Abstract

As users, we depend on the Internet in our daily life for simple tasks such as checking emails, but also for managing private and financial information. However, entrusting such information to the Internet also means that the network has become an alluring place for hackers. To this threat, the research community has answered with an increased interest in intrusion detection. With the number of attacks almost exponentially increasing, and the attackers' motivations moving from ideological to economical, the researchers' attention is focused on developing new techniques to timely detect intruders and prevent damage. Our studies in the field of intrusion detection, however, made us realize that additional research is needed.

The contribution of this thesis is that it develops an hybrid approach to intrusion detection that focuses on both the packet level and flow level so that the attacker cannot enjoy the leakage of any one. One can conclude that network meta information as provided by NetFlow is sufficient for detecting certain types of attacks. Surely, security incidents which deposit solely in network payload and at the same time look benign on flow level cannot be detected by flow-based intrusion detection systems. So we develop our approach by focusing on network flows and packets at the same time. Flows offer an aggregated view of network traffic, by reporting on the amount of packets and bytes exchanged over the network, so it is important to consider in case of high speed networks in packet loss is very high. And packet based is important in case of intruders who use the payload field of the packets to launch attack as flow cannot gather any payload information.

Hence we aim at detecting anomalies in flow-based time series as well as packet based. And to make this possible we have generated the real life flow level dataset and packet level dataset from the captured traffic information. So one outcome of our research work is a publicly released flow-based labeled data set and a packet based labeled dataset. To the best of our knowledge, no such data set already exists

Secondly we have developed a Similarity Graph Based approach for detecting anomalous traffic, and we have applied this technique in both the packet level and flow level at the same time.

Finally we make a union of all the records for which alarms are generated in packet based and/or the records for which alarms are generated in flow based, so that the attacks that are not detected by packet based can detected in flow based and vice-versa. For these we have hybridized both the techniques into a single one.