Characteristics based Detection of Internet worms using Combined Machine Learning Methods and Worm Containment

By

S. Divya
(12PHCSF001)

Supervisor
Dr. G.Padmavathi

A Thesis Submitted to
Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore - 641 043

In partial fulfilment of the requirements for the degree of Doctor of philosophy in Computer Science

June 2015

CERTIFICATE

This is to certify that the thesis entitled **"Characteristics based Detection of Internet worms using Combined Machine Learning Methods and Worm Containment"** submitted to the Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for the award of the degree of **Doctor of Philosophy in Computer Science**, is a record of original research work done by **S. Divya,** during the period of her study in the Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, under my supervision and guidance and the thesis has not formed the basis for the award of any Degree / Diploma / Associateship / Fellowship or similar title to any candidate of any University.

Signature of the Head of the Department

Signature of the Supervisor

Signature of the Dean

DECLARATION

I hereby declare that the thesis titled **"Characteristics based Detection of Internet worms using Combined Machine Learning Methods and Worm Containment"** is the result of investigations carried out by me in the Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, under the supervision and guidance of **Dr. G. Padmavathi,** Professor and Head, Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, University, Coimbatore, and that it has not been submitted for the award of any Degree / Diploma / Associateship / Fellowship or similar title of any University or Institute.

Signature of the Supervisor **Signature of the Candidate**

ACKNOWLEDGEMENT

First and foremost, I am extremely thankful to the **GOD ALMIGHTY** for his graces and blessings bestowed on me.

I would like to place on record my reverential gratitude to **Late Ayya Dr. T. S. Avinashilingam Avl.**, Founder, President and First Chancellor of Avinashilingam University for Women, Coimbatore for providing the temple of learning and I owe my sincere and humble gratitude to **Late Amma Dr. Rajammal P. Devadas Avl., M.A., M.Sc., Ph.D.** Former Chancellor, Avinashilingam University for Women, Coimbatore for their heavenly blessings.

I record my sincere thanks to **Dr. T. S. K. Meenakshi Sundaram,** Chancellor, Avinashilingam University for Women, Coimbatore for providing the infrastructure for the conduct of the research.

I express my sincere and heartfelt thanks to **Dr. Sheela Ramachandaran, M.Sc., PG. Dip., Ph.D.**, Vice Chancellor, Avinashilingam University for Women, Coimbatore for providing necessary facilities and resources for the successful completion of this research work.

I am gratefully indebted to **Dr. (Mrs.) A. Venmathi**, M.Sc., Dip. Ed., M.Phil., Ph.D., Registrar (i/c), Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for continuous help to carry out this research programme successfully.

I also record my thanks to **Dr. Mrs. Parvathi,** Dean, Faculty of Science, University for Women, Coimbatore for motivation, constant encouragement, care and advice throughout the research.

I feel extremely privileged and fortunate to have worked under the able guidance and professional supervision of my guide, **Dr. G.Padmavathi, M.Sc., M.Phil., Ph.D.,** Professor and Head, Department of Computer Science, Avinashilingam University for Women, Coimbatore. Her constant encouragements, motivation, valuable and constructive suggestions and untiring support have guided me to gain and explore deep knowledge in the research field.

She helped me to define my research goals and showed the way to achieve them. Her able guidance, systematic approach guided me to put my best possible efforts in completing my work and documentation. Her sympathetic friendly nature, timely counseling, willingness to help at anytime, anywhere and at any situation during the entire period of the study have molded this research into a reality.

My special thanks to **Dr. E. George Dharma Prakash Raj,** Department of Computer Science and Engineering, Bharathidasan University, Trichy, for providing encouraging and constructive feedback for my research work.

I record my gratitude to all the **Faculty members** of Department of Computer Science and **Staff Members** of Computer Center, Avinashilingam University for Women, Coimbatore for their encouragement and support.

I also thank **my entire family** for their patience, wishes, prayers, encouragement and constant support extended to me at all the times throughout my career. I also thank all my **Friends** for their support and encouragement.

S. Divya

ABSTRACT

With the rapid growth of Internet today, many Internet based applications are evolving. Internet is an open network accessed by all. The major challenge in Internet is security. Many attacks and vulnerabilities affect the network. Among the various attacks, Internet worms are vulnerable because they infect a large number of hosts within a short period of time and from that infected hosts they further initiate attacks like distributed denial-of-service, phishing and spyware through their propagation. Internet worms like "Code-Red" in 2001, "Slammer" in 2003, "Witty"/ "Sasser" in 2004, Storm in 2007, Conficker in 2008 and StuxNet in 2010-2012 have created prominent damages to the hosts. Within the period of five years, 4,00,000 computers got infected due to Blaster worm in 2003. Conficker worm damaged approximately 13 million IP addresses. This number may increase year by year. To overcome these damages and to defend against these attacks, effective defense mechanism is necessary. Therefore, detection and containment of Internet worms are the need of the day. For Internet Worm detection, there are two main approaches existing namely, signature based and anomaly based. Out of the two, anomaly based detection schemes provide better detection on newly appearing worms. There are various anomaly based detection approaches exist in the literature such as Probabilistic modeling, Spectrum based, Statistical estimation, Game theory, Epidemic spreading and Machine learning methods. Among these approaches, Machine Learning methods provide faster detection accuracy for rapidly changing Internet worms. Containment methods are applied to prevent the network from further infection after detection.

Based on the challenges created by Internet worms, the objectives of this research work are formulated after studying significant literatures. Though Internet worms are detected using different Machine Learning Approaches, the detection based on the characteristics of worms provides for better detection of new unknown worms. The characteristics refer to the nature of worms and it makes the detection effective at the initial stages of the propagation itself.

A Three-Step Methodology is proposed with four contributions to meet the objectives of the thesis. The Internet worm detection is done using the combined Machine Learning Methods based on anomaly detection schemes and Containment based on

blocking schemes. The proposed **P**rincipal Component Analysis with **M**ulticlass Support Vector Machine and **R**abin Footprint Algorithm ***(PMR)*** detects the Malcode existence in the downloaded programs based on unknown signatures. The detected and classified Malcode programs are contained to prevent from further infection. The proposed **D**eterministic Finite Automata with **F**uzzy Logic Classifier and **F**ilter-Ary Sketch (***DFF)*** performs detection and containment of malicious contents in packets based on payload. The proposed **E**nhanced **C** 4.5 Algorithm and **B**lacklist ***(ECB)*** detects and contains the unused addresses based on illegal traffic. The proposed **k**ernelized **E**xtreme Learning Machine with **A**utomated Worm Containment Algorithm ***(kEA)*** is used for detection and containment of malicious traffic from non-existing IP addresses based on connection attempt failures.

The proposed methods are implemented using Java NetBeans IDE 7.4 and Microsoft SQL Server. The parameters used for evaluation are Memory Utilization, Time Consumption, Precision Value, Recall Value, Accuracy, Detection Rate and Containment Rate. In contribution one, the detection accuracy achieved by the proposed ***PMSVM*** is improved by 13.57% and all the detected Malcode programs are blocked using ***PMR*** with 100% containment rate. The time taken to contain the detected programs is 200ms. In contribution two, the proposed ***DDF*** method achieved better detection accuracy with improved 0.23% and proposed ***DFF*** method achieved containment rate with 100% and the time consumed to block is 1300ms. In contribution three, the proposed ***CPC*** method for detection of illegal traffic achieved detection accuracy improved by 14.47%, and proposed ***ECB*** method provides containment with 100% of blocking all detected malicious traffic within 20 ms. In contribution four, the proposed ***kELM*** method achieved detection accuracy improved by 23.67%. Finally, the proposed ***kEA*** method blocks all the detected malicious IP addresses with 100% containment at the time span of 33ms. The four contributions based on combined Machine Learning Methods provide better detection and containment of newly appearing Internet worms entering the networks.

The proposed research methodology can be applied for other characteristics of Internet worms like hitlist, topological and web search target finding worms with polymorphic and metamorphic payload schemes. The proposed methodology can be integrated with the hardware devices at the Network Intrusion Detection System to handle real attacks affecting the network.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AL	Active Learning
ANN	Artificial Neural Network
ADS	Automated Defense System
AWC	Automated Worm Containment
BDDC	Basic Delayed Dictionary Compression
BFWC	Bloom Filters With Counters
CL	Connected List
CPC	C 4.5 with Pearson's correlation Coefficient
DDF	Delayed Dictionary Compression and Fuzzy Logic Classifier
DFF	Deterministic Finite Automata with Fuzzy Logic Classifier and Filter-Ary Sketch
DFA	Deterministic Finite Automaton
DDC	Delayed Dictionary Compression
DDF	Delayed Dictionary compression and Fuzzy Logic Classifier
DLAL	Detection and Location Algorithm against the Local-worm
DDOS	Distributed Denial of Service
ECB	Enhanced C 4.5 Algorithm and Blacklist
FAS	Filter-Ary Sketch
GFGS	General Frequent-common Gram Searching
IOT	Internet Of Things
IM	Instant Messaging
IFSEng	Iterative Feature Selection Engine
JSD	Jensen Shannon Divergence
kEA	kernelized Extreme Learning Machine with Automated Worm Containment Algorithm
kELM	kernelized Extreme Learning Machine
LPS	Local Preference Scanning
MDM	Mahalanobis Distance Map
MD	Mahalanobis Distance

McPAD	Multiple classifier Payload-based Anomaly Detector
MSE	Mean Square Error
MSVM	Multiclass Support Vector Machine
NFA	Non-deterministic Finite Automaton
OS	Operating System
OADS	Orchestration-Oriented Anomaly Detection System
PCC	Pearson Correlation Co-efficient
PL	Pending List
PMSVM	Principal Component Analysis with Multiclass Support Vector Machine
PCA	Principal Component Analysis
PMR	Principal Component Analysis with Multiclass Support Vector Machine and Rabin Footprint Algorithm
PMSVM	Principal component analysis with Multiclass Support Vector Machine
PSD	Power Spectral Density
RFA	Rabin Footprint Algorithm
RBF	Radical Basis Function
REP Tree	Reduced Error Pruning Tree
RePIDS	Real-time Payload-based Intrusion Detection System
RST	Rough Set Theory
RSWD	Rough Set Worm Detection
SLFN	Single Hidden Layer Feed Forward Network
SVM	Support Vector Machine
SFM	Spectral Flatness Measure
SEIS-V	Susceptible – Exposed – Infectious – Susceptible with Vaccination
SIDQV	Susceptible, Infected, Delayed, Quarantined, Vaccinated
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
UNIDS	Unsupervised Network Intrusion Detection System
VEISV	Vulnerable – Exposed – Infectious – Secured – Vulnerable
WAW	Worm-Anti-Worm
ZASMIN	Zeroday-Attack Signature Management Infrastructure