CHAPTER 1 INTRODUCTION

1.1 Motivation

Internet has rapidly become one of the main communication methods in our society. More and more types of internet applications and usage are available. The more usage of network applications, the more security risks the internet users may face.

To prevent unwanted or dangerous threats, we have to detect them first. Therefore, developing an intrusion detection method is a challenging research issue. There are four challenging issues about designing IDS. The first is the high accuracy and low false alarm rates, especially, the false positive rate (which should be less than 1%) and the false negative rate. Second, the IDS should be able to detect new/unknown attacks because new threats evolve every day. In addition, the performance of classification algorithm in the IDS should be good enough for real-time detection, such as computation speed, memory consumption, etc., because there are a lot of data packets over the real network. A bad performance can cause the system clash. Finally, the IDS should provide more information about the attacks in order to prevent the malicious activities such as attack type, target computer, etc.

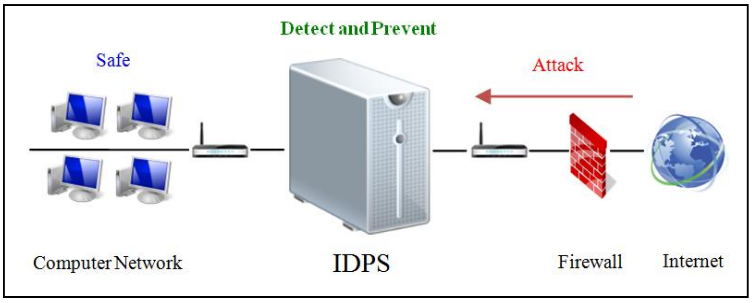

Figure 1.1 Network Environments and Intrusion Detection System

1.2 Objective and Research Scope

In this thesis, we focus on real-time and unknown detection. The algorithm is able to handle the attack and send an alarm message with useful information within three seconds after the packet arrives to the system. There are two output classes from the system including the normal class and the attack class. We are interested in designing an IDS algorithm using fuzzy genetic algorithm. The fuzzy rule is a supervised learning technique and a genetic algorithm which help the system find the best rule from a training dataset. This technique has a high detection rate, a low false alarm rate, fast processing and is robust for unknown data. Therefore, we develop the fuzzy genetic algorithm approach to implement our real-time intrusion detection system where the input network data is captured in the online network, and it will respond to the attack within 2-3 seconds. We evaluate our IDS in terms of the detection speed, CPU consumption, memory consumption, false alarm rate and overall detection rate.

1.3 General Approach

KDD99 dataset is a benchmark dataset that is used in various research studies. However, there were some research groups which generated their own datasets because the KDD99 dataset was now too old. Moreover, it did not contain the present network activities and present attack.

In general, IDS can be classified into two types which are host-based (HIDS) and network-based (NIDS). The HIDS analyzes information that is available on individual computers, such as system calls and log file while the NIDS monitors information in network traffic. The IDS can be further classified into misuse-based and anomaly-based. The misuse-based is a pattern matching. When the packets are matched with the patterns, they will be classified as the attacks. This technique has high accuracy and a low false alarm rate; however, it is not robust for new attacks. The anomaly-based IDS is designed to detect new/unknown attacks. However, it has a low detection rate when comparing with the misuse-based IDS. The general techniques for the unknown detection are as follows:

- Clustering is the algorithm that clusters input data into groups without training data (unsupervised-learning) such as k-means, k-nearest neighbors. However, these techniques have low accuracy.
- Neural network is a group of nodes which are associated with each other. The algorithm will create a neural network structure to recognize the given information. The neural network can work well with noisy data and incomplete data. However, it uses high computation time.
- Fuzzy set is used to create a rule (s). The behavior which agrees with the rule will be considered as an attack. There are many researchers using the fuzzy method because of its robustness and efficiency in detecting unknown data.
- Artificial immune system is a concept of simulating immunology which is inspired by a biological immune system. The intrusion detection system can be considered as an immune system and the attack packet is pathogens. The algorithm only creates a model of normal behavior. When the matching behavior is not found, it will be labeled as an attack.

1.4 Research Contribution

In summary, we make the following contribution:

1. We develop a real-time ID that detects both known and unknown attacks.

2. We improve performance of unknown detection with our proposed approach and compare the results with those from the existing methods.

3. We demonstrate the real time in a real-time network environment.

1.5 Report Overview

This research proposal is organized as follows: in chapter two, there are background study and literature review. Then, chapter three describes the detection approach. The experimental designs and results will be presented in chapter four, and chapter five is the conclusion.