**Title:** Online Detection of Anomalies over Distributed Streams

New Proposal
Researcher Name: Srikantha Tirthapura
University: Iowa State University
Date(s): 01/01/2012 – 6/31/2013

**Long Term Goals.** Develop methods for tracking anomalies over geographically distributed streams of network traffic. Different nodes in a network may observe their own local streams of traffic, and it is often necessary to observe and monitor patterns over the union of multiple streams. Our goal is to devise methods for detecting patterns/anomalies in an *online* manner, and using as little communication as possible.

**Background for Long Term Goals.** 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.

**Intermediate Term Objectives.** 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.

**Schedule of Major Steps.**

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

The expected times for different steps are: 7 months for steps 1 and 2 together, 4 months for steps 3 and 4, and 7 months for steps 5 and 6.

**Dependencies:**
Steps (1) and (2) and (3) have to be performed in sequence. Step (4) can be performed in parallel with Steps (1), (2) and (3), and metrics for evaluation can also be developed in parallel with Steps (1), (2), (3). Steps (5) and (6) will be done after the previous steps have completed.

**Major Risks:**
Monitoring certain anomalies in an online manner may require inherently large amounts of memory/communication in a distributed setting.
**Budget:**
1 graduate student for 18 months, including stipend and tuition support.

**Staffing:**

**Category of Current Stage:**
New proposal

**Contacts with Affiliates:**

**Publications and Research Products:**
Prior work by the PI in a related direction includes the following.


**References:**
Frédéric Giroire, Jaideep Chandrashekar, Nina Taft, Eve M. Schooler, Dina Papagiannaki: Exploiting Temporal Persistence to Detect Covert Botnet Channels. Recent Advances in Intrusion Detection (RAID) 2009: 326-345