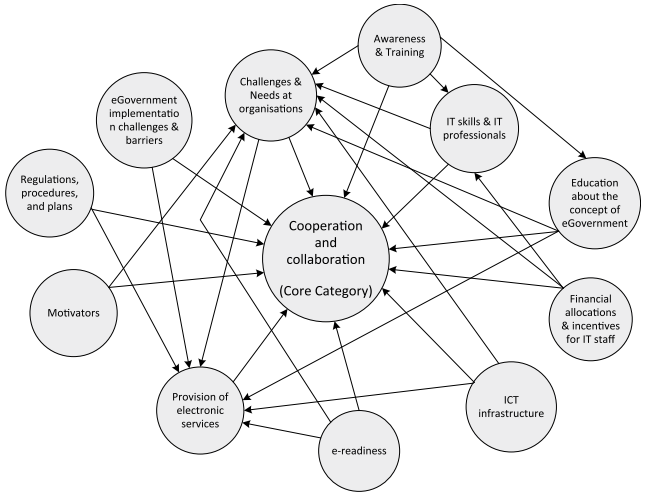


Figure 1. Core category and relationships

C. Selective Coding

Selective coding or focused coding is closely similar to the axial coding but here it is on more abstract level [15]. The aim of this step of analysis is to find out the central category among created categories which will become the central of the research phenomenon and other categories will be the causal conditions which basically the factors that influencing and caused the core phenomenon [8, 17, 18].

Determining the core phenomenon which will be the central and core category is based on showing the stress of the concept in the data through finding out how frequently the concept appears in the data. However, determining the frequencies based on the number of participants who mentioned particular concept rather than the number of times a concept appears in the data [19].

The concept of 'cooperation and collaboration' was mentioned and stressed by sixteen participants and determined in this study to be the core concept (core category) as indicted in Table 2 and illustrated in Figure 1.

TABLE II. Determining the most frequent concept in the data

No.	Main concepts / Core codes	Number of participants
1	Cooperation an collaboration	16 out of 21
2	Challenges & Needs at organisations	15 out of 21
3	IT skills & IT professionals	14 out of 21
4	eGovernment implementation challenges & barriers	14 out of 21
5	Awareness & Training	14 out of 21
6	Provision of electronic services	14 out of 21
7	Education about the concept of eGovernment	13 out of 21
8	Financial allocations & incentives for IT staff	10 out of 21
9	Regulations & procedures	10 out of 21
11	e-readiness	9 out of 21
12	ICT infrastructure	8 out of 21
13	Decision-makers & top management	8 out of 21
14	Delay in applying eGovernment	7 out of 21
15	Motivations	7 out of 21
16	Roles & responsibilities	6 out of 21
17	Strategies and Plans	6 out of 21
18	Support of Yesser	5 out of 21
19	Utilizing the experiences of eGovernment	4 out of 21

V. The use of grounded theory procedures to analyze the Data

This section will discuss the factors that been found in the empirical data to have an influence the implementation of eGovernment at government organisations in Saudi Arabia according to the results of analysis.

A. Cooperation and collaboration

The category of cooperation and collaboration is found in this study as the core category/ the central phenomenon which has relationships with all other surrounding categories as shown in Figure 1. The category of cooperation and collaboration contains sub-categories and core codes where placed into this category as they all relate to the same concepts of cooperation and collaboration.

The cooperation and collaboration in its all aspects as mentioned by the most of participants in this study are the main and important factors that have an influence on the implementation of eGovernment initiatives at government sectors and especially the first factors that contributing to cause the delay in the implementation of eGovernment. Discussing the factors of cooperation and collaboration will be done through the following sub-sections.

1) Cooperation and collaboration between government sectors.

Cooperation and collaboration between government sectors/agencies in terms of sharing data, services, experiences in eGovernment, and developing eServices are importantly needed for eGovernment implementation in Saudi Arabia

as the proper implementation for eGovernment projects cannot be performed without the cooperation and help of government sectors with each other. An IT expert working in Al-Elm Company explained the need for cooperation between government sectors by saying that '*Cooperation between government sectors to develop the services is needed, because in most cases offering any service requires obtaining and collecting information from more than a government body*'. Another participant from the same company showed an example on the importance of cooperation between government sectors in exchanging the required data for providing eServices by saying that '*one of eServices that we are currently developing is informing marriage, when someone gets married he needs to certify and authenticate that at the ministry of justice and getting a family card issued by agency of civil affairs. So, ministry of justice supposes to exchange the information of marriages electronically with the agency of civil affairs to ensure accuracy of the data and complete the transaction*'. Therefore, the extent of collaboration between the government sectors is very important because If there is one of the government sectors not happy and desire to provide such information then the transaction and the service won't be complete it.

2) Lack of cooperation between government sectors.

The previous point showed the importance of cooperation between government sectors to enhance the implementation of eGovernment projects because some of these sectors do not properly cooperate with each other for the same regard as indicated by some participants. An IT expert in Al-Elm Company discussed the experience of his sector about the cooperation with other government sector by saying that '*We have a cooperation with an important sector in the country and actually we are a part of that sector and we are the only sector authorized to communicate with them to supply and support with the needed information for building electronic services for other government sectors in the country but unfortunately we have an inactive cooperation with them and sometimes we need to wait lots of time to get their response in things that we need*'. Another IT manager at King Saud University also showed the lack of response for cooperation by another government sector in making the electronic link that will facilitate the exchange of data between the two sectors by saying that '*The preparation for linking in our environment is done now and tested to make sure is ready but still waiting to test the connection in their side –another government sector-*'.

3) Lack of cooperation with Yesser program.

it is one of the main factors among the cooperation and collaboration factors that influencing the implementation of eGovernment and caused the delay in its initiatives as indicated by the empirical data. One of e-services project managers in Yesser program was referring the very low process in connecting and linking the government sectors with Yesser to the lack of cooperation by saying that '*The linking process of government sectors with Yesser was very slow caused by the lack of cooperation from some of the government sectors. So, I am neither optimistic nor pessimistic, but we only achieved around 40% of what we have planned to reach*'.

From the government sectors side, one of the IT managers at General Directorate of Education in Riyadh mentioned the cooperation with Yesser by saying that '*There is no direct cooperation with the Program of Yesser and if so then it should be via the Ministry of Education*'. Another IT manager at ministry of justice expressed their relationship with Yesser by saying that '*our relationship with Yesser has only started few months ago*'. This means the cooperation of that ministry with Yesser has started late for unknown reasons.

4) Plans, strategies and changing procedures for cooperation.

Setting up strategies and plans for cooperation between government agencies is another (cooperation and collaboration) factor that influencing the implementation of eGovernment projects. Such strategies and plans can draw the roadmap for government sectors in relation to the cooperation in implementing of eGovernment and what is needed for that. One of the IT experts from Al-Elm Company was expressing their relationship of cooperation with Yesser by saying that '*Let's say that there is no direct relationship in a clear model with Yesser program*'. Yesser program is acting as a controller and enabler for eGovernment at government sectors while Al-Elm Company is working as a developer for eServices for both government and private sectors. These two sectors are currently the main ones needed for helping the government sectors to implement eGovernment projects. Yesser as the one that has the full responsibility for eGovernment implementation and authorised directly by the government should have plans of cooperation with Al-Elm Company in terms of drawing plans and setting up strategies that can help the implementation of eGovernment at government sectors.

5) Understanding the cooperation concept for eGovernment implementation.

The cooperation and its purpose to implement eGovernment have to be understood by all government sectors. An e-Services project manager at Civil Affairs Agency discussed the importance of understanding the concept of cooperation by showing an example as he stated that '*in the case of death if it is recorded directly then other related sectors can do their duties towards this died person and in the same time can prevent any kind of misusing for his/her identity in undesirable purposes. So, cooperation of Health Ministry will benefit other sectors and cooperation of other sectors will benefit the Ministry of Health. This concept of cooperation has to be understood by all sectors because it is a collaborative work more than an individual work*'. He also described the status of government sectors without the cooperation in implementing eGovernment by saying '*The concept of electronic government*

is a collaborative work between government sectors... the government sectors before moving towards e-government they were like islands every sector works for itself.. Meant doing only its responsibilities'.

6) Cooperation of top management.

Top management plays a great role on accelerating or delaying the implementation process of eGovernment at government sectors. An IT manager (e-services project manager) at the ministry of higher education stated that '*Top management has a very important role in the process of accelerating or delaying the implementation eGovernment. So, it has to be supportive*'. He also mentioned that '*The main factors in the success of e-transformation is the commitment for senior management ... if the leader (Manager - Chairman - Minister) has a background knowledge of the true benefit of electronic transformation then the influence will be noted on the entire sector. For instance, here in the ministry –ministry of higher education- we are supported by the minister*'.

The support of top management is very important and especially the support of top management given to IT department to complete eGovernment projects. An IT manager at ministry of justice mentioned that '*Direct support and confidence from the minister given to the manager of information and communication department at the sector is very important to help removing some of the obstacles that can face the development and implementation processe*s'. The projects of eGovernment would not be possible to implement without support of the top management within the organisation.

7) Cooperation of financial departments at government sectors.

Funding the eGovernment projects at government sectors is the responsibility of the financial departments within organisations and these departments in most cases do not give the IT projects a proper care. An IT manager at the ministry of education stated that '*IT projects in government sectors take long time to get approved by top management and finance department compare to the private sectors where the IT projects get a high priority*'. Another IT manager at King Saud University mentioned that '*Last few years, King Abdullah has command to allocated huge budgets to government sectors in order to help them in E-transformation however, some financial departments within government sectors are not helping the IT projects to be implemented in specified time regardless the poor equipment that provided for such projects due to they looking for lowest offers provided by IT businesses*'.

8) Cooperation with researchers.

Cooperation of government sectors and Yesser program with universities and academic centers is important. It can enhance the implementation process of eGovernment through finding out the challenges that can affect the implementation as well as the efficient ways to introduce new concepts to the workplace such as eGovernment. One of the IS academics at King Faisal University stated that '*I can say that there is a lack of benefit from academic research regarding eGovernment implementation and making a partnership with universities*'. Another IS academic at King Abdulaziz University mentioned that '*Sometimes we send many emails to some government sectors like the ministry of commerce asking for a statistical info to help us in doing research but unfortunately we usually do not get response from them*'. An IT manager at ministry of defence thought that '*Government agencies have to employ and host researchers to find the good ways for applying technologies such as eGovernment*'.

B. Challenges and needs at government organisations

The category of challenges and needs at government sectors is referred to the issues and concerns that been found in the empirical data to have an influence on the implementation of eGovernment projects within government organisations. These issues are considered to be challenges and needs facing government organisations while implementing eGovernment projects. This category is one of the major categories that surrounded the core category of 'cooperation and collaboration' as illustrated in Figure 1.

1) Understanding the concept of eGovernment.

As discussed in previous point (V.A.5) the concept of eGovernment has to be understood by both employees and top managements within organisations because most of the participants indicated that there is a wrong understanding for the concept of eGovernment across government sectors.

2) Change management.

Resisting the change to eGovernment at government sectors was indicated by some participants. One of the e-services project managers at Yesser program stated that '*there is a resistance noted from some government sectors regarding the implementation of eGovernment*'. Another e-services project manager at Yesser program mentioned that '*Some of the expected reasons that led to eGovernment delay include the change resistance and effectiveness of the efforts regarding eGovernment implementation*'.

An IT manager at ministry of Islamic affairs was referring the delay on the implementation of eGovernment projects to the change management issue as one of the main issues influencing the eGovernment implementation in Saudi Arabia. He stated that '*It is the issue of change management within the government sectors more than any something else. We do not have problem with financial resources or technical aspects because they are available*'. He continued saying that '*The change management is really needed for change strategy because it is the most difficult aspect as I said the financial and technical aspects are available and remains the human aspect which is the hardest part of the equation*'.

3) Technical departments at government sectors.

The implementation of eGovernment projects or any IT projects within government organisations would not possible without the help of the IT qualified staff working at IT technical departments. Therefore, the absence of the active role for these IT departments would have an influence on the implementation of eGovernment and its initiatives within government organisations and in general leads to the delay in the implementation process for eGovernment projects.

An IT manager at the ministry of education stated that '*Technical departments in all different government agencies are supposed to be more developed than what they are now. This is due to the lack of qualified IT staff as well as poor training in the Information sector at different government departments*'. Another IT manager at justice ministry mentioned that '*In the past there was no IT department at the ministry that can rely on for electronic transactions, but before five months ago we have established one*'. As it can be seen, some government organisations were not having IT departments and even more some of these sectors lack for appropriate IT equipment. The IT manager at justice ministry said that '*Currently we do not communicate electronically with other government sectors because the ministry does not have a data center that can be relied on*'.

4) The ownership & reputation.

Another issue that been identified in the empirical data as to have an influence on the implementation of eGovernment projects at government sectors is the ownership of eServices and the gained reputation credits for developing the eServices. Usually, the development of eServices is a shared task between more than a government body as to have cooperation between sectors in designing the service, obtaining the required data for building the service, and implementing the service. through the implementation and development of an eService each government sector involved within the development process look for owning the eService that been developed and gaining the credits in front of public for doing such thing by referring the development of that eService to them. It has happened only with few sectors but still be considered as a hinder for those sectors to cooperation again to develop other services. An IT expert at Al-Elm Company stated that '*The implementation of eGovernment in Saudi Arabia is very slow and this is due to several problems, including the ownership of the service and the reputation credit of doing this should goes to whom? especially if several sectors and departments are involved in the development and designing of an electronic service*'. He also mentioned that '*Participation in the development of eServices to get the credit and the ownership of eServices are the main problem that we are facing right now in our eGovernment projects with government sectors especially the services which need for cooperation of different parties*'.

5) Sincerity in work.

Another influencing issue on the implementation of eGovernment projects at government sectors as been found in the empirical data is the absence of sincerity in work or in other words, the lack of feeling about the country benefit from the implementation of eGovernment projects by employees and top managements within government organisations. An IT expert in Al-Elm Company stated that '*There is a lack of feeling about citizens' pains by responsible managers and decision-makers at vital government sectors as they think of that as normal things that happen every day so, their view of e-transformation is something that not urgent and it does not deserve to be given a high priority*'. An IT consultant at ministry of higher education mentioned that '*I think government sectors do not lack for budgets and money rather than feeling of doing this for country benefit. We need sincerity in work*'.

6) Electronic Systems.

The absence and weakness of electronic systems is another issue that been identified in the empirical data that have an influence on transferring data from government sector to another which affecting the implementation of eGovernment projects. An IT expert at Al-Elm Company said that '*One of the problems that we are facing is the weakness and the absence of electronic systems in government sectors that can be used for linking these sectors with each other to electronically transfer the data*'. Another sub-issue is the differences in databases used for electronic system across departments of a one sector. An IT manager at ministry of justice mentioned the same issue as one of the challenges that justice ministry has faced during the implementation of eGovernment projects by saying that '*We had the second problem in distributed databases among ministry departments across the country which has a different encryption*'. Moreover, the lack of integration between electronic systems used within organisation is another sub-issue that been identified to have an influence on the implementation of eGovernment at government sectors. Another IT consultant at the ministry of higher education stressed on the integration between the system used within the organisation by saying that '*There is something important has to be done at government sectors which are the integration between systems used in the same organization*'.

Understanding the used systems within government organisations by employees is also another sub-issue. An eServices manager and IT expert at at the ministry of higher education stated that '*Another problem is that understanding systems by employees ... some employees they do not understand why some procedures for doing services have been changed within electronic transactions because they used to do paper transactions in different way and different procedures*'.

C. IT professionals and IT skills

This category of IT professionals and IT skills is one of the major categories that surrounded the core category of

'cooperation and collaboration'. It is referred to the issues such as IT skills required for government employees and qualified IT staff that can help in the implementation of IT project within organisations.

1) IT skills.

Having IT skills and knowledge about IT by employees and top managements within government sectors is important to help them contributing on the implementation of eGovernment projects through using and utilizing the new technologies that been introduced to the work environment. Some participants indicated that government employees need for training on IT as they lack for IT skills. An IT manager at the electronic services unit at general directorate of education in Riyadh stated that '*In our sector we have a lack of IT skills to some employees who need for comprehensive training sessions*'. Moreover, another IT manager at the ministry of commerce stressed on the issue of IT skills to government employees and top managers by saying that '*We have a big number of employees who do not know how to deal with technologies even computers and unfortunately the majority of them are managers*'.

2) IT professionals and human resources.

The lack in IT qualified staff at government sectors were indicated by some participants and they mentioned that it has an influence on the implementation of eGovernment. One participant from Yesser Consulting Group stated that '*There is a significant lack of human resources and expertise necessary for the transition to electronic transactions at government sectors beside the lack of readiness of these sectors*'. An IT manager at higher education ministry mentioned the impact of the absence of IT qualified staff on the implementation of eServices by saying that '*I also expect the absence of specialists in information technology had a major impact on the implementation of electronic government*'. More participants indicated the lack of IT qualified staff working at government sectors. An IT manager and the e-services manager at Civil Affairs Agency stated that '*There is no doubt we have a lack of human resources as they are very few*'. An e-services project manager at Yesser mentioned the same issue by saying that '*IT departments in government sectors lack for IT people as we saw some departments have only two people and that is not enough*'.

3) IT professionals and financial departments.

Some participants believe that financial departments need to have IT professionals who can help in taking right decisions about the IT projects because most of working staff at these departments are not fully aware about technology. An IT manager at the ministry of higher education stated that '*Normally people who work in financial departments within government sectors are not aware in technology so sometimes they take incorrect decisions regarding IT projects or even they take long time to accept funding the IT projects*'. Another IT manager at King Saud University mentioned that '*Financial departments within government sectors need to have IT professionals who have knowledge and experience in IT projects and who can decide on what is needed for their sectors*'.

Having IT professionals working at financials departments in government sectors would help the decisions taken regarding IT projects by the organisations as such people have a good background and knowledge to decide on good offers. IT projects are not like other projects which look for the lowest prices from offers however, it need for a high features and requirements. Having a poor equipment would affect negatively on the implementation.

D. eGovernment implementation challenges and barriers

The category of eGovernment implementation challenges and barriers is referred to the issues that been found in the data to have an influence on the projects of eGovernment at government sectors in cooperation with Yesser program. This category is one of the major categories that surrounded the core category of 'cooperation and collaboration' as presented in Figure 1.

1) Following up the IT projects.

Following up the projects of eGovernment at government sectors is essential because it can help in maintaining the cooperation of government sectors with Yesser as well as determining easily the level of readiness at these government sectors. An e-services project manager at Yesser program stated that '*monitoring the process of projects at government sectors or what so enrolled is important*'. In order to monitor the status of eGovernment projects at government sectors, an eGovernment projects manager should be appointed in each government sector whether by the government sector itself or by Yesser program. This action would help Yesser to easily communicate with these sectors through those representatives (project managers) in regard to the projects of eGovernment. An e-business analyst working in Yesser program stated that '*if there is a project manager from Yesser or from any government sector that can observe the integration of such projects especially IT projects it will be better*'.

2) Believe in change.

Belief in electronic services by government sectors is pointed out by some participants as one of the current obstacles that facing the implementation of eGovernment. One of the IT experts at Al-Elm Company stated that '*In my view, one of the current obstacles that facing the implementation of electronic government is the belief in electronic services and its importance by government agencies*'. He also mentioned that '*the second problem that facing the implementation of eGovernment is the extent of government agencies willingness to change because some sectors in the country initially accept the change, but when you tell them that implementation*

requires a set of procedural changes they reject because they fear of change'.

Willingness to change towards eGovernment by government sectors is required to effectively cooperate with Yesser program and other related sectors to implement the projects of eGovernment. This willingness to change can be obtained through believing these government sectors in electronic services can enhance the work and reduce the workload.

3) The 150 selected government services.

It is one of factors that been identified in the empirical data as affecting and influencing Yesser plans and strategies during the implementation of eGovernment. Specifying the electronic services with 150 services that Yesser wants to start implementing at Government sectors without the engagement of the government sectors is the issue here. One participant from Yesser mentioned that one of the main problems that delay our projects with government sectors regarding eGovernment is depending on previous plan which contained developing 150 services for public sectors. An eServices project manager at Yesser program mentioned that '*The program of Yesser started with a plan consists of 150 services which was a mistake from my view of point. Because how come you limit and specify the services with 150 service without consulting and engaging the government sectors to see whether these services are main ones to them and are important to start with or there is something else more important*'. Moreover, an e-business analyst at Yesser stressed on the same issue by saying that '*I think we have 1000 services but we have not reached the 150 that Yesser has specified within plan. Also, a lot of these 1000 services are informative services & enquires*'.

Yesser after five years since the beginning of eGovernment program has noticed that this plan was not successful anymore as it goes wrong with desire of government sectors. Because the majority of government sectors did not see the proposed and suggested services are main ones and important for them to implement because they got other services which have more priority to start with. An eServices project manager stated that '*The good thing in the new five-year plan from 2010-to-2015, we are not going to depend and stick with 150 services .. We are going to work with each government sector and see which services they want to develop by now and start with*'.

4) Partnership strategy with private sectors.

The implementation of eGovernment projects which basically include planning, designing, and implementing electronic services as well as preparing the ICT infrastructure needs the involvement of private sectors and especially technical sectors to assist in this regard. An IT consultant at the ministry of higher education asserted on this issue by saying that '*I think it is important to have the idea of partnership with the private sector to complete eGovernment projects*'. An e-business analyst at Al-Elm Company showed the need for involving the private sectors within the implementation of eGovernment by saying that '*Al-Elm Company has a direct relationship with some government sectors in the country for the implementation and provision of electronic services which, as I said previously it is a profitable company which looks at the profit at the first place because there is no alternative, meaning that the government by itself can not launch the electronic services without getting technical sectors such as Al-Elm company involved in such projects*'

The involvement of private sectors is important especially at the current time where no enough IT qualified staff existed within government sectors as discussed in section (V.C.2) and the current weakness at the IT departments within organisations as discussed in section (V.A.1).

5) Documentation of procedures and processes.

It is very important point that most of the government sectors while shifting to eGovernment do not document the processes and procedures that have been done regarding the implementation of eGovernment initiatives for future development. An e-business analyst at Yesser program mentioned that '*some government sector or almost all of them they do not document their processes and nothing regarding procedures*'. Another participant from Yesser who is an eServices projects manager stressed on this issue by saying that '*It is important to document the procedures that have been done within eGovernment projects to make sure the progress of such projects would not be affected with a leave of individuals who were responsible for projects*'.

E. Awareness and training

This category of awareness and training encompasses factors that been found in the data to be related to the issue of awareness and training at government sectors. This category is one of the major categories that been identified through the empirical data as illustrated in the Figure 1.

1) Awareness for employees.

Providing training and awareness through running the workshops and educational sessions are essentially required by both Yesser to all government sectors and government sectors to their employees.

Most of the interviewed participants indicated that there is a lack of awareness and training about IT at government sectors. One of the IT managers in the electronic unit at General Directorate of Education in Riyadh mentioned the lack in the awareness about using technologies at government sectors by saying that '*We have a lack in the awareness about the use of technology for public employees across government sectors*'. Another IT manager at Planning and Information Affairs division at the ministry of higher education stressed on the issue of awareness for employees by saying that '*Lack of employees' awareness of the expected benefits for*

eGovernment is also one of the difficult issues that facing eGovernment implementation'. An IT manager at the ministry of Islamic affairs added that '*Government sectors should focus on aspects of IT awareness and training*'. Providing awareness and training helps employees to understand the electronic systems they used and the purpose for its use.

2) Awareness for managers and decisions-makers.

As an extension of what has been discussed in the previous section regarding the need for providing awareness and training to government employees, this section will highlights the need of awareness and training for managers and decision makers at government.

One of the academics in IS field at the King Faisal University indicated the need for providing awareness and training especially to managers because they play a great role on affecting on the entire sector therefore, they must be educated. He stated that '*There must be awareness and education programs about what is the eGovernment for staff and managers because some managers do not want the change which can reflect on staff*'. Another participant from the electronic unit at General Directorate of Education in Riyadh mentioned that '*Decisions makers need for a lot of awareness about the usefulness of e-services and eGovernment*'.

F. Provision of electronic services

This category includes all concepts connected to the provision of electronic services that been found in the empirical data to have an influence on the implementation of eGovernment projects. This category is one of the major categories that created in the axial coding phase in the analysis. Specifically, it is one of the categories that surrounding the core category of 'cooperation and collaboration' as illustrated in Figure 1.

1) Designing of eServices and electronic systems.

The design of electronic systems or electronic services needs to be easy to use and much understandable for government employees especially with current lack in IT skills for employees at government sectors as discussed in section (V.C). One of IT managers at the ministry of education mentioned this issue which the complexity in the design of electronic services as one of the reasons caused the delay in the implementation of eGovernment. He stated that '*The complexity in the design and delivery of electronic services is a real reason to delay its implementation*'.

2) Linking government sectors.

It is another affecting factor that been identified in the data to have an influence on providing electronic services through the lack of linking government sectors with each other to exchange the required data to build and offer electronic services. One of important steps that government sectors need to do while implementing eGovernment is making and establishing the link with GSB (Government Service Bus) to start electronically communicate with other sectors in such easy and secure way. It is like an integration channel developed by Yesser to connect all government sectors through.

Some participants indicated the lack in process of linking government sectors with GSB which is associated with the lack of cooperation from the government sectors side. An eServices projects manager at Yesser mentioned that '*The linkage process of government sectors with Yesser was very slow caused by lack of cooperation from some of the government sectors*'. Another eServices project manager at Civil Agency mentioned the importance and the benefits for linking the agency of Civil Affairs sector with the ministry of health. He mentioned that '*Second project that we are going to implement is fallen under the umbrella of G2G, which is the linkage with the Ministry of Health ... of course the link with the Ministry of Health for registering two things which are the birth and death*'. He continued that '*Creating the link with the ministry of health helps to raise the efficiency of information, speed up the registration of the information and inform related sectors instantly to do their roles*'.

3) Privacy and security.

Privacy and security are important issues that always associated with developing electronic services. As indicated by some participants in this study that there is a need to have a privacy officer in each government sector that can review the electronic service before offering it to ensure it would not breach the privacy of others. An IT expert at Al-Elm Company explained the reason for the presence of a privacy officer in government sectors by saying that '*The importance of having a privacy officer in all government sectors is to ensure reviewing and studying the eServices before offering them*'. Moreover, awareness about privacy and security while designing and offering electronic services should be disseminated across government sectors because there some sectors do not believe in such thing as indicated by the same participant. He stated that '*There is some government sectors are aware about privacy policy while others do not believe in such thing*'.

G. Education about the concept of eGovernment

The category is one of the major categories that been created at the axial coding phase during the analysis as shown in Figure 1. Understanding the concept of eGovernment by the top managements and employees within government organisations is very important need to enhance the implementation process of eGovernment projects.

1) Lack of education about eGovernment.

Education about eGovernment is importantly needed at both organizational and national levels because there is a lack of knowledge about eGovernment program Yesser as

indicated by a participant. An IS academic at King Faisal University stated that '*Last year, I did a research and I asked the people a question if whether they know Yesser and if they visit the national portal website for Saudi Arabia and the responses of 120 people were no which form 85% of the total sample*'. In particular, some other participants indicated that there is a misunderstanding about the eGovernment and what is really about at government sectors. An IT manager at the ministry of education explained this misunderstanding in terms of implementation of eGovernment by saying that '*some government agencies are misunderstood the meaning of eGovernment and they thought that their duty is only uploading the application forms online and let citizens download them*'. Another IT expert at the Al-Elm Company explained another misunderstanding for eGovernment in terms of impact by government employees by saying that '*there is a misunderstanding by government sectors employees for the aim of eGovernment as in the most cases they thought that it means reduce the number of staff* '. The real meaning of eGovernment should be clearly disseminated across government sectors because some employees at these sectors have a misunderstanding for such concept.

H. Financial allocations and incentives for IT staff

Financial allocations & incentives for IT staff category refers to the issues and concerns of financial allocations and incentives specified to IT staff that been found in the empirical data to have an influence on the existence of IT staff at government sectors. This category is one of the major categories that been identified at the axial coding phase during the analysis as illustrated in Figure 1.

1) Provision of financial incentives.

According to what has been found in the empirical data, this factor play great role in motivating those IT staff at government sectors to work forward with Yesser program and other government sectors in regard to the implementation of eGovernment.

The existence of IT staff at IT departments within government sectors are important because they are the key elements to help in the implementation of eGovernment projects. Therefore, IT staff needs to be motivated financially and morally. One of the IT managers at King Saud University explained this issue by saying that '*Actually, these incentives are whether financial or even appreciations motivate government officials to work effectively and learn new things*'. An IT manager at the ministry of Islamic affairs mentioned that '*Normally, employees in the public sector lack for incentives which can motivate them to learn new things and make an effort to work*'.

Some participants indicated the lack of IT staff working at government sectors and they referred that to the lack of financial incentives and allocations specified to the IT staff. An IT manager at the electronic services unit at general directorate of education in Riyadh stated '*We have a lack of Saudi IT people in our sector and the reason is due to shortages of incentives and financial allocations*'. Another IT manager at King Saud University asserted the same issue by saying that '*The lack of experienced people in government organisations comes from the rarity of incentives and allocations specified for those IT specialists working in government sectors*'.

2) The need for increasing the salary scale for IT staff.

As extension to the discussion done in previous point, the existence of IT staff working at IT departments within government organisations is influenced by another factor which is the lack of salary scale specified to such qualified staff compared to the private sectors. An IT consultant at the ministry of higher education explained the reason behind the lack in qualified IT staff at government sectors by saying that '*We have lack of qualified IT people in the government sectors and the reason is due to the lack of the salary scale for such people*'. A participant from Yesser who is an eServices project manager referred the problem of existing of IT staff at government sector to the same reason by saying that '*Qualified IT people would not come to work in government sectors with salary of 6000 or 7000 Riyal while they can gain the double in private sectors*'.

I. Regulations, procedures and plans

Regulation, procedures and plans category refers to the issues and concerns that been found in the empirical data to have an influence on the delivery of electronic services.

The category is one of the major categories that been created in the axial coding phase and developed in the selective coding during the analysis as presented in Figure 1.

1) The complexity in procedures & the need for change.

Designing and offering electronic services is required to change or modify some traditional procedures that associated with providing services at government sectors in order to suit the new direction of eGovernment. An eServices project manager at the agency of civil affairs asserted the same issue by saying that '*Procedures and existing systems in the agency of civil affairs were built on a paper-based structure. So, we need to re-formulate these procedures in order to suit e-government direction*'. An IS academic at King Abdulaziz University mentioned the need for changing procedures before designing and offering electronic services as he stated that '*The work is automated and become electronic but, the procedures are still as in the traditional way. I mean administrative procedures must be changed and modified to facilitate providing electronic transactions*'. Another participant indicated the complexity in procedures for doing services with government sectors as an important issue affecting the implementation of eGovernment. An IT manager at the ministry of higher

education stated that the current procedures for doing transactions with government sectors is the problem because it is complicated and complex. He mentioned that '*The problem is the current administrative procedures for existing transactions with government sectors which seem to be very complicated. So, the complexity of these procedures makes it difficult to join eGovernment in such a quick way*'.

2) The need for unifying procedures.

Unifying procedures for doing the same service/transaction at different places is one of the problems that facing Yesser with some government sectors during the implementation of eGovernment. One of the eServices project managers at Yesser program mentioned this issue by saying that '*We are trying with Ministry of Municipal and Rural Affairs to unify the procedures at Amanahs and municipalities because some procedures are different from municipal to another*'. An IT manager at Planning and Information Affairs division at the ministry of higher education asserted the same problem with the cultural missions offices and he has considered as an challenge and obstacle by saying that '*We have a problem with cultural attaches which is each attaché office has its own culture in terms of work procedures*'.

3) The lack of strategic plans.

It is one of the factors that been found on the data to have an influence the implementation of eGovernment. One of IT manager at King Saud University mentioned the lack in plans by saying that '*There is no strategic plan is adopted by high authorities in the country for E-transformation and can clarify the targets and objectives*'. He also indicated the lack of clarity in the national strategic plans for eGovernment by saying that '*The plan of the national strategic for E-Government is not clear .. Even Yesser has changed its program target this year from e-services providing to become supporting infrastructure projects*'.

Another IT manager in the electronic unit at General Directorate of Education in Riyadh stressed on the impact of the absence of clear regulations and plans on the implementation of eGovernment projects especially for decision makers by saying that '*The absence of clear regulations and plans for some of the leaders in the decision-making make the implementation of e-government more difficult*'.

J. E-readiness

E-readiness category refers to the ICT capabilities in government organisations that can assist these organisations in delivering of electronic services as well as enhancing the communication channels between the sectors involved in the implementation of eGovernment projects.

1) Provision of electronic communication channels.

As indicated by some participants that some government sectors lack for simple tools of communication that can assist employees to easily communicate with each other such as the email. One of the IT managers in the electronic unit at General Directorate of Education in Riyadh mentioned that '*There is a weakness in the internal communication between the staff of the sector through email as there are no official email accounts created by the sector for its employees*'. Another IT manager at the ministry of commerce mentioned the same issue which is the lack in providing email accounts to employees by saying that '*In our ministry we lack for official email accounts for employees and only few people in this sector using their personal hotmail accounts to communicate and complete the work sometimes*'.

2) The lack of e-readiness at government sectors.

It is one of the main influencing factors that affecting the implementation of eGovernment projects at government agencies. One of the IT experts at Yesser Consulting Group mentioned the variation in the e-readiness between government agencies and the reflected impact of that in the cooperation of government sectors with Yesser by saying that '*There is a variation between government sectors regarding the e-readiness which definitely reflect on the performance of these sectors and the extent of cooperation with Yesser*'. Another IT manager at the ministry of education asserted on the variation in e-readiness between government agencies by saying that '*Also the disparity of e-readiness between government sectors has another impact on eGovernment implementation in the whole country*'. Some participants from Yesser indicated this variation between government sectors in e-readiness as a problem currently existed that impedes Yesser to implement the eGovernment project in the specified time. An eServices project manager at Yesser stated that '*the problem here is dealing with different government sectors and the majority of these sectors are not ready to implement eGovernment*'.

K. ICT infrastructure

The category of ICT infrastructure included concepts that are related to the issues and concerns that been identified through the empirical data to have an influence on the ability and readiness of government sectors to provide electronic services and implement the projects of eGovernment. The category of ICT infrastructure is one of the major categories that been created at the axial coding phase during the analysis as shown in Figure 1.

1) Provision of ICT infrastructure.

Having high standards ICT equipment helps any government organization to successfully offer e-services with less or even no problems. According to one of the IT experts at Al-Elm Company said that '*There is a problem within providing electronic Services when the system is down you can doing nothing ... and this makes it essential to supply the appropriate equipment such as servers and others equipment in order to offer electronic Services with less or even no*

problems'. The role of providing the infrastructure is mainly the responsibility of government sectors however; Yesser is the one responsible to provide the shared infrastructure that can be used to link government sectors with each other such as GSB (Government Service Bus). An IT expert at Yesser Consulting Group mentioned that '*Yesser has a key role in the provision of a shared infrastructure, but the responsibility remains on the government sectors to develop their infrastructure at their agencies*'.

2) The weakness in the infrastructure.

The lack in the infrastructure was indicated by some participants. An IT manager at King Saud University mentioned the lack in the infrastructure for some government sectors by saying that '*Yesser is facing the problem of lacking in the infrastructure for many government sectors which impede it to link these sectors with each other as I know this from colleagues working in Yesser*'. An IT expert at Yesser Consulting Group stressed on the same problem by saying that '*There are problems in the infrastructure and they are deep, but these things are much easier than things that required the human side and change management comes under this*'. Moreover, some participants indicated the weakness of infrastructure at their organisations or at other sectors they deal with. One of the IT managers in the electronic unit at General Directorate of Education in Riyadh at the ministry of education stated that '*the current infrastructure is not ready yet and it needs for more development*'. Such delay in the development of the infrastructure has an influence the implementation of eGovernment projects on the same ministry and sharing data with other government sectors.

L. Motivators

The category of motivators refers to the all issues that been found in the data to have a positive influence on the implementation of eGovernment projects at government sectors. This category is one of the major categories that been created at the axial coding phase and developed in selective coding during the analysis as presented in Figure 1.

1) Having the intension to work towards eGovernment.

Having intention to shift to eGovernment by government sectors is essential in order to effectively and continuously cooperate with Yesser and other related sectors to implement the projects of eGovernment. An IT manager at King Saudi University mentioned the importance of having intention to work forward to implementing eGovernment implementation as to do more than what is in the plans. He explained that '*The University has taken strong steps, but we are still not satisfied because if you reach satisfaction then you are not working for development. We believe that we can do more ... we can do more. Actually, we are not looking for reputation in marketing and ranking however, we have a roadmap of initiatives that we are planning to achieve*'. Having the intention and desire to implement of eGovernment projects by government sectors leads to efficient cooperation of that sector with Yesser program and other government sectors involved in the same regard.

2) Engaging beneficiaries within decisions-making.

Engagement of beneficiaries within decisions-making in relation to the implementation of electronic services as well as the offered electronic services is important in order to meet the needs of targeted users from such services. Some participants indicated the efforts made at their organisations to obtain the feedback from the users regarding the offered electronic services for the purpose of development. An IT manager at King Saud University mentioned that '*Regarding students, university has implemented a special system for students called 'eRegister' which enable students to access various interactive services via online. Also, we are really keen to get students opinions regarding the system through asking students to fill a survey*'. Another eServices project manager at Planning and Information Affairs division at the ministry of higher education mentioned their experience in getting the feedback from students regarding electronic services offered by the ministry of higher education by saying that '*I can not assume that all students and employees are happy with electronic services provided by higher education … so we have designed a survey and planned to send it to all students overseas in order to measure their satisfaction about electronic services provided over the 'student portal' and asking for suggestions*'.

3) The support of Yesser.

According to the results, it is noted that the support and effort that Yesser has made in relation to the implementation of Government is in the right direction however, it lacks for the cooperation of government sectors as discussed in section (V.A). One of the IT managers at the ministry of defense mentioned the effective support of Yesser during the implementation of the eGovernment projects at the ministry. He stated that '*We contacted Yesser to ask for consultation and support in building a website and designing some electronic services, indeed Yesser's representatives came and helped us. I am trying to say that Yesser has an obvious efforts and good contribution in applying the eGovernment*'. He continued saying that '*Actually, Yesser and its team are really supportive and work well towards applying the eGovernment across government agencies and it has a big role in this regard*'.

4) Utilizing the experiences of eGovernment.

One of the good practices during and before the implementation of eGovernment projects is to benefit from the advanced experiences of others in eGovernment whether internal or external experiences. Internal experiences are from inside the country and external is the ones borrowed from the international experience under the condition of taking what only is suit the implementation environment. One of the IT

managers at the ministry of higher education stressed on getting the benefit from the local experience in eGovernment implementation and activate it in other sectors where needed to. He stated that '*Benefit from the experiences of the advanced sectors in the field of electronic government in the country and activate it into other sectors*'. He continued saying that '*I think that benefit from the expertise and international experiences in the field of e-government is something crucial and needs to be done*'. The results also show one of the successful local experiences in eGovernment which is the experience of higher education ministry. Such experience can be taken as a good example to be followed by other government sectors and Yesser has to play active role in this. An IT manager at the ministry of higher education stated that '*One of the good examples for eGovernment application is the ministry of higher education through its electronic portal designed for its students who studying overseas*'.

Conclusion

This study explored the factors that been found in the empirical data to have an influence on the implementation of eGovernment and contributing to cause the delay of its initiatives at government organisations in Saudi Arabia.

The results indicate that cooperation and collaboration factors are the main and important factors that currently influencing the implementation of eGovernment as well as contributing to cause the delay of its initiatives at government organisations in Saudi Arabia. The factors of cooperation are the main influencing factors that affecting the implementation of eGovernment projects at government sectors beside other identified factors such as lack of e-readiness at government sectors, lack of IT staff at government sectors, lack of financial allocations and incentives specified to IT staff, lack of strategic plans, lack of awareness and education about electronic services and the real benefits, lack of understanding the concept of eGovernment, and others more.

Most of the identified factors to have an influence on the implementation of eGovernment projects at government sectors can be overcome and solved through the effective cooperation between all government sectors involved in the implementation of eGovernment projects such as government sectors, Yesser program, and Al-Elm Company.

References

[1] Yesser Program Team 2010, *Kingdom of Saudi Arabia e-Government Program (Yesser)*, The Ministry of Communication and Information Technology, Riyadh, KSA, http://www.yesser.gov.sa.

[2] Al-Soma, A 2008, *Saudi e-Government Yesser Plans and Achievements*, Beirut.

[3] AL-Shehry, A, Rogerson, S, Fairweather, NB & Prior, M 2006, 'The Motivations for Change Towards E-Government Adoption: Case Studies from Saudi Arabia', paper presented to eGovernment Workshop '06 (eGOV06), Brunel University, West London, UK.

[4] Al-Shehry, A 2008, 'Transformation Towards E-Government in the Kingdom of Saudi Arabia: Technological and Organisational Perspectives', Doctoral thesis, De Montfort University, Leicester, UK.

[5] Altameem, T 2007, 'The critical factors of eGovernment adoption: An Emprical study in the Saudi Arabia public sectors', Doctor of Philosophy thesis, Brunel University.

[6] Al-Fakhri, MO, Cropf, RA, Kelly, P & Higgs, G 2008, 'E-Government in Saudi Arabia: Between Promise and Reality', *International Journal of Electronic Government Research (IJEGR)*, vol. 4, no. 2, pp. 59-85.

[7] Alharbi, S 2006, 'Perceptions of Faculty and Students toward the Obstacles of Implementing E-Government in Educational Institutions in Saudi Arabia', Master thesis, West Virginia University.

[8] Strauss, AL & Corbin, J 1990, *Basics of qualitative research: Grounded theory procedures and techniques*, Sage Newbury Park, California.

[9] King, N & Horrocks, C 2009, *Interviews in qualitative research*, Sage Publications Ltd.

[10] Cutcliffe, JR 2000, 'Methodological issues in grounded theory', *Journal of Advanced Nursing*, vol. 31, no. 6, pp. 1476-84.

[11] Matavire, R & Brown, I 2008, 'Investigating the use of Grounded Theory in information systems research', in *Annual Conference of the South African Institute of Computer Scientists and Information Technologists*, Wilderness, South Africa, pp. 139-47.

[12] Furniss, D, Blandford, A & Curzon, P 2011, 'Confessions from a grounded theory PhD: experiences and lessons learnt', in *roceedings of the 2011 annual conference on Human factors in computing systems*, Vancouver, BC, Canada, pp. 113-22.

[13] Charmaz, K 2006, *Constructing grounded theory: A practical guide through qualitative analysis*, Sage Publications Ltd.

[14] Saldaña, J 2009, *The coding manual for qualitative researchers*, Sage Publications Ltd.

[15] Niekerk, JC & Roode, J 2009, 'Glaserian and Straussian grounded theory: similar or completely different?', in *Proceedings of the 2009 Annual Conference of the South African Institute of Computer Scientists and Information Technologists*, Vanderbijlpark, South Africa, pp. 96-103.

[16] Dey, I 1999, *Grounding grounded theory*, Academic Press San Diego, CA.

[17] Creswell, JW 2008, *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*, 3rd edn, Pearson Education International.

[18] Creswell, J 2012, *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*, Third edn, Sage Publications, Inc.

[19] Namey, E, Guest, G, Thairy, L & Johnson, L 2008, *Data reduction techniques for large qualitative data sets*, eds G Guest & KM MacQueen, Altamira Press, UK.

DOI: 10.5769/C2013010 or http://dx.doi.org/10.5769/C2013010

Computer Forensic Laboratory: Aims, Functionalities, Hardware and Software

Paulo Quintiliano[1,2], João Paulo C. Lustosa da Costa[2,3],
Flavio Elias de Deus[2,3], and Rafael Timóteo de Sousa Jr[2,3]
[1]Brazilian Federal Police Department (DPF)
[2]Laboratory of Array Signal Processing
[3]Laboratory for Decision-Making Technologies (LATITUDE)
Department of Electrical Engineering
University of Brasilia (UnB)
URL: www.pgea.unb.br/~lasp

Abstract. The Internet and other computer resources *became also means for* ***criminal activities****.* ***Regardless the*** *types of* ***crimes investigated, law enforcers seize from crime scene*** *a* ***large quantity of equipments and media containing evidences for investigation****. After a forensic examination of the evidences, a forensic report is issued by computer forensic examiners.* ***In order to recovery and analyze the evidences seized with higher accuracy,*** *a well structured computer forensic laboratory is essential.*

In this paper, we present the basic elements of a standard computer forensic laboratory as well as the tools associated with it, not only in terms of functionalities and aims but also in terms of hardware and software. For the software part, we present commercial and free solutions.

Keywords. *Computer forensic laboratory; computer crime.*

I. Introduction

Computer forensics deals with computer-based evidence investigation, and consists of preparation, identification, preservation, and analysis of data stored, transmitted, or produced by a computer system or computer network [1], and also of issuing a forensic report. In other words, computer forensics is a process of identifying, preserving, analyzing, and presenting digital evidence in a manner that is legally acceptable via the application of computer technology to the investigation of computer-based crime [2].

Computer forensics is becoming so usual that computer forensic laboratories are not only found in law enforcement agencies, but also in private companies, research centers, and universities. Furthermore, private forensic investigation is done by many specialized private companies and universities offer computer forensic courses.

The main goal in computer forensics is to prove the materiality, authorship and circumstances of an investigated fact by assigning responsibility of an event by processing the evidences found on computer resources. Therefore, the forensic experts and investigators are required to understand very well the principles behind their technologies and solutions. Note that the computer data itself only forms the evidence, and not any conclusion the forensic examiner may have [1].

The research on computer forensics is mainly conducted by law enforcement organizations, industry, and academia [4]. A computer forensic laboratory is necessary to develop their research, to make forensic analysis, to train professionals, and to teach computer forensic techniques.

According to Francia III [5], computer forensic tools can be divided into three main categories: imaging, analysis, and visualization [5]. Imaging tools are used to imaging volatile memory, disk and file imaging, write blocking, and for integrity code generators and checkers; analysis tools are used for data and file recovery, searching for text strings, file conversion, searching, and for data mining; and visualization tools are used for time-lining, and link analysis [5].

Both private companies and government security agencies are striving to improve computer forensic tools [6], since ability to cover the criminal tracks [6], using anti-forensics techniques and tools are also being enhanced. In this context, the importance of a laboratory with specifics tools, in terms of hardware and software, is very important and necessary.

In this paper, we present the basic elements of a computer forensic laboratory not only in terms of functionality and aims, but also in terms of hardware and software.

This paper is organized into 5 sections, including this introduction. The methodology of computer forensic examination is overviewed in Section 2. In Section 3, the environment and hardware for a computer forensic laboratory are described, while in Section 4, the softwares for a computer forensic laboratory are mentioned. Conclusions are drawn in Section 5.

II. Methodology of Computer Forensic Examinations

In this section, we overview the methodology used on computer forensic examinations, which usually consists of 5 steps: (a) identification, (b) preparation, (c) imaging forensic

data, (d) forensic data analysis, and (e) forensic report issuing. Forensic examinations only get started under a forensic request, made by investigators who are dealing the case. After this request, forensic examiners start to work, and usually follow the five steps sequentially. Forensic request should contain questions regarding the case. Forensic examiners have to answer all forensic request questions based on the analysis of the seized material.

In the first step (a) identification, digital evidences sized on a crime scene need to be identified. Responsible for the seizure have to write a report, in order to specify everything captured, with type of evidence, trademark, serial number, memory capacity, exact place where it was, and people to whom it belongs. Based on that report, forensic examiners write these evidences characteristics on the forensic report.

During the second step (b) preparation, all forensic examination is performed based on a document, called "forensic request". This request is analyzed, in order to verify if there is sufficient information to start the examinations. Forensic examiners can coordinate the investigation and examination with requesters in order to determine additional steps. They also setup and validate forensic hardware and software. Afterwards, based on the forensic request and on the data to be analyzed, forensic tools are selected, for instance: tools for evidence analysis, data recovery, decryption and password cracking, steganography analysis and mobile forensics examination.

In the third step c) imaging forensic data, forensic examiners have to make a forensic imaging of the evidence data, since they are not allowed to work directly on the original evidences due to the risk to modify and/or to damage evidences. Hence, after this step, the forensic examiners work just based on the copies of the evidences.

The fourth step d) forensic data analysis is the process to discover evidentiary information in the computer evidences based on the forensic request. In some cases, this information is not apparent to the investigators or may be protected by passwords or encryption. Forensic examiners may use specific softwares, such as FTK, EnCase, SleuthKit and others, in order to locate, undelete, and put available all user's files, for instance, .pdf, Microsoft Office files and email. Note that the exemplified files are usually the most important for the investigation.

In the last step e) forensic report issuing, a forensic report may be issued based on the forensic request, and on the forensic data analysis. All forensic procedures done during the examination may be written in this report, as well as all important evidence discovered. The questions of the forensic request need to be answered in this report.

III. Environment and Hardware for a Computer Forensic Laboratory

A Computer Forensic Laboratory needs at least an infrastructure with no-break, Gigabit Network, Access Point, scanner, printer, digital camera, multimeter, symmetrical power supply DC, DVD printers, and some other tools. A room with locked door for suspected material is also needed, so that all sized computer, media, and other computer resource are stored on this room. Only two or three people from the laboratory have password and access to this room, in order to limit and to control the access to this place.

Usually forensic examinations need virtual machines, such as malicious code analysis and packet sniffing. Forensic examiners and investigators use to perform packet sniffing, in order to investigate network activities, and they need a virtual machine, since they may need to use a different OS or their data are infected with malicious code.

Both Computer Forensic Laboratories – based on commercial and on free software - need many equipments in order to process forensic data, to make the examination, and to issue the forensic report. Computer media change very quickly, and many times old medias are seized from crime scene to be investigated. Consequently in a computer forensic laboratory old devices are needed to read these medias.

Next, we present the hardware of a Computer Forensic Laboratory.

3.1. Forensic Examiner Workstation and Server Cluster

Forensic Examiners need to have a workstation with RAM DDR3 ECC 12GB, two 20" monitors, USB mouse, two 256GB SSD, two 3TB HD, webcam, headset, flash memory readers, Windows 8 Pro, Microsoft Office, Firewall, antivirus, external HD connection, blu-ray recorder, and four 3TB SATA II HD. With this workstation, forensic examiners can make their examination and issue the forensic report.

A load-balancing cluster of server with a group of linked computers is needed in a computer forensic laboratory. This server cluster shares the decryption analysis execution workload, working as they were a single virtual machine, in order to recover passwords from data under examination.

3.2. Data imaging

As mentioned in Section 2, forensic examiners never can do their examinations directly on the seized original media, but just on an imaged hard drive. In order to make copies of the suspect data seized, forensic imaging tools are needed. We suggest the Image MASSter Solo 4 (see Figure 1) and Logicube Dossier to capture data from IDE, SATA, laptop drives, SCSI drives, as well as, from Flash Cards.

Figure 1. Forensic imaging tool (Image MASSter – Solo 4).

3.3. Media readers and portable computer laboratory

In some investigations, many medias are seized, such as diskettes, zip disks, and backup tapes (LTO, DAT, SDLT). Our forensic computer laboratory has these media readers in order to access and to image the suspect data and then to make forensic examinations.

A portable computer laboratory is an equipment with a bag size (see Figure 2), is portable for live forensic analysis in a crime scene, and is reliable for hard drive imaging and forensic analysis. These equipments are built to be used outside the laboratory and have tools to acquire data from IDE, SATA, USB, and SCSI interfaces.

Figure 2. Forensic Portable Lab (RoadMASSter-3).

3.4. Mobile forensics tools

Since mobile phones have a considerable memory, for instance, 64 GB or more, and a significant processing power, these phones may also contain evidences of illicit activities. The branch of Computer Forensic known as Mobile Phone Forensics is responsible for extracting data from mobile phones, including deleted data, to be used as criminal evidences.

Considering all devices seized as crime evidence, cell phone devices are already ranked on second place, just after the Hard Drives on first place.

On this context, some companies, such as Cellebrite [15], AccessData [16], Paraben [17], and others, developed many forensic mobile tools, that are both software and hardware components (see Figure 3). Mobile phones used to have a great number of connectors, so that forensic examiners need a hardware device with all kind of cables, in order to access data from all mobile cell phones, as well as to charge them.

In a computer forensic laboratory, we recommend the Cellebrite UFED Touch Ultimate, Cellebrite UFED Chinex, Logicube CellXtract, and Logicube CellDECK Forensic. These technological solutions represent what is best in the market and enable the extraction and recovery of data from cell phones, securely and friendly, ensuring high productivity of forensic experts, and also high quality of forensics reports.

Figure 3. Example of a cell phone forensic tool.

IV. Softwares for a Computer Forensic Laboratory

In this section, we present the commercial and free solutions of softwares for computer forensic.

4.1. Forensic laboratory based on commercial software

A computer forensic laboratory should be based on commercial software and the market offers many commercial computer forensic softwares. We indicate the AccessData FTK and Guidance EnCase for evidence analysis; VM Ware Workstation to run virtual machines; OfficeRecovery

Ultimate, EasyRecovery and GetDataBack for data recovery; AccessData DNA/PRTK for decryption and password cracking; StegAlyzerAS and StegAlyzerSS for steganography analysis; WinHex Specialist and Xways Forensics for hex editor; and also Mobile Master Corporate Edition, and AccessData Mobile Phone Examiner for mobile forensics examination.

Although these products are expensive, the performance of forensic examiners grows considerably, when these softwares are employed. Moreover, forensic reports quality are enhanced, since appropriate tools helps to find deleted or hidden important files for the investigation.

4.2. Forensic laboratory based on free software

Some free and open source forensic softwares are available in the market, including data analysis tools. One of them is the "Sleuth Kit" [18], a collection of tools to support forensic analysis in suspect digital data. Autopsy is a user friendly interface to make forensic data analysis easier, based on Sleuth Kit.

Free and open source code tools are not as good as commercial ones. Sleuth Kit and Autopsy are good free solutions, but they still need improvements with respect to new features.

There are some other open source forensic softwares, for data carving, such as Foremost, PhotoRec, Scalpel, Recover My Files; for wipe, such as KillDisk, CMRR and Secure Erase.

V. Conclusion

In this paper we presented the elements of a standard computer forensic laboratory. Moreover, we presented the forensic examination methodology, the laboratory's functionalities and aims, some computer forensic tools, and also an approach for a laboratory with commercial and with open source forensic software.

A well structured computer forensic laboratory is needed not only at law enforcement agencies but also at private companies, research centers, and universities. Specific forensic hardware, such as server clusters for decryption analysis, data imaging, media readers, portable computer lab and mobile forensic tools are also required. Moreover, forensic softwares for evidence analysis, mobile phone data processing and decryption analysis are also crucial.

A computer forensic laboratory cannot depend only on free software, since the quality of free tools provided by the community is not comparable to commercial solutions.

References

[1] Anderson, A.; Collie, B.; De Vel, O.; McKemmish, R.; and Mohay, G.; "Computer and Intrusion Forensics", Artech House, 2003.

[2] McKemmish, R.; "What is forensic computing", *Trends and Issues in Crime and Criminal Justice*, 118, 1999.

[3] Gong, Tony; and Gaertner, Mathias, "Case-Relevance Information Investigation: Binding Computer Intelligence to the Current Computer Forensic Framework", in: International Journal of Digital Evidence, Spring 2005, Volume 4, Issue 1.

[4] Yasinsac, Alex; Erbacher, Robert; Marks, Donald; Pollitt, Marc; and Sommer, Peter; "Computer Forensics Education", in: Forensics Education, july/august 2003, pages 15 to 23.

[5] Francia, Guillermo; and Clinton, Keion; "Computer Forensics Laboratory and Tools", in: Journal of Computing Sciences in Colleges, Vol. 20, Issue 6, June 2005, pages 143-150.

[6] Kuchta, Kelly; "Computer Forensics Today", in: Law, Investigations, and Ethics, 2000, CRC Press LLC.

[7] Kathirvel, Ayyaswamy; and Srinivasan, R.; "Double Umpiring System for Ad Hoc Wireless Mobile Network Security"; in The International Journal of Forensic Computer Science (IJoFCS), 2010, Vol. 5, Number 1, pages 22-29, DOI: 10.5769/J201001003.

[8] Nogueira, José; and Celestino Júnior, Joaquim; "Autonomic Forensics a New Frontier to Computer Crime Investigation Management"; in The International Journal of Forensic Computer Science (IJoFCS), 2009, Vol. 4, Number 1, pages 29-41, DOI: 10.5769/J200901003.

[9] Nogueira, José; and Vasconcelos, Wamberto; "Ontology for Complex Mission Scenarios in Forensic Computing", in The International Journal of Forensic Computer Science (IJoFCS), Vol. 3, Number 1, pages 42-50; DOI: 10.5769/J200801004.

[10] Milsan, Richard; "Creating laboratories for undergraduate courses in mobile phone forensics"; In Proceedings of the 2010 ACM Conference on Information technology Education, 2010, PAGES 111-116, ISBN: 978-1-4503-0343-9; DOI: 10.1145/1867651.1867680.

[11] Allen, W.H., "Computer Forensics", IEEE Security & Privacy, IEEE, vol. 3, Issue 4, Pages 59-62, 2005,

[12] R. Oppliger and R. Rytz, "Does Trusted Computing Remedy Computer Security Problems?" *IEEE Security & Privacy,* vol. 3, no. 2, Mar./Apr., 2005, pp. 16-19.

[13] S.L. Garfinkel and A. Shelat, "Remembrance of Data Passed: A Study of Disk Sanitization Practices," *IEEE Security &, Privacy,* vol. 1, no. 1, 2003, pp. 17-27

[14] Howard, Richard ; Thomas, Ralph; Burstein, Jeff; and Roxanna Bradescu, "Cyber Fraud Trends and Mitigation", in The International Journal of Forensic Computer Science (IJoFCS), 2008, Vol. 3, Number 1, pages 9-24; DOI: 10.5769/J200801001.

[15] Cellebrite website at www.cellebrite.com.

[16] AccessData website at www.accessdata.com.

[17] Paraben website at www.paraben.com.

[18] Sleuthkit website at www.sleuthkit.org.

DOI: 10.5769/C2013011 or http://dx.doi.org/10.5769/C2013011

O princípio da autonomia da perícia oficial no âmbito da Lei 12.030/2009

Paulo Quintiliano
Departamento de Polícia Federal (DPF)

Resumo. Este artigo discute os alcances da Lei 12.030, que dispõe sobre as perícias oficiais de natureza criminal e dá outras providências. Faz estudo comparado das autonomias dos peritos criminais de natureza criminal com as de outros servidores públicos. Discute questões relativas aos princípios jurídicos, bem como propõe o princípio da autonomia da perícia oficial, no âmbito da Lei 12.030/2009.

Palavras chave: Perícia criminal, princípios jurídicos, Lei 12.030/2009.

I. Introdução.

A Lei nº 12.030, de 17 de setembro de 2009, que entrou em vigência em 16/12/2009, poderá trazer avanços significativos para a perícia da Polícia Federal e das Polícias Civis dos Estados. *Ex vi* do artigo 2º dessa lei, os peritos oficiais de natureza criminal gozam de autonomia técnica, científica e funcional, *ipsis verbis*: "*Art. 2º No exercício da atividade de perícia oficial de natureza criminal, é assegurado autonomia técnica, científica e funcional, exigido concurso público, com formação acadêmica específica, para o provimento do cargo de perito oficial.*".

Neste artigo discutimos e esclarecemos os alcances das autonomias técnica, científica e funcional, no âmbito dos órgãos técnico-científicos, encarregados da elaboração dos laudos periciais criminais, no âmbito da Lei Nº 12.030/2009. Seguramente, essas autonomias são de importância fundamental para os órgãos técnico-científicos, encarregados da elaboração das perícias oficias e da produção dos respectivos laudos periciais criminais.

São feitos estudos comparados sobre as diversas autonomias dos peritos oficiais de natureza criminal com essas prerrogativas de outros servidores públicos, buscando o entendimento sobre as mudanças na atuação dos peritos oficiais, a partir dos benefícios concedidos pela lei. Busca-se também o entendimento sobre a abrangência da lei, com relação aos órgãos de gestão da perícia oficial de natureza criminal, nos âmbitos federal e estadual.

II. Primórdios da busca da autonomia científica

Segundo Silva [12], o astrônomo Galileu Galilei, no século XVI, foi o cientista pioneiro em pleitear a "autonomia científica", para que fosse possível o real desenvolvimento da ciência. Naquele momento histórico, pleiteava-se a autonomia da ciência em relação aos dogmas presentes na teologia. Ainda segundo os estudos de Silva [12], as investigações realizadas por Galileu, tendo em vista a comprovação do movimento terrestre, não apontam apenas para uma tomada de posição com respeito à astronomia ou em direção à construção de uma nova cosmologia. Mais do que isso, elas indicam uma postura intimamente voltada para a defesa de um valor cognitivo extremamente caro a Galileu: a autonomia científica. Ao defender a autonomia científica, o grande astrônomo reclamava a obediência a dois componentes importantíssimos: as experiências sensíveis e as demonstrações necessárias.

Segundo Mariconda [5, 6, 7, 8], o pleito de autonomia científica, formulado por Galileu, baseia-se principalmente na defesa vigorosa da liberdade da pesquisa científica, baseada na idéia da suficiência do método científico: as experiências sensíveis e as demonstrações necessárias são suficientes para decidir acerca das questões naturais, em particular, para determinar a escolha de uma entre várias explicações astronômicas rivais.

III. Estudo comparado das autonomias dos servidores públicos

Segundo Mazzilli [9], autonomia funcional consiste na liberdade de se exercer o ofício em face de outros órgãos e instituições do Estado. Em decorrência da autonomia funcional, técnica e científica, o perito oficial não estará sujeito a qualquer tipo de ingerência, quer seja de natureza científica, técnica ou administrativa, sobre o seu trabalho, quando no exercício de suas atividades fins.

Segundo Moraes [11], o servidor público que goza da autonomia funcional, em virtude de lei, quando no cumprimento de seus deveres funcionais, submete-se unicamente aos limites determinados pela Carta Magna, pelas leis e pela sua própria consciência, não estando subordinados a nenhum outro Poder. Isto é, não se subordinam nem ao Poder Executivo, nem ao Poder Legislativo, nem ao Poder Judiciário. Assim, em decorrência da autonomia funcional, concedida pela Lei Nº 12.030/2009, o perito oficial de natureza criminal, quando no exercício de suas atividades fins, não sofrerá ingerências de outros órgãos ou servidores públicos, independentemente do cargo ou da posição hierárquica ocupada.

Assim, o legislador assegurou ao perito oficial de natureza criminal, por meio da Lei nº 12.030/2009, a autonomia técnica, científica e funcional, garantindo-se a defesa vigorosa da liberdade da pesquisa científica, baseada na idéia da suficiência do método científico. Por meio dessas autonomias,

o perito oficial tem a prerrogativa de se submeter unicamente aos limites determinados pela Carta Magna, pelas leis e pela sua própria consciência, não tendo que se subordinar a nenhum outro poder, quando no exercício de suas funções, em busca da verdade real.

Outros servidores públicos e órgãos públicos também gozam de autonomia funcional e científica. A Constituição Federal assegura, de forma explícita, autonomia administrativa e financeira ao Poder Judiciário, autonomia funcional e administrativa ao Ministério Público e às Defensorias Públicas Estaduais e autonomia didático-científica às universidades, *in verbis*:

> "Art. 99. Ao Poder Judiciário é assegurada autonomia administrativa e financeira.
>
> Art. 127. § 1º São princípios institucionais do Ministério Público a unidade, a indivisibilidade e a independência funcional.
>
> § 2º Ao Ministério Público é assegurada autonomia funcional e administrativa, podendo, observado o disposto no art. 169, propor ao Poder Legislativo a criação e extinção de seus cargos e serviços auxiliares, provendo-os por concurso público de provas ou de provas e títulos, a política remuneratória e os planos de carreira; a lei disporá sobre sua organização e funcionamento.
>
> Art. 134. § 2º Às Defensorias Públicas Estaduais são asseguradas autonomia funcional e administrativa e a iniciativa de sua proposta orçamentária dentro dos limites estabelecidos na lei de diretrizes orçamentárias e subordinação ao disposto no art. 99, § 2º.
>
> Art. 207. As universidades gozam de autonomia didático-científica, administrativa e de gestão financeira e patrimonial, e obedecerão ao princípio de indissociabilidade entre ensino, pesquisa e extensão."

Ressalte-se que a Lei Suprema, além de assegurar aos membros do Ministério Público a "autonomia funcional", que lhes concede a liberdade de exercerem o seu ofício em face de outros órgãos e instituições do Estado, também lhes assegura a "independência funcional", concedendo-lhes, portanto, a liberdade de exercerem o seu ofício em face de outros órgãos da própria instituição do Ministério Público. Observe-se que a Lei nº 12.030/2009 não assegurou a independência funcional aos peritos oficiais, certamente porque o princípio da independência funcional se contrapõe ao princípio da hierarquia, sendo esse último típico dos órgãos policiais.

Segundo a doutrina de Mazzilli [9], os membros do Ministério Público e os seus órgãos colegiados e individuais, quando no exercício de sua atividade fim, só estão adstritos ao cumprimento da Constituição e das leis e que não estão obrigados a observar portarias, instruções, ordens de serviço ou quaisquer comandos nem mesmo dos órgãos superiores da administração, no que diz respeito ao que devem ou não fazer, em decorrência da autonomia e independência funcionais. Excetuados os casos expressamente previstos na lei, na sua atividade fim, os membros e órgãos do Ministério Público não podem receber ordens para proporem ou não proporem determinada ação, para recorrerem ou não, para sustentarem determinada tese e não outra.

Há, ainda, outros servidores públicos que buscam a autonomia e a independência funcional, para melhor cumprirem seus misteres, *verbi gratia,* os Membros da Advocacia Geral da União e os Procuradores dos Estados e do Distrito Federal. Com efeito, há doutrinadores que defendem a necessidade de se assegurarem a independência e a autonomia funcionais a esses servidores públicos.

Conforme discutido alhures, a autonomia técnica, científica e funcional, assegurada aos peritos oficiais pela Lei nº 12.030/2009, é prerrogativa de raros servidores públicos, notadamente dos membros do Ministério Público e das Defensorias Públicas Estaduais e das Universidades. Note-se que uns têm autonomia funcional e outros científica, sendo que os peritos oficiais têm cumulativamente a autonomia técnica, a científica e a funcional. Em consonância com o art. 99 do Texto Supremo, ao Poder Judiciário somente é concedida a autonomia administrativa e financeira. Ressalte-se que outros servidores públicos, *exempli gratia,* os membros da Advocacia Geral da União, os Procuradores dos Estados e do Distrito Federal, os Delegados de Polícia, dentre outros, não gozam de autonomia funcional.

Dessa forma, a partir da vigência da Lei nº 12.030/2009, *mutatis mutandis*, a atuação dos peritos criminais federais, quando no exercício de sua atividade fim, poderá ser semelhante à dos membros do Ministério Público e das Defensorias Públicas Estaduais, haja vista que todos esses servidores públicos gozam igualmente da autonomia funcional. Nesse diapasão, os peritos criminais federais, quando no exercício de suas atividades fins, submeter-se-ão unicamente aos limites determinados pela lei e pelas suas próprias consciências, não podendo receber ordens para direcionarem os exames periciais para esse ou aquele rumo, excetuados os casos expressamente previstos em lei. Ressalte-se que, em decorrência de não gozarem da independência funcional, estão os peritos oficiais sujeitos ao princípio da hierarquia, existente em seus órgãos de lotação, ao contrário do que ocorre com os membros do Ministério Público, que gozam cumulativamente da independência funcional.

IV. Abrangência da Autonomia Concedida Pela Lei Nº 12.030/2009

Com efeito, a Lei nº 12.030/2009 não é explícita com relação aos beneficiários da autonomia concedida, não deixando claro se tais beneficiários são somente os peritos oficiais de natureza criminal ou se também os órgãos de gestão da perícia oficial estão contemplados com a autonomia. Entende-se que o legislador quis ser abrangente e conceder a autonomia técnica, científica e funcional ao perito oficial de natureza

criminal e aos órgãos públicos encarregados da gestão da perícia oficial, em nível federal e estadual. Em nível federal, os órgãos agraciados pela autonomia são o Instituto Nacional de Criminalística (INC) e as unidades de criminalística nos Estados, no âmbito da Polícia Federal. No âmbito estadual, os beneficiados são os órgãos estaduais de natureza criminal, encarregados da gestão da perícia oficial, em todas as unidades da federação, sendo que algumas estão vinculadas diretamente à estrutura administrativa das polícias civis e outras têm estrutura administrativa independente.

Buscando-se uma interpretação teleológica do art. 2º da lei, visando revelar o escopo desse dispositivo legal, recorrendo-se a elementos jurídicos, sistemáticos, políticos e sociológicos, como será demonstrado, pode-se concluir que os destinatários da autonomia concedida são: o perito oficial de natureza criminal e os órgãos públicos responsáveis pela gestão da perícia oficial, em nível federal e estadual. Para reforçar essa tese, observe-se que, ao se eliminar a parte acessória do texto do dispositivo legal aludido, resta claro que o legislador não excluiu os órgãos de gestão da perícia, *ipsis literis*:

> "Art. 2º No exercício da atividade de perícia oficial de natureza criminal, é assegurado autonomia técnica, científica e funcional, para o provimento do cargo de perito oficial.".

Ademais, seguramente os efeitos da lei seriam extremamente limitados e muitas vezes inócuos se o benefício da autonomia fosse concedido somente ao perito e não aos órgãos, hipótese que certamente poderia tornar a lei uma mera "letra morta". Nessa hipótese, os órgãos encarregados da gestão da perícia oficial ficariam em posição fragilizada diante da autonomia dos peritos oficiais de natureza criminal, não poderiam estabelecer padrões de laudos ou de exames periciais e tampouco esperar o cumprimento desses critérios por parte dos peritos oficiais, o que certamente prejudicaria em muito a gestão da perícia oficial.

Resta claro que o texto legal não estabelece que "é assegurada autonomia técnica, científica e funcional (somente) ao perito oficial". A *contrario sensu*, estabelece que "é assegurado autonomia técnica, científica e funcional, para o provimento do cargo de perito oficial.". Como a lei não limitou a concessão da autonomia somente ao perito oficial, entende-se que a autonomia também foi concedida aos órgãos públicos encarregados da gestão da perícia oficial de natureza criminal, em nível federal e estadual, visto que, como o legislador não restringiu essa abrangência, não pode outro isso fazer. Assim, entende-se que, no âmbito federal, a Lei nº 12.030/2009 assegurou ao Instituto Nacional de Criminalística (INC) da Polícia Federal e às unidades de criminalística da Polícia Federal nos estados a autonomia técnica, científica e funcional, de forma implícita.

V. Princípio da Autonomia da Perícia Oficial

Segundo Carrazza [2] e Fazoli [4], princípio jurídico é um enunciado lógico, implícito ou explícito, que por sua grande generalidade, ocupa posição de preeminência nos vastos quadrantes do Direito e, por isso mesmo, vincula de modo inexorável o entendimento e a aplicação das normas jurídicas que com ele se conectam. Segundo Mello [10], princípio é, por definição, mandamento nuclear de um sistema, verdadeiro alicerce dele, disposição fundamental que se irradia sobre as diferentes normas, compondo-lhes o espírito e servindo de critério para sua exata compreensão e inteligência, exatamente por definir a lógica e a racionalidade do sistema normativo, no que lhe confere tônica e lhe dá sentido harmônico.

Como ensina Borges [1], o princípio explícito não é necessariamente mais importante que o princípio implícito. Tudo depende do âmbito de abrangência de um e de outro, e não do fato de um estar melhor ou pior desvendado no texto jurídico. Carvalho [3] esclarece que entre os princípios implícitos e os expressos não se pode falar em supremacia ou hierarquia. Ambos retiram fundamento de validade do mesmo texto jurídico. Com efeito, os princípios podem ser encontrados em todos os escalões da "pirâmide jurídica", não havendo óbices para a existência de princípio no âmbito da perícia oficial.

Com base na interpretação teleológica do texto legal em apreço, em que se recorreu aos elementos jurídicos, sistemáticos, políticos e sociológicos, enxerga-se o "princípio da autonomia da perícia oficial", insculpido de forma implícita na Lei 12.030/2009, no âmbito da perícia oficial. Entende-se que o legislador quis que a abrangência da lei fosse maior e que os beneficiários da autonomia técnica, científica e funcional sejam o perito oficial de natureza criminal e os órgãos públicos encarregados da gestão da perícia oficial, em nível federal e estadual, estabelecendo-se, dessa forma, o "princípio da autonomia da perícia oficial".

VI. Conclusões

Em decorrência da autonomia técnica, científica e funcional, assegurada pela Lei nº 12.030/2009 ao perito oficial de natureza criminal e aos órgãos públicos encarregados da gestão da perícia oficial de natureza criminal, em níveis federal e estadual, garante-se que o perito oficial, no exercício de suas atividades fins, submeta-se unicamente aos limites determinados pela lei e pela sua própria consciência, não podendo receber ordens para direcionarem os exames periciais para esse ou aquele rumo, excetuados os casos expressamente previstos em lei. A lei assegura ao perito oficial a autonomia científica, haja vista a suficiência do método científico. Também os órgãos públicos encarregados da gestão da perícia oficial de natureza criminal, em decorrência da autonomia a esses assegurada pela lei, estão submetidos unicamente aos limites determinados pela lei, quando na gestão de suas atividades fins, não podendo aceitar ingerências de outros órgãos, independentemente de suas posições hierárquicas.

Como a Lei nº 12.030/2009 assegurou aos peritos criminais, de forma explícita, a autonomia técnica, científica e funcional, e, de forma implícita, as mesmas autonomias ao

Instituto Nacional de Criminalística (INC) da Polícia Federal e às unidades de criminalística da Polícia Federal nos estados, esse fato pode ser interpretado como um princípio jurídico implícito: o "princípio da autonomia da perícia oficial".

Referências

[1] Borges, José Souto Maior. "O Princípio da Segurança Jurídica na Criação e Aplicação do Tributo". In: Revista Diálogo Jurídico, Número 11, fevereiro de 2002, Salvador, BA, Brasil.

[2] Carrazza, Roque Antonio. "Curso de Direito Constitucional Tributário". 17ª Ed. São Paulo: Malheiros, 2002.

[3] Carvalho, Paulo de Barros. "O Princípio da Segurança Jurídica em Matéria Tributária". In: Revista Diálogo Jurídico, Número 16, 2007, Salvador, BA, Brasil.

[4] Fazoli, Carlos Eduardo de Freitas. "Princípios jurídicos". In: Revista Uniara, n.20, páginas 13-29, 2007.

[5] Mariconda, P. R. "A contribuição filosófica de Galileu ". In: CARNEIRO, F. L. (Org.). *350 anos dos "Discorsi intorno a due nuove scienze" de Galileu Galilei*. Rio de Janeiro, Marco Zero/Coppe, 1989. p. 127-37.

[6] Mariconda, P. R. "Duhem e Galileo: uma reavaliação da leitura duhemiana de Galileu ". In: ÉVORA, F. R. (Org.). *Século XIX: o nascimento da ciência contemporânea*. Campinas, CLE/Unicamp, 1993. p. 123-60.

[7] Mariconda, P. R. "O Diálogo de Galileu e a condenação . *Cadernos de História e Filosofia da Ciência"*, 10, 1, p. 77-160, 2000. Republicado como Introdução. O Diálogo e a condenação. In: GALILEI, G. *Diálogo sobre os dois máximos sistemas do mundo ptolomaico e copernicano*. Trad., introd. e notas de P. R. Mariconda. São Paulo, Discurso/Fapesp, 2001. 2a. edição: São Paulo, Discurso/Imprensa Oficial, 2004. p. 15-70.

[8] Mariconda, P. R. "Lógica, experiência e autoridade na carta de 15 de setembro de 1640 de Galileu a Liceti". *Scientiae Studia*, 1, 1, p. 63-73, 2003. Tradução e notas da Carta de Galileu Galilei a Fortunio Liceti em Pádua.

[9] Mazzilli, Hugo Nigro. "A independência do Ministério Público". In: Revista dos Tribunais, 1996.

[10] Mello, Celso Antônio Bandeira de. "**Curso de direito administrativo".** 12ª ed. São Paulo: Malheiros Editores, p. 748, 2000.

[11] Moraes, Alexandre de. "Direito Constitucional". 21ª ed. São Paulo: Atlas, 2007.

[12] Silva, Paulo Tadeu, "Copernicanismo, autonomia científica e autoridade religiosa em Marin Mersenne", Scientiæ Zudia, São Paulo, v. 2, n. 2, p. 239-50, 2004.


ABEAT
ASSOCIAÇÃO BRASILEIRA DE ESPECIALISTAS EM ALTA TECNOLOGIA

ISBN 978-85-65069-09-C

9 788565 069090 >