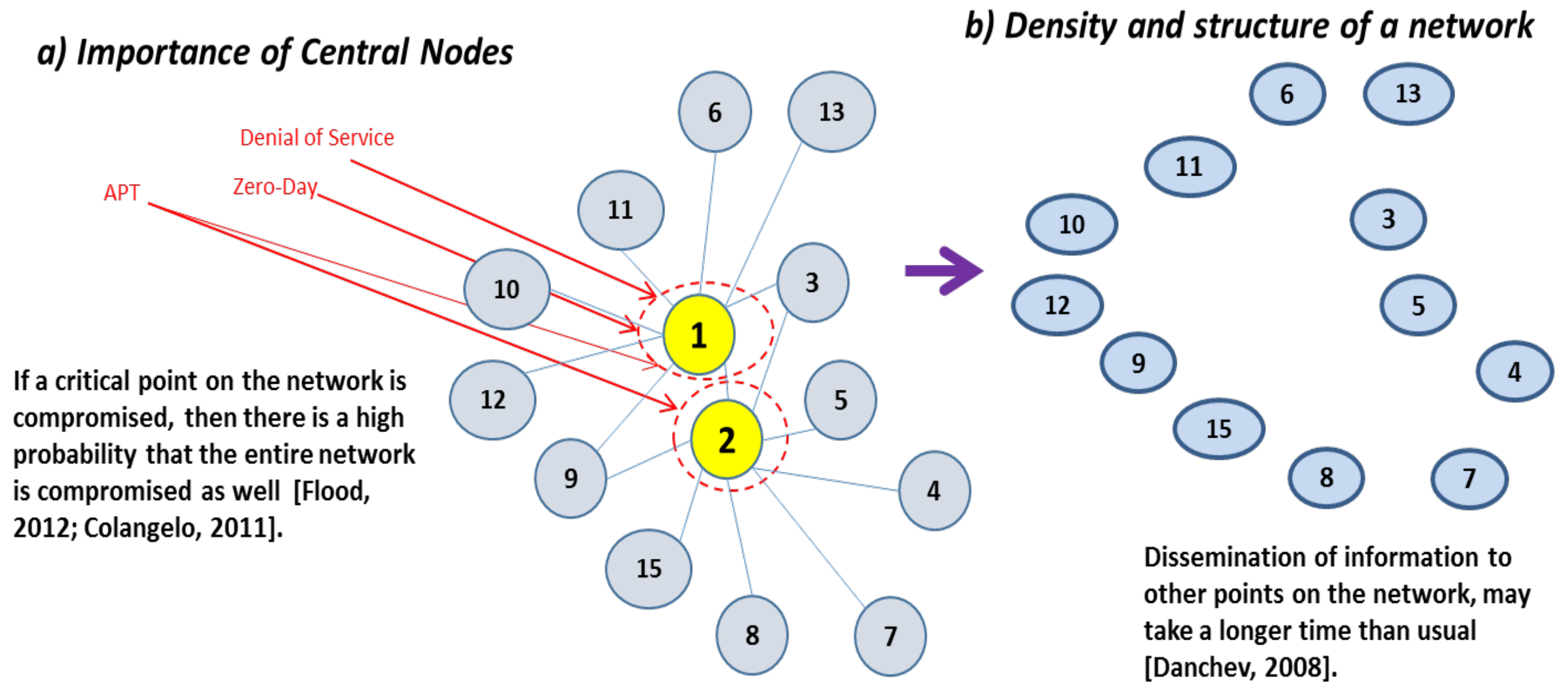
Big Data analytics for Cyber security: A case study in change detection for evolving networks

Josephine Namayanja, Akshay Grover, Jay Gholap, Vandana P. Janeja

University of Maryland, Baltimore County

Motivation

Temporally Evolving Computer Network


a) Importance of Central Nodes
b) Density and structure of a network
Denial of Service
Zero-Day
APT
1
2
3
4
5
6
7
8
9
10
11
12
13
15

If a critical point on the network is compromised, then there is a high probability that the entire network is compromised as well [Flood, 2012; Colangelo, 2011].

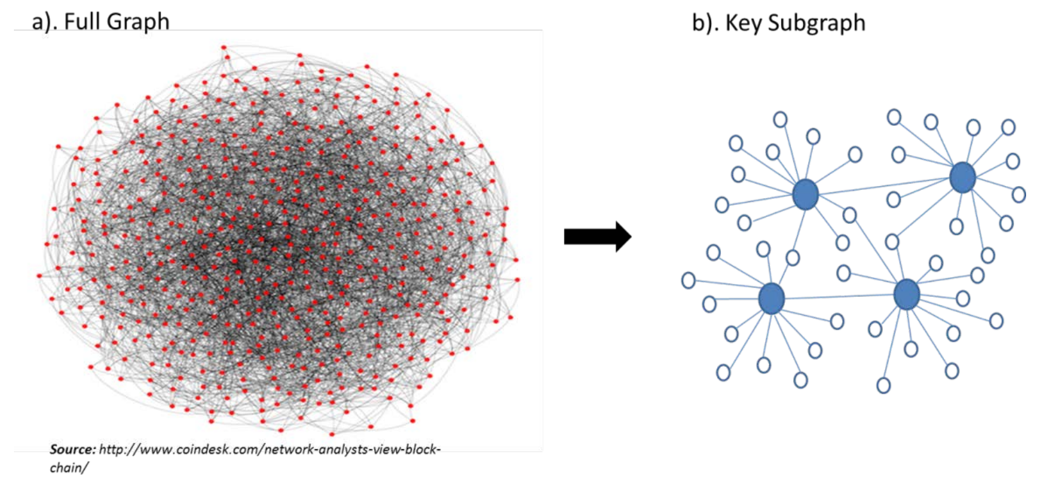
Dissemination of information to other points on the network, may take a longer time than usual [Danchev, 2008].

Research Objective

To identify pertinent changes in temporally evolving computer networks by evaluating the behavior of central(key) nodes and their impact on the network over time while utilizing an efficient data processing framework.

Challenges

- **Numerous nodes and edges.**
- **Traffic is captured at multiple time intervals.**
- **Understanding Holistic Change is challenging.**
- **Computationally costly.**

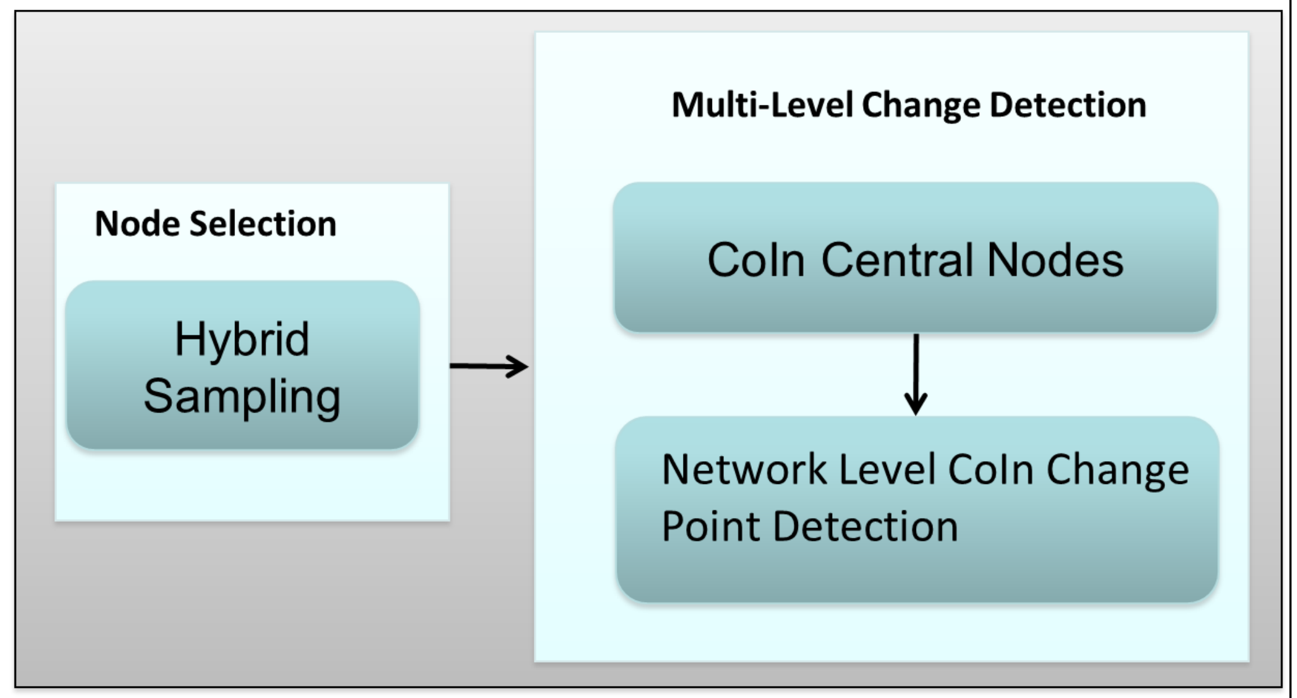


Source: http://www.coindesk.com/network-analysts-view-block-chain/

Contributions

- **Node Selection** *Namayanja and Janeja , IJCMAM “11, JMS “12, IEEE ISI “13, IEEE Big Data “14, IEEE ISI “15*
  - Hybrid Sampling
- **Multi-Level Change Detection** *Namayanja and Janeja, IEEE ISI “13, IEEE Big Data “14* *Namayanja and Janeja, IEEE ISI “15*
  - Consistent and Inconsistent (CoIn) Central Nodes *Namayanja, Grover, Gholap and Janeja, Technical Report “15, IEEE Big Data “14*
  - Network Level Change Points due to CoIn Central Nodes (NL-CoIn)
- **Big Data Framework**
- **Extensive experimental and comparative results using real world internet traces.** *Namayanja and Janeja, IEEE ISI “15*
- **Validate results with real world cyber attacks.**

Approach



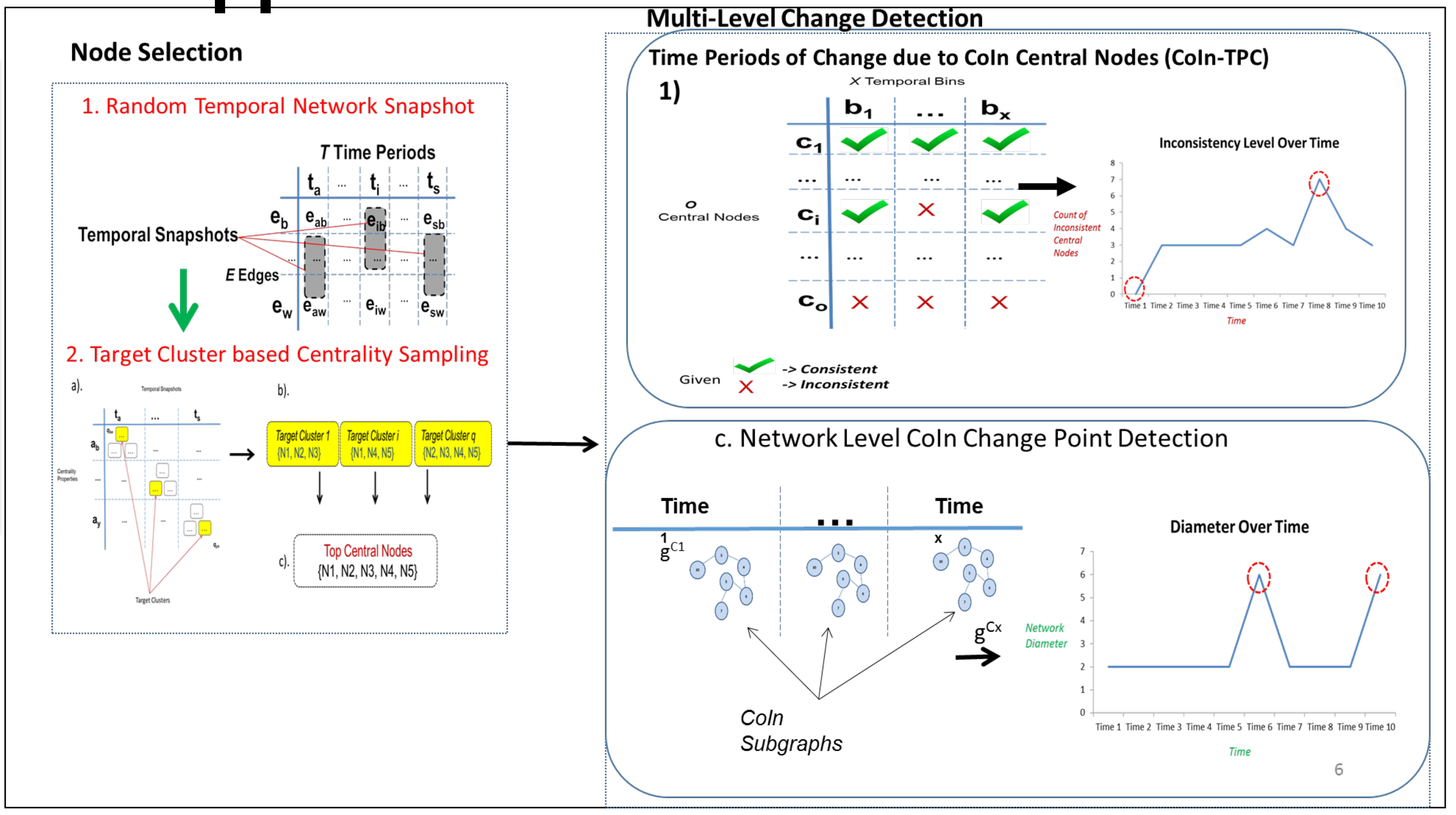
Node Selection

Hybrid Sampling

Multi-Level Change Detection

CoIn Central Nodes

Network Level CoIn Change Point Detection



Node Selection

1. Random Temporal Network Snapshot

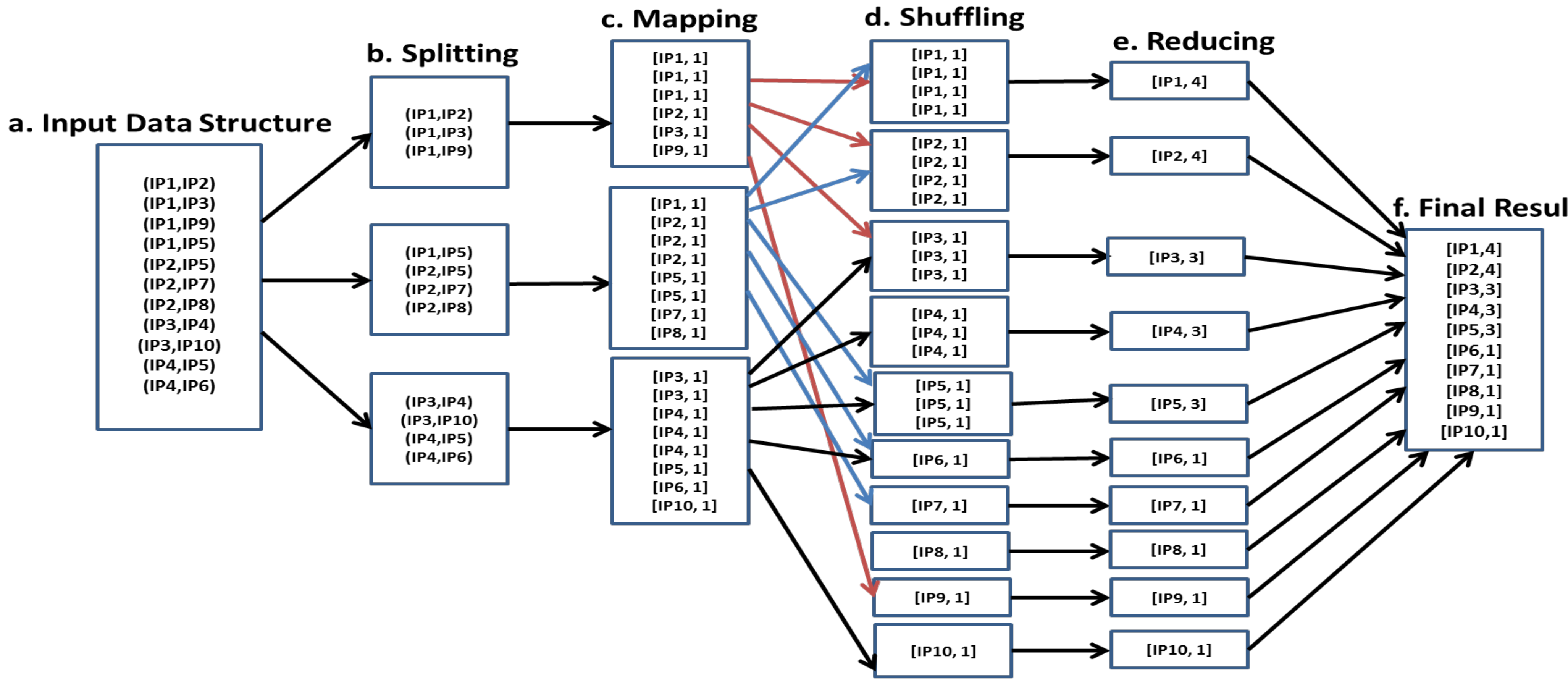
2. Target Cluster based Centrality Sampling

Multi-Level Change Detection

Time Periods of Change due to CoIn Central Nodes (CoIn-TPC)

c. Network Level CoIn Change Point Detection

Big Data Framework



a. Input Data Structure

(IP1,IP2)
(IP1,IP3)
(IP1,IP9)
(IP1,IP5)
(IP2,IP5)
(IP2,IP7)
(IP2,IP8)
(IP3,IP4)
(IP3,IP10)
(IP4,IP5)
(IP4,IP6)

b. Splitting

(IP1,IP2)
(IP1,IP3)
(IP1,IP9)

(IP1,IP5)
(IP2,IP5)
(IP2,IP7)
(IP2,IP8)

(IP3,IP4)
(IP3,IP10)
(IP4,IP5)
(IP4,IP6)

c. Mapping

[IP1, 1]
[IP1, 1]
[IP1, 1]
[IP2, 1]
[IP3, 1]
[IP9, 1]

[IP1, 1]
[IP2, 1]
[IP2, 1]
[IP2, 1]
[IP5, 1]
[IP5, 1]
[IP7, 1]
[IP8, 1]

[IP3, 1]
[IP3, 1]
[IP4, 1]
[IP4, 1]
[IP4, 1]
[IP5, 1]
[IP6, 1]
[IP10, 1]

d. Shuffling

[IP1, 1]
[IP1, 1]
[IP1, 1]
[IP1, 1]

[IP2, 1]
[IP2, 1]
[IP2, 1]
[IP2, 1]

[IP3, 1]
[IP3, 1]
[IP3, 1]

[IP4, 1]
[IP4, 1]
[IP4, 1]

[IP5, 1]
[IP5, 1]
[IP5, 1]

[IP6, 1]

[IP7, 1]

[IP8, 1]

[IP9, 1]

[IP10, 1]

e. Reducing

[IP1, 4]
[IP2, 4]
[IP3, 3]
[IP4, 3]
[IP5, 3]
[IP6, 1]
[IP7, 1]
[IP8, 1]
[IP9, 1]
[IP10, 1]

f. Final Result

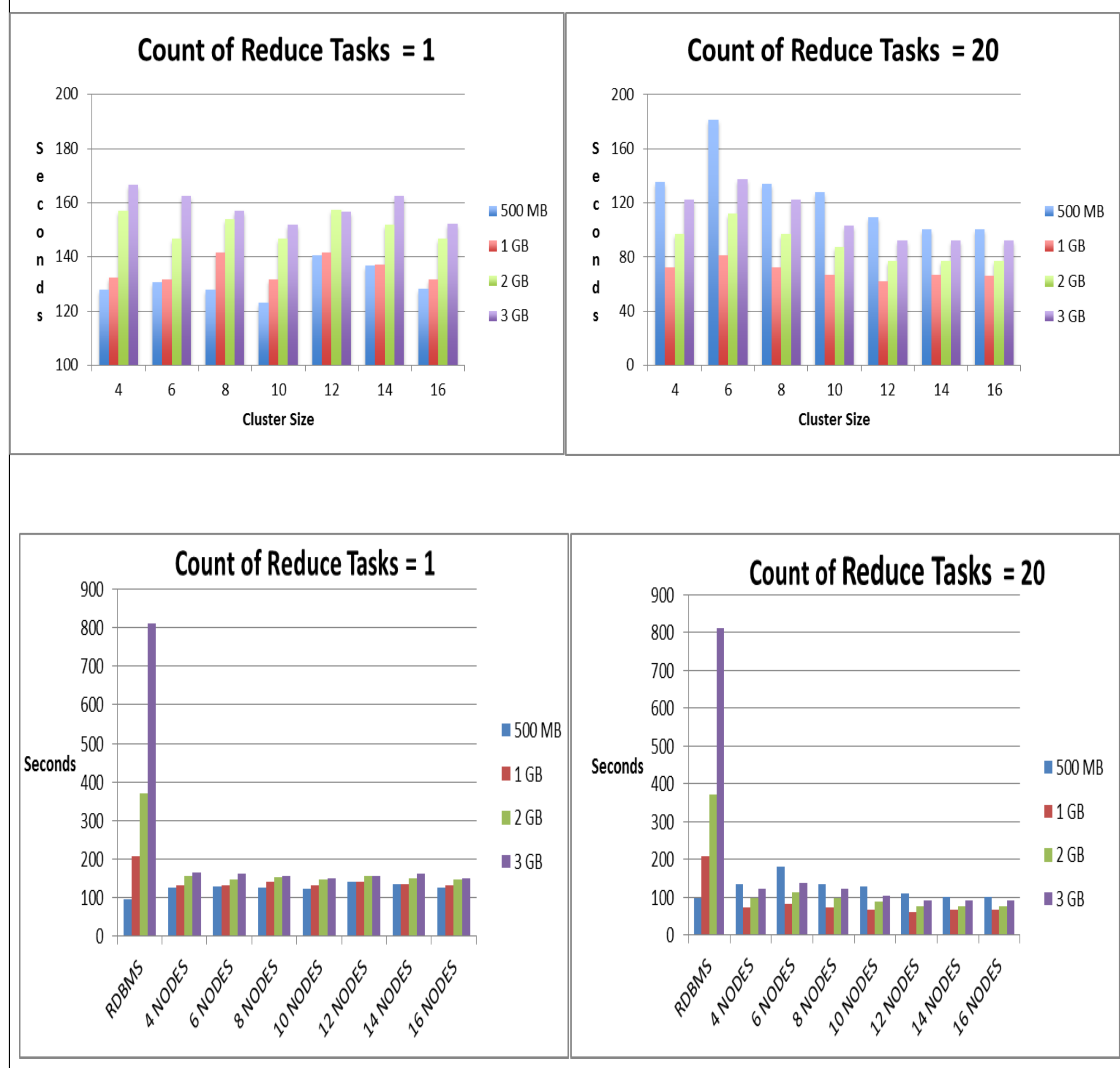
[IP1,4]
[IP2,4]
[IP3,3]
[IP4,3]
[IP5,3]
[IP6,1]
[IP7,1]
[IP8,1]
[IP9,1]
[IP10,1]

Map Reduce tasks for change detection in evolving network using degree centrality

Cluster Specifications: **Maya** HPCF (NSF, UMBC)

- The hardware used in the computational studies is part of the UMBC High Performance Computing Facility (HPCF)
- 240 Nodes Cluster [16 used for Hadoop]
- Each node consists of two quad-core 2.8 GHz Intel Nehalem X5560 CPUs and 24 GB memory are designed for fastest number crunching and connected by a dual-data rate (DDR) InfiniBand network.

Results



Execution Time for Pig Scripts on Maya Cluster

Related Publications

- *J. M. Namayanja*, V.P. Janeja, Change Detection in Evolving Computer Networks: Changes in Densification and Diameter Over Time, IEEE International Conference on Intelligence and Security Informatics., 2015, Baltimore MD
- J. M. Namayanja, V.P. Janeja, Change Detection In Temporally Evolving Computer Networks: A Big Data Framework, First International Workshop on High Performance Big Graph Data Management, Analysis, and Mining, co-located with the IEEE BigData 2014
- V.P. Janeja, A. Azari, J. M. Namayanja, B. Heilig, B-DIDS: Mining Anomalies In A Big-Distributed Intrusion Detection System, 2014 IEEE International Conference on Big Data October 27-30, 2014, Washington DC, USA
- J. M. Namayanja, V.P. Janeja, Discovery of Persistent Threat Structures through Temporal and Geo-Spatial Characterization in Evolving Networks, IEEE Intelligence and Security Informatics (ISI) 2013
- J. M. Namayanja, V.P. Janeja, An Assessment of Patient Behavior Over Time–Periods: A Case Study of Managing Type 2 Diabetes Through Blood Glucose Readings and Insulin Doses, Journal of Medical Systems, DOI: 10.1007/s10916-012-9894-3, Oct 2012
- J. M. Namayanja, V. P. Janeja, Subspace Discovery for Disease Management: A Case Study in Metabolic Syndrome. IJCMAM 2(1): 38-59 (2011)