CHARACTERISTICS BASED DETECTION OF INTERNET WORMS USING COMBINED MACHINE LEARNING METHODS AND WORM CONTAINMENT

CHAPTER 3

PROPOSED METHODOLOGY

This chapter discusses the proposed research design to handle the detection and containment of Internet Worms based on their characteristics using combined Machine Learning Methods.

3.1. Steps involved in the Proposed Methodology

The main focuses of the thesis are

i. Improved Detection and Classification Accuracy of Internet Worms
ii. Minimized Memory Utilization and Time Consumption while detecting worms
iii. Enhanced Containment Rate for all detected anomalies

To achieve the above mentioned objectives, a three-step methodology is followed. The three significant steps are: *Analysis and Preprocessing, Detection and Classification of Internet Worms and Containment of Internet Worms*. The methodology proposed is shown in figure.3.1.

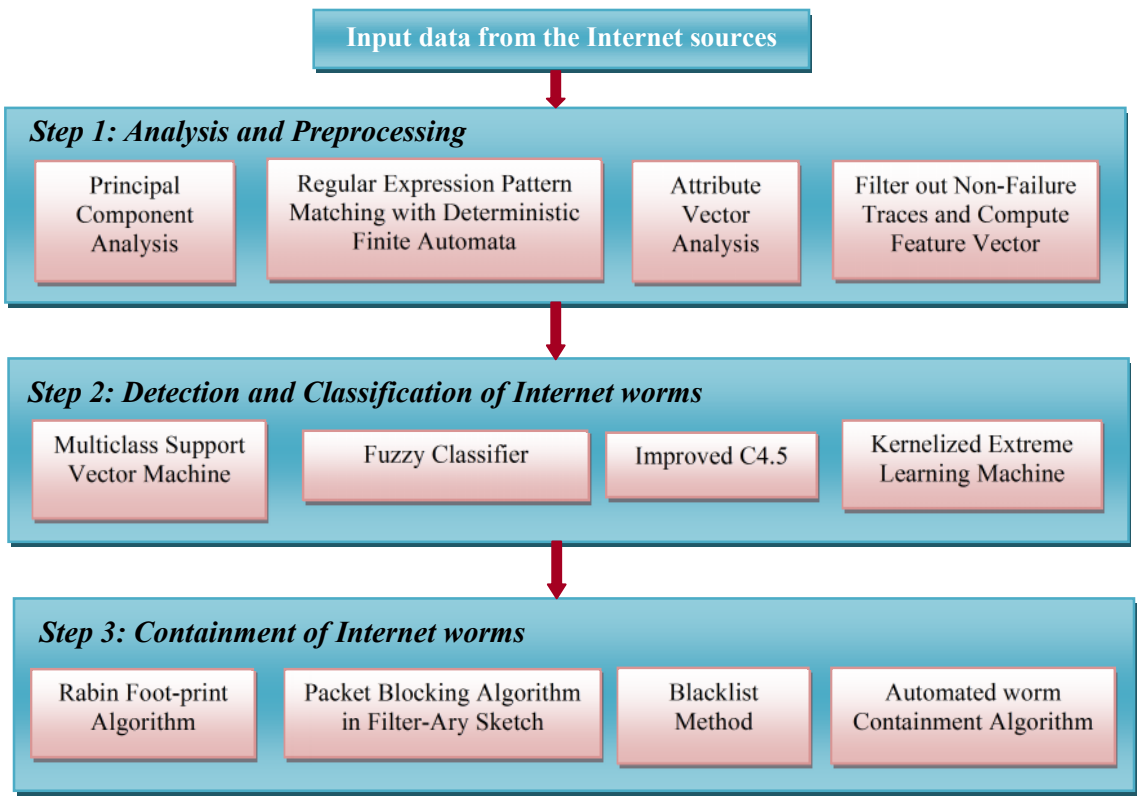

Figure.3.1. The Proposed Three-Step Methodology

Initially the real traces of data are collected from the different Internet web sources. The different techniques applied in the three-step methodology are explained below.

Step 1: Analysis and Preprocessing

The Malcodes or traffic in the dataset is analyzed using the following methods:

i. Principal Component Analysis

ii. Regular Expression Pattern Matching with Deterministic Finite Automata

iii. Attribute Vector Analysis

iv. Filter out Non-Failure Traces and Computes Feature Vector

Step 2: Detection and Classification of Internet Worms

Internet worms are detected and classified using the following methods

i. Multiclass Support Vector Machine

ii. Fuzzy Classifier

iii. Improved C 4.5

iv. Kernelized Extreme Learning Machine

Step 3: Containment of Internet Worms

Containment of Internet worms is done using the following methods:

i. Rabin Foot-print Algorithm

ii. Packet blocking algorithm in Filter-Ary Sketch

iii. Blacklist Method

iv. Automated worm containment algorithm

3.2. Specific Contributions of the Thesis

The entire research work is based on the three-step methodology. There are four research contributions following the three-step process. The significant contributions of the thesis is shown in figure.3.2.

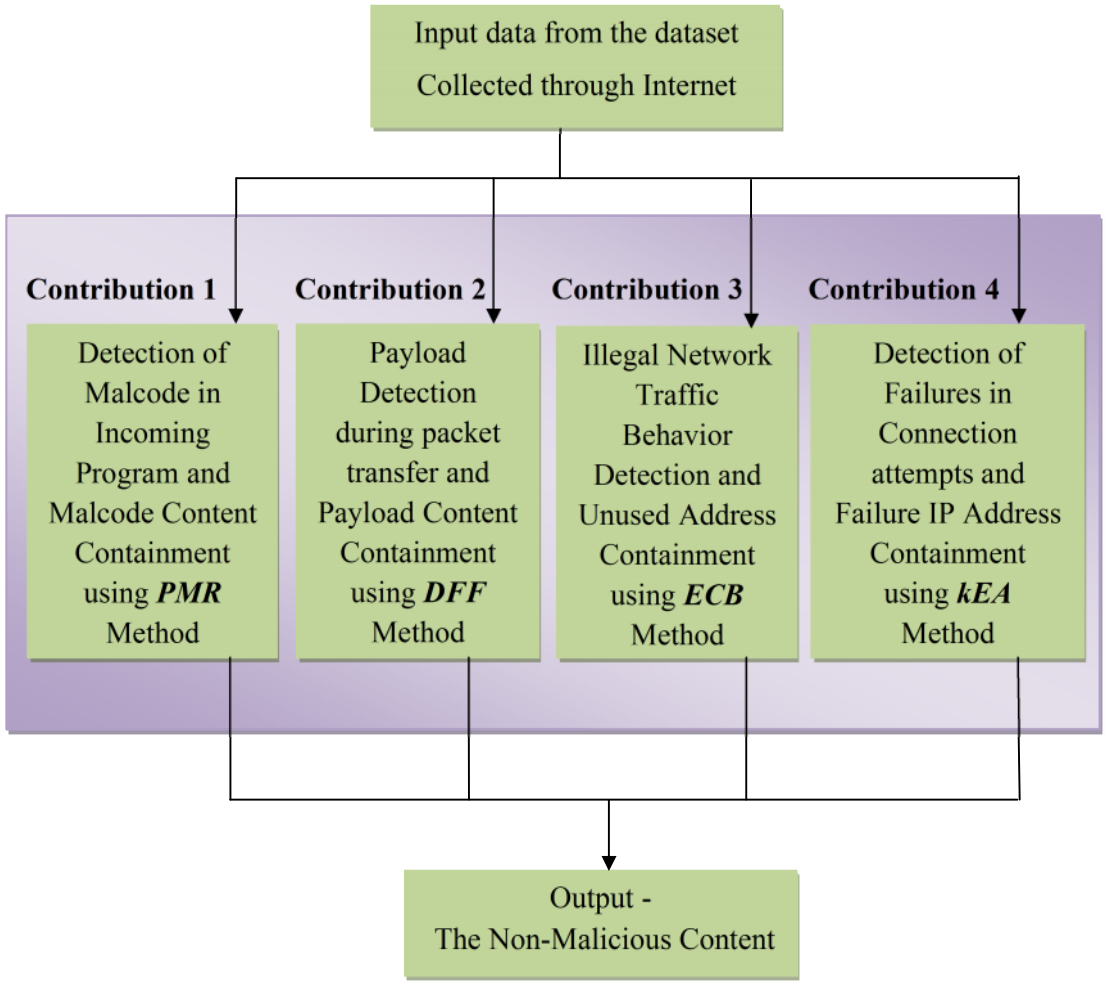

Figure.3.2. Contributions of the Thesis

The proposed work in this thesis is designed to achieve better detection accuracy and containment of detected Internet worms. The proposed combined Machine Learning methods detect and classify the presence of malicious activities in the executable files, payloads, illegal traffic and connection failures. The classified malicious activities are blocked from further propagation using containment methods. The consolidated view of the proposed methodology with the techniques applied and their outcomes is shown in figure.3.3

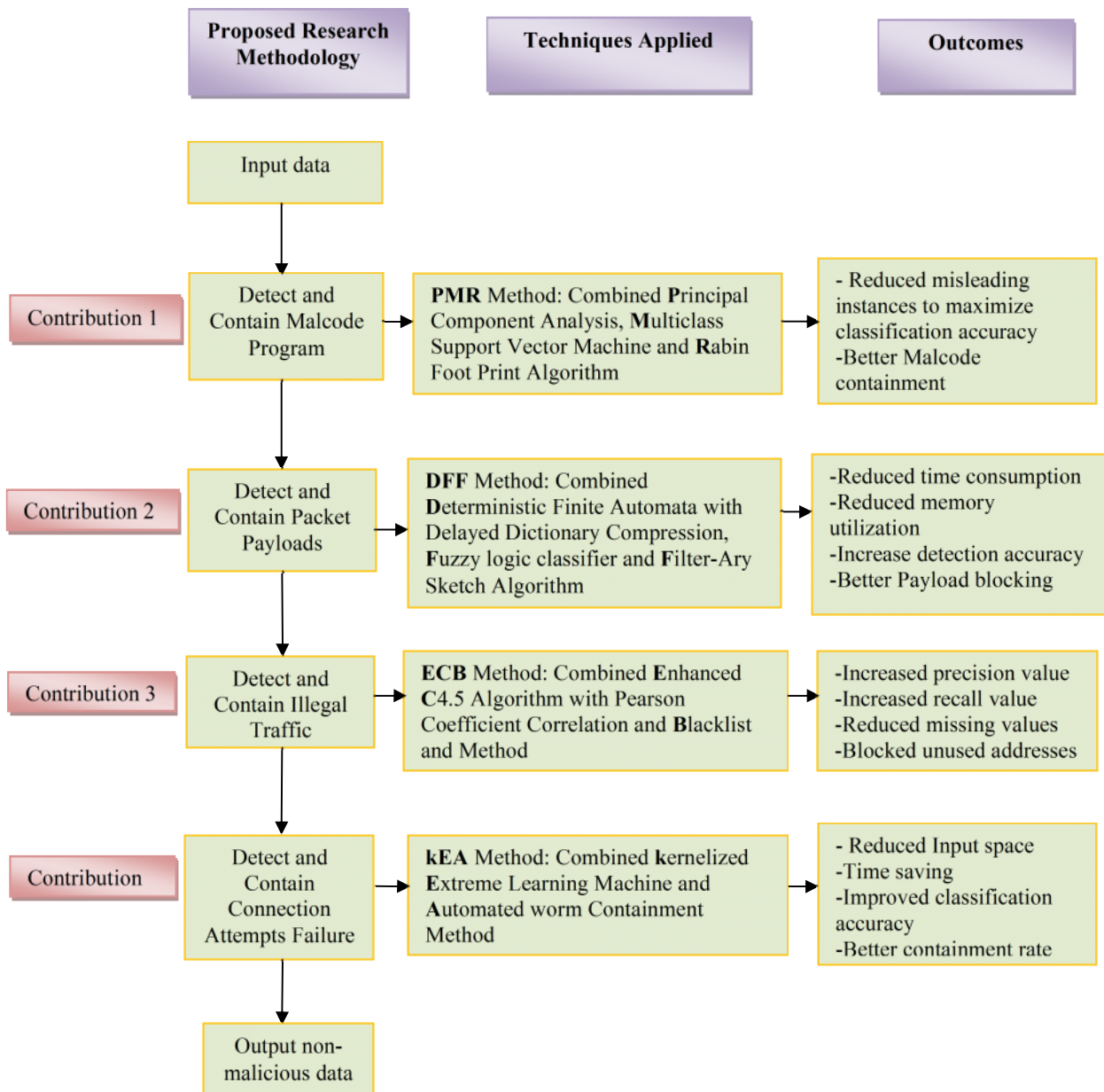

Figure.3.3. Consolidated view of the Proposed Methodology

The expanded view of the four major contributions of the thesis is shown in figure.3.4.

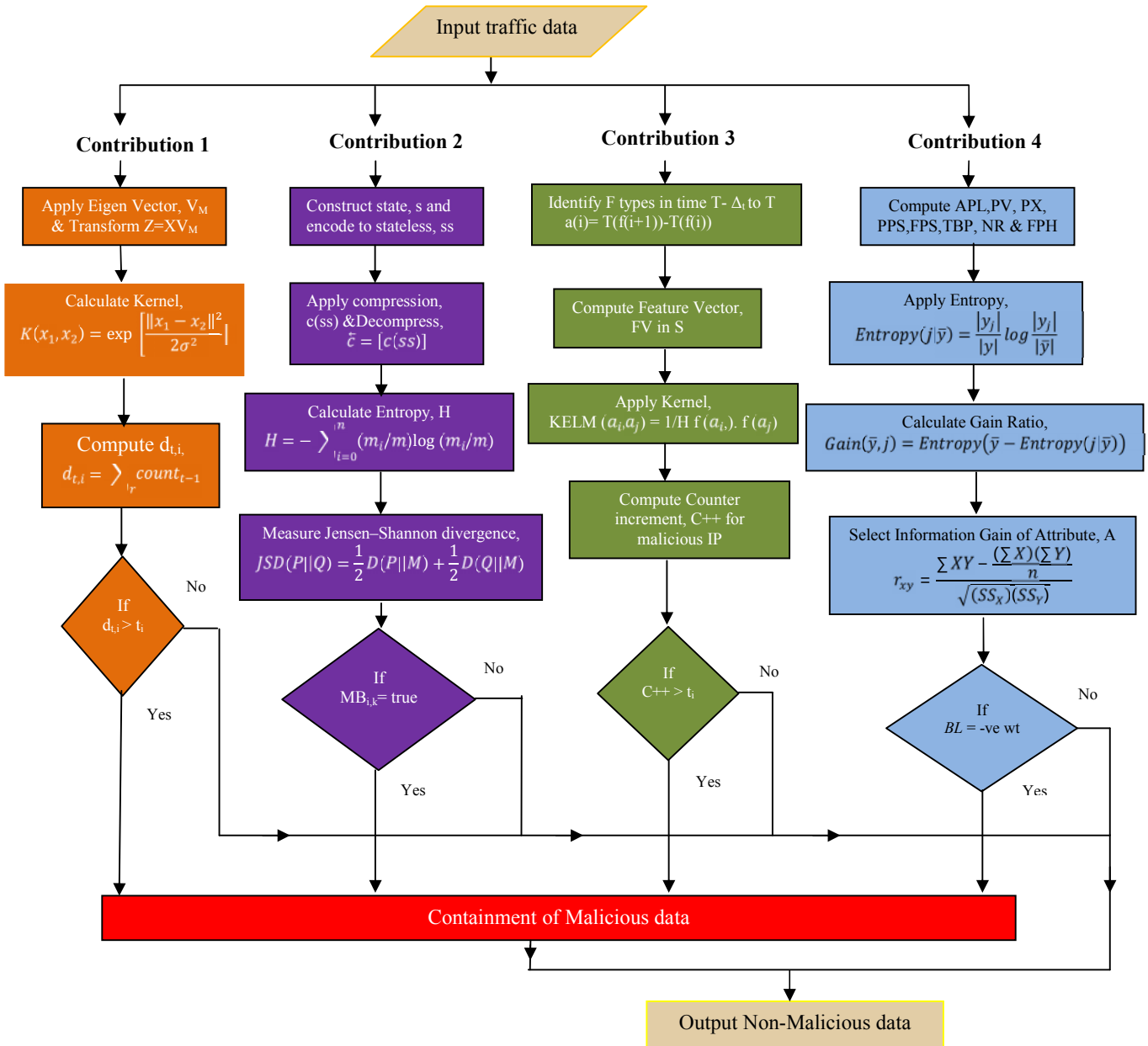

Figure.3.4. Technical details of four research contributions based on the Three-Step Methodology

3.3. Chapter Summary

This chapter briefly discussed about the proposed Research Design. A Three-Step Methodology is proposed to meet the objectives of the thesis. The entire research contribution is discussed in four phases based on the three-step methodology. The three major steps followed are: Analysis and Preprocessing, Detection and Classification and Containment. In the first contribution namely the ***PMR*** Method, the unknown Malcodes propagating through network level emulations are detected and classified using improved Multiclass Support Vector Machine and blocked. In the second contribution namely the ***DFF*** Method, enhanced Deterministic Finite Automata is used to detect the packet payload and the detected payloads are classified and blocked. In the third contribution namely the ***ECB*** Method, unused addresses creating illegal traffic are classified using enhanced C 4.5 and blocked. In the fourth contribution namely the ***kEA*** Method, improved Extreme Learning Machine detects and classifies the failures in connection attempts and blocks from further infection.

All the four contributions are explained in detail in the forth coming chapters.