An Interactive Web Based Platform for Modeling and Analysis of Large Scale Argus Network Flow Data

Angel Kodituwakku
J.T. Liso
Dr. Jens Gregor
Jan 10, 2018

This material is based upon work supported by the National Science Foundation under Grant No. IRNC-1450959
InSight2 Foundational Work

- GLORIAD: World wide network for research & education
  Global Ring Network for Advanced Applications Development
  NSF sponsored project 2006-2015, Greg Cole (PI)
- InSight: Visualization of GLORIAD Argus flow-data Development ended with GLORIAD
- InSight2: Newly developed, completely redesigned tool
InSight2 Motivation

- Open-source Argus flow data analytics platform that provides:
  - Performance metrics
  - Threat detection
  - Advanced analytics
  - Web based visualization

- Modular architecture that supports large scale data, real-time processing, and site-specific requirements
InSight2 Features

- Core functionality: Performance metrics
- Plug-in extensions: Advanced analytics
  - Markov chain: Behavior prediction
  - Tensor analysis: Anomaly detection
  - Community plugins: TBD
- Data enrichment: Value-added knowledge
  - Geo-IP, Global Science Registry (IP-org mapping)
  - Threat lists, Blacklists (botnets, ransomware etc.)
InSight2 Capabilities 1/2

- Measurements
  - **Network statistics** (load, packets dropped, retransmitted)
  - **Usage statistics** (countries, organizations, ISPs)
  - **Diagnostics** (jitter, packet size, hops, delay)

- Visualizations
  - Critical **activity gauges**
  - Overlaid **advanced metrics**
  - **Connections graphs** of top users
Capabilities 2/2

• Intuitive filtering by UI interactions
  • Click UI elements to add/remove filters by country, ISP etc
  • Click and drag to filter time range in timeline
  • Click and drag define visual geo-location bounds in geo-maps

• Geo-location mapping: MaxMind Geo-IP database
• Threat detection: Misc. on-line databases
• Utilization prediction: Markov chain modeling
• Anomaly detection: Tensor based data analysis
InSight2 Traffic Overview

- Main Dashboard
- Activity Gauges
- Country Tag Cloud
- Geo Map
- Intuitive filters
- Traffic ratio and PCR
- Setup time and hops
- Packet size
- Jitter and inter-packet arrival time
InSight2 Argus Flow Data

Network Mirror

NetFlow Source

Radium Server

ra* clients pipeline

ONLINE

InSight2

OFFLINE

ra* clients pipeline

Flow data storage (Argus files)

Powered by QoSient argus

Argus is used by

Carnegie Mellon University
Software Engineering Institute
Stanford University
Oak Ridge National Laboratory
NSF
CERT
ArcSight
MySQL
SONY
MaxMind

01/10/2017
InSight2 uses elastic

elasticsearch  kibana

- Highly scalable
- NoSQL database
- Full-text search engine
- Distributed

- Visualization platform
- Intuitive dashboards
- Native integration with ES
- Geo-map tile service
Plug-in: Markov Chain 1/2

- State transition model
- Stochastic: $\text{Prob}(s_{i+1}|s_i)$
- Inferred from training data
- Model analysis
  - Steady-state
  - First-transitions
- Live data processing
• Usage: Network utilization prediction

Actual Usage

Predicted Usage

State Transition Probabilities
Plug-in: Tensor Analysis 1/3

- Tensor: multidimensional matrix of real numbers
- Each mode is $n$-dimensional matrix (called slice)
Plug-in: Tensor Analysis 2/3

- Tensor energy
  - Average sum of squares per slice given mode

- Data sparsification
  - Low energy change data discarded during update

- Event detection
  - High energy change data indicates new trend that may warrant investigation (anomalous behavior?)

S. Papadimitriou et al, Streaming Pattern Discovery in Multiple Time-Series, Proc. VLDB, Trondheim, Norway, 2005
InSight2  Plug-in: Tensor Analysis 3/3

Observed source traffic

Observed destination traffic

Slice Energy

Anticipated Energy

Actual Energy

Energy Ratio
InSight2 Summary

- Argus flow-data modeling and analysis
- Interactive web based platform
- Open-source modular software (release TBD)

- Partners
  - QoSient, Cisco ASIG
  - Stanford University, KISTI (South Korea)

- Work supported by NSF: IRNC-1450959