Data Streaming Algorithms for Rapid Cyber Attack Detection

Yang Liu

PHD

Major Professor: Yong Guan

Internet has become an essential part of the daily life for billions of users worldwide. However, every day we are still reading news stories about major security breaches, new polymorphic worm/virus spreading, identity theft, DDoS or phishing emails. Attackers are using botnets as an effective tool to conduct a wide range of criminal activities even more efficiently than before. It becomes very challenging to detect/defend cyber-attacks in high-speed networks nowadays. Firstly, the Internet traffic exhibits huge fluctuations and long range dependence, which makes the attack traffic often be hidden by large volumes of normal traffic. In order to get reliable information for the malicious traffic, we must be able to process each packet in a wire speed. Secondly, ISPs want to detect the attacks when they are still at a low-profile volume in order to reduce the damage as much and early as possible. Therefore, the detection algorithm should be able to correlate multiple data sources and identify the root of causes efficiently.

In this dissertation, we propose data streaming algorithms to detect/defend cyber attacks in high-speed networks. We propose a novel data structure to detect click frauds in the online advertising network, and track the command-and-control communications in the botnets. Next, we introduce a fast sketch for the aggregate queries in high-speed network traffic, which can preserve some critical information for root cause analysis, like IP addresses, port numbers, etc. This sketch can also be extended to track super spreaders. Lastly, we apply the low-rank matrix approximation to monitor network-wide traffic anomalies and traffic activity graphs.