VDI Center for Visual & Decision Informatics

Visual Analytics Sandbox: A big data platform for processing network traffic

Raju Gottumukkala, Ph.D.

Director of Research, Informatics Research Institute

Site Director, NSF Center for Visual and Decision Informatics

Assistant Professor, College of Engineering

University of Louisiana at Lafayette

2017 Internet2 Global Summit (04/26/2017)

Funded by NSF award No.1429526

A National Science Foundation Industry/ University Cooperative Research Center

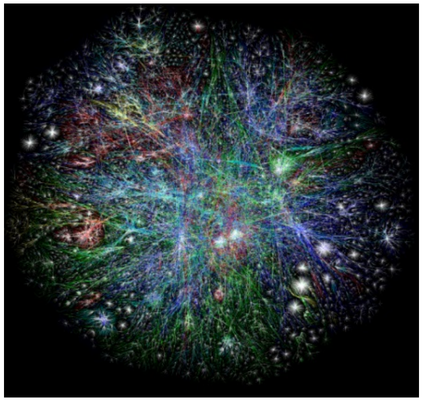
University of Louisiana Lafayette | Informatics Research Institute

University of Louisiana Lafayette

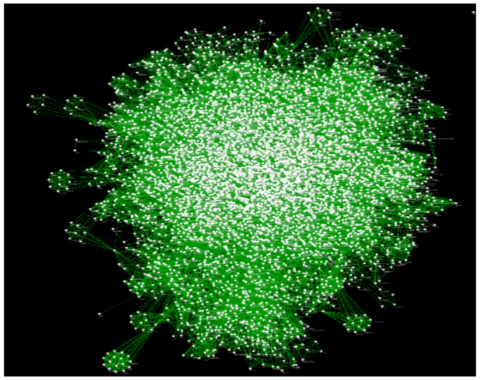
Center for Visual & Decision Informatics

Motivation

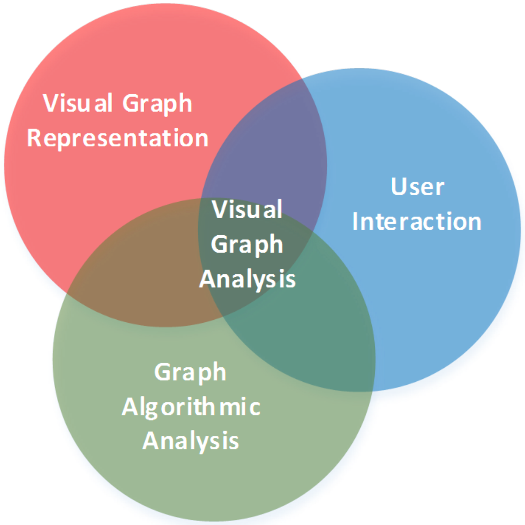
- Cyber environment is increasingly getting complex
- Existing tools do not support interactive analysis of dynamic graphs

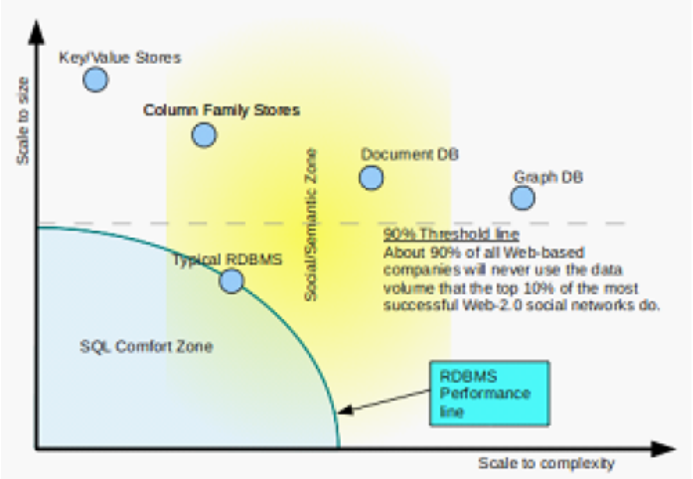


Web graph [©GW3BI]



Ip Flow Graph [©CVDI]





- **Graphs are complex data stores**

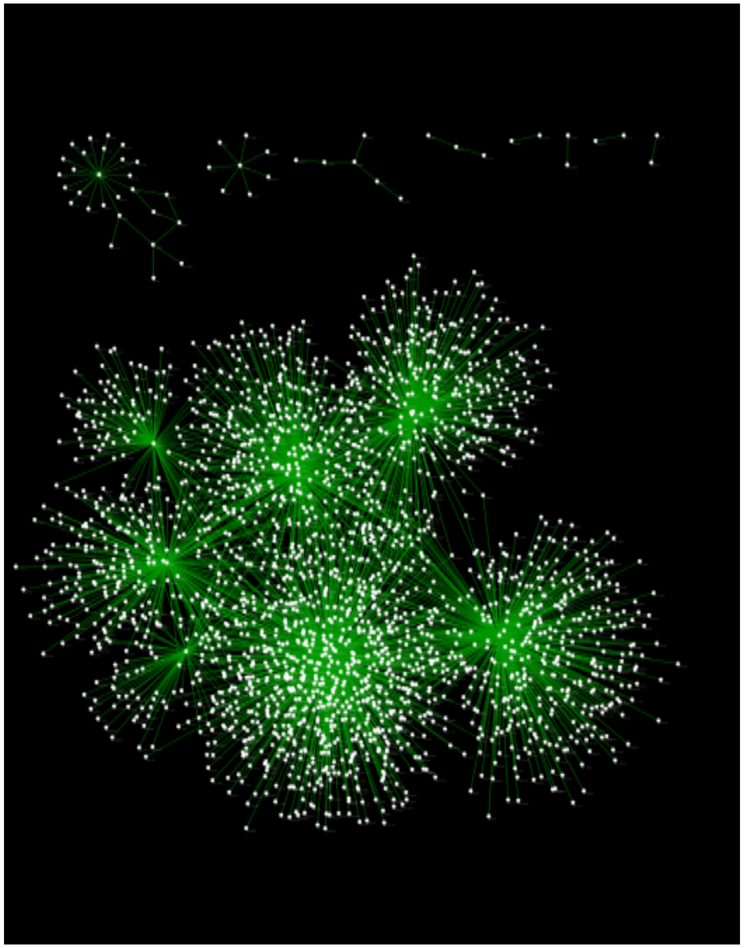
A National Science Foundation Industry/ University Cooperative Research Center

UNIVERSITY of LOUISIANA LAFAYETTE

Application of TVG in cyber-security

- Extraction of Traffic Dispersion Graph (TDG) from IP-Flows
- Graph structural properties indicate abnormal traffic patterns (malware)

TDG Properties	Malware Indication (TDG Class)
Average In-Degree > 2.8 In & Out Edges > 1	P2P traffic
In & Out > 1 Diameter > 11	P2P traffic (BitTorrent)
Sink Vertices > 95% GWCC < 5	MyDoom Malware
Sink Vertices > 95% Average Out Degree > 150	Bobax Malware
Sink Vertices > 95% Average Out Degree 50 to 150	Slammer Malware
Otherwise	Normal Traffic (HTTP, FTP, etc.)

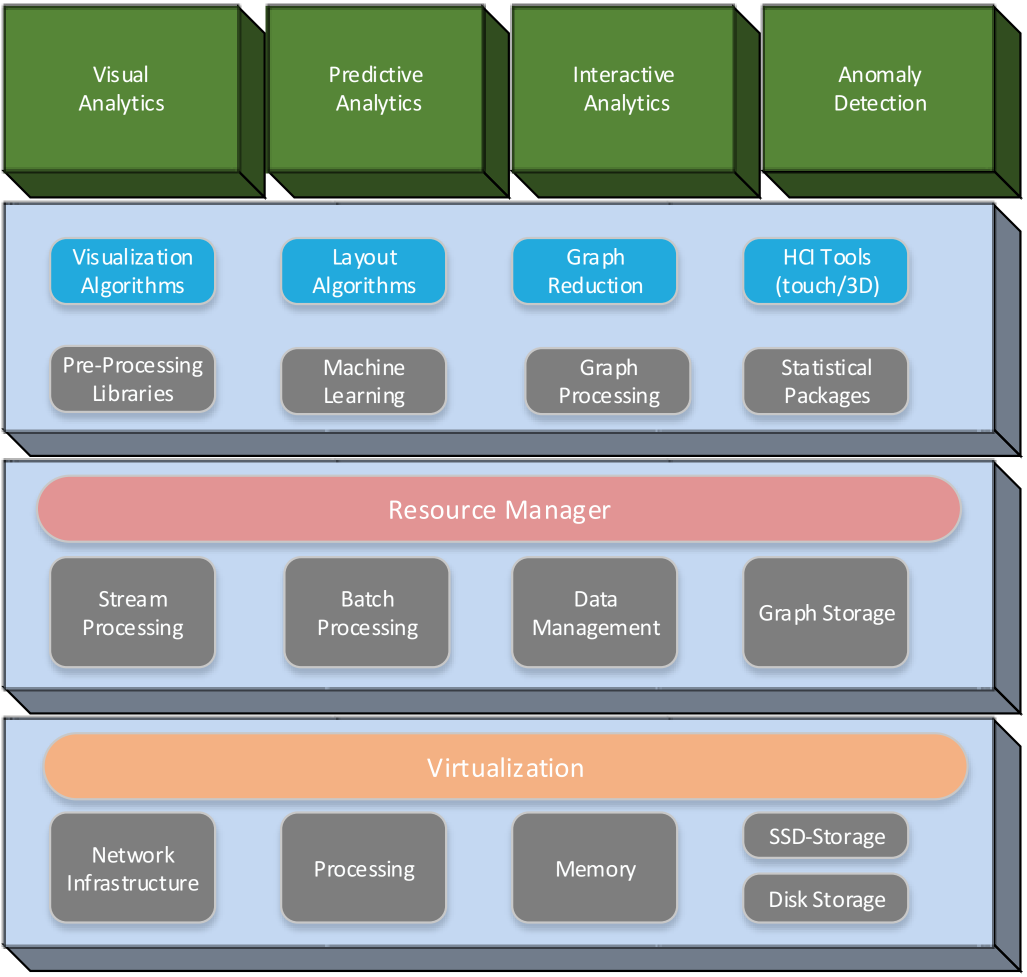


A TDG for soribada P2P

VDI Center for Visual & Decision Informatics

What is Visual Analytics Sandbox?

- A unique analytics environment for processing high-volume, high-velocity data streams
  - IoT, IP flow graphs, click streams, social media, etc.
- Experimental infrastructure to develop next-generation decision support tools



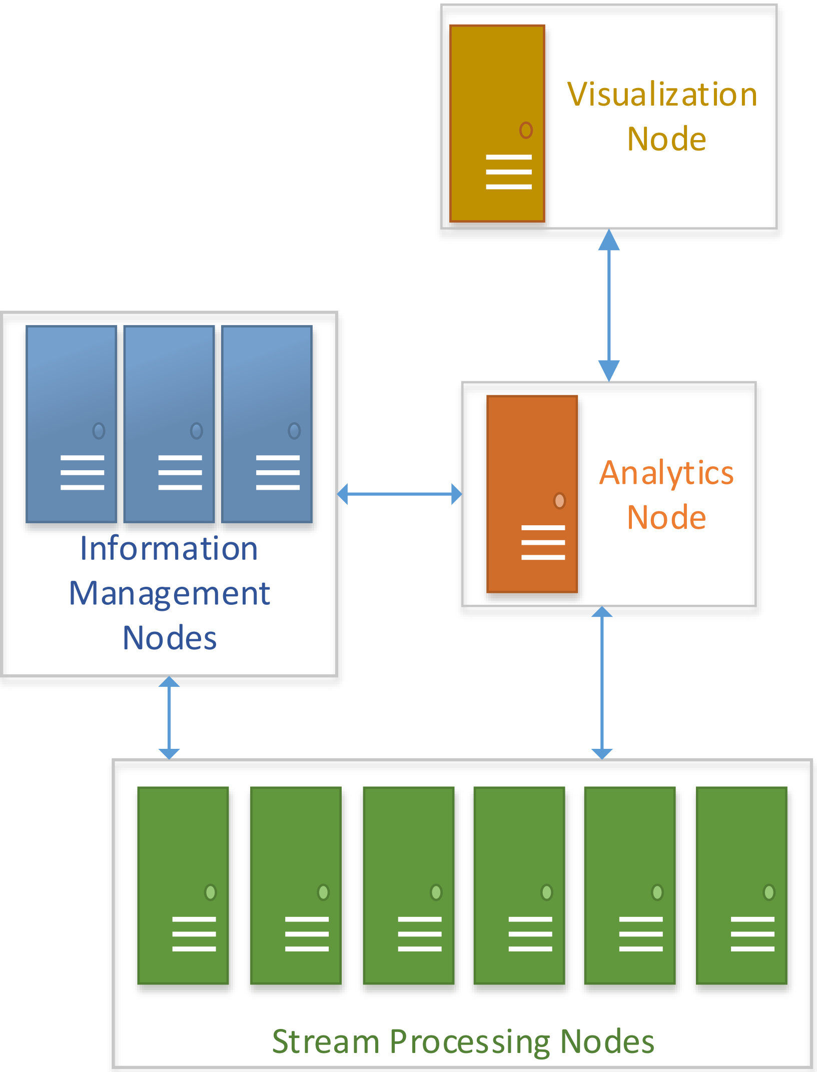
A National Science Foundation Industry/ University Cooperative Research Center

UNIVERSITY of LOUISIANA LAFAYETTE

Center for Visual & Decision Informatics

Hardware

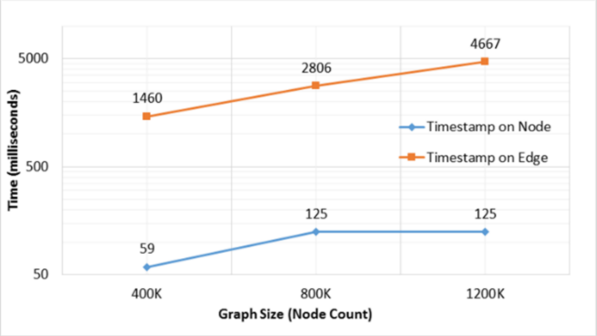
- 112 T Flops of computing
  - 22 Intel Xeon E5 processors (308 cores)
  - 12 NvDIA P100 GPU’s
- Storage
  - 1.8 TB RAM
  - 25.6 TB NvME SSD
    - 25X faster than HDD, 5X faster than SSD
  - 20 TB HDD
- Networking
  - 2X10G Ethernet Cards per Node



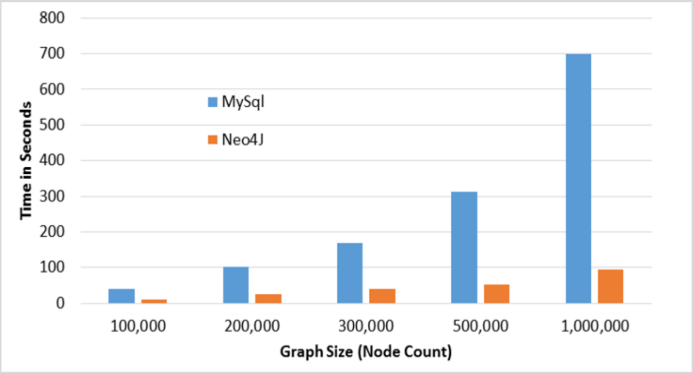
Center for Visual & Decision Informatics

Benchmarking Graph Operations

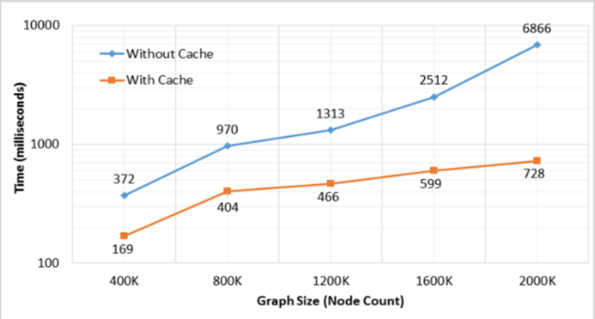
Type of Temporal Characteristic	Graph Operations
Temporal network topology	Degree, connectivity, density
Reachability analysis	Paths, walks, trails
Detecting outliers	Node or edge clustering
Node neighborhoods	Persistent patterns & motifs



Execution times for retrieving shortest path between two selected nodes (Neo4j schemes)



Execution times for retrieving cumulative edge weights of a node (Neo4J vs MySQL)



Execution times for retrieving weighted shortest path between two nodes (Neo4j with timestamp on node and caching vs no caching)

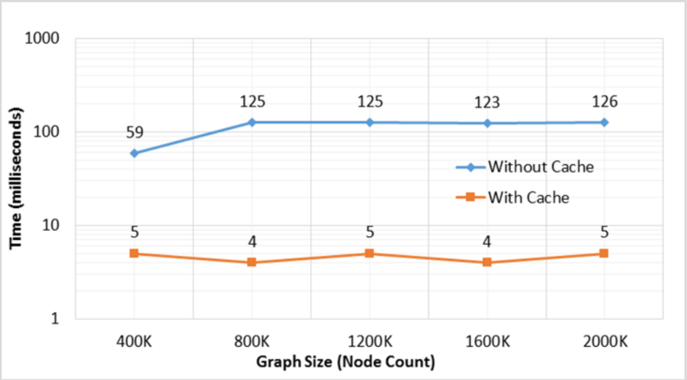


Figure 9. Execution times for retrieving shortest path between two nodes (Neo4j with timestamp on node and caching vs no caching)

A National Science Foundation Industry/ University Cooperative Research Center

UNIVERSITY of LOUISIANA LAFAYETTE

Touch interactions

(a) Node history graphs with timestamp browsing

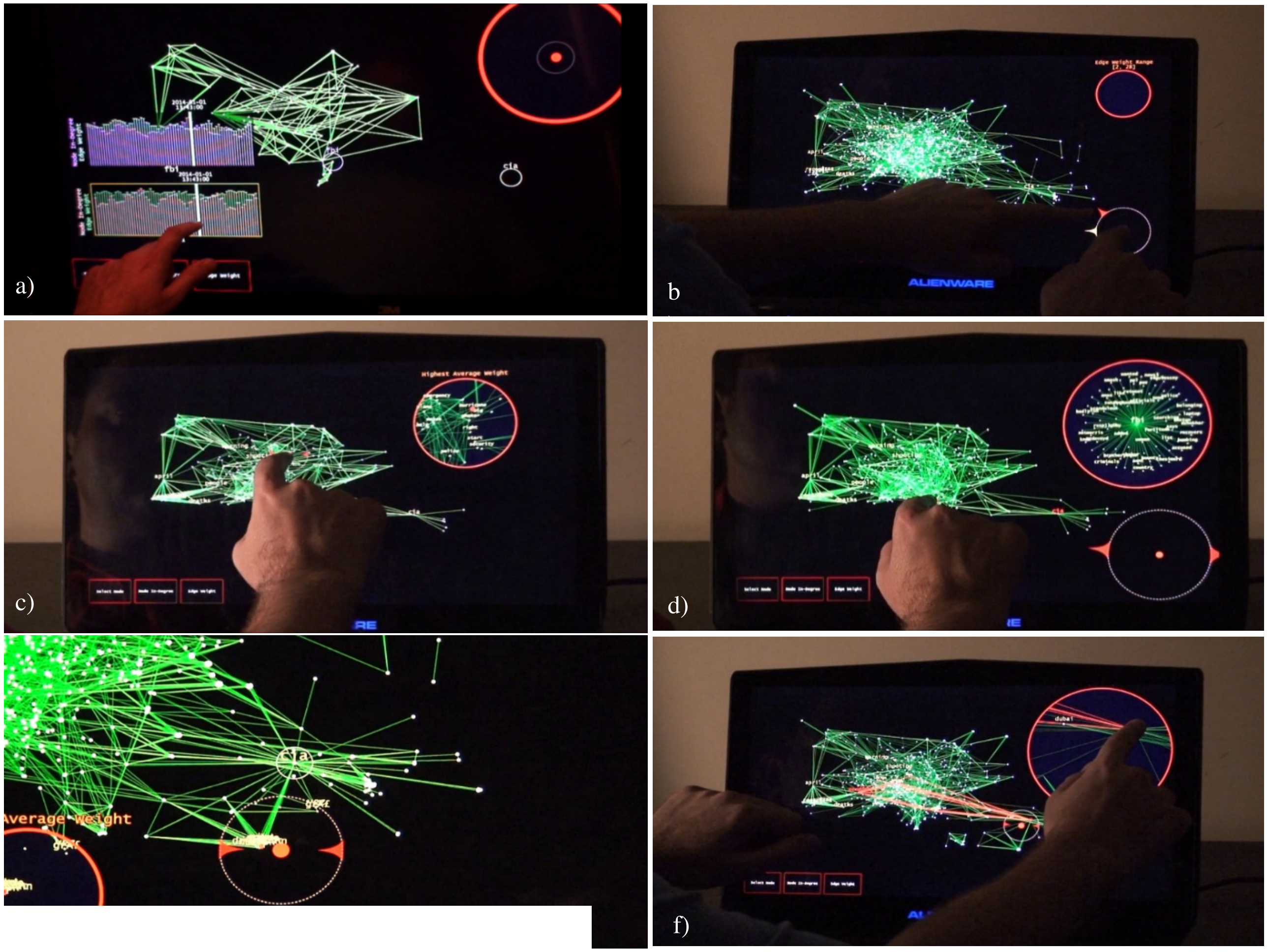
(b) Graph filtering based on edge weight with multitouch controller (lower-bound highlighted yellow).

(c) Control widget (around finger) and zoom window with detached source.

(d) Local neighborhood of a selected node arranged and rendered in detached window (upper-right).

(e) Progress indicator circle provides feedback to users while analytics jobs are processed.

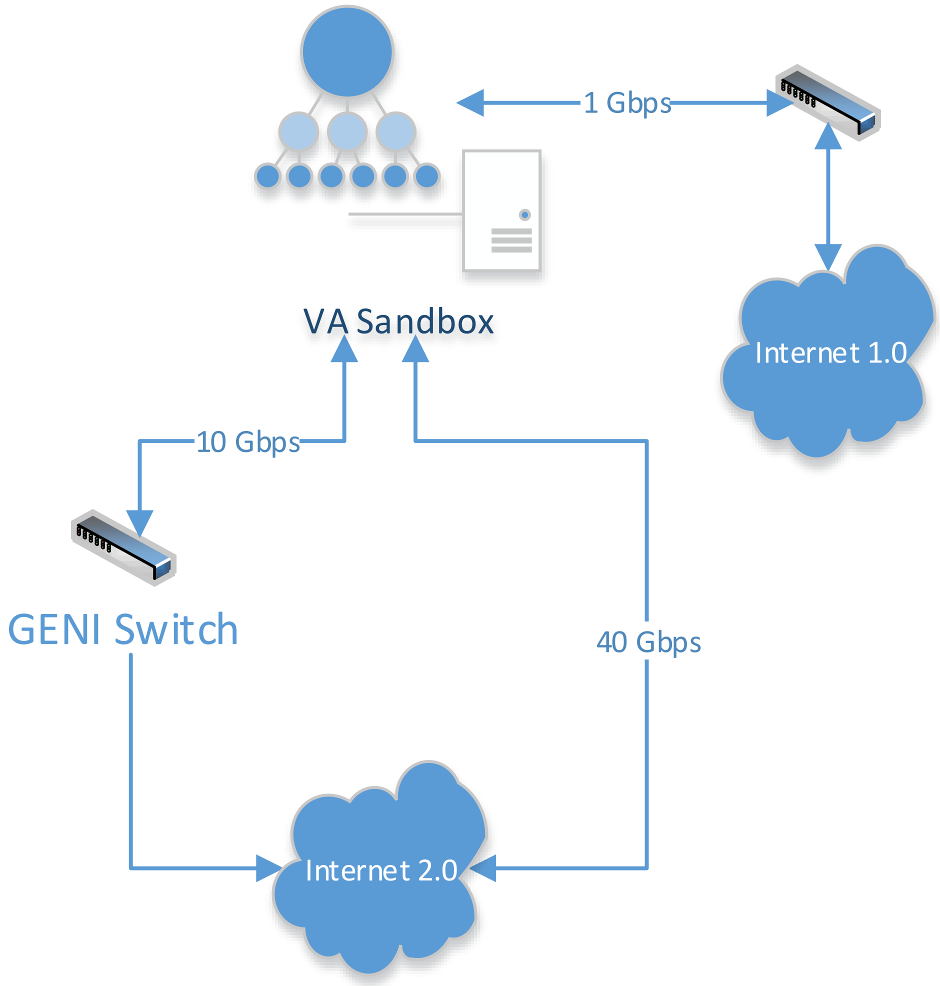
(f) Shortest paths between nodes



VDI Center for Visual & Decision Informatics

Connectivity

- Connected to LONI Network and Science DMZ
- Connected to Internet 2.0 through GENI switch
- 40 Gbps connections for faster data transfer



A National Science Foundation Industry/
University Cooperative Research Center

UNIVERSITY of LOUISIANA LAFAYETTE

Expectations & Timeline

- How can you use this system?
  - Virtual big data environment for stream processing, graph analytics
  - Leverage stream processing, graph mining & visualization tools that are part of the sandbox
- Plan for deployment
  - Software developed in a pilot environment for (a) intelligent levee surveillance, (b) event detection in social media and (c) influenza forecasting
  - Looking for collaboration for one use case in cybersecurity use case for network analysis (Aug 2017)
  - System will be made available for users on Internet 2 (Spring 2018)

Thank You

Contact: raju@louisiana.edu

Homepage: www.ucs.louisiana.edu/~nrg0821

Center for Visual & Decision Informatics

Real Time Event Detection

A National Science Foundation Industry/ University Cooperative Research Center