

# InSight2

**Angel Kodituwakku & Jens Gregor, UTK**

**Alex Keller & John Gerth, Stanford University**

**Buseung Cho, KISTI**

**Carter Bullard, Qosient**

IRNC: AMI: Advanced Measurements and  
Instrumentation



THE UNIVERSITY OF  
TENNESSEE  
KNOXVILLE

# What is InSight2 ?

- Interactive situational awareness and analytics platform for real-time network traffic modeling and analysis.
- Argus flow data enriched with GeoIP, bad actor, and Global Science Registry (GSR) information.
- Multi-threaded, scalable, extendible architecture.
- Simple virtualized deployment.
- Plugin-based analytics modules.

# Who is it for?

- Network managers and operators
  - Make proactive planning decisions
  - Determine optimum times for large data transfers
- Network analysts
  - Live and historic data enrichment
  - Real-time data visualization
  - Anomaly detection
  - Intuitive dashboards for detailed drill-down to flow level

# InSight2 Dashboards

## InSight2

Performance

Security

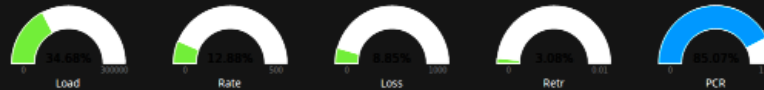
Tutorial

9.3  
Peta Bytes sent

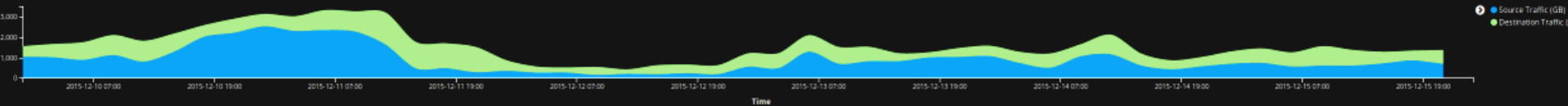
75.3  
Billion packets sent

3.6  
Billion packets lost

10.2  
Trillion retransmissions



This is the performance dashboard. It provides information relevant to traffic, packets transmitted, packet loss and retransmissions, flow duration as well as cumulative metrics regarding unique IP addresses and organizations. First row of visualization indicate the system health. The 5 gauges load, package rate, PCR, packet loss and retransmissions provide the system capacity at a glance. Each gauge is tuned to the capacity of the network end represents the capacity for the last 5 days which is the default time period each dashboard is loaded when it is first loaded. Load gauge indicates Bytes transmitted per second as a percentage of the total capacity of the links. This is particularly useful to understand how bandwidth is used to last and the head room left. Packet rate gauge indicates packets transmitted per second. Higher packet rate combined with low system load indicates anomaly that may correspond to



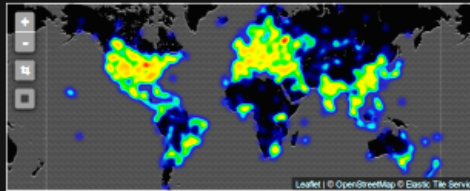
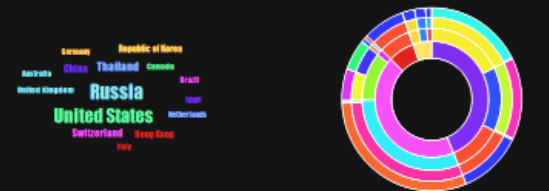
233  
Countries

32,986  
Cities

11,864  
Organizations

3,295,745  
IPs

This row of visualizations provides information aggregated by source country. First visualization visually represents the countries that transmit data as flow originator in a tag cloud format, the size of the label is proportional to the amount of data transmitted. To filter by a country simply click the country name, the segmented pie chart to the right of that incorporates compact representation of the geographical information of the countries shown in the tag cloud. By hovering your mouse over each segment location information from country, province, city zip code can be found. As each ring in the pie chart gives higher granularity at each step. By clicking on each of the rings you can apply one or more filters



Top Organizations	Traffic (GB)
Fermi National Accelerator Laboratory (Fermilab)	2,156,964
Joint Institute for Nuclear Research	758,761
Institute for Theoretical and Experimental Physics	597,753
National Electronics and Computer Technology Center	578,443
Kurchatov Institute	396,948
Vanderbilt University	301,219
Russian Space Science Internet	273,074
Institute for Nuclear Research	257,451

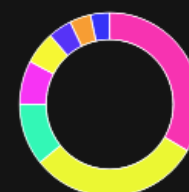
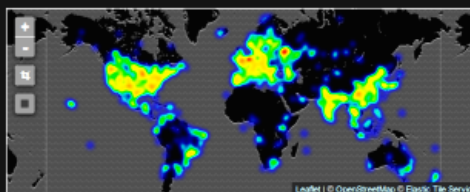
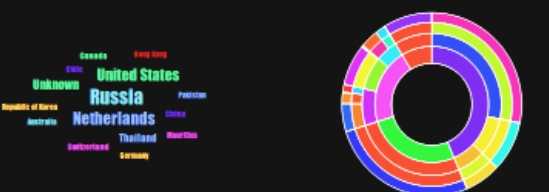
112  
Countries

3,680  
Cities

7,028  
Organizations

1,320,136  
IPs

Following dashboards visualize the information aggregated by destination country. The visualizations are the same except they represent two floor receiving end. The geographical map also supports double click to zoom in. The counters on the left hand side show the unique number of destination countries, cities, organizations and IP addresses. Note that this information is related to the time frame selected either by default or by user by clicking and dragging the time series graph on the top. On the right hand side the scrollable list of organizations is sorted by the amount of the year they have transmitted and at the end of the list are two links to download the raw data or the formatting data format.



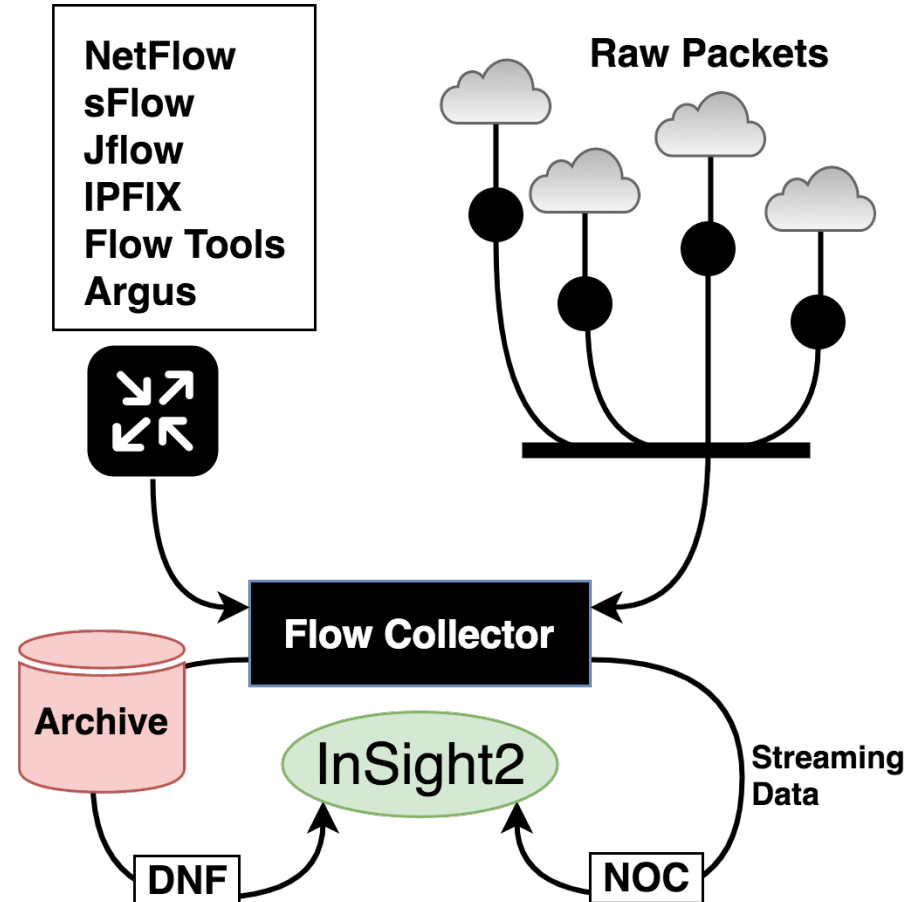
Top Organizations	Traffic (GB)
Jump Management SRL	2,229,269
Joint Institute for Nuclear Research	2,120,748
Unknown	701,674
Institute for Nuclear Research	516,892
Institute for Theoretical and Experimental Physics (ITEP)	400,489
Institute for Theoretical and Experimental Physics	282,190
National Electronics and Computer Technology Center	263,552
The University of Tennessee Health Science Center	236,074

# Capabilities

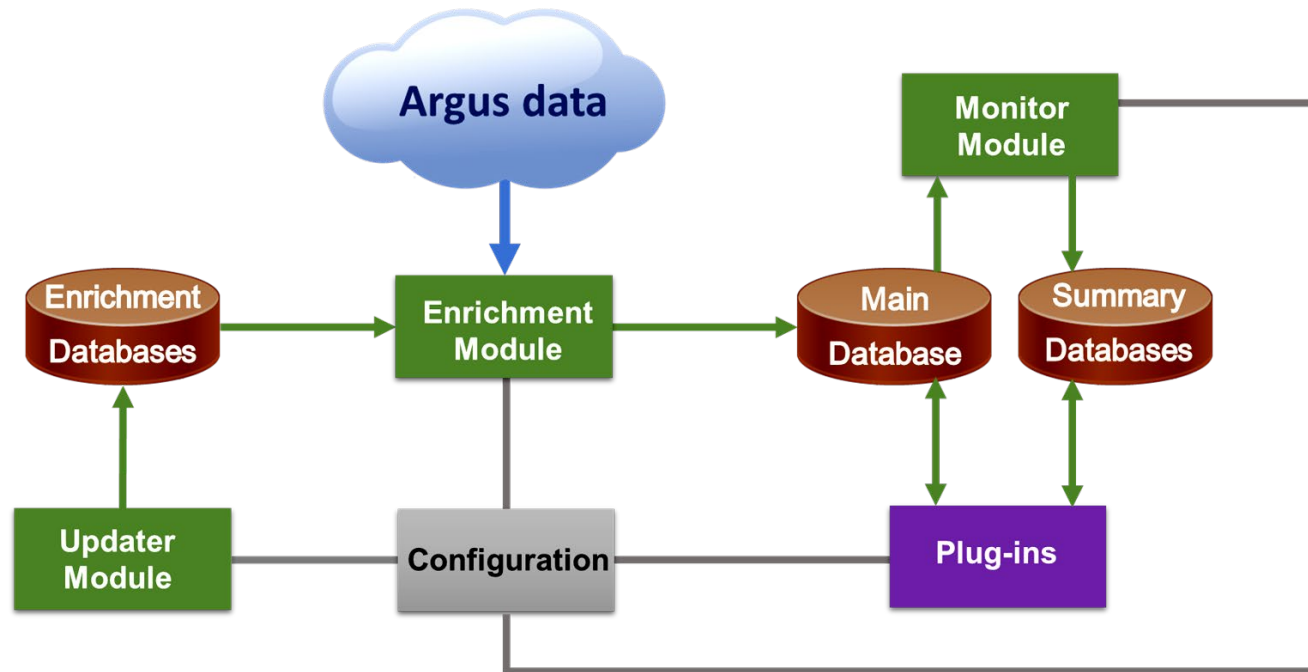
- Flow Data Measurements
  - Network statistics (load, packets dropped, retransmitted)
  - Usage statistics (countries, organizations, ISPs)
  - Diagnostics (jitter, packet size, hops, delay)
- Advanced Analytics
  - Traffic prediction
  - Event detection (automatic reporting)
- Visualizations
  - Critical activity gauges
  - Advanced metrics
  - Connection graphs (top users)

# Flow Data Ingestion

- Multiple flow standards supporting existing infrastructure
- SPAN / mirror port support for direct live data



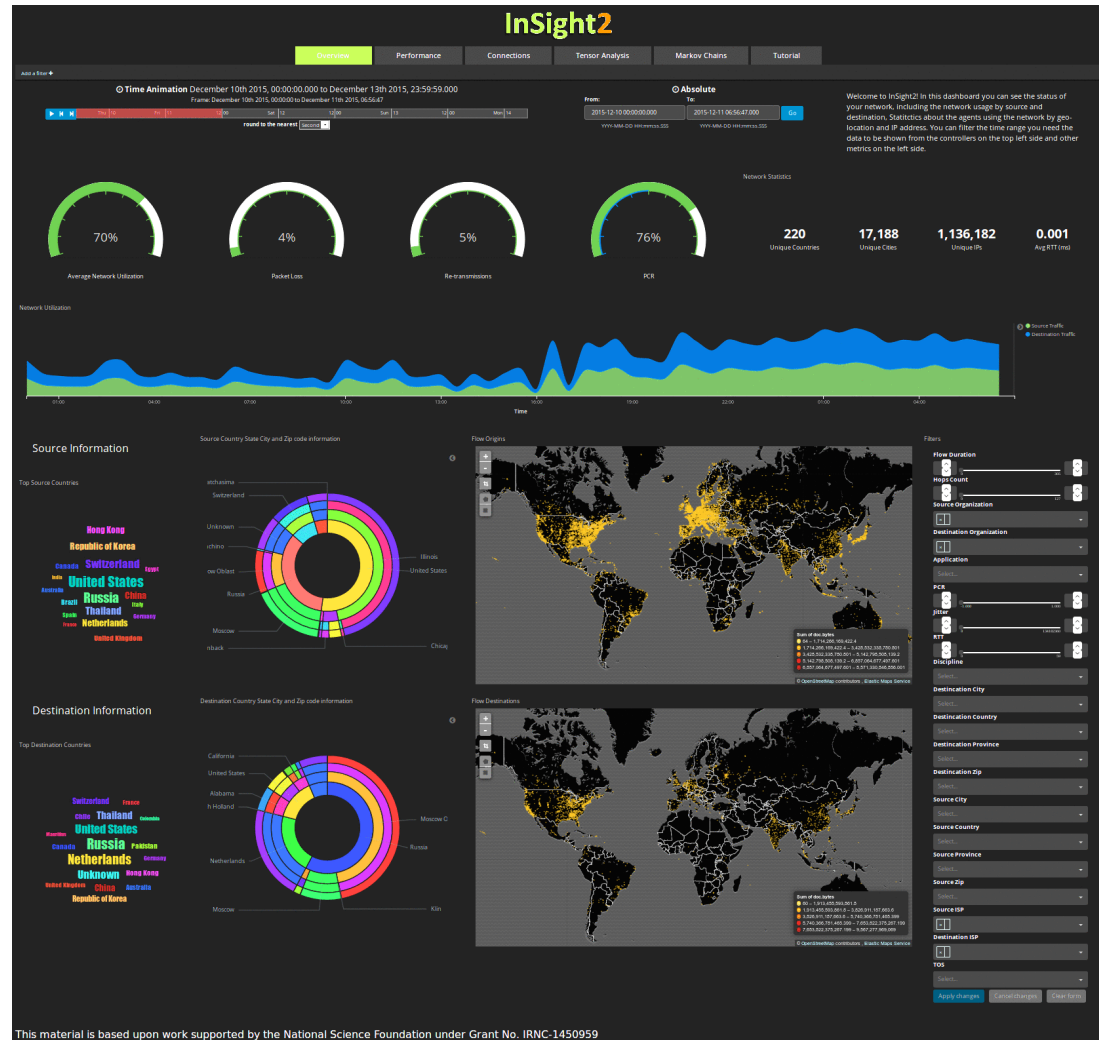
# Software Architecture



- Robust, extensible system architecture
- Supporting modular collaborative development
- Development by academia, deployment by everyone

# Performance Measurements

- Main Dashboard
- Activity Gauges
- Country Tag Cloud
- Geo Map
- Intuitive filters



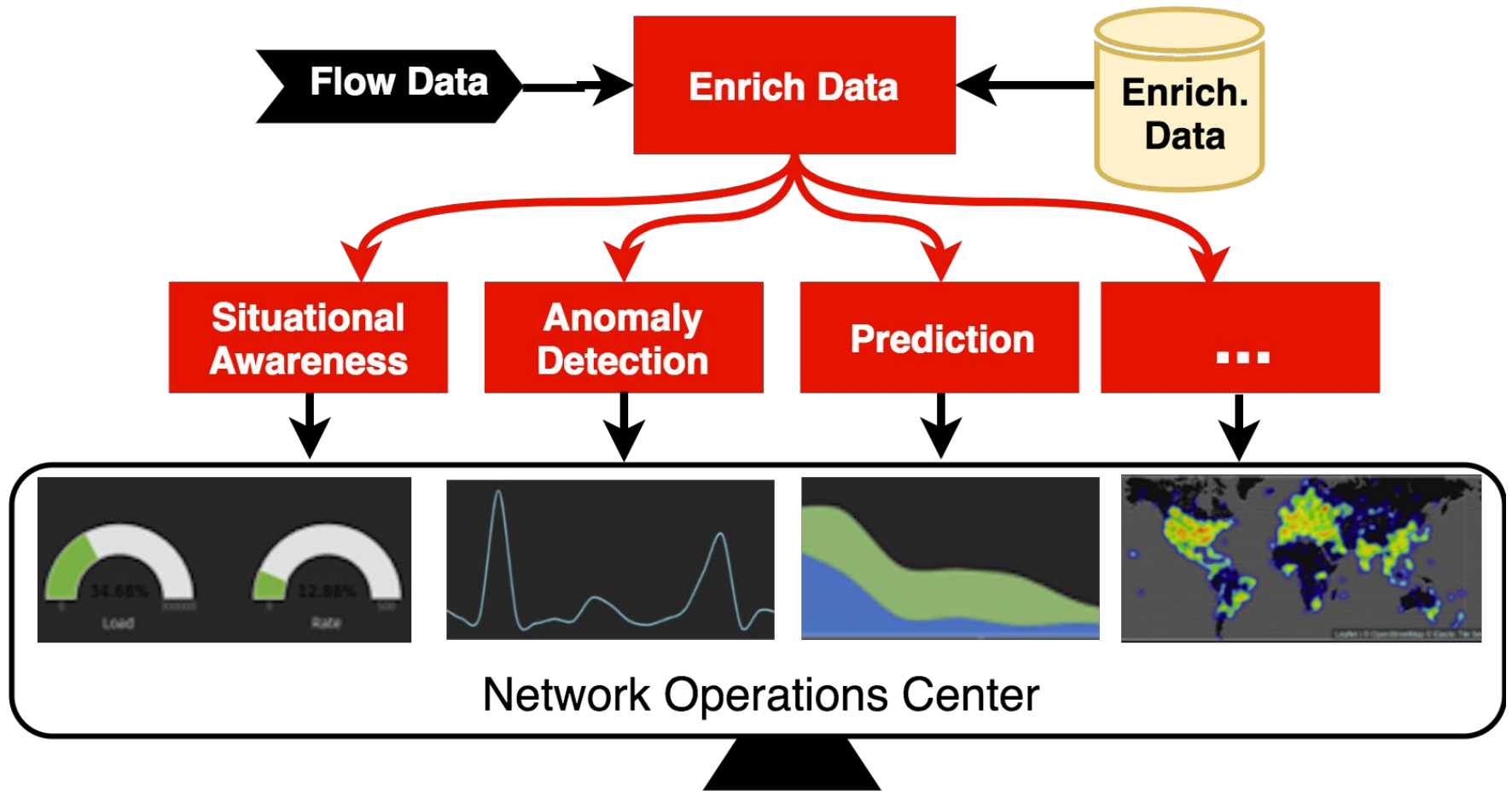


# Performance Metrics

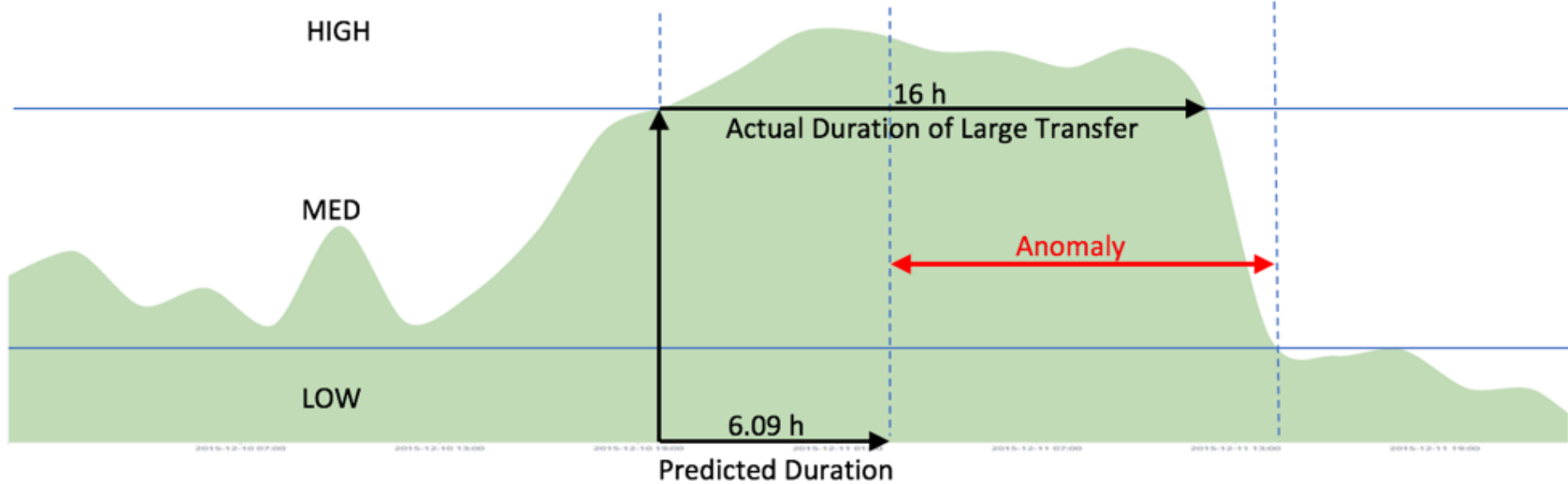
- Traffic ratios
- PCR
- TCP timers
- Path hops
- Packet sizes
- Jitter
- Inter-packet arrival time



# Modular Analytics



# Prediction of Large Data-transfers



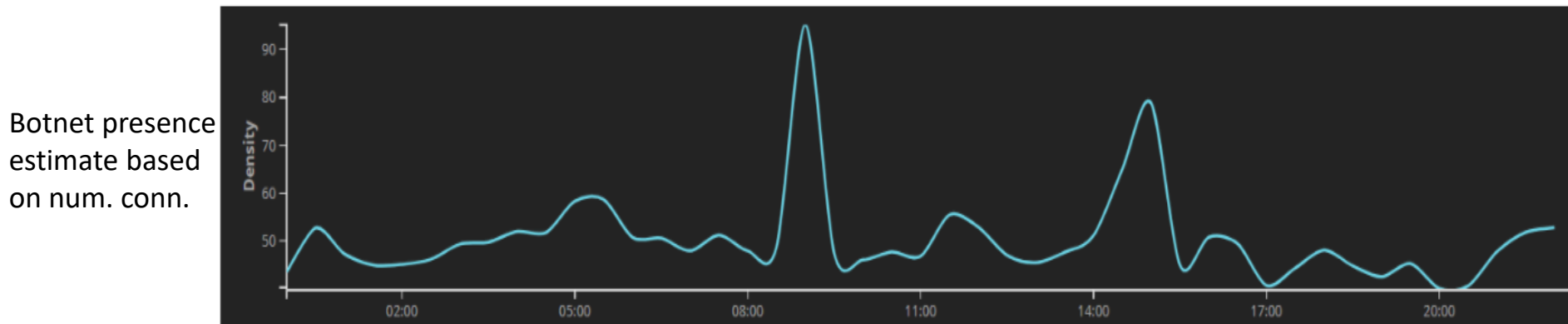
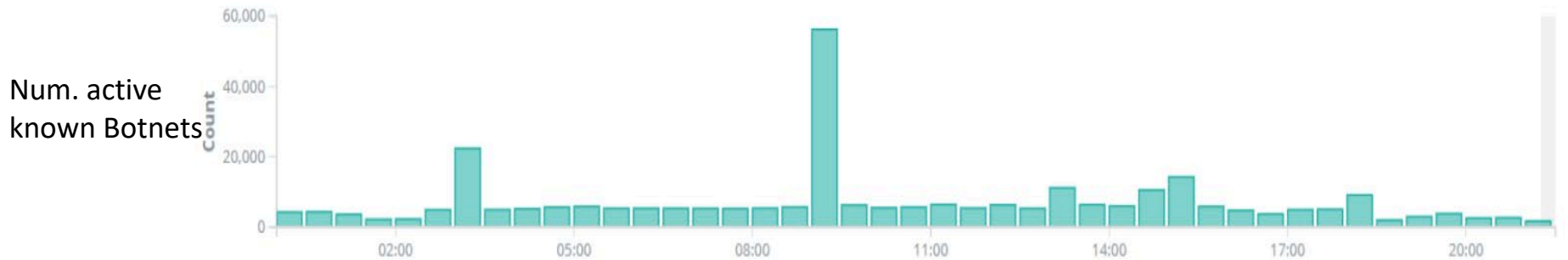
Demonstration using flow data from GLORIAD network  
from 2012 - 2015

# Prediction of Large Data-transfers

- Markov chain
  - Deterministic finite-state machine where, given the present state, future transitions only depend on the current state:  $P(i_n | i_0, \dots, i_{n-1}) = P(i_n | i_{n-1}) \equiv P_{i,j}$
- Steady-state probabilities
  - Expected number of times each state contributes to infinitely long realization: Solve  $\pi P = \pi$
- Mean first passage time
  - Expected number of transitions from given state till another state is reached:  $m_{i,j} = 1 + \sum_{k \neq j} P_{i,k} m_{k,j}$

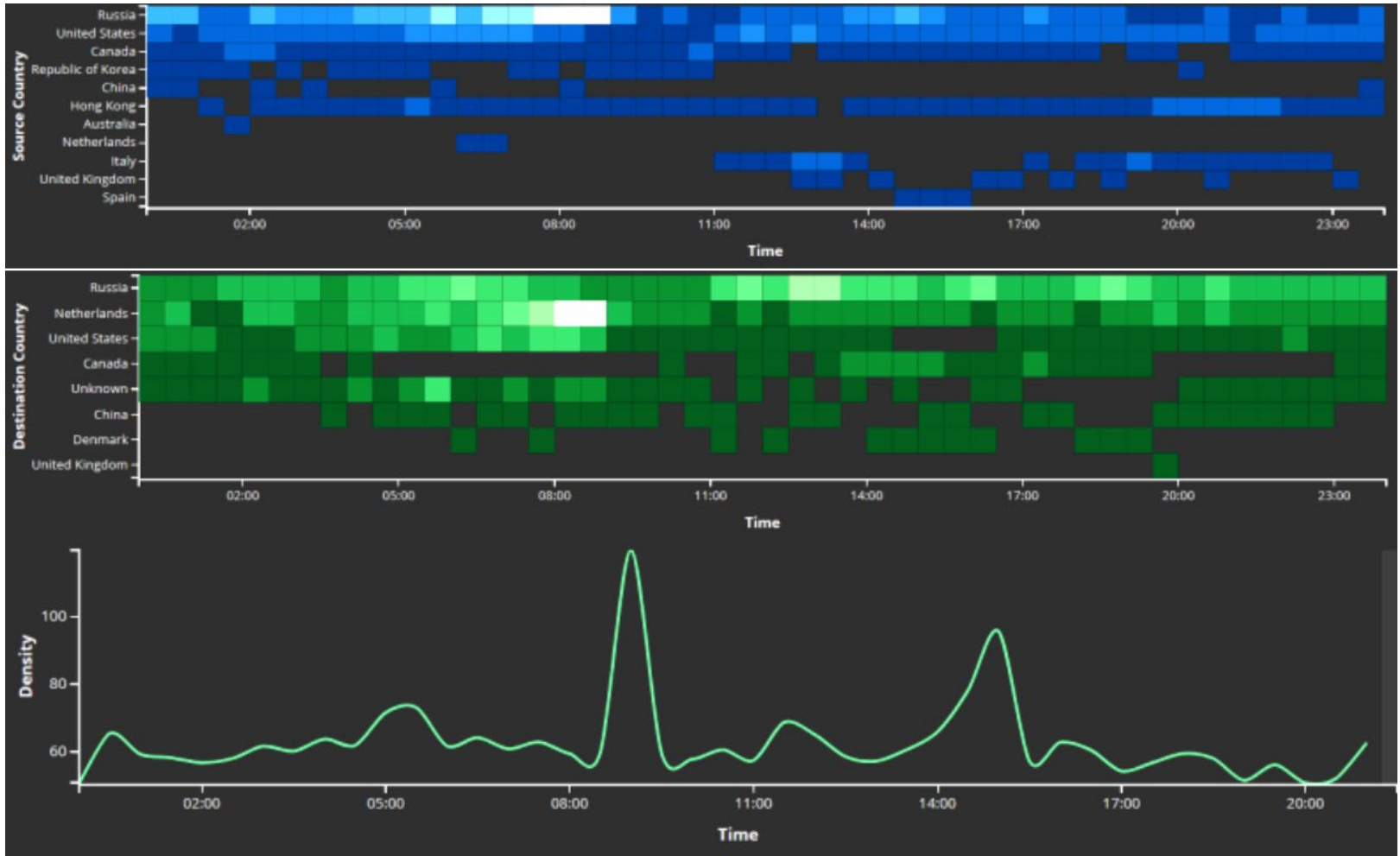
# Botnet Detection

- RED Alert Algorithm (Recursive Event Detection).
- Uses tensors as storage containers for data.
- Based on multi-linear algebra theory.



Host IPs identified by automatically filtering data

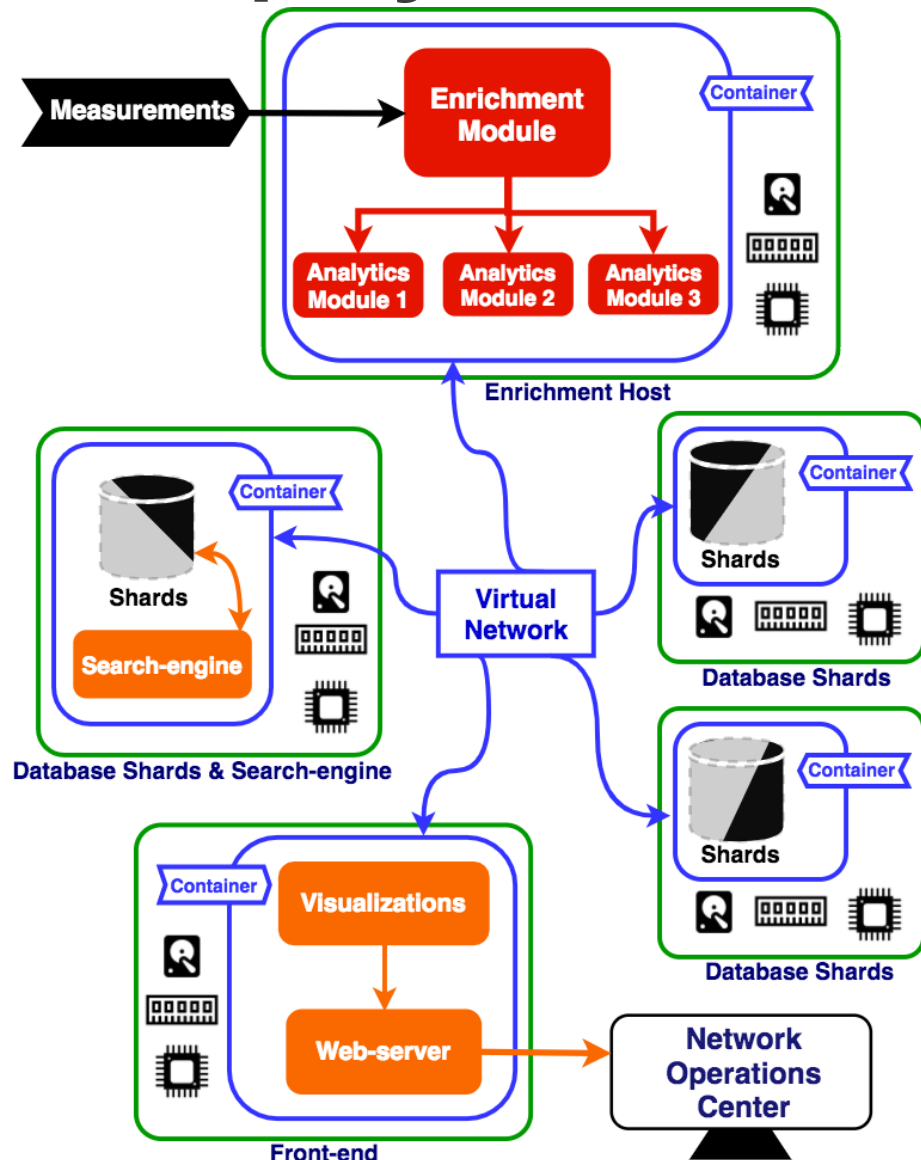
# Botnet Detection



Botnet activity detected by RED Alert

# Virtualized Deployment

- Extensibility via modularized deployment
- Third-party plugin support
- Docker based distribution
- Available via GitHub



# More Information

- Contact information:
  - Angel Kodituwakku
    - Email: [angelk@utk.edu](mailto:angelk@utk.edu)
    - LinkedIn: **angelkdev**
  - Jens Gregor
    - Email: [jgregor@utk.edu](mailto:jgregor@utk.edu)
- Download at GitHub
  - <https://github.com/angelkdev/InSight2>
- InSight2 usage demo  
<https://youtu.be/jcc7Bk9BHpM>