



Ain Shams University
Faculty of Engineering
Department of Computer and Systems Engineering

Defending Attacks on Cloud Computing

By

Eman Ahmed Abd El-Azim

Bachelor of Science Degree majoring in Computer Engineering – College
of Engineering – Qatar University, 2008

A THESIS

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE IN ELECTRICAL
ENGINEERING

DEPARTMENT OF COMPUTER AND SYSTEMS ENGINEERING

Supervised By

Prof. Dr./ Ayman Mohammad Bahaa-Eldin

Dr./ Mohammed Ali Sobh

Dr./ Amin Abdul Wahab Sorrou

Cairo, Egypt

2017

© Eman Ahmed Abd El-Azim, 2017



Faculty of Engineering

Department of Computer and Systems Engineering

Name : Eman Ahmed Abd El-Azim
Thesis : Defending Attacks on Cloud Computing
**Degree : Master of Science in Electrical Engineering –
Computer and Systems Engineering**

Examiners Committee

Name, Title, and Affiliate	Signature
Prof. Dr./ Samy Sayed Abdo Ghoniemy Computer Systems British University, Cairo, Egypt (Examiner)
Prof. Dr./ Hoda Korashy Mohamed Ismail Computer and Systems Engineering Ain Shams University, Cairo, Egypt (Examiner)
Prof. Dr./ Ayman Mohammad Bahaa-Eldin Computer and Systems Engineering Ain Shams University, Cairo, Egypt (Supervisor)
Dr./ Mohammed Ali Sobh Computer and Systems Engineering Ain Shams University, Cairo, Egypt (Supervisor)

Date: / /

Abstract

Eman Ahmed Abd El-Azim

Defending Attacks on Cloud Computing

Master of Science in Electric Engineering – Computer and Systems

Engineering Dissertation

Ain Shams University, 2017

Malwares are increasing rapidly. The nature of distribution and effects of malwares attacking several applications requires a real-time response. Therefore, a high performance detection platform is required. In this thesis, Hadoop is utilized to perform static binary search and detection for malwares and viruses in portable executable files deployed mainly on the cloud. Hadoop was chosen as it is a software platform that allows designing applications capable of handling huge data amounts in a parallel manner in large clusters. The thesis presents an approach used to map the portable executable files to Hadoop compatible files. The Boyer–Moore-Horspool Search algorithm is modified to benefit from the distribution of Hadoop. The performance of the proposed model is evaluated using a standard virus database and the system is found to outperform similar platforms.

Keywords:

Cloud computing, Security issues, Malware, Static Binary Search, BMH, Hadoop.

Acknowledgements

"In the name of Allah, the Most Gracious and the Most Merciful."

Foremost, Alhamdulillah, all praises to Allah, the Almighty for the strengths and the blessings for the completion of the master's thesis. This thesis is accomplished in its current form due to the assistance and guidance of my supervisors: Dr. Ayman Mohammed Bahaa, Dr. Mohammed Ali Sobh and Dr. Amin Abdul Wahab. I would like to express my sincere gratitude to their efforts, patience, advices and continuous support which helped me a lot throughout this research. Sincere thanks to all my family, it would not have been done without your support, encouragement and prayers. Eventually, special thanks to my father, I know your spirit is always with us.

Statement

This dissertation is submitted to Ain Shams University for the degree of Master of Science in Electrical Engineering in Computer and Systems Engineering

Department of Computer and Systems Engineering

The work included in this thesis was out by the author at Computer and Systems Engineering, Ain Shams University.

No part of this thesis has been submitted for a degree or qualification at other university or institution.

Date : / / 2017

Signature :

Name : Eman Ahmed Abd El-Azim

Table of Contents

Abstract	i
Acknowledgements	ii
Statement	iii
Table of Contents.....	iv
List of Publications	vii
List of Tables.....	viii
List of Figures and Illustrations	ix
Chapter 1: Introduction	1
1.1 Overview and Thesis Main Concern	1
1.2 Methodology.....	2
1.3 Thesis Organization.....	3
Chapter 2: Background Information for Malwares Cloud Issues ..	5
2.1 Cloud Computing Environment Overview	5
2.1.1 Cloud Computing Emerging.....	5
2.1.2 Cloud Service Delivery Models	7
2.1.3 Cloud Deployment Models.....	11

2.2	Malwares	13
2.2.1	Common types of Malware	13
2.2.2	Malware Detection Techniques	15
2.3	Malware in Cloud Environments.....	16
2.4	Handling the Different Malware Detection Techniques by Other Researchers	23
2.4.1	A Framework for Behaviour-Based Malware Analysis in the Cloud	23
2.4.2	A Retrospective Detection of Malware Attacks by Cloud Computing	33
2.4.3	Towards a Distributed, Self-Organizing approach to malware detection in cloud computing	47
2.4.4	Malware-Detection and Kernel-Rootkit-Prevention in Cloud Computing Environments.....	54
2.4.5	Malware Analysis on the Cloud: Increased Performance, Reliability, and Flexibility	64
2.4.6	Hard-Detours: a Dynamic Code Analysis new Technique.....	75
Chapter 3: Related Work and The Proposed System Design		86
3.1	State-of- the-Art and Related Work in Malware Signature Based Detection.....	86
3.1.1	Malware Analysis Using Hadoop and MapReduce.....	86
3.1.2	BinaryPig: Scalable Static Binary Analysis over Hadoop.....	88
3.2	The Proposed System Environment	91
3.2.1	The Environment Used.....	91
3.2.2	Hadoop	92
3.2.3	Infected Files and Database Used.....	98

3.3	System General Architecture.....	103
3.4	Factors Affecting Performance during Testing Phase.....	104
3.4.1	Resources and Configurations.....	104
3.4.2	Determining a Search Algorithm.....	105
3.4.3	Hadoop File-Size.....	112
3.4.4	Database organization and location.....	115
Chapter 4: Results and Discussion		117
4.1	The System Architecture Details and Results	117
4.2	Comparing the Performance of the Three Tests in Pseudo-Distributed Mode.....	120
4.3	Running on Virtual Multi-node Cluster and Results.....	122
Chapter 5: Conclusion and Future Work.....		129
References.....		131
Appendix A.....		141
A.1	General steps for running the program in Both Modes	141
مستخلص.....		1
شكر.....		5

List of Publications

1. Cloud Computing and Malware Issues: A survey paper
2. A Cloud-based Malware Detection Framework: A paper

List of Tables

<i>Table3. 1: Test1 - Naïve Search Performance for 4 Portable Executable Files</i>	<i>107</i>
<i>Table3. 2: Test2 - BMH Search Performance for 4 Portable Executable Files</i>	<i>111</i>
<i>Table4. 1: System Performance Using Sequence files and BMH Searching for 10 Portable Executable Files.....</i>	<i>119</i>
<i>Table4. 2: Comparing the Performance of Three Tests in Pseudo-Distributed Mode</i>	<i>121</i>
<i>Table4. 3: System Performance on the Virtual Multi-node Cluster</i>	<i>126</i>

List of Figures and Illustrations

<i>Fig2. 1: Cloud Service delivery Models</i>	<i>7</i>
<i>Fig2. 2: Pseudo-code example for BANCOS malicious program</i>	<i>24</i>
<i>Fig2. 3: Executing the malicious program in L and forcing it to act as if it is in U</i>	<i>25</i>
<i>Fig2. 4: Block diagram for executing multiple instances of the malicious program in multiple environments. The central L aggregates the results</i>	<i>28</i>
<i>Fig2. 5: Interception and remote execution of system calls (The analyzed-program is P).....</i>	<i>30</i>
<i>Fig2. 6: The design block diagram.....</i>	<i>35</i>
<i>Fig2. 7: High level working architecture</i>	<i>36</i>
<i>Fig2. 8: A log file.....</i>	<i>37</i>
<i>Fig2. 9: a PDF attack.....</i>	<i>38</i>
<i>Fig2. 10: Building file-index using MapReduce.....</i>	<i>41</i>
<i>Fig2. 11: Building Relation-Index using MapReduce</i>	<i>43</i>
<i>Fig2. 12: MapReduce for searching suspicious file chains.....</i>	<i>45</i>
<i>Fig2. 13: System architecture block diagram</i>	<i>48</i>
<i>Fig2. 14: The flow of information and hierarchy of engines within a CRM</i>	<i>51</i>
<i>Fig2. 15: Malware Scanner Software Modules.....</i>	<i>59</i>
<i>Fig2. 16: The state transition diagram of authorized module loading</i>	<i>63</i>
<i>Fig2. 17: Malware detection techniques</i>	<i>66</i>
<i>Fig2. 18: The general analysis flow (left) and the intra-category flow (right).....</i>	<i>71</i>
<i>Fig2. 19: Win32 Architecture for Win-OS.....</i>	<i>77</i>
<i>Fig2. 20: The Function Original Code Vs its Detoured Code</i>	<i>79</i>

<i>Fig2. 21: Calling Sequence before and After Interception</i>	<i>80</i>
<i>Fig2. 22: The Function Ms-Detoured Code Vs its Anti-Detoured Code..</i>	<i>82</i>
<i>Fig2. 23: Calling Sequence after applying Anti-Detouring</i>	<i>83</i>
<i>Fig2. 24: The Function Original Code Vs its Hard-Detoured Code.....</i>	<i>84</i>
<i>Fig3. 1: Malware detection flow</i>	<i>87</i>
<i>Fig3. 2: BinaryPig General Architecture</i>	<i>89</i>
<i>Fig3. 3: Hadoop Pseudo Distributed Mode Vs Cluster Mode</i>	<i>94</i>
<i>Fig3. 4: HadoopV1 Vs HadoopV2.....</i>	<i>96</i>
<i>Fig3. 5: EICAR text</i>	<i>99</i>
<i>Fig3. 6: EICAR TEST File Hexadecimal Values.....</i>	<i>100</i>
<i>Fig3. 7: EICAR-SIGNATURE in ClamAV DB</i>	<i>100</i>
<i>Fig3. 8: EICAR detailed by disassembler tool</i>	<i>101</i>
<i>Fig3. 9: EICAR in simple assembly.....</i>	<i>102</i>
<i>Fig3. 10: System General Architecture</i>	<i>103</i>
<i>Fig3. 11: Naïve Brute Force Algorithm</i>	<i>106</i>
<i>Fig3. 12: Test1 Detailed Jobs' Architecture.....</i>	<i>106</i>
<i>Fig3. 13: Performance of Test1 Searching Phase.....</i>	<i>108</i>
<i>Fig3. 14: Boyer–Moore-Horspool Algorithm</i>	<i>109</i>
<i>Fig3. 15: Test2 Detailed Jobs' Architecture.....</i>	<i>109</i>
<i>Fig3. 16: Performance of Test2 Searching Phase.....</i>	<i>111</i>
<i>Fig3. 17: Scanner JOB of the Three Tests in Pseudo-Distributed Mode</i>	<i>114</i>

<i>Fig4. 1: A detailed System Architecture is shown in Figures Fig4.1a and Fig4.1b.....</i>	118
<i>Fig4. 2: System's Output Report for the Scanned Files.....</i>	118
<i>Fig4. 3: Performance of System Searching Phase</i>	120
<i>Fig4. 4: Cluster nodes</i>	122
<i>Fig4. 5: HDFS instances</i>	123
<i>Fig4. 6: Yarn Instances in the Virtual Multi-node Cluster</i>	124
<i>Fig4. 7: YARN Running Architecture</i>	125
<i>Fig4. 8: Performance of System Searching Phase</i>	126
<i>Fig4. 9: Pseudo-Distributed Vs. Virtual Multi-Node Cluster</i>	127

Chapter 1: Introduction

1.1 Overview and Thesis Main Concern

Malware stands for "malicious software." It is any software program that is created to perform harmful actions or unwanted by a computer's legitimate user. It has a variety of forms as: computer viruses, which are the most familiar type, worms, spyware, adware, trojans, ransomware and Botnets.

Having Malwares spreading widely on the internet, every host faces the risks of malware attacks. Since cloud environment inherited internet properties, cloud environment is vulnerable to Malwares. Most users transfer or share a vast amount of small files across the cloud as PE files, images. Many of these files could be infected by different Malware types. Hence, a demand to make researchers study and improve different techniques for scanning of files across clouds in a fast manner. Malware-detection techniques can be commonly classified into three categories: signature-based-detection, anomaly-based-detection and code-emulation [1] [2] [3]. A further explanation is presented in the next chapter for the cloud environment, difference between Malware types, detection techniques and some efforts carried using these techniques by other researchers.

The problem of detecting Malware in cloud environment is a performance and detection accuracy problem. This thesis handled this issue by utilizing Hadoop framework to speed up the detection rate. It presents the problems faced that affect detection performance and accuracy and how it was handled. In the next section, a brief description for the detection technique chosen, the type of files used in testing, the framework used and previous efforts done by other researchers using same framework.

1.2 Methodology

The detection technique handled in this thesis is signature-based detection for binary-executable files (e.g. in MSDOS: the .exe files and in Windows: the PE-files, ...etc.). Considering the assumption that there are many of them and there is a need to process them in fast and parallel manner, hadoop platform is utilized to perform this detection. Apache Hadoop is "a free software framework for having a distributed storage and processing for huge datasets of computer-clusters" [4]. A further explanation for Hadoop architecture and versions is represented in the third chapter and its practical tests.

Researchers as [5] handled this problem using hadoop environment. But they proposed architecture of how hadoop framework uses its daemons to cooperate in scanning without further details. Other Researchers [6] emphasized how malwares have spread widely in the past few years. Where Endgame received 20M samples of malwares, McAfee receives 10M malware samples in 2012 and VirusTotal receives almost 600k unique files per day. These numbers proves how malware samples increased rapidly in the internet which needs making more studies on