Online Detection of Anomalies over Distributed Streams

PI: Srikanta Tirthapura
Research Assistant: Sneha Amam Singh
Company: Northrup Grumman (VT)

The goal of this project is to design a method for online monitoring of patterns of stealthy but “persistent” behavior in the union of multiple distributed streams. Such a method can lead to identification of botnet traffic, and other fraudulent activity in the system.

Recent data (Giroire et al. 2009) has shown that a pattern of traffic known as “persistent” items often implies certain types of botnet-related communication over the network. Tracking such patterns over distributed network streams will lead to timely identification of botnet related traffic. In general, there are many other interesting patterns in traffic whose detection helps identify network threats, but we do not have the tools yet to track these in an online manner, especially over distributed streams.

Our work aims to detect persistent traffic patterns in an online fashion as opposed to prior work which could detect these patterns offline, after the fact. To develop algorithms and software for tracking persistent elements in distributed data streams. The CPU, memory, and especially the communication overhead of such monitoring will be measured carefully, and techniques will be devised to keep this as small as possible. Such tracking may lead to false alarms (false positives) or missed anomalies (false negatives). The goal is to develop algorithms where we can bound the rate of false positives and false negatives.

1. Formulation of the precise metrics and properties to monitor on distributed streams. To begin, we will consider persistence and entropy.
2. Design of online distributed monitoring algorithms, analysis of memory, CPU, communication overhead.
3. Implementation over a streaming system. We plan to implement on top of IBM Infosphere Streams, with which we have prior experience.
4. Collection of appropriate trace data for evaluation
5. Evaluation of effectiveness of proposed approach
6. Collection of results, writing

This can help detect botnet-related anomalous traffic in an online manner, and also lead to a smooth tradeoff between detection accuracy and resource consumption, in a distributed setting. In such cases, distributed monitoring maybe a communication-intensive activity.