AN ENHANCED PERFORMANCE MODEL FOR METAMORPHIC COMPUTER VIRUS CLASSIFICATION AND DETECTION

BABAK BASHARIRAD

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy (Computer Science)

Faculty of Computing
Universiti Teknologi Malaysia

OCTOBER 2013

To My MOTHER, who supported me as a firm mountain overall duration of my life, Thanks for your cares and inspirations, Mom

To My FATHER, for his love, cares and encouragement

With Special Thanks

To My BELOVED WIFE, NAHID, who accommodated me in all circumstances, helped me to grow spiritually and intellectually, Thanks for your love and patience

To My WONDERFUL SON, MANI, who is all my life, His love inspired me to work hard and finish this work

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to my dear supervisor Assoc. Prof. Dr. Maslin Masrom for her valuable guidance, patience, and encouragements. This research would never have been completed without her advices and supports. She always was ready to spend her time to counsel me, and to improve the thesis report, patiently and cordially.

My special thanks to my co-supervisor Assoc. Prof. Dr. Suhaimi Ibrahim for his supports, suggestions and valuable comments.

ABSTRACT

Metamorphic computer virus employs various code mutation techniques to change its code to become new generations. These generations have similar behavior and functionality and yet, they could not be detected by most commercial antivirus because their solutions depend on a signature database and make use of string signature-based detection methods. However, the antivirus detection engine can be avoided by metamorphism techniques. The purpose of this study is to develop a performance model based on computer virus classification and detection. The model would also be able to examine portable executable files that would classify and detect metamorphic computer viruses. A Hidden Markov Model implemented on portable executable files was employed to classify and detect the metamorphic viruses. This proposed model that produce common virus statistical patterns was evaluated by comparing the results with previous related works and famous commercial antiviruses. This was done by investigating the metamorphic computer viruses and their features, and the existing classifications and detection methods. Specifically, this model was applied on binary format of portable executable files and it was able to classify if the files belonged to a virus family. Besides that, the performance of the model, practically implemented and tested, was also evaluated based on detection rate and overall accuracy. The findings indicated that the proposed model is able to classify and detect the metamorphic virus variants in portable executable file format with a high average of 99.7% detection rate. The implementation of the model is proven useful and applicable for antivirus programs.

ABSTRAK

Virus komputer metamorfik menggunakan pelbagai teknik mutasi kod untuk menukarkan kod menjadi generasi baru. Generasi ini mempunyai tingkah laku dan fungsi yang serupa namun tidak dapat dikesan oleh kebanyakan antivirus komersial kerana penyelesaian bergantung kepada pengkalan data yang menggunakan tandatangan dan menggunakan kaedah pengesanan berasaskan tandatangan-rentetan. Walau bagaimanapun, enjin pengesanan antivirus dapat dielakkan dengan teknik metamorfik. Tujuan kajian ini adalah untuk membangunkan sebuah model prestasi berdasarkan klasifikasi virus komputer dan pengesanan. Model ini juga juga dapat mengenal pasti fail-fail boleh laksana mudah alih yang akan mengelaskan dan mengesan virus komputer metamorfik. Model Markov Tersembunyi digunakan pada fail-fail boleh laksana mudah alih untuk mengelaskan dan mengesan virus metamorfik. Model yang dicadangkan ini menghasilkan bentuk statistik biasa yang dinilai dengan membandingkan keputusannya dengan hasil-hasil yang berkaitan dengan virus komersial yang terkenal sebelum ini. Ini dilakukan dengan menyelidiki virus komputer metamorfik dan ciri-cirinya dan pengelasan sedia ada serta kaedah pengesanan. Khususnya, model ini telah digunakan pada format binari fail-fail boleh laksana mudah alih dan mampu membuat klasifikasi jika fail-fail tergolong dalam keluarga virus. Selain itu, prestasi model, pelaksanaan secara praktikal dan diuji, juga dinilai berdasarkan kadar pengesanan dan ketepatan keseluruhan. Dapatan kajian menunjukkan bahawa model yang dicadangkan mampu untuk mengelaskan dan mengesan varian virus metamorfik fail-fail boleh laksana mudah alih dengan kadar purata pengesanan yang tinggi yakni 99.7%. Pelaksanaan model ini terbukti berguna dan boleh diaplikasikan untuk program-program antivirus.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CFG	-	Control Flow Graph
DFA	-	Deterministic Finite Automaton
DOS	-	Disk Operating System
DR	-	Detection Rate
EPO	-	Entry Point Obfuscation/Obscuring
FN	-	False Negative
FP	-	False Positive
FPR	-	False Positive Rate
GD	-	General Decryption
HMM	-	Hidden Markov Model
LLPO	-	Log Likelihood Per Opcode
NGVCK	-	Next Generation Virus Creation Kit
OA	-	Overall Accuracy
OS	-	Operating System
PE	-	Portable Executable
RPME	-	Real Permutation Engine
TN	-	True Negative
TP	-	True Positive
VCL	-	Virus Creation Lab
Sens	-	Sensitivity
Spec	-	Specificity

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

These days, data are stored and shared in the digitally linked storage devices in the modern electronic world. The way we study, buy, play, work, earn money and live has become very different. The large majority of business transactions, which include extremely important and sensitive information, are performed via computers and over the digital networks and the internet. Therefore, it is imperative to care for information safety as a concern of dominant significance. Some digital applications, data processor and electronic storage devices, which can be connected through computer networks, are displayed in Figure 1.1.

Figure 1.1 Storage and Sharing of Information in Modern Digital World

Hence, we expose the safety and secrecy of private and personal information to danger in this way. We can see, even in this new kind of data storage and accessing system, information is even being stolen. We hear all the time, people’s personal information and money are stolen. Even worse, malicious programs wipe out valuable information and wealth of companies and organizations. Because the current digital world comprises multi-faceted vulnerabilities, the important issue is to protect data from being damaged or removed by malware programs. The destruction induced by malicious codes is more dangerous in today’s modern society, where personal and social communications and commercial business strongly depend on digital networks. In the latest study conducted by Symantec Corp., it is shown that the security is the top issue for 42 percent of organizations. In the past 12 months, 75 percent of organizations have suffered from computer network attacks that cost the enterprise commerce an average of 2 million dollars yearly. The result of the study is presented in Figure 1.2 (Symantec, 2010).

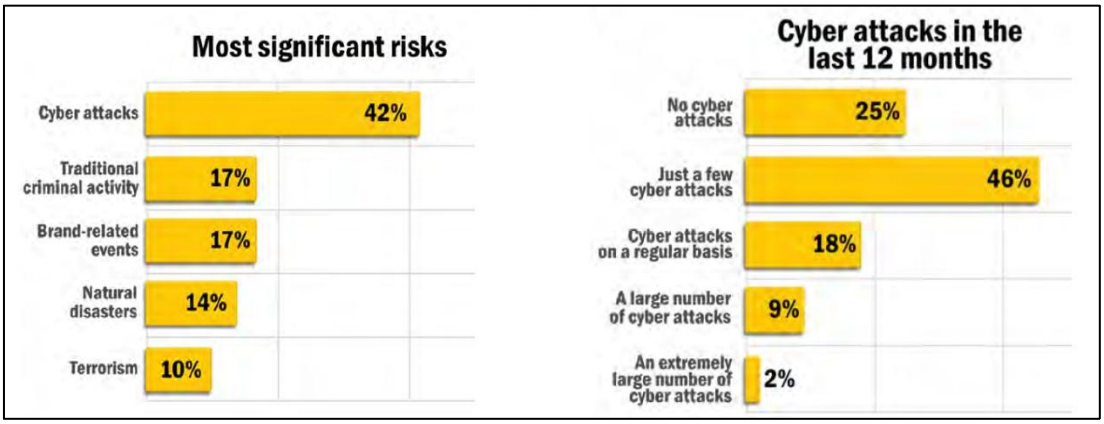

Figure 1.2 Symantec Corp. Status of Enterprise Security in 2010: “Most Significant Risks and Cyber Attacks in Last Year” (Symantec, 2010)

The security problem has contributed to the development of antivirus production and it is now practically mandatory for all computer users to have antivirus softwares on their personal computers. Nevertheless, we still hear much news on recent malware and cautions against them daily, so therefore we need to update database files of our antivirus software to stay away from corruption and prevent wider proliferation of malware. The most important reason is that existing antivirus products mostly work based on byte-to-byte comparison of files, where binary string sequences extracted from the analyzed viral files are exploited as signatures (Wang *et al.*, 2003; Zhang, 2008). If a given file contains corresponding signatures existed in the patterns database, then it is accepted as a viral program, Therefore, there is a weakness in finding previously non-accurately analyzed viruses (Mori, 2004; Xu *et al.*, 2013).

On the other hand, research and practical investigation in computer virology is still a contentious issue (Filiol *et al.*, 2006; Marpaung *et al.*, 2012). There is an extensively misconceived standpoint, which argues that research on computer virus and related areas not only is not productive, but also is potentially dangerous. The reason for this view is that the proponents of this opinion posit that it may lead to the growth of more overwhelming and harmful techniques of viral infection. They also believe that it is only waste of time, since fighting against the computer viruses is restricted to the found, analyze, extracted and identified signature of virus, the usual cycle of the antivirus software production (Filiol *et al.*, 2006).

For this reason, only a few research groups and laboratories in organizations and universities worldwide investigate malware and study computer virology (Bonfante *et al.*, 2006). Lack of knowledge in this subject has resulted in small number of significant scientific findings in this field have been obtained for this belief, while a logical and accurate view on the problem shows that more researches and practical examinations on the issue of computer virology are really vital and crucial (Filiol *et al.*, 2006; Mansfield-Devine, 2013).

In order to protect our systems in an efficient way and predict computer malware risks before they practically appear as tools for attackers and virus creators, we need to profoundly, understand the threat we are facing (Mansfield-Devine, 2013).

As far as computer virology is concerned, unfortunately, there are several open problems remained unsolved, in aspects of modeling and implementation. Numerous new problems will certainly appear soon in the future, because of the progresses of malware creators. Whilst upcoming computer systems grow swiftly to be complicated and sensitive, outdated defense and protection methods and models become even more inadequate (Filiol *et al.*, 2006).

When we deal with metamorphic viruses, the problem becomes even worse (Singhal and Raul, 2012). Metamorphic viruses propagate by creating new transformed instances of its code (Leder *et al.*, 2009). In other words, for a signature-based detector, metamorphic variants are considered as new viruses (Han *et al.*, 2011b). Metamorphic viruses are created, propagated and spread more and more, because with virus creation kits and automatic metamorphic engines it is not necessary to be an expert and spend much time to create new malware. Consequently, the time needed for analyzing and reverse engineering of viral codes to extract new signatures is even more prolonged, and the size of database is becoming bigger.

In this study, a machine learning-based method, which uses an enhanced Hidden Markov Model is developed and implemented to analyze, classify and detect metamorphic computer viruses. This chapter concentrates on the background of the study, statement of the problem, research questions and objectives of the study. Then, the objectives of the research are presented, in more details. Next, the scope of the study, and in the last section, the significance of the research is explained.

1.2 Background of the Study

Early computer viruses were broadly spread in the 1980s, as a sequence of wide usage of personal computers and Microsoft's new operating system (Sanok Jr., 2005; Makowsky, 2009). In late 1992, the reported number of malware spread in the cyber world was about 2,300 maximally, while this number reached to 60,000 known viruses until 2002, and in 2008, more than 100,000 computer viruses were detected and reported (Al Daoud *et al.*, 2008). Unfortunately, the number of computer viruses and other varieties of malware are growing, tremendously.

In today's digital world, although the reasons to create the new generations of computer viruses and malware programs have changed, we still hear about them in digital news daily. There are new ways and causes, which motivate computer experts to work on new kinds of these programs, and consequently, information security researches should concentrate on this area to find the new methods to be able and to improve defensive systems against them. Computer networks, Internet and web-based applications are the main streams in spreading such codes over the electronic world and they help virus authors to share their malware programs, computer by computer.

To be safe from harms of various malicious codes, antivirus software producers utilize different methodologies to protect computer systems. They employ scanning methods to detect the file signature. Scanning methods usually consist of scanning email attachments, scanning downloaded files, and static file scanning. In addition, they make use of scanning by heuristic methods and General Decryption to fight against with modern and more complicated and advanced computer viruses (Sanok Jr., 2005).

A computer virus is usually a part of a saved program, which when it is run is able to generate a duplication of itself in another saved program (Cohen, 1987; Agapow, 1993; Cohen and Cohen, 1994; Johansson, 1994; Colombell, 2002; Rajala, 2004; Aycock, 2006; Khosrow Pour, 2007; Alsagoff, 2008). During the procedure of reproduction, virus can modify its code in many various methods. It is important to mention that the term “virus” is sometimes incorrectly used to describe different kinds of computer malware. A genuine virus is able to replicate itself and spread from a computer to another, usually via different forms of executables, but this ability is not contained in other kinds of malware, such as worm, trojans, and spywares. In this report, these terms are employed.

In computer virology, metamorphic code is a piece of program, which is able to rewrite itself with new format. Usually, it performs this action through converting its own program into a provisional version, revise this intermediate copy, and finally rewrite its code back to standard code, another time. This process is carried out on the virus by its own body, and consequently the metamorphic engine becomes different (Al Daoud *et al.*, 2008). Some viruses use this approach while they are infecting new victims; as a result, the offspring are never similar to its original producer. Actually, this technique is utilized by the malware with the purpose of keeping away from the signature detection of static scanning engines used in antivirus softwares. Therefore, the main function and purpose of the malware do not vary by this conversion, but all others may change.

In practice, metamorphic computer viruses effectively change some sequences of their instructions with syntactic or semantic corresponding instructions sequences, in consecutive offspring (Webster and Malcolm, 2006). Thus, although their code format actually looks different, the behaviors of all generations are similar. Typically, it is performed to prevent detection methods based on static analysis, which are normally used by antivirus software engines. Thus, static code scanning method to detect metamorphic viruses is not applicable; instead, heuristic analysis can be exploited to find unknown variants of a metamorphic virus.

Several researches have deployed different techniques of machine learning to apply heuristic analysis in detection and classification of metamorphic viruses. One of the newest methods, which have achieved heuristic detection of metamorphic variants, is Hidden Markov Model (HMM), which is an appropriate technique in statistical pattern recognition. In the early 1970s, firstly HMM was applied to speech recognition problem (Rabiner, 1989). Until now, it has been used in many other areas, especially in biological sequence problems (Krogh, 1998). In recent years, many attempts are carried out to apply HMM in the problem of metamorphic virus detection. Wong and Stamp (2006) trained their models on disassembled executable codes of viruses. The pre-process of disassembly needs a lot of time and output of the process depends upon the capability of the disassembler. In (Attaluri *et al.*, 2009), authors used profile HMM and applied the method to analyze the metamorphic variants. Their findings proved that Profile Hidden Markov Models is successfully applicable to model the virus families. However, she also performed the implementation on assembly source codes. Govindaraj (2008), tried to implement the HMM algorithm on PE files. She followed the same methodology of (Wong, 2006), but she extended it into code segment of Portable Executable files.

1.3 Problem Statement

Since signature detection is the most commonly used detection strategy, virus writers have developed many techniques to evade such detection (Shanmugam *et al.*, 2013). Metamorphic computer viruses change their code as they spread, with the intention of preventing detection by static signature-based virus scanners (Konstantinou, 2008; Leder *et al.*, 2009; Murad *et al.*, 2010; Santos *et al.*, 2010; You and Yim, 2010; Xu *et al.*, 2013). In order to achieve this goal, metamorphic viruses employ different code mutation methods to contest intense static analysis (Tabish *et al.*, 2009; Runwal *et al.*, 2012). They are able to defeat dynamic analyzers, as well, when they sense they are running and examining in an environment, like an emulator, which is monitoring their behaviors (O'Kane *et al.*, 2011).

Detection of metamorphic computer viruses is not easy, because their authors have the knowledge of the feeblenesses of antivirus scanners. Static and dynamic analysis methods bring the limitations for antivirus scanners (Lee *et al.*, 2011). Metamorphic and obfuscation techniques make virus finding by means of signature scanning practically impossible (Santos *et al.*, 2010; Han *et al.*, 2011b; O'Kane *et al.*, 2011; Toderici and Stamp 2012). To detect a metamorphic computer virus, some other more complicated techniques such as inspecting the structure of the file, testing the behavior of the program, or machine learning methods must be used (Desai and Stamp, 2010). In other words, antivirus software should use heuristic techniques rather than string scanning to analyze and detect this kind of viruses (Konstantinou, 2008; Kasina *et al.*, 2010). Given this situation, it is highly required that researchers in information security area make serious efforts in studying the metamorphism (Xufang *et al.*, 2011). Today, a few researches are carried out to develop and implement new methods in order to improve the weakness of antivirus software's against the modern techniques which are exploited by metamorphic virus authors (Santos *et al.*, 2011). In response to this concern, this study was set up to investigate this issue further.

In this study, the following research questions will be answered:

1 What are the metamorphic computer viruses and their features?
2 What are the metamorphic computer virus classification and detection methods?
3 How metamorphic computer viruses can be reliably and effectively classified and detected?
4 How to improve the existing classification and detection methods?
5 How the performance of proposed model can be tested and evaluated?

1.4 Purpose of the Study

The purpose of this research is to propose and develop an enhanced model to classify and detect metamorphic viruses in format of portable executable. The performance of proposed model would also be evaluated and justified.

1.5 Objectives of the Study

The specific objectives of this study are:

i. To investigate the metamorphic computer viruses and their features.
ii. To analyze the existing metamorphic computer viruses classification and detection methods.
iii. To propose a new model for metamorphic computer virus classification and detection.
iv. To evaluate the performance of the proposed model.

1.6 Significance of the Study

The result of this research presents a noticeable knowledge in the computer security and virology science area. In addition, the result of this study can be exploited by antivirus software vendors. They can use the proposed model to improve antivirus products in defending against metamorphic computer viruses.

1.7 Scope of the Study

The focus of this study is on applying an enhanced form of Hidden Markov Model, particularly on the Portable Executable files. This study is to train the proposed model on the binary format of PEs, directly. The proposed model is applied on metamorphic virus family to involve the statistical features of the family.

The metamorphic viruses chosen for the data set used in the experiments in this study include Next Generation Virus Creation Kit (NGVCK) virus family. The NGVCK virus family have been chosen, because based on the study done by Wong (Wong, 2006), the NGVCK virus family is able to create viruses that share only a few percent of similarity. It means the NGVCK is the most powerful morphing engine to create the metamorphic virus.

1.8 Structure of the Thesis

In Chapter 1, an introduction to digital security and computer virology is given. A brief description of the metamorphic computer virus is provided. The problem statement, purpose and objectives of the study are presented.

In Chapter 2, a literature review on evolution of the computer virus concealment strategies, code obfuscation techniques, metamorphic computer virus detection techniques are given. In addition, some definitions of classification are introduced and Hidden Markov Model is explained, in details. In the last part of chapter 2, some of more related recent works are reviewed and gap of the study is presented.

Chapter 3 begins with an introduction to different research paradigms, research framework, and research design. Then, the proposed method and its evaluation are explained in details. At the end of chapter 3, instrumentation, assumptions and limitations of the study are presented.

Chapter 4 focuses on the experimental design. This chapter contains experimental setup, feature extraction, data set and implementation of the model.

In Chapter 5, the results of the experiments are given; the analysis of the proposed model and the threshold value are presented. Moreover, evaluation and justification of the proposed model based on a comparative study between the proposed model and four more related recent works are also given and discussed. The results of the proposed model are also compared with 44 famous commercial antiviruses.

Chapter 6 contains discussion and conclusions. Finally, the contributions and importance of this study, and the some recommendations for the future studies are given.

REFERENCES

Agapow, P.-M. (1993). *Computer viruses: the inevitability of evolution*. In Green, D. and Bossomaier, T. (Ed.) *Complex systems: from biology to computation*. (pp. 46-54). Amsterdam: IOS Press.

Agrawal, R., Grosky, W. and Fotouhi, F. (2009). Virus detection and removal service architecture in digital ecosystems. *Proceedings of the 3rd IEEE International Conference on Digital Ecosystems and Technologies (DEST '09).* 1-3 June 2009. Istanbul, pp. 301-305.

Al Daoud, E., Jebril, I. H. and Zaqaibeh, B. (2008). Computer Virus Strategies and Detection Methods. *International Journal of Open Problems in Computer Science and Mathematics*. 1(2), pp. 29-36.

Alon, J., Sclaroff, S., Kollios, G. and Pavlovic, V. (2003). Discovering clusters in motion time-series data. *Proceedings of the Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on.* 18-20 June 2003. pp. I-375-I-381 vol.371.

Alsagoff, S. N. (2008). Malware self protection mechanism. *Proceedings of the Information Technology, 2008. ITSim 2008. International Symposium on.* 26-28 Aug. 2008. pp. 1-8.

Amaral, J. N., Buro, M., Elio, R., Hoover, J., Nikolaidis, I., Salavatipour, M., Stewart, L. and Wong, K. (2011). About Computing Science Research Methodology.

Arlot, S. and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*. 4, pp. 40-79.

Arnold, W. and Tesauro, G. (2000). Automatically generated Win32 heuristic virus detection. *10th Virus Bulletin International Conference (VB2000).* 28-29 September 2000. *Orlando, FL, USA*, pp. 51-60.

Attaluri, S., McGhee, S. and Stamp, M. (2009). Profile hidden Markov models and metamorphic virus detection. *Journal in Computer Virology*. 5(2), pp. 151-169.

Aycock, J. (2006). *Computer Viruses and Malware.* New York, NY, USA: Springer.

Bailey, M., Oberheide, J., Andersen, J., Mao, Z., Jahanian, F. and Nazario, J. (2007). *Automated Classification and Analysis of Internet Malware*. In Kruegel, C., Lippmann, R. and Clark, A. (Ed.) *RAID'07 Proceedings of the 10th international conference on Recent advances in intrusion detection*. (pp. 178-197). Springer-Verlag Berlin, Heidelberg.

Bailey, M. W., Coleman, C. L. and Davidson, J. W. (2008). Defense against the dark arts. *SIGCSE Bulletin.* 40(1), pp. 315-319.

Baldi, P., Brunak, S., Chauvin, Y., Andersen, C. A. F. and Nielsen, H. (2000). Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics.* 16(5), pp. 412-424.

Basili, V. R. (1993). The Experimental Paradigm in Software Engineering. *Proceedings of the International Workshop on Experimental Software Engineering Issues: Critical Assessment and Future Directions.* pp. 3-12.

Bates, J. (1990). WHALE... A Dinosaur Heading For Extinction. *Virus Bulletin,* pp. 17-19.

Beaucamps, P. (2007). Advanced Polymorphic Techniques. *International Journal of Computer Science.* 2(3), pp. 194-205.

Benny (1999). Theme: Metamorphism. *29A E-Magazine,* No. 4.

Bicego, M., Cristani, M., Murino, V., Pękalska, E. and Duin, R. W. (2009). *Clustering-Based Construction of Hidden Markov Models for Generative Kernels*. In Cremers, D., Boykov, Y., Blake, A. and Schmidt, F. (Ed.) *Energy Minimization Methods in Computer Vision and Pattern Recognition*. Series of *Lecture Notes in Computer Science* (pp. 466-479). Springer Berlin Heidelberg.

Bicego, M., Murino, V. and Figueiredo, M. A. T. (2004). Similarity-based Classification of Sequences using Hidden Markov Models. *Pattern Recogn.* 37(12), pp. 2281-2291.

Bicego, M., Murino, V. and Figueiredo, M. T. (2003). *Similarity-based Clustering of Sequences using Hidden Markov Models*. In Perner, P. and Rosenfeld, A. (Ed.) *Machine Learning and Data Mining in Pattern Recognition*. Series of *Lecture Notes in Computer Science* (pp. 86-95). Springer Berlin Heidelberg.

Birney, E. (2001). Hidden Markov Models in biological sequence analysis. *IBM Journal of Research and Development.* 45(3.4), pp. 449-454.

Bishop, M. (2012). *Computer Security: Art and Science.* (1). Addison-Wesley.

Bolan, C. and Mende, D. (2004). Computer Security Research: Approaches and Assumptions. *Proceedings of the 2nd Australian Information Security Management Conference*. p. 115.

Bonfante, G., Kaczmarek, M. and Marion, J. Y. (2005). *Toward an Abstract Computer Virology*. In Hung, D. V. and Wirsing, M. (Ed.) *Proceedings of Second International Colloquium on Theoretical Aspects of Computing – ICTAC 2005*. (pp. 579-593). Hanoi, Vietnam: Springer Berlin / Heidelberg.

Bonfante, G., Kaczmarek, M. and Marion, J. Y. (2006). On Abstract Computer Virology from a Recursion Theoretic Perspective. *Journal in Computer Virology*. 1(3), pp. 45-54.

Bontchev, V. (1996). Possible Macro Virus Attacks and how to prevent them. *Computers & Security*. 15(7), pp. 595-626.

Bontchev, V. (1998). Macro Virus identification problems. *Computers & Security*. 17(1), pp. 69-89.

Bontchev, V. (1999). The problems of wordmacro virus upconversion. *Computers & Security*. 18(3), pp. 241-255.

Borello, J., Filiol, É. and Mé, L. (2009). Are current antivirus programs able to detect complex metamorphic malware? An empirical evaluation. *Proceedings of the 18th Annual EICAR Conference*. 11-12 May 2009. Berlin, Germany, pp. 45–63.

Borello, J. M. and Me, L. (2008). Code obfuscation techniques for metamorphic viruses. *Journal in Computer Virology*. 4(3), pp. 211-220.

Bosch, A. v. d. (2010). *Hidden Markov Models* In Sammut, C. and Webb, G. I. (Ed.) *Encyclopedia of Machine Learning*. (pp. 493-495). New York: Springer US.

Bramer, M. (2007). *Principles of Data Mining*. London: Springer.

Breon, R. and Katano, K. (1999). Microsoft Office 97 Executable Content Security Risks and Countermeasures. *Architectures and Applications Division of the Systems and Network Attack Center (SNAC), National Security Agency*.

Bridwell, L. and Tippett, P. (2001). ICSA Labs 7th Annual computer virus prevalence survey 2001. *ICSA Labs, A Division of TruSecure Corporation*. Herndon, VA, USA.

Bruschi, D., Martignoni, L. and Monga, M. (2006). *Detecting Self-mutating Malware Using Control-Flow Graph Matching*. In Büschkes, R. and Laskov, P. (Ed.) *Detection of Intrusions and Malware & Vulnerability Assessment*. Series of *Lecture Notes in Computer Science* (pp. 129-143). Springer Berlin / Heidelberg.

Bruschi, D., Martignoni, L. and Monga, M. (2007). Code normalization for self-mutating malware. *IEEE Security & Privacy,* 5/2, pp. 46-54.

Burset, M. and Guigó, R. (1996). Evaluation of gene structure prediction programs. *Genomics.* 34(3), pp. 353-367.

Cheetancheri, S. G. (2004). *Modelling a Computer Worm Defense System*. Master, Office of Graduate Studies, University of California Davis.

Chen, C., Wang, Y., Chang, Y. and Ricanek, K. (2012). Sensitivity analysis with cross-validation for feature selection and manifold learning. *Proceedings of the 9th international conference on Advances in Neural Networks - Volume Part I. Shenyang, China*, pp. 458-467.

Chen, T. (2003). Trends in viruses and worms. *The Internet Protocol Journal.* 6(3), pp. 23-33.

Chen, T. and Robert, J. (2004). The Evolution of Viruses and worms. *Statistical methods in computer security.*

Chouchane, M. R. and Lakhotia, A. (2006). Using engine signature to detect metamorphic malware. *Proceedings of the Proceedings of the 4th ACM Workshop on Recurring Malcode, WORM'06. Co-located with the 13th ACM Conference on Computer and Communications Security, CCS'06.* pp. 73-78.

Christodorescu, M. and Jha, S. (2003). Static analysis of executables to detect malicious patterns. *Proceedings of the 12th USENIX Security Symp. (SSYM'03).* pp. 169-186.

Christodorescu, M. and Jha, S. (2004). Testing malware detectors. *Proceedings of the ISSTA 2004 - Proceedings of the ACM SIGSOFT International Symposium on Software Testing and Analysis.* pp. 34-44.

Cohen, F. (1987). Computer Viruses: Theory and Experiments. *Computer & Security.* Vol. 6, pp. 22-35.

Cohen, F. and Cohen, F. (1994). *A short course on computer viruses.* New York, NY, USA: John Wiley & Sons, Inc.

Cohen, L., Manion, L. and Morrison, K. (2000). *Research methods in education.* Routledge.

Collberg, C., Thomborson, C. and Low, D. (1997). A taxonomy of obfuscating transformations. *Department of Computer Science, The University of Auckland, New Zealand*.

Collis, J. and Hussey, R. (2009). *Business Research: A Practical Guide for Undergraduate and Postgraduate Students.* Palgrave Macmillan.

Colombell, M. (2002). The legislative response to the evolution of computer viruses. *Rich. JL & Tech.* 8(3).

Cooil, B., Winer, R. S. and Rados, D. L. (1987). Cross-validation for prediction. *Journal of Marketing Research.* 24(3), pp. 271-279.

Costa, E., Lorena, A., Carvalho, A. and Freitas, A. (2007). A review of performance evaluation measures for hierarchical classifiers. *Proceedings of the Evaluation Methods for Machine Learning II: papers from the AAAI-2007 Workshop.* pp. 1-6.

Creswell, J. W. (2008). *Research design: Qualitative, quantitative, and mixed methods approaches.* SAGE Publications, Incorporated.

Daume, H. (2012). A Course in Machine Learning.

Desai, P. (2008). *Towards An Undetectable Computer Virus*. Master Thesis, Faculty of the Department of Computer Science, San Jose State University, San Jose, CA.

Desai, P. and Stamp, M. (2010). A highly metamorphic virus generator. *International Journal of Multimedia Intelligence and Security* 1(4), pp. 402 - 427.

Deshpande, S. (2012). *Eigenvalue Analysis for Metamorphic Detection*. Master Thesis, Faculty of the Department of Computer Science, San Jose State University, San Jose, CA.

Dietterich, T. G. (2002). Machine Learning for Sequential Data: A Review. *Proceedings of the Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition.* pp. 15-30.

Dodig-Crnkovic, G. (2002). Scientific methods in computer science. *Proceedings of the Conference for the Promotion of Research in IT at New Universities and at University Colleges in Sweden, Skövde, Suecia.* pp. 126-130.

Dugad, R. and Desai, U. (1996). A tutorial on hidden markov models. *Techn. Report SPANN-96-1, Signal Processing and Artificial Neural Networks Laboratory Department of Electrical Engineering Indian Institute of Technology—Bombay Powai, Bombay.* 400, p. 076.

Dunne, P. (2008). *Experimental Methods in Computing*. Unpublished note, Department of Computer Science, University of Liverpool.

Durand, J. and Atkison, T. (2012). Applying random projection to the classification of malicious applications using data mining algorithms. *Proceedings of the 50th Annual Southeast Regional Conference. Tuscaloosa, Alabama*, pp. 286-291.

Durbin, R., Eddy, S. R., Krogh, A. and Mitchison, G. (1998). *Biological sequence analysis: probabilistic models of proteins and nucleic acids.* Cambridge university press.

Dwan, B. (2000). The Computer Virus -- From There to Here.: An Historical Perspective. *Computer Fraud & Security.* 2000(12), pp. 13-16.

Elliott, R., Aggoun, L. and Moore, J. (1995). *Hidden Markov models: estimation and control.* Springer.

Ellis, T. J. and Levy, Y. (2009). Towards a guide for novice researchers on research methodology: Review and proposed methods. *Issues in Informing Science and Information Technology.* 6, pp. 323-337.

Feitelson, D. G. (2006). *Experimental computer science: The need for a cultural change*. Unpublished note, The Hebrew University of Jerusalem.

Feitelson, D. G. (2007). Introduction to Experimental Computer Science. *Communications of the ACM.* 50(11), pp. 24-26.

Fieguth, P. (2011). *Hidden Markov Models*. In *Statistical Image Processing and Multidimensional Modeling*. Series of *Information Science and Statistics* (pp. 215-239). Springer New York.

Filiol, E. (2005). *Computer viruses: from theory to applications.* Paris: Springer.

Filiol, E. (2007). Metamorphism, formal grammars and undecidable code mutation. *International Journal of Computer Science.* 2(1), pp. 70-75.

Filiol, E., Helenius, M. and Zanero, S. (2006). Open Problems in Computer Virology. *Journal in Computer Virology.* 1(3), pp. 55-66.

Fine, S., Singer, Y. and Tishby, N. (1998). The Hierarchical Hidden Markov Model: Analysis and Applications. *Machine Learning.* 32(1), pp. 41-62.

Finones, R. G. and Fernandez, R. T. (2006). Solving the metamorphic puzzle. *Virus Bulletin*, pp. 14-19.

Flach, P. (2012). *Machine learning: the art and science of algorithms that make sense of data.* Cambridge University Press.

Ford, R. (2004). The future of virus detection. *Information Security Technical Report.* 9(2), pp. 19-26.

Freeman, E. A. and Moisen, G. G. (2008). A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological Modelling.* 217(1–2), pp. 48-58.

Gao, P. (2012). A Measurement Approach to Binary Classifications and Thresholds. *Chicago Booth Research Paper.* (12-51).

Gavrilut, D., Cimpoesu, M., Anton, D. and Ciortuz, L. (2009). Malware detection using machine learning. *Proceedings of the Computer Science and Information Technology, 2009. IMCSIT '09. International Multiconference on.* 12-14 Oct. 2009. pp. 735-741.

Govindaraj, S. (2008). *Practical Detection of Metamorphic Computer Viruses*. Master Thesis, Faculty of the Department of Computer Science, San Jose State University, San Jose, CA.

Govindaraju, A. (2010). *Exhaustive Statistical Analysis for Detection of Metamorphic Malware*. Master Thesis, Faculty of the Department of Computer Science, San Jose State University, San Jose, CA.

Han, J. and Kamber, M. (2006). *Data mining: concepts and techniques.* (Second Edition). San Francisco: Morgan Kaufmann Publishers.

Han, J., Kamber, M. and Pei, J. (2011a). *Data Mining: Concepts and Techniques Third Edition.* Elsevier.

Han, K. S., Kang, B. and Im, E. G. (2011b). Malware classification using instruction frequencies. *Proceedings of the 2011 ACM Symposium on Research in Applied Computation. Miami, Florida*, pp. 298-300.

Hand, D., Mannila, H. and Smyth, P. (2001). *Principles of data mining.* The MIT press.

Highland, H. (1997). A history of computer viruses--Introduction. *Computers & Security.* 16(5), pp. 412-415.

Hruska, J. (2002). Virus detection. *Proceedings of the European Conference on Security and Detection, 1997 (ECOS 97).* 28-30 Apr 1997. London, UK, pp. 128-130.

Hughes, L. and DeLone, G. (2007). Viruses, worms, and trojan horses: Serious crimes, nuisance, or both? *Social science computer review.* 25(1), pp. 78-89.

Ian H. Witten, Eibe Frank and Hall, M. A. (2011). *Data Mining: Practical machine learning tools and techniques.* (3rd). Morgan Kaufmann.

Jacobs, R. (1991). Tequila - A Cocktail of Viral Tricks. *Virus Bulletin,* June 1991, pp. 16-17.

Jain, A. K., Duin, R. P. W. and Jianchang, M. (2000). Statistical pattern recognition: a review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on.* 22(1), pp. 4-37.

Jensen, R. S. (2002). *Immune system for virus detection and elimination*. Master Thesis, Department of Informatics and Mathematical Modelling, Technical University of Denmark (DTU).

Johansson, K. (1994). *COMPUTER VIRUSES: The Technology and Evolution of an Artificial Life Form.*

Jordan, M. (2002). Dealing with Metamorphism. *Virus Bulletin,* October 2002, pp. 4-6.

Josse, S. (2007). Secure and advanced unpacking using computer emulation. *Journal in Computer Virology.* 3(3), pp. 221-236.

Karlof, C. and Wagner, D. (2003). *Hidden Markov Model Cryptanalysis*. In *Cryptographic Hardware and Embedded Systems - CHES 2003*. Series of *Lecture Notes in Computer Science* (pp. 17-34). Springer Berlin / Heidelberg.

Karnik, A., Goswami, S. and Guha, R. (2007). Detecting obfuscated viruses using cosine similarity analysis. *AMS 2007: First Asia International Conference on Modelling & Simulation Asia Modelling Symposium, Proceedings.* pp. 165-170.

Kasina, A., Suthar, A. and Kumar, R. (2010). *Detection of Polymorphic Viruses in Windows Executables*. In Ranka, S., Banerjee, A., Biswas, K. K., Dua, S., Mishra, P., Moona, R., Poon, S.-H. and Wang, C.-L. (Ed.) *Contemporary Computing*. Series of *Communications in Computer and Information Science* (pp. 120-130). Springer Berlin Heidelberg.

Kaspersky, E. (1996). Zhengxi: Saucerful of Secrets. *Virus Bulletin,* April 1996, pp. 8-10.

Keller, B. and Lutz, R. (2002). *Improved Learning for Hidden Markov Models Using Penalized Training*. In O'neill, M., Sutcliffe, R., Ryan, C., Eaton, M. and Griffith, N. (Ed.) *Artificial Intelligence and Cognitive Science*. Series of *Lecture Notes in Computer Science* (pp. 153-166). Springer Berlin / Heidelberg.

Kephart, J., Sorkin, G., Chess, D. and White, S. (1997). Fighting computer viruses. *Scientific American.* 277(5), pp. 56-61.

Khosrow Pour, M. (2007). *Dictionary of Information Science and Technology*. Hershey, PA, USA: Idea Group Reference.

Kiltz, S., Lang, A. and Dittmann, J. (2008). *Malware: Specialized Trojan Horse*. In Janczewski, L. and Colarik, A. (Ed.) *Cyber warfare and cyber terrorism*. (pp. 154-160). IGI Global.

Kim, M. and Pavlovic, V. (2010). Sequence classification via large margin hidden Markov models. *Data Mining and Knowledge Discovery*. pp. 1-23.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th international joint conference on Artificial intelligence (IJCAI'95).* 20-25 August 1995. Montréal, Québec, Canada, pp. 1137-1143.

Konstantinou, E. (2008). *Metamorphic Virus: Analysis and Detection*. Master Thesis, Department of Mathematics, Royal Holloway, University of London, Egham, England.

Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Proceedings of the 2007 conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies.* pp. 3-24.

Krogh, A. (1998). An introduction to hidden Markov models for biological sequences. *Computational Methods in Molecular Biology*. pp. 45-63.

Kumar, R. (2010). *Research methodology: a step-by-step guide for beginners.* Sage Publications Limited.

Lakhotia, A., Kapoor, A. and Kumar, E. (2004). Are metamorphic viruses really invincible? *Virus Bulletin,* December 2004, pp. 5–7.

Lakhotia, A. and Singh, P. K. (2003). Challenges in getting 'formal' with viruses. *Virus Bulletin,* September 2003, pp. 15-19.

Lammer, P. (1990). Jonah's Journey. *Virus Bulletin*, November 1990, p. 20.

Leder, F., Steinbock, B. and Martini, P. (2009). Classification and detection of metamorphic malware using value set analysis. *Proceedings of the 4th International Conference on Malicious and Unwanted Software (MALWARE).* 13-14 Oct. 2009. pp. 39-46.

Lee, J., Im, C. and Jeong, H. (2011). A study of malware detection and classification by comparing extracted strings. *Proceedings of the 5th International Conference* *on Ubiquitous Information Management and Communication. Seoul, Korea*, pp. 1-4.

Leedy, P. D. and Ormrod, J. E. (2009). *Practical research: planning and design.* (9th). Pearson

Li, C. and Biswas, G. (1999). Clustering sequence data using hidden Markov model representation. *Proceedings of the AeroSense'99.* pp. 14-21.

Li, Z., Wu, Z., He, Y. and Fulei, C. (2005). Hidden Markov model-based fault diagnostics method in speed-up and speed-down process for rotating machinery. *Mechanical Systems and Signal Processing.* 19(2), pp. 329-339.

Lin, D. (2009). *Hunting for Undetectable Metamorphic Viruses*. Master Thesis, Faculty of the Department of Computer Science, San Jose State University, San Jose, CA.

Lin, D. and Stamp, M. (2010). Hunting for undetectable metamorphic viruses. *Journal in Computer Virology*. pp. 1-14.

Liu, L. and Zsu, M. T. (2009). *Encyclopedia of Database Systems.* Springer Publishing Company, Incorporated.

Ludwig, M. (1995). *The Giant Black Book of Computer Viruses.* Arizona: American Eagle Publications.

Lyne, J. (2013). Security Threat Report 2013, New platforms and changing threats. *SOPHOS.*

Makowsky, J. (2009). *Computer Risks and Insurability of Software*. Unpublished note, Technion - Israel Institute of Technology.

Mamitsuka, H. (1997). Supervised learning of hidden Markov models for sequence discrimination. *Proceedings of the first annual international conference on Computational molecular biology. Santa Fe, New Mexico, USA*, pp. 202-208.

Mansfield-Devine, S. (2013). Security review: the past year. *Computer Fraud & Security*. 2013(1), pp. 5-11.

Marinescu, A. (1999). ACG in the Hole. *Virus Bulletin*, July 1999, pp. 8-9.

Marpaung, J. A., Sain, M. and Lee, H.-J. (2012). Survey on malware evasion techniques: state of the art and challenges. *Proceedings of the Advanced Communication Technology (ICACT), 2012 14th International Conference on.* pp. 744-749.

MathWorks (2008). *MATLAB - The Language Of Technical Computing*, Ver. R2008a, 'More information on online website:' http://www.mathworks.com/

Maxion, R. A. (2009). Experimental Methods for Computer Science Research. *Proceedings of the Dependable Computing, 2009. LADC '09. Fourth Latin-American Symposium on.* 1-4 Sept. 2009. pp. 136-136.

McBrewster, J., Miller, F. and Vandome, A. (2009). *Computer Virus: Timeline of computer viruses and worms, Computer program, Malware, Adware, Spyware, Computer worm, Trojan horse (computing), Elk Cloner,... Polymorphic code, Metamorphic code.* Alphascript Publishing.

Medlock, B. (2008). *Investigating classification for natural language processing tasks.* VDM Verlag.

Michie, D., Spiegelhalter, D. and Taylor, C. (1994). *Machine learning, neural and statistical classification.* Ellis Horwood.

Microsoft Corporation (2013). Microsoft Portable Executable and Common Object File Format Specification.

Miller, L. (1994). Stealth, Polymorphism and Other Strange Words. *CHIPS Magazine.* p. 48.

Mishra, P. (2003). *Taxonomy of Uniqueness Transformations*. Master Thesis, The Faculty of the Department of Computer Science, San Jose State University, San Jose, CA.

Morar, J. and Chess, D. (2000). Can Cryptography Prevent Computer Viruses? *10th Virus Bulletin International Conference (VB2000).* 28-29 September 2000. *Orlando, FL, USA*, pp. 127-138.

Mori, A. (2004). Detecting Unknown Computer Viruses - A New Approach. *Proceedings of the International Symposium, ISSS 2003.* Tokyo, Japan, pp. 226-241.

Mouret, M., Solnon, C. and Wolf, C. (2009). Classification of Images Based on Hidden Markov Models. *Proceedings of the Content-Based Multimedia Indexing, 2009. CBMI '09. Seventh International Workshop on.* 3-5 June 2009. pp. 169-174.

Murad, K., Shirazi, S., Zikria, Y. and Ikram, N. (2010). *Evading Virus Detection Using Code Obfuscation*. In Kim, T.-H., Lee, Y.-H., Kang, B.-H. and Slezak, D. (Ed.) *Future Generation Information Technology*. Series of *Lecture Notes in Computer Science* (pp. 394-401). Springer Berlin / Heidelberg.

Murphy, K. P. (2007). Performance evaluation of binary classifiers. *University of British Columbia*.

Myers, M. D. (1997). Qualitative research in information systems. *Management Information Systems Quarterly*. 21, pp. 241-242.

Myers, M. D. (2008). *Qualitative Research in Business & Management.* SAGE Publications Limited.

Nachenberg, C. (1997). Computer Virus-Coevolution. *Communications of the ACM.* 50(1), pp. 46-51.

Nachenberg, C. (1998). Understanding heuristics: Symantec's bloodhound technology.

Nikishin, A. and Pavluschick, M. (1999). pOLEmorphism. *Virus Bulletin,* June 1999, pp. 14-15.

O'Kane, P., Sezer, S. and McLaughlin, K. (2011). Obfuscation: The Hidden Malware. *Security & Privacy, IEEE.* 9(5), pp. 41-47.

Orlikowski, W. J. and Baroudi, J. J. (1991). Studying information technology in organizations: Research approaches and assumptions. *Information systems research.* 2(1), pp. 1-28.

Panuccio, A., Bicego, M. and Murino, V. (2002). A Hidden Markov Model-Based Approach to Sequential Data Clustering. *Proceedings of the Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition.* pp. 734-742.

Paquette, J. (2000). A history of viruses. *SecurityFocus, January.* 16, p. 2004.

Pearce, S. (2003). *Viral polymorphism*. Unpublished note, SANS Institute.

Peisert, S. and Bishop, M. (2007). *How to Design Computer Security Experiments*. In Futcher, L. and Dodge, R. (Ed.) *Fifth World Conference on Information Security Education*. Series of *IFIP — International Federation for Information Processing* (pp. 141-148). Springer US.

Perriot, F. and Ferrie, P. (2004). Principles and practise of x-raying. *14th Virus Bulletin International Conference (VB2004).* 29 September - 1 October 2004. *Fairmont Chicago, Illinois, USA*, pp. 51–56.

Perriot, F., Ször, P. and Ferrie, P. (2003). Striking similarites: Win32/simile and metamorphic virus code. *Symantec Corporation.*

Pietrek, M. (1994). Peering Inside the PE: A Tour of the Win32 Portable Executable File Format. *Microsoft Systems Journal*.

Pietrek, M. (2002). An In-Depth Look into the Win32 Portable Executable File Format. *MSDN Magazine*, 17/2.

Polk, W. T., Bassham, L. E., Wack, J. P. and Carnahan, L. J. (1995). *Anti-virus tools and techniques for computer systems.* Park Ridge, NJ, USA: Noyes Publications.

Poritz, A. (2002). Hidden Markov models: A guided tour. *Proceedings of the 1988 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1988. ICASSP-88.* 11-14 Apr 1988 New York, NY pp. 7-13.

Priyadarshi, S. (2011). *Metamorphic Detection via Emulation*. Master Thesis, The Faculty of the Department of Computer Science, San Jose State University, San Jose, CA.

Rabiner, L. and Juang, B. (1986). An introduction to hidden Markov models. *IEEE ASSp Magazine.* 3(1), pp. 4-16.

Rabiner, L. R. (1989). A Tutorial on Hidden markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE.* 77(2), pp. 257-286.

Rad, B. B. and Masrom, M. (2010). Metamorphic Virus Variants Classification Using Opcode Frequency Histogram. *Proceedings of the 14th WSEAS International Conference on COMPUTERS.* July 23-25, 2010. Corfu Island, Greece, pp. 147-155.

Rad, B. B., Masrom, M. and Ibrahim, S. (2011). Evolution of Computer Virus Concealment and Anti-Virus Techniques: A Short Survey. *International Journal of Computer Science Issues (IJCSI).* 8(1), pp. 113-121.

Rajala, J. B. (2004). Computer Virus Protection. *THE Journal (Technological Horizons In Education).* 31(9), pp. 22-23.

Rodriguez, J. D., Perez, A. and Lozano, J. A. (2010). Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* 32(3), pp. 569-575.

Roll-Hansen, N. (2009). *Why the distinction between basic (theoretical) and applied (practical) research is important in the politics of science.*

Runwal, N., Low, R. M. and Stamp, M. (2012). Opcode graph similarity and metamorphic detection. *Journal of Computer Virology.* 8(1-2), pp. 37-52.

Russell, S. and Norvig, P. (2009). *Artificial intelligence: a modern approach.* Prentice Hall.

Salah, D., Aslan, H. K. and El-Hadidi, M. T. (2002). A Detection Scheme for the SK Virus. *Proceedings of the IFIP TC11 17th International Conference on*

Information Security: Visions and Perspectives (SEC2002). 7-9 May 2002. Cairo, Egypt, pp. 171-182.

Salim, N., Shamsuddin, S. M. H., Deris, S., Alias, R. A., Ibrahim, S., Sallehuddin, R., Hashim, S. Z. M., Rahman, A. A., Jawawi, D. N. A., Rahim, N. Z. A. and Salim, J. (2010). *Handbook of Research Methods in Computing.* (1st Edition). Skudai, Johor, Malaysia: Faculty of Computer Science and Information Systems Universiti Teknologi Malaysia.

Salomon, D. (2006). *Examples of Malware*. In *Foundations of Computer Security*. (pp. 125-137). London: Springer.

Sanok Jr., D. J. (2005). An Analysis of How Antivirus Methodologies Are Utilized in Protecting Computers from Malicious Code. *Proceedings of the 2nd annual conference on Information security curriculum development (InfoSecCD '05).* 23-24 September 2005. Kennesaw, GA, USA, pp. 142-144.

Santos, I., Brezo, F., Nieves, J., Penya, Y., Sanz, B., Laorden, C. and Bringas, P. (2010). *Idea: Opcode-Sequence-Based Malware Detection*. In Massacci, F., Wallach, D. and Zannone, N. (Ed.) *Engineering Secure Software and Systems*. Series of *Lecture Notes in Computer Science* (pp. 35-43). Springer Berlin Heidelberg.

Santos, I., Brezo, F., Ugarte-Pedrero, X. and Bringas, P. G. (2011). Opcode sequences as representation of executables for data-mining-based unknown malware detection. *Information Sciences.* (0).

Satitsuksanoh, P., Sophatsathit, P. and Lursinsap, C. (2009). *A Hybrid Technique for Complete Viral Infected Recovery*. In Papasratorn, B., Chutimaskul, W., Porkaew, K. and Vanijja, V. (Ed.) *Advances in Information Technology*. Series of *Communications in Computer and Information Science* (pp. 147-159). Springer Berlin Heidelberg.

Schultz, M. G., Eskin, E., Zadok, E. and Stolfo, S. J. (2001). Data Mining Methods for Detection of New Malicious Executables. *2001 IEEE Symposium on Security and Privacy.* 14-16 May 2011. *Oakland, California* p. 38.

Seeley, D. (1989). A Tour of the Worm. *Proceedings of the 1989 Winter USENIX Conference.* February 1989. Berkeley, CA, pp. 287–304.

Seliya, N., Khoshgoftaar, T. M. and Van Hulse, J. (2009). Aggregating performance metrics for classifier evaluation. *Proceedings of the Information Reuse &*

Integration, 2009. IRI '09. IEEE International Conference on. 10-12 Aug. 2009. pp. 35-40.

Shanmugam, G., Low, R. and Stamp, M. (2013). Simple substitution distance and metamorphic detection. *Journal of Computer Virology and Hacking Techniques.* pp. 1-12.

Shevchenko, A. (2007). The evolution of self-defense technologies in malware. *Kaspersky LAB Reading Room.*

Singh, A. (2009). *Portable Executable File Format*. In *Identifying Malicious Code Through Reverse Engineering*. Series of *Advances in Information Security* (pp. 1-15). Springer US.

Singhal, P. and Raul, N. (2012). Malware detection module using machine learning algorithms to assist in centralized security in enterprise networks. *International Journal of Network Security & Its Applications (IJNSA).* 4(1), pp. 61-67.

Skulason, F. (1990a). 1260 - The Variable Virus. *Virus Bulletin,* March 1990, p. 12.

Skulason, F. (1990b). Virus Encryption Techniques. *Virus Bulletin,* November 1990, pp. 13-16.

Snyder, L. (1994). *Academic careers for experimental computer scientists and engineers.* National Academies Press.

Spafford, E. (1991). Computer viruses: A form of artificial life. *Artificial Life II, Studies in the Sciences of Complexity.* pp. 727–747.

Spafford, E. (1994). Computer viruses as artificial life. *Artificial Life.* 1(3), pp. 249-265.

Srinivasan, R. (2007). *Protecting Anti-Virus Software Under Viral Attacks*. Master Thesis, ARIZONA STATE UNIVERSITY.

Srivastava, P. K., Desai, D. K., Nandi, S. and Lynn, A. M. (2007). HMM-ModE–Improved classification using profile hidden Markov models by optimising the discrimination threshold and modifying emission probabilities with negative training sequences. *BMC bioinformatics.* 8(1), p. 104.

Stallings, W., Brown, L., Bauer, M. and Howard, M. (2008). *Computer Security: Principles and Practice.* (1st Edition). Prentice Hall.

Stamp, M. (2012). *A revealing introduction to hidden Markov models*. Unpublished note, San Jose State University.

Sukamolson, S. (2007). Fundamentals of quantitative research.

Sun, L., Versteeg, S., Boztaş, S. and Yann, T. (2010). *Pattern Recognition Techniques for the Classification of Malware Packers*. In Steinfeld, R. and Hawkes, P. (Ed.) *Information Security and Privacy*. Series of *Lecture Notes in Computer Science* (pp. 370-390). Springer Berlin Heidelberg.

Symantec (2001). Virus-Antivirus Co-evolution. *Symantec Research Labs*.

Symantec (2010). State of Enterprise Security. *Symantec Corp.*

Szor, P. (1997a). Coping with Cabanas. *Virus Bulletin*, November 1990, pp. 10-12.

Szor, P. (1997b). Junkie Memorial? *Virus Bulletin,* September 1997, pp. 6-8.

Szor, P. (1998a). Attacks On Win32. *8th Virus Bulletin International Conference (VB'98).* 22-23 October 1998. *Munich, Germany*, pp. 57-84.

Szor, P. (1998b). The Marburg Situation. *Virus Bulletin,* November 1998, pp. 8-10.

Szor, P. (2000). The new 32-bit medusa. *Virus Bulletin,* December 2000, pp. 8-10.

Szor, P. (2005). *The Art of Computer Virus Research and Defense.* Addison-Wesley Professional.

Szor, P. and Ferrie, P. (2001a). Hunting for Metamorphic. *11th Virus Bulletin International Conference.* 27-28 September 2001. *Prague, Czech Republic*, pp. 123-144.

Szor, P. and Ferrie, P. (2001b). Zmist opportunities. *Virus Bulletin,* March 2001, pp. 6-7.

Tabish, S. M., Shafiq, M. Z. and Farooq, M. (2009). Malware detection using statistical analysis of byte-level file content. *Proceedings of the ACM SIGKDD Workshop on CyberSecurity and Intelligence Informatics. Paris, France*, pp. 23-31.

Theodoridis, S., Pikrakis, A., Koutroumbas, K. and Cavouras, D. (2010). *Hidden Markov Models*. In *Introduction to Pattern Recognition*. (pp. 147-157). Boston: Academic Press.

Tichy, W. F. (1998). Should Computer Scientists Experiment More? *Computer.* 31(5), pp. 32-40.

Tichy, W. F., Lukowicz, P., Prechelt, L. and Heinz, E. A. (1995). Experimental evaluation in computer science: A quantitative study. *Journal of Systems and Software.* 28(1), pp. 9-18.

Toderici, A. and Stamp , M. (2012). Chi-squared distance and metamorphic virus detection. *Journal in Computer Virology*. pp. 1-14.

Trochim, W. M. and Donnelly, J. P. (2008). *Research methods knowledge base.* Atomic Dog/Cengage Learning.

Udupa, S., Debray, S. and Madou, M. (2005). Deobfuscation: Reverse engineering obfuscated code. *Proceedings of the 12th Working Conference on Reverse Engineering (WCRE'05).* November 07-November 11. Pittsburgh, Pennsylvania, p. 10.

Vaseghi, S. V. (2007). *Multimedia signal processing: theory and applications in speech, music and communications.* Wiley.

Vecna (1998). Miss Lexotan 8. *29A E-Zine.*

Venkatachalam, S. (2010). *Detecting Undetectable Computer Viruses*. Master Thesis, The Faculty of the Department of Computer Science, San Jose State University, San Jose, CA.

Venkatesan, A. (2008). *Code Obfuscation And Virus Detection*. Master Thesis, The Faculty of the Department of Computer Science, San Jose State University, San Jose, CA.

Vinod, P., Laxmi, V., Gaur, M. S., Phani Kumar, G. V. S. S. and Chundawat, Y. S. (2009). Static CFG analyzer for metamorphic malware code. *Proceedings of the SIN'09 - Proceedings of the 2nd International Conference on Security of Information and Networks.* pp. 225-228.

Vrabec, J. and Harley, D. (2010). Real Performance? *19th EICAR Annual Conference.* 8-9 May 2010. *Paris, France*.

VXHeavens. 2009. *VX Heavens - Computer Virus Information, Library, Collection, and Sources* [Online]. VXHeavens. Available: http://vx.netlux.org/vl.php [Accessed].

Wahyuni, D. (2012). The Research Design Maze: Understanding Paradigms, Cases, Methods and Methodologies. *Journal of Applied Management Accounting Research.* 10(1), pp. 69-80.

Walenstein, A., Mathur, R., Chouchane, M. and Lakhotia, A. (2007). The design space of metamorphic malware. *Proceedings of the Proceedings of the 2nd International Conference on Information Warfare and Security (ICIW 2007).* 8-9 March 2007. Monterey, California, USA, pp. 241-248.

Walenstein, A., Mathur, R., Chouchane, M. R. and Lakhotia, A. (2006). Normalizing metamorphic malware using term rewriting. *Proceedings of the Sixth IEEE*

International Workshop on Source Code Analysis and Manipulation, Proceedings. pp. 75-84.

Wang, J.-H., Deng, P. S., Fan, Y.-S., Jaw, L.-J. and Liu, Y.-C. (2003). Virus detection using data mining techinques. *Proceedings of the IEEE 37th Annual 2003 International Carnahan Conference on Security Technology*. 14-16 Oct. 2003 Taipei, Taiwan, pp. 71-76.

Webster, M. and Malcolm, G. (2006). Detection of metamorphic computer viruses using algebraic specification. *Journal in Computer Virology*. 2(3), pp. 149-161.

Wiggins, G. (2001). Living with malware. *SANS Institute*.

Wong, W. (2006). *Analysis and detection of metamorphic computer viruses*. Master Thesis, The Faculty of the Department of Computer Science, San Jose State University, San Jose, CA.

Wong, W. and Stamp, M. (2006). Hunting for metamorphic engines. *Journal in Computer Virology*. 2(3), pp. 211-229.

Wroblewski, G. (2002). General method of program code obfuscation. *Proceedings of*, pp. 153-159.

Wu, Z., Gianvecchio, S., Xie, M. and Wang, H. (2010). Mimimorphism: a new approach to binary code obfuscation. *Proceedings of the 17th ACM conference on Computer and communications security. Chicago, Illinois, USA*, pp. 536-546.

Xing, K., Deng, R., Wang, J., Feng, J., Huang, M. and Wang, X. (2005). Analysis and prediction of baculovirus promoter sequences. *Virus research.* 113(1), pp. 64-71.

Xing, Z., Pei, J. and Keogh, E. (2010). A brief survey on sequence classification. *SIGKDD Explor. Newsl.* 12(1), pp. 40-48.

Xu, M., Wu, L., Qi, S., Xu, J., Zhang, H., Ren, Y. and Zheng, N. (2013). A similarity metric method of obfuscated malware using function-call graph. *Journal of Computer Virology and Hacking Techniques.* pp. 1-13.

Xufang, L., Loh, P. K. K. and Tan, F. (2011). Mechanisms of Polymorphic and Metamorphic Viruses. *Proceedings of the 2011 European Intelligence and Security Informatics Conference (EISIC).* 12-14 Sept. 2011. pp. 149-154.

You, I. and Yim, K. (2010). Malware Obfuscation Techniques: A Brief Survey. *Proceedings of the Fifth International Conference on Broadband, Wireless Computing, Communication and Applications (BWCCA 2010).* 4-6 November 2010. Fukuoka, Japan, pp. 297-300.

Yount, W. R. (2006). *Research Design & Statistical Analysis in Christian Ministry*. (4th). W.R. Yount (1988).

Yusoff, M. and Jantan, A. (2011). *A Framework for Optimizing Malware Classification by Using Genetic Algorithm*. In Zain, J., Wan Mohd, W. and El-Qawasmeh, E. (Ed.) *Software Engineering and Computer Systems*. Series of *Communications in Computer and Information Science* (pp. 58-72). Springer Berlin Heidelberg.

Zenkin, D. (2001). Fighting against the invisible enemy - Methods for detecting an unknown virus. *Computers & Security*. 20(4), pp. 316-321.

Zhang, Q. (2008). *Polymorphic and metamorphic malware detection*. Ph.D. Thesis, Graduate Faculty, North Carolina State University, Raleigh, NC, USA.

Zhang, Y., Li, T. and Qin, R. (2008). *Computer Virus Evolution Model Inspired by Biological DNA*. In Huang, D.-S., Wunsch, D., Levine, D. and Jo, K.-H. (Ed.) *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence*. Series of *Lecture Notes in Computer Science* (pp. 943-950). Springer Berlin / Heidelberg.

Zobel, J. (1998). Reliable research: Towards experimental standards for computer science. *Proceedings of the Proceedings of the Australasian Computer Science Conference*. February 1998. Perth, Western Australia, pp. 217-229.