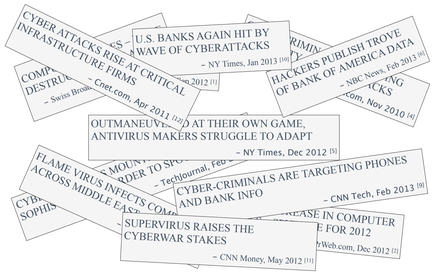
A Modular Multi-Location Anonymized Traffic Monitoring Tool for a WiFi Network

Justin Hummel, Andrew McDonald, Vatsal Shah, Riju Singh, Bradford D. Boyle, Tingshan Huang, Nagarajan Kandasamy, Harish Sethu and Steven Weber

Department of Electrical and Computer Engineering
Drexel University, Philadelphia, Pennsylvania 19104

Introduction



- Increasing sophistication of malware makes early detection difficult
- Signature-based detection rates can be as low as 5% [Perlroth 2012]
- Real- time monitoring for anomalous behavior in the traffic data over a network offers a surer approach to protecting networks
- Traffic data collected at a single location is usually insufficient to infer an anomaly
- Volume of collected traffic data can be prohibitively large

N. Perlroth. Outmaneuvered at Their Own Game, Antivirus Makers Struggle to Adapt. *The New York Times*, December 31, 2012.

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
- Generate 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 complementary set of tools to gather traffic data from multiple locations
- Not able to detect anomalies whose signature is not yet known

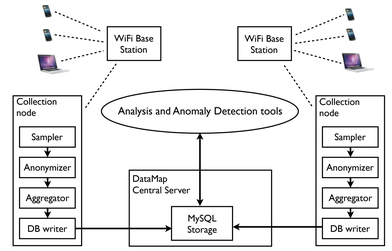
Security Onion

- Linux distro for intrusion detection, network security monitoring, and log management
- Based on Ubuntu 12.04
- Contains Snort, Suricata, Bro, Sguil, Squert, Snorby, ELSA, Xplico, NetworkMiner
- Requires 1GB memory per interface

Suricata

- 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



A DataMap traffic monitoring system showing two collection nodes (also called traffic sensors) and the central server

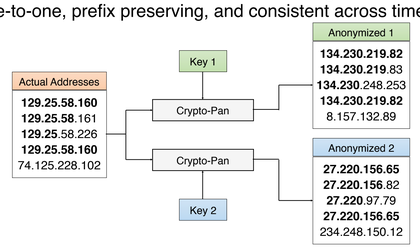
Collection Nodes

Sampler

- 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



J. Fan, J. Xu, M. H. Ammar, and S. B. Moon. Prefix-Preserving IP Address Anonymization: Measurement-based Security Evaluation and a New Cryptography-based Scheme. Computer Networks, 2004.

Aggregator

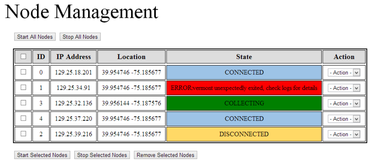
- Aggregates raw network data into configurable width time-slices
- Based on any combination of packet header fields
- Data is passed via shared memory from the ***Sampler*** through the ***Anonymizer*** to the ***Aggregator*** to keep identifying information from ever being written to disk

DB Writer

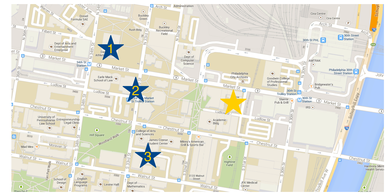
- Sends filtered, anonymized, and aggregated data to central server
- Data is labelled with location of collection node before being stored to allow the investigation of spatial correlations in the collected data

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
- Manages collection node configurations for sampling and aggregation
- Control functionality exposed through included web interface

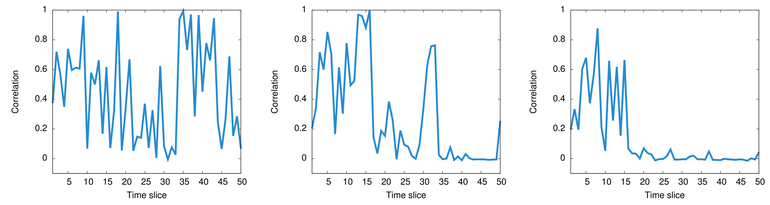


An Example of DataMap Usage

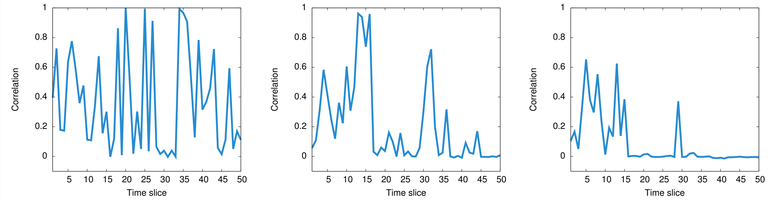


Collection Node Locations
1. **Athletic Center**
2. **University Library**
3. **University Bookstore**

Central Server Location
1. Bossone Research Enterprise Center



Correlation between source address at locations 1 and 2 (left), at locations 1 and 3 (middle), and locations 2 and 3 (right)



Correlation between destination address at locations 1 and 2 (left), at locations 1 and 3 (middle), and locations 2 and 3 (right)

- Data collected over 1800 seconds
- Divided into 50 non-overlapping windows 36 seconds in duration
- Computed the coefficient of correlation between the IP addresses at different locations
- Each data point represents the correlation computed over a 36 second period

- There is a sustained shift in the pattern of correlation between traffic at location 3 and the other two locations beginning somewhere between 540–720 seconds
- May indicate an anomalous event
- May be normal event with legitimate reason for the deviation
- Investigation of the correlations between additional features can be used to determine if the event should be flagged

Typical DataMap Collection Node Hardware

- alix3D3 single board computer from PC Engines
- AMD x86 compatible processor
- 256 MB DDR DRAM
- Atheros AR5414 wireless chipset (supported by ath5k driver)
- Ubuntu 12.04 (requires special kernel for support of non-PAE capable CPU)

`http://pcengines.ch/alix3d3.htm`

Fork on GitHub

https://github.com/DataMap13/DataMap/

Acknowledgment

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