Introduction

DataMap

- Modular open-source tool for the collection and real-time analysis of network traffic data
- Configurable sampling, anonymization, filtering, and aggregation of collected data
- Collects data from multiple WiFi locations in a network
- Easily adaptable to new sampling and aggregation techniques
- Can be utilized for real-time monitoring for anomalous behavior using a combination of signature-based and anomaly detection
- Facilitates the study of newer techniques for anomaly detection

Related Work

Snort

- Lightweight sniffer, packet logger and an intrusion detection system
- Generates alerts when it observes specific types of probes or attacks that indicate a potential intrusion attempt
- Detection based on custom or included rule sets that are updated daily
- Runs on a single machine, requiring a complimentary set of tools to gather traffic data from multiple locations
- Not able to detect anomalies whose signature is not yet known

Sguil

- Open source next generation intrusion detection and prevention engine
- Similar to Snort—signature-based detection
- Designed for better performance on multi-core CPUs
- Runs on single machine, without multiple data collection nodes at different locations

Overview of DataMap Components

Collection Nodes

- Uses Versatile Monitoring Toolkit (Vermont) for traffic capture
- Supports configurable sampling and filtering schemes
- Places wireless NIC in monitor mode to capture data from all access points
- Configurable network identification and channel hopping

Anonymizer

- Discards MAC and application layer data
- Anonymizes IP addresses with Crypto-PAn while preserving network topology information
- Crypto-PAN is one-to-one, prefix preserving, and consistent across time and location

Aggregator

- Aggregates raw network data into configurable time-slices
- Based on any combination of packet header fields
- Data is passed via shared memory from the Sampler through the Anonymizer to the Aggregator to keep identifying information from ever being written to disk

DB Writer

- Sends filtered, anonymized, and aggregated data to central server
- Data is labelled with location of collection node before being stored to allow the investigation of spatial correlations in the collected data

Central Server

- At start-up, a collection node sends a Hello message containing its ID and location
- Keeps track of collection node state with periodic Heartbeat messages
- Sends commands to individual nodes to begin/end data collection
- Sends filtered, anonymized, and aggregated data to central server
- Scope: 1 to 3 locations

Acknowledgment

This work was partially funded by the National Science Foundation Award #1228847.