CHAPTER 7

CONCLUSION

This thesis has made contributions to two key research areas namely Intrusion Detection and ML. The contributions apply specifically to the application of ML to intrusion detection. Several factors were found to affect significantly the results and further investigation demonstrated that this is indeed a critical challenge to in designing an IDS. SLFN was capable of detecting a particular class of intrusion. CP algorithm was proposed to optimize the detection of outliers. This approach was found to be successful and able to detect the class of intrusion. Since single objective optimization is performed, there was no control on the classification. To address this limitation, RVM was proposed and class imbalance has been identified as a significant challenge for intrusion detection.

The overall performance and scope of detection of the IDS directly depends on the feature selection stage. The main focus of this thesis is on mining the most useful network features for attack detection. In order to do this a network feature classification schema is proposed and a deterministic feature evaluation procedure that helps to identify the most useful features that can be extracted from network packets. The difference however is in the time of collection, size of the network, throughput, and also the type of users that the networks have.

The proposed method uses mathematical, statistical and RVM techniques to rank the participation of individual features into the detection process. The presented experimental results empirically confirm that the proposed model can successfully be applied to mine new features in the detection process.
An ideal data for IDS should be labeled at packet level and there must be a considerable number of attacks. The current work does not differentiate the final results based on the speed of the attacks. We believe that an interesting further study would be to analyze the set of features that are appropriate for fast or slow attacks. However, to do that, the dataset that will be used needs to have an equal number of attacks in each of the attack categories for study. Multiple datasets can be considered and data sets need to be extracted from a set of diverse networks.

An extensive review of Artificial Intelligence (AI) applied to intrusion detection was conducted and the findings are reported in the literature survey. Various research studies have adopted the ML techniques and evaluated them on the KDD Cup '99 data set. The results thus obtained are and also contradictory. This made to investigate the causes of the discrepancies and it was found that a critical challenge to intrusion detection is the collection of data set and to detect a particular class of intrusion in real time.

SLFN was proposed to optimize the weights and selection of layers in MLP to better learn and was found to be successful. The system was able to detect the previously unknown class of intrusion. The data set selected posed several challenges during the selection of ML algorithms such as:

1. Working with high dimensional data and large memory requirements.
2. Learning speed that gets affected because of very large data set.
3. Feature selection from the large data set.
4. Implementation learning that is incremental / continuous.
5. Detecting new unknown intrusions.

As explained before CP and OD methods with the use of RVM was carried out without too much loss of accuracy. The system has proved good with respect to architecture, data processing, alert aggregation, and reporting mechanisms. However, the methods proposed here have addressed a critical challenge of learning from large data set and providing the user with a set of
solutions. User can select and can then incorporate in an IDS framework as a detection module. Furthermore, there is always an improvement required from the proposed methods, concerning scalability and performance that are discussed further in Section 7.2.

CP and OD combination was a successful method for improving on the performance of IDS. The results obtained in this thesis support the observations that combination of different methods can improve the performance of IDS. Current approaches of creating hybrid are prone to succeed because they may yield a solution with a good classification trade-off. It has been demonstrated in this thesis that FPR were comparatively and outperformed other methods.

7.1 MAJOR CONTRIBUTIONS AND NOVELTY

To the field of intrusion detection domain the main contribution of this thesis is the suitability and application of RVM in building robust and efficient IDS. A novel framework is developed that addressed three critical issues which affect the performance of anomaly and hybrid IDS in high speed networks.

The following three issues are addressed:

1. Attack detection coverage
2. Generating less false alarms
3. Efficiency in operation.

As a result of this research, a framework is built to develop efficient IDS. The framework offers customization and ease of detecting different variety of attacks. The system can identify the type of attack and specific intrusion response mechanism can be initiated by the user so that the impact of the attack is minimized.
CP and OD are efficient methodologies available for building robust and
efficient IDS. Integrating the framework with these two technologies can be used to
build effective IDS. Using CP and OD as intrusion detectors resulted in very few
false alarms and the attacks can be detected with very high accuracy.

The logging framework developed using JPCAP, WINPCAP and JAVA
can capture network data that are significant to detect attacks. The framework can be
used for a variety of applications that requires IDS as plug-in.

Network session needs to be in order to detect attacks with high accuracy
and Feature Extractor can be effectively used to model the events and select required
features. Using CPOD attacks can be detected with smaller window size and good
selection of threshold. A range of experiments are performed and in order to detect
intrusions effectively, it is critical to model the correlations between multiple
features. Since feature sets are independent it makes the model complex and
inefficient as it affects the attack detection capability. The framework developed can
easily define and specific features are extracted, which enables for building effective
intrusion detectors. Our framework is customizable and can be used to build
efficient network IDS which can detect a wide variety of attacks. Experimental
results and comparison with other well known methods for intrusion detection such
as Decision Trees, Naive Bayes and Support Vector Machines has proved better in
terms of accuracy and detection rate without affecting the overall system
performance.

The notable part of our research work is the improvement in attack
detection accuracy. Statistical tests using CPOD demonstrates a higher assurance in
detection accuracy. As the system developed is not based on signatures of attacks it
is capable of detecting novel attacks. Experimental result confirms that our system,
based on CPOD, RVM methods can detect attacks at an early stage by analyzing
only a small number of data set resulting in an efficient system which can block
attacks in real-time.
7.2 FUTURE WORK

The task of detecting intrusions in networks is very critical and leaves no margin for errors. Developing successful attack detection is to identify the best possible approach which is an extremely difficult task. To develop a single solution that can work for every network and application is a real challenge. In my research work a novel framework is developed using different methods which perform better than previously known approaches. In order to improve the overall performance the domain knowledge is used for selecting better features for training which is justified because of the critical nature of the task of intrusion detection. An interesting direction for future research is to develop completely automatic IDS. Another area of work is to develop a faster implementation by employing our approach on multi core processors.

IPS/IRS which aim at preventing attacks rather than simply detecting them are another area that can be explored. This can be achieved by integrating IDS with the known security policy of individual networks. This would help by minimizing the false alarms raised by the IDS.

But one of our objectives in this work was to detect and classify network attacks. Future research in this area is definitely needed and other DM methods can be incorporated. Studies could also be conducted with more attack types which are totally new, or variations on existing types. In this vein, other studies could address the problem of classifying rarely seen attack types like the U2R and R2L. Future work will research on the possibility of expanding the output of the individual classifiers so that it would be easier to identify the exact source of a given attack. By adding a prediction layer it is possible to reduce the complexity of carefully tuning the thresholds and window sizes. Idea is to develop a layer that predicts next data point with some probability. The layer would also learn to readjust probabilities from current deviations and enhance accuracy. Although this study makes a contribution to the IDS classification, there are other DM methods, such as memory-based systems, logistic regression, and discriminant analysis that can be further explored.
7.3 SIGNIFICANT CHALLENGES AND OPEN ISSUES

1. It is really very hard to trace the true source of attack. Hence if a reliable method is developed that can trace back the packets to their actual source then it is possible to prevent many of the attacks. Though some solutions are available, a global effort is required which is a real challenge ahead. It is not only to identify the true source but the overall performance of the system should not be affected.

2. New methods based on user profiling can be developed which will learn the normal user activity and the same can be used to detect deviations if any from the model that is learnt. Most of the works related to this are based on thresholds. Hence a detailed empirical analysis can be performed and develop IDS.

3. In the current Internet era to keep pace with the rapid and ever changing networks and applications is still a major task. The research in developing IDS must synchronize with the present network that supports wireless technologies, ADHOC networks and mobile devices. IDS must be developed in such a way that they can integrate with such networks and devices. They should also provide support for advances in a comprehensible manner.