Chapter 9

Conclusions and Future Directions

This chapter summarizes the thesis with discussion of (a) the findings and the contributions to the state-of-the-art in the disciplines covered by this work, and (b) future work, those directions that our research will undertake addressing open issues that deserve further attention.

9.1 Summary of Research Contributions

With the explosive growth of the Internet during last two decades, Internet-based attacks on large scale enterprise networks have grown rapidly. It is important to keep secure enterprise networks from these threats. The main motivation of the research conducted during this thesis is to monitor and analyze network traffic for anomaly detection.

This thesis has mainly focused on applying data mining techniques in network traffic monitoring and analysis to address the problem of efficient anomaly detection and has evaluated their performance using real world benchmark network intrusion datasets. Summarized, the most important contributions of this thesis are the following:

- In Chapter 2, we present the anomalies in networks, taxonomy of network based attacks, anomaly detection, and evaluation criteria. This chapter also discusses definitions, causes, sources, types of anomalies in networks or hosts and detection approaches for anomaly detection with generic architecture for each of them.
9.1. Summary of Research Contributions

- A structured and comprehensive review on network traffic anomaly detection, methods, systems and tools has been reported in Chapter 3. It includes detection methods and systems under six different categories, viz., statistical, classification, clustering and outlier-based, soft computing, knowledge-based, and combination learners. We also list common and relevant tools used by both attackers and network defenders during live network traffic capture, visualization and analysis.

- In Chapter 4, we present a systematic approach towards generation of benchmark network intrusion datasets. Three separate datasets are prepared using the TUIDS testbed architecture. They are (i) TUIDS intrusion dataset, (ii) TUIDS coordinated scan dataset, and (iii) TUIDS DDoS dataset. Out of several real world attacks, we have chosen and incorporated 28 attacks in preparing our datasets. We have been able to provide a path and a template that ultimately leads to generate a dataset that reflects the appropriate amounts of normality, anomalousness as well as realism. Our datasets demonstrate several features including that (i) they are built by incorporation of latest network scenarios with real network traffic, (ii) We extraction maximum amounts of packet and flow level features during generation of final datasets, and (iii) They are large in size to support effective validation of the performance of detection method.

- In Chapter 5, we initially examine the state-of-the-art of modern port scans and detection methods. Next we introduce an adaptive outlier-based method for coordinated scan detection. In contrast to exploring features for clustering and visualization, AOCD uses random sample selection using a linear congruential generator for distinct profile generation. We also propose an outlier score function to test each candidate object to identify coordinated port scans using score values. The method reports each candidate object as normal or coordinated port scan with respect to a user defined threshold. It is capable of detecting coordinated scans that have a stealthy and horizontal or strobe footprint across a contiguous network address space. We test this method using several real world datasets including the TUIDS coordinated
scan dataset and KDDcup99 probe dataset. Due to non availability of similar datasets from other sources, we use the KDDcup99 probe dataset to evaluate the method. Coordinated scans are performed in an isolated environment, combining network traffic traces with those collected from live networks. This method achieves high detection accuracy and low false positive rate on various real life datasets in comparison to existing coordinated scan detection methods.

- In Chapter 6, we introduce an effective outlier detection technique to identify anomalies in high dimensional network traffic datasets. It introduces a tree-based subspace clustering technique to cluster large high dimensional datasets. The clustering technique arranges the data in depth-first manner before applying our algorithm for network anomaly detection. It also uses the outlier score function to rank each candidate data object based on scores. The key features of this technique are (i) It is able to successfully detect all outlier cases that we consider and (ii) It can use any proximity measure for the computation of anomaly score. But choosing the threshold value is a difficult task during network traffic anomaly detection. We choose this threshold based on a heuristic approach. The performance of the proposed technique is assessed using several datasets, viz., (i) synthetic, (ii) UCI ML repository datasets, (iii) TUIDS intrusion dataset, (iv) TUIDS coordinated scan dataset, and (v) KDDcup99 and NSL-KDD datasets. Our technique performs well compared to similar algorithms.

- Chapter 7 presents a tree-based subspace clustering technique for unsupervised network anomaly detection in high dimensional large datasets. It generates the approximate number of clusters without having any prior knowledge of the domain. We analyzed cluster stability for each cluster by using an ensemble of cluster indices. We also introduce a multi-objective cluster labelling technique to label each stable cluster as normal or anomalous. The major attractions of this method are: (i) TreeCLUSS does not require the number of clusters apriori, (ii) It is free from the restriction of using a specific proximity measure, (iii) CLUSSLab is a multi-objective cluster labelling technique
containing an effective unsupervised feature clustering technique to identify a
dominant feature subset for each cluster, and (iv) TreeCLUSS exhibits a high
detection rate and a low false positive rate, especially in case of probe, U2R,
and R2L attacks. Thus, we are able to establish the proposed method to be
superior compared to competing network anomaly detection techniques. We
also demonstrate that the results produced by our method are statistically
significant.

- Finally, in Chapter 8, we discuss an overview of DDoS attacks, generic archi-
tectures, detection schemes and tools. We use information entropy metrics
for DDoS flooding attack detection. We propose an extended entropy metric-
based victim-end scheme for detecting classes of DDoS flooding attacks by
measuring the difference of the metric between legitimate traffic and attack
traffic. The method exploits a generalized entropy metric with packet inten-
sity computation over sampled traffic within a time interval. We also extend
the mechanism to an ensemble of extended entropy metrics for increasing
detection rate in near real-time. The proposed scheme is evaluated using sev-
eral real world DDoS datasets and it outperforms competing schemes when
detecting classes of DDoS flooding attacks, viz., constant rate, increasing rate,
pulsing rate and subgroup attack.

\section*{9.2 Future Work}

Despite being well-investigated fields, the topics covered in this thesis are far from
being dead-ends. This final section is devoted to discuss possible continuations for
the research carried out in this thesis, some being part of our ongoing work.

- Even though several network intrusion datasets are available for the research
community, they lack comprehensiveness and completeness, and are not avail-
able in the public domain. Therefore, we provide a template toward generation
and preparation of benchmark network intrusion datasets. It is possible to
further extend the work by incorporating both low rate and high rate attacks
for all categories of datasets.
Chapter 9. Conclusions and Future Directions

- Coordinated scan represent a community effort to reduce network bandwidth when attempting to quickly gain the vulnerability information, helpful to attackers. We introduce an adaptive outlier-based technique to detect coordinated scans. But its observations of scan activities are limited to the network layer. So, it needs to be extended further to detect address resolution protocol (ARP) based coordinated scanning, low-rate coordinated scans, and high rate coordinated scans within stipulated time periods.

- An outlier-based network anomaly detection technique can play an important role in identifying types of attacks. We develop a distance-based outlier detection technique and apply it to anomaly detection. There is ample scope still to develop a parameter free hybrid outlier detection technique for mixed type data to efficiently detect a larger number of attacks that combine both distance and density features.

- If a method can detect network traffic anomalies without using any domain knowledge, it is known as an unsupervised method. Such methods always generate large amounts of false alarms because they do not use appropriately labeled data for training. We introduce a completely unsupervised network anomaly detection method. Developing a real-time unsupervised network anomaly detection method for mixed-type data remains a challenging task.

- Though we develop an information metric-based scheme to detect DDoS flooding attacks, there are several open challenges to achieve real-time performance. Hence, we are planning to apply our outlier-based technique to detect DDoS flooding attacks and also aim to extend our scheme with an effective IP traceback mechanism. Finally, we note that the development of low-rate DDoS attack detection with an appropriate IP traceback technique is another open problem.