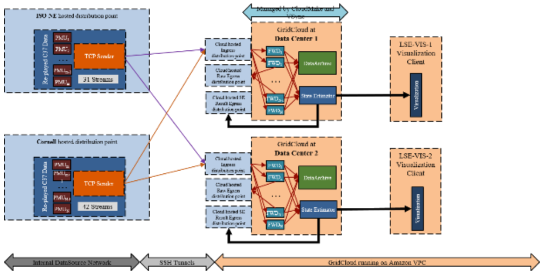
A Cloud Data Sharing Platform for Real-time PMU Data

Carl Hauser

Washington State University

(hauser@eecs.wsu.edu)



PSERC Webinar
September 13, 2016

PSERC Project S-67G
Cloud Data Sharing Platform

Washington State University

Carl Hauser, Anjan Bose

Ming Meng, Dave Anderson

Cornell University

Ken Birman

Theodoros Gkountouvas, Weijia Song

ISO-NE

Eugene Litvinov

Xiaochuan Luo, Qiang Zhang

ADVISORS: Jay Giri (GE), George Stefopoulos (NYPA)

ALSO: Thanks to the US Dept of Energy ARPA-E GENI Program that supported the initial GridCloud research

Presentation Outline

Cloud opportunities and challenges

GridCloud concept and architecture

Performance

Cyber-security

For further investigation

Cloud Opportunity

- Resource elasticity: pay for what you need, only for as long as you need it
  - Low constant cost
  - Massive computation available if needed (event analysis, etc.)
- Geographic flexibility
  - data and computation located close to where needed
  - move to (or backup in) distant location for disasters (Hurricane Sandy experience)
- Neutral ground for data sharing
  - Data sharing platform need not be under physical control of one utility, ISO, etc.

Cloud-specific Challenges

Clouds such as EC2 are surprisingly hostile for real-time work

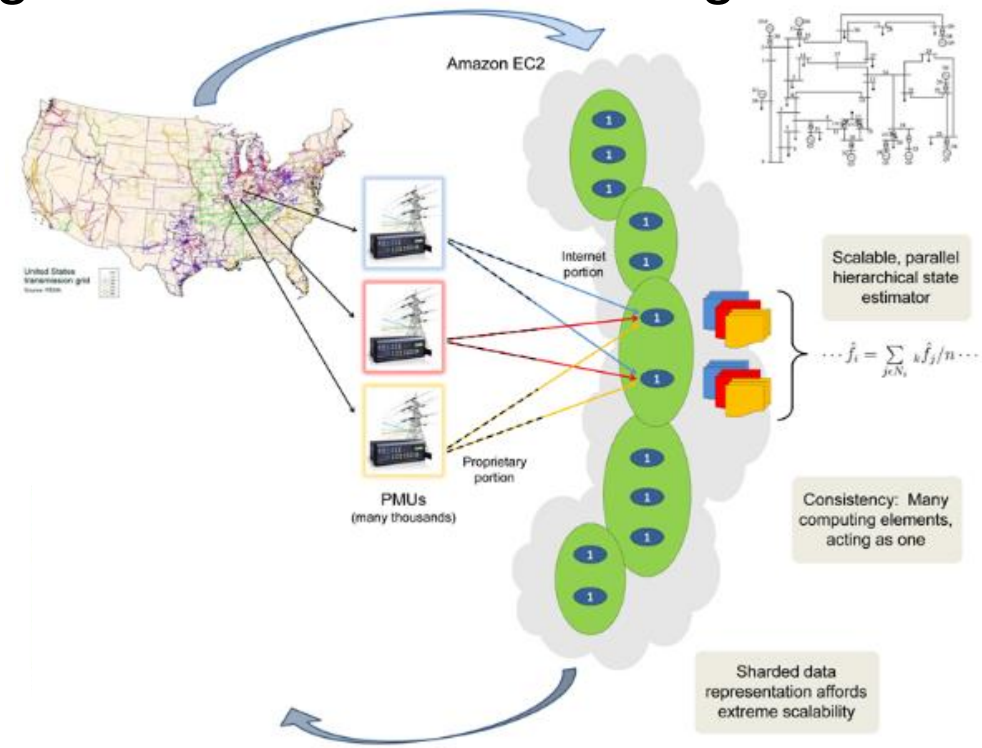
- Underlying scheduler and network layer are unreliable
- Strange timing problems, bursts of message loss, other anomalies
- Overcoming this is made difficult by Amazon's unwillingness to document the AWS infrastructure
- But we've never encountered a problem that we couldn't eventually pin down and solve

Business Environment Challenges

- Distinct owners: peers &hierarchy (ISO)
- Owners control data flow: entities have different security & sharing policies
- ISOs integrate data ... but as we get further from sources, quality of information is a potential concern
- How valuable is shared PMU data for operations?
  - Is sharing unthinkable due to technical barriers? We can help with that
  - Due to business barriers? That’s harder!

System Concept

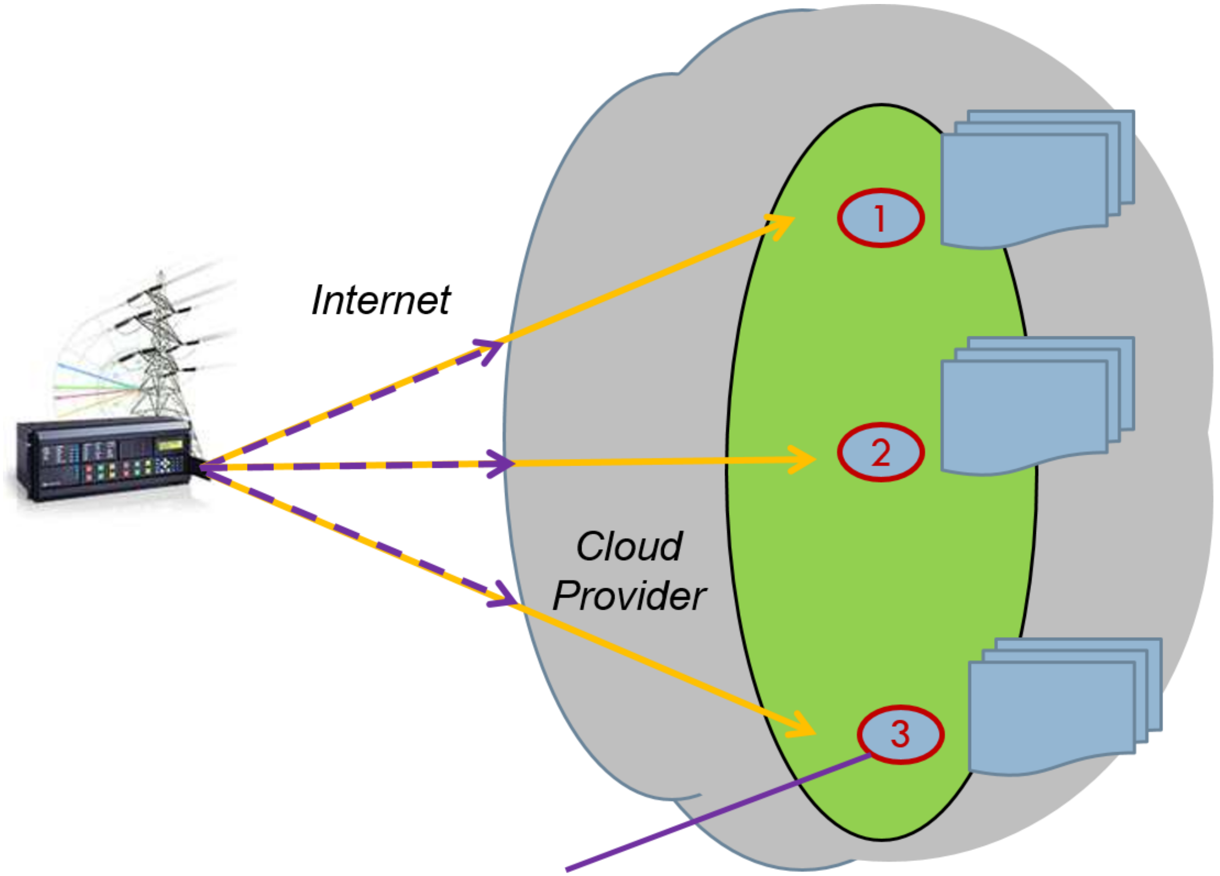
A distributed platform for real-time data collection, storage, processing and dissemination using Cloud Computing



- Use redundancy to overcome real-time disruptions and failures.
- Use proven techniques from distributed computing to manage issues of consistency and availability

Redundant Communication

Reliability and Performance

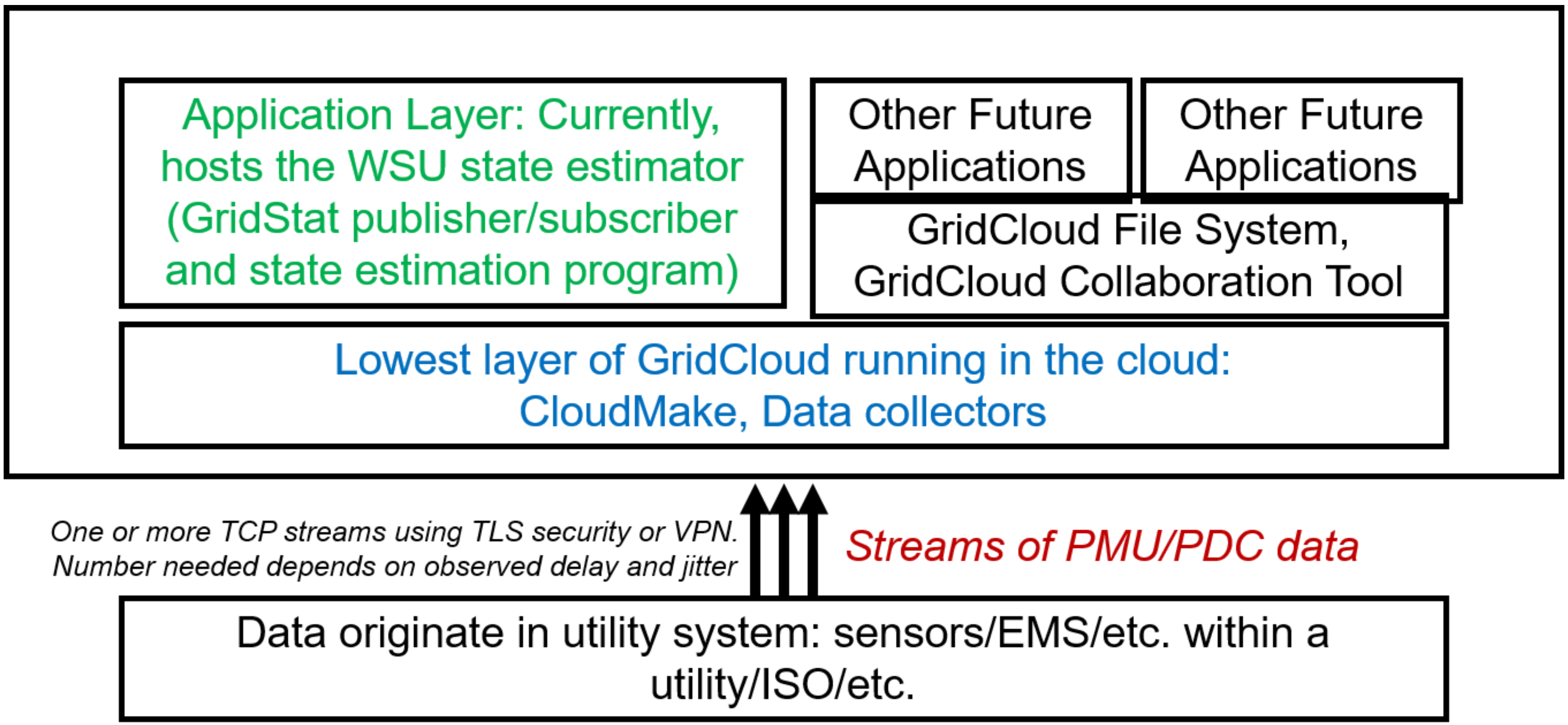


Different network paths and delays cause shards to receive the same data at different times

Delay is an issue inside the Cloud as well as in the Internet

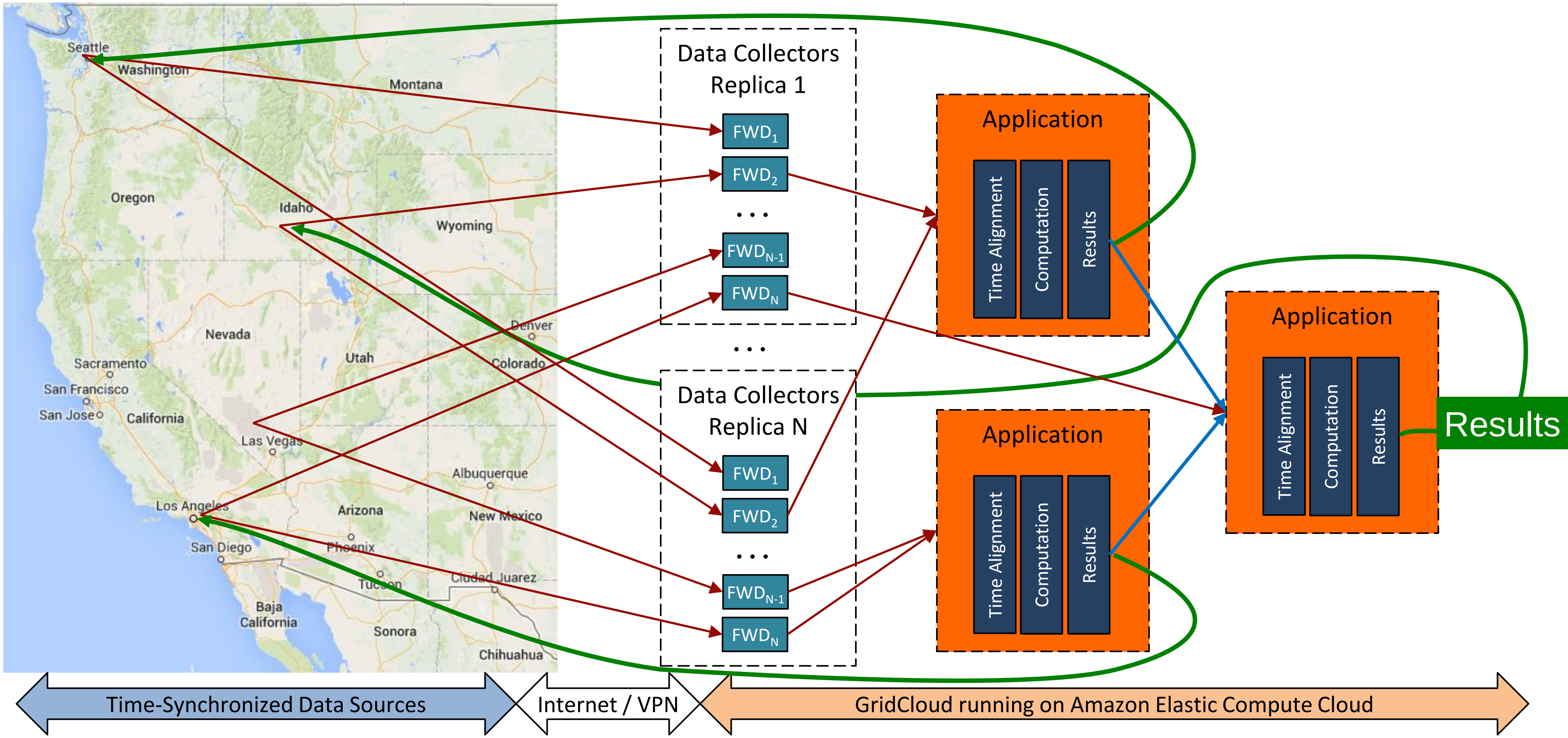
Data Collectors forward data and can (optionally) store them as received in disk files using our real-time snapshotting file system.

System Architecture



GridCloud – the ARPA-E Demonstration

Scalability and Fine-grained Replication



Dave Anderson, Theodoros Gkountouvas, Carl Hauser

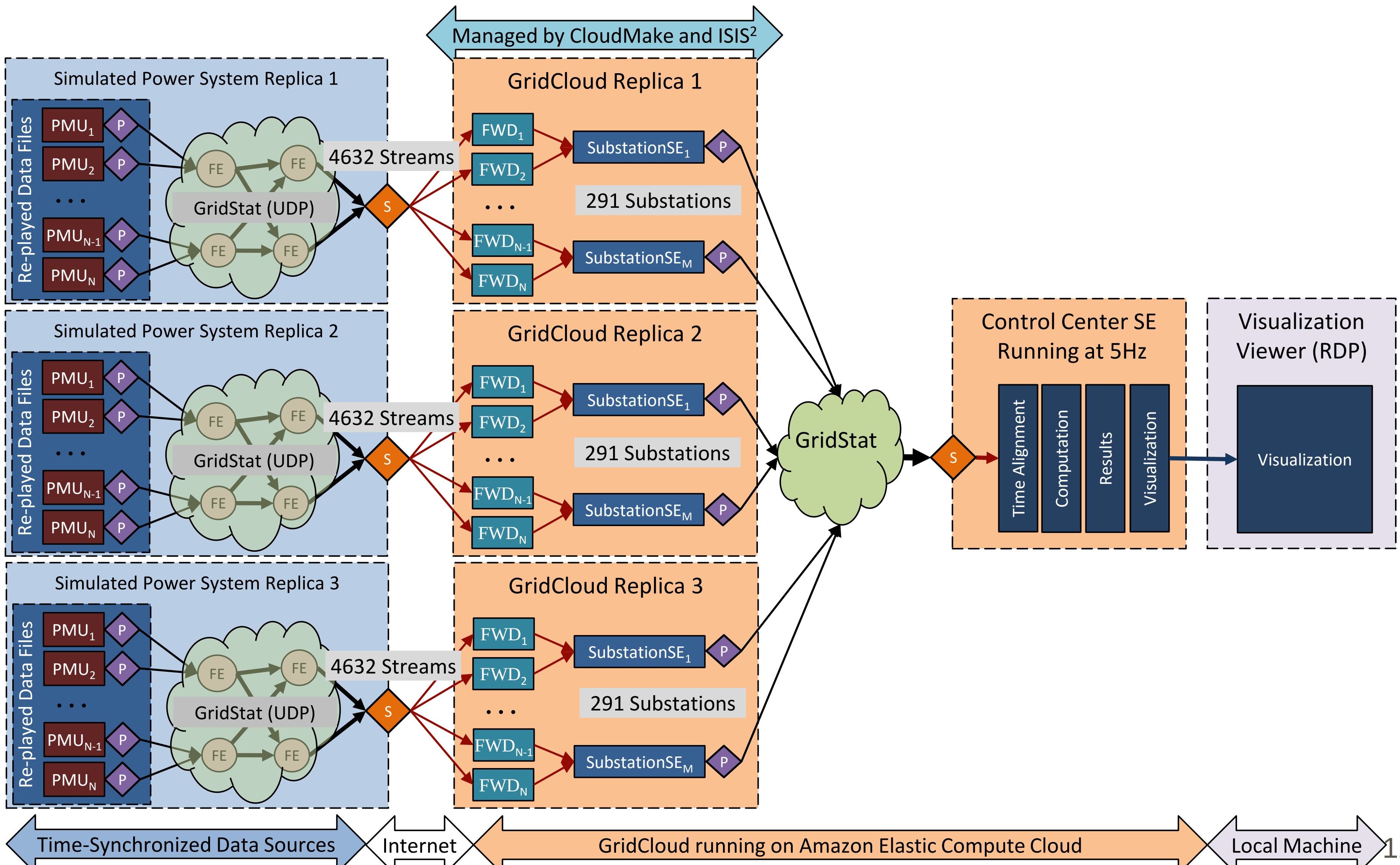
GridCloud for ARPA-E Demonstration
Simulated 6K Bus WECC System

- Comparison of 6K to 179 Bus system:

	Old	New	Scale
Substations	127	**291**	2.29x
Busses	179	**6,000**	33.52x
Streams (PMUs)	1,577	**4,632**	2.94x

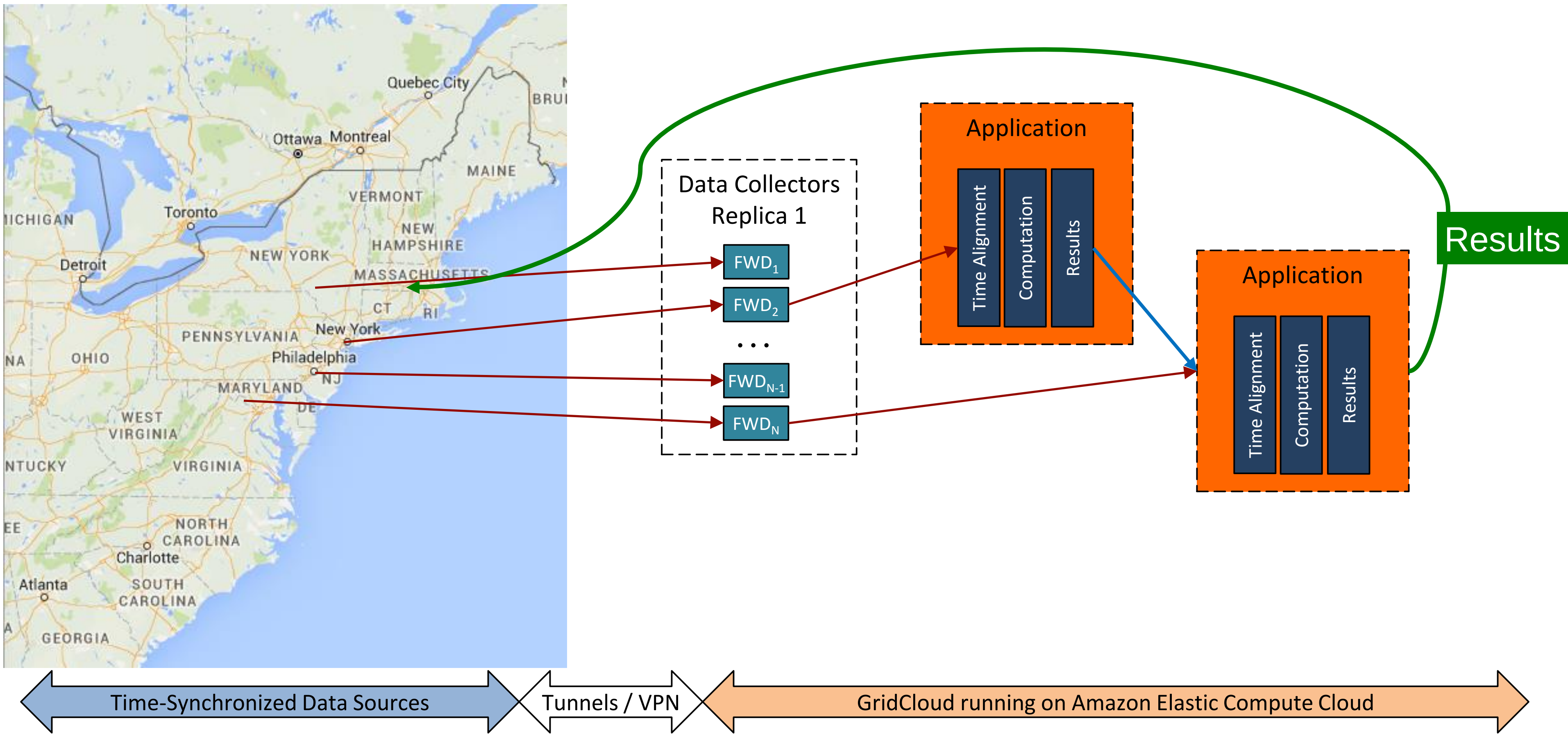
- Power System Description:
  - 6,000 busses
  - A simplified model (~1/3rd number of busses) of the entire WECC system
  - This is the primary model used by industry and academia for studying the July 2nd 1996 blackout
  - All power components (busses, branches, etc.) in the system above and including 230kv are monitored

ARPA-E GridCloud Demonstration
(6K bus, 3 Replicas)



ISO-NE GridCloud Demonstration

Smaller-scale, Geographic Redundancy, Security, Multiple Data Sources

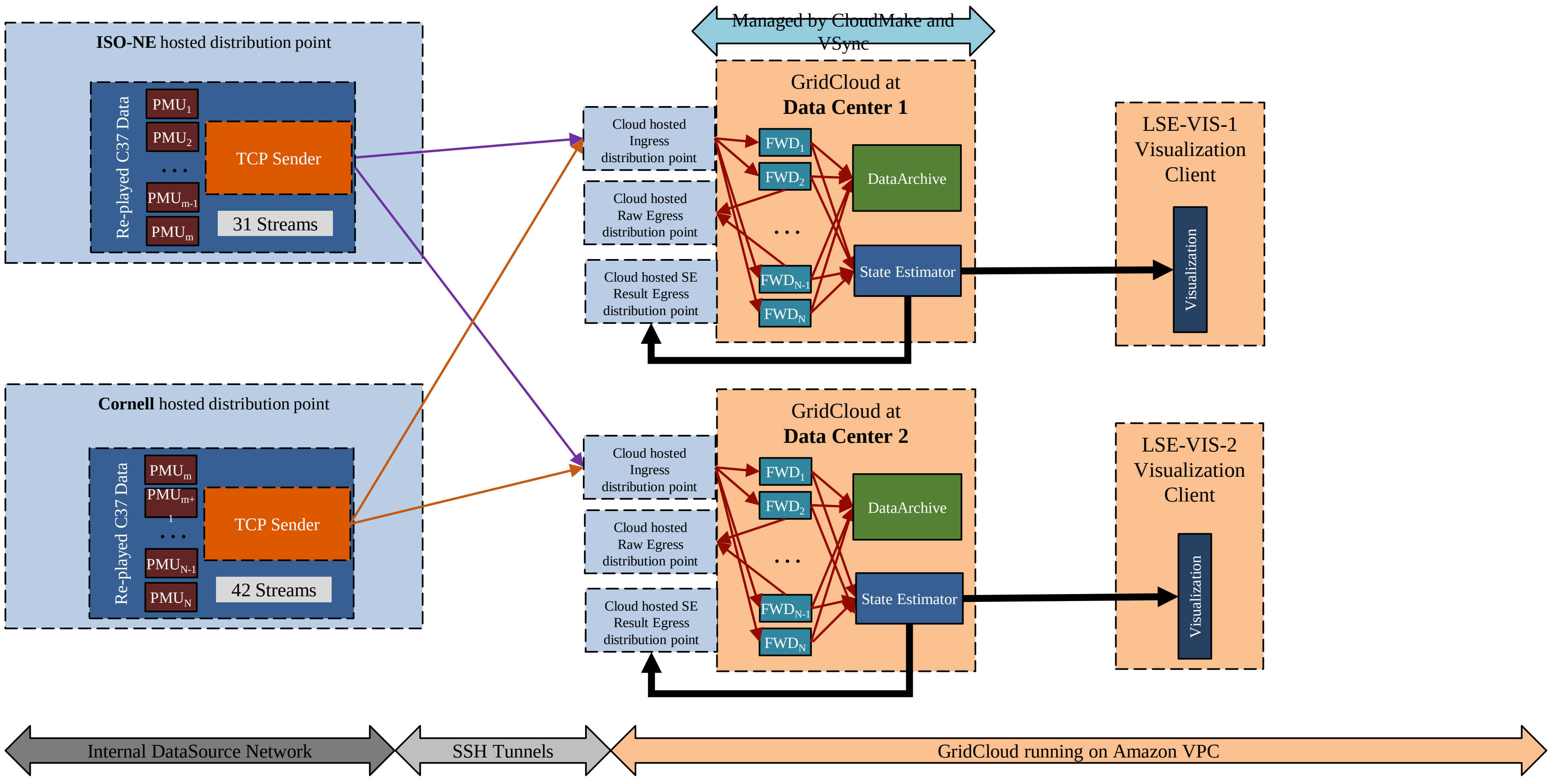


ISO-NE SYSTEM

New England System

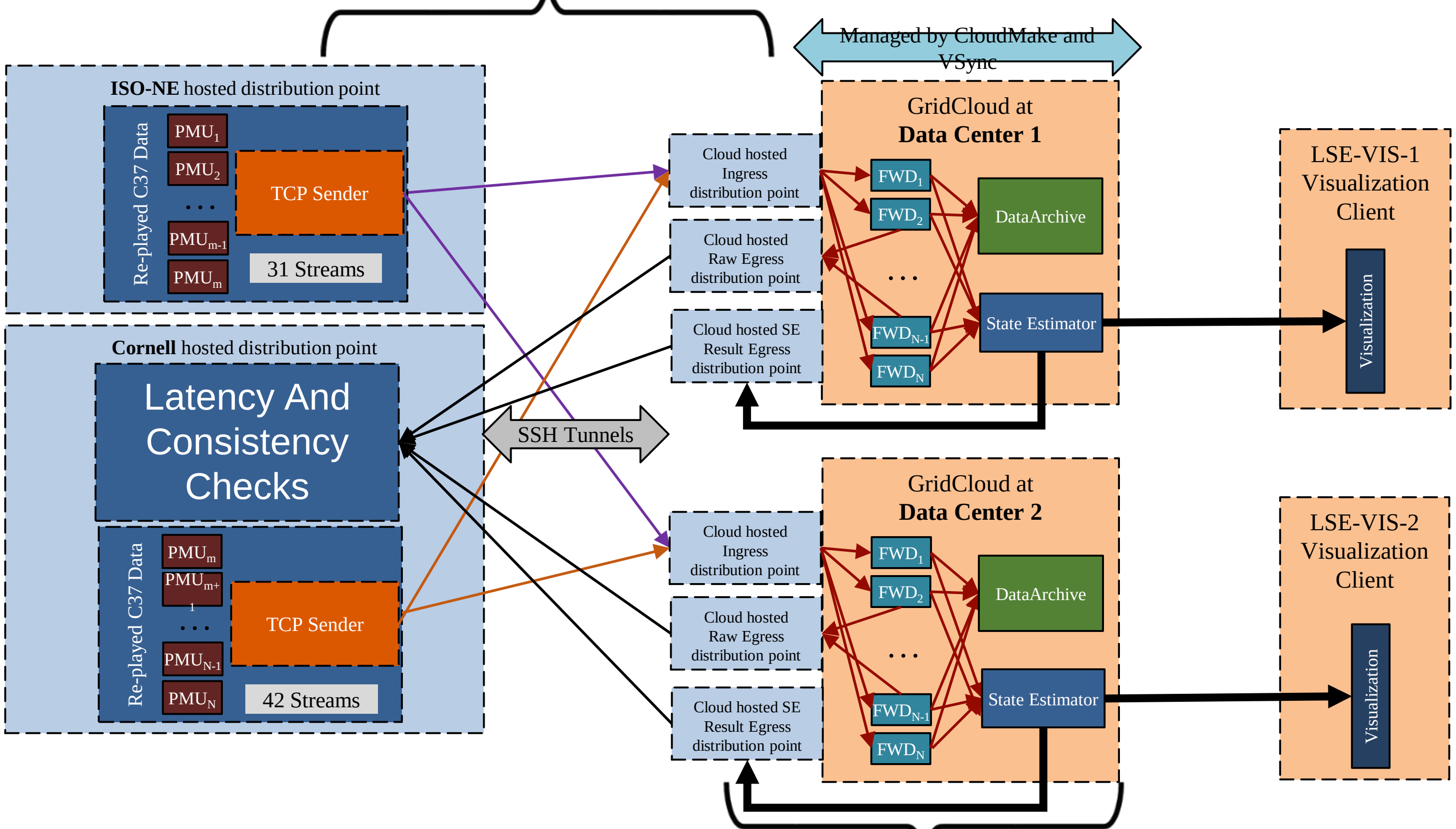
- 761 buses (planning model)
- 73 PMUs
  - 96 voltage phasors
  - 127 current phasors
- 93 buses observable (including all 345kV)
- 11 seconds of recorded real time data
  - PMU data @ 30Hz
  - PMU data is run in a loop to obtain longer runs
- LSE solution @ 5Hz
  - Returned as C37.118 data stream

ISO-NE Demonstration



ISO-NE Demonstration Monitoring

L3raw: Round trip from datasource to CloudRelay to datasource



L2: Round trip from Ingres Relay to Egress relay

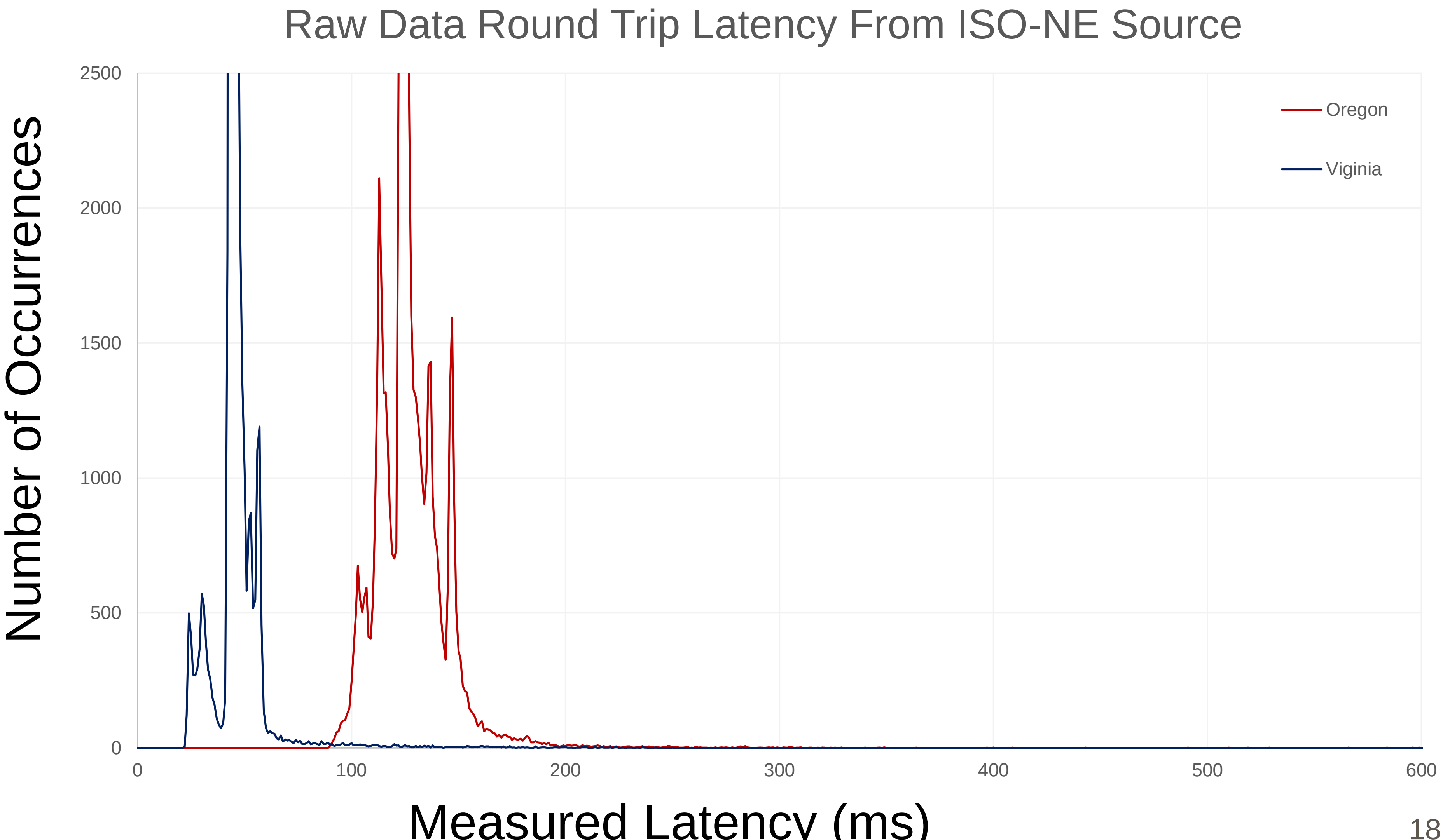
L3se: Round trip from datasource to SE to datasource

Performance
L2 and L3 Latency Tests

- Sampled over 4 hours
- Tests performed from Cornell and ISO-NE datasource machines over SSH tunnels
- Sampled 4 raw feeds and two SE feeds from each datacenter
  - Lowest numbered PMU from each datasource (ISO-NE and Cornell)
  - Highest numbered PMU from each datasource
    - PMUs send to the cloud in order from the datasource; this helps show us the spread of data from first to last measurement sent per round
  - Lowest and Highest latency SE result
- Tests presented in the following slides as histograms and table of overall statistics
  - Histograms only cover highest numbered PMU/SE as they have the highest variability

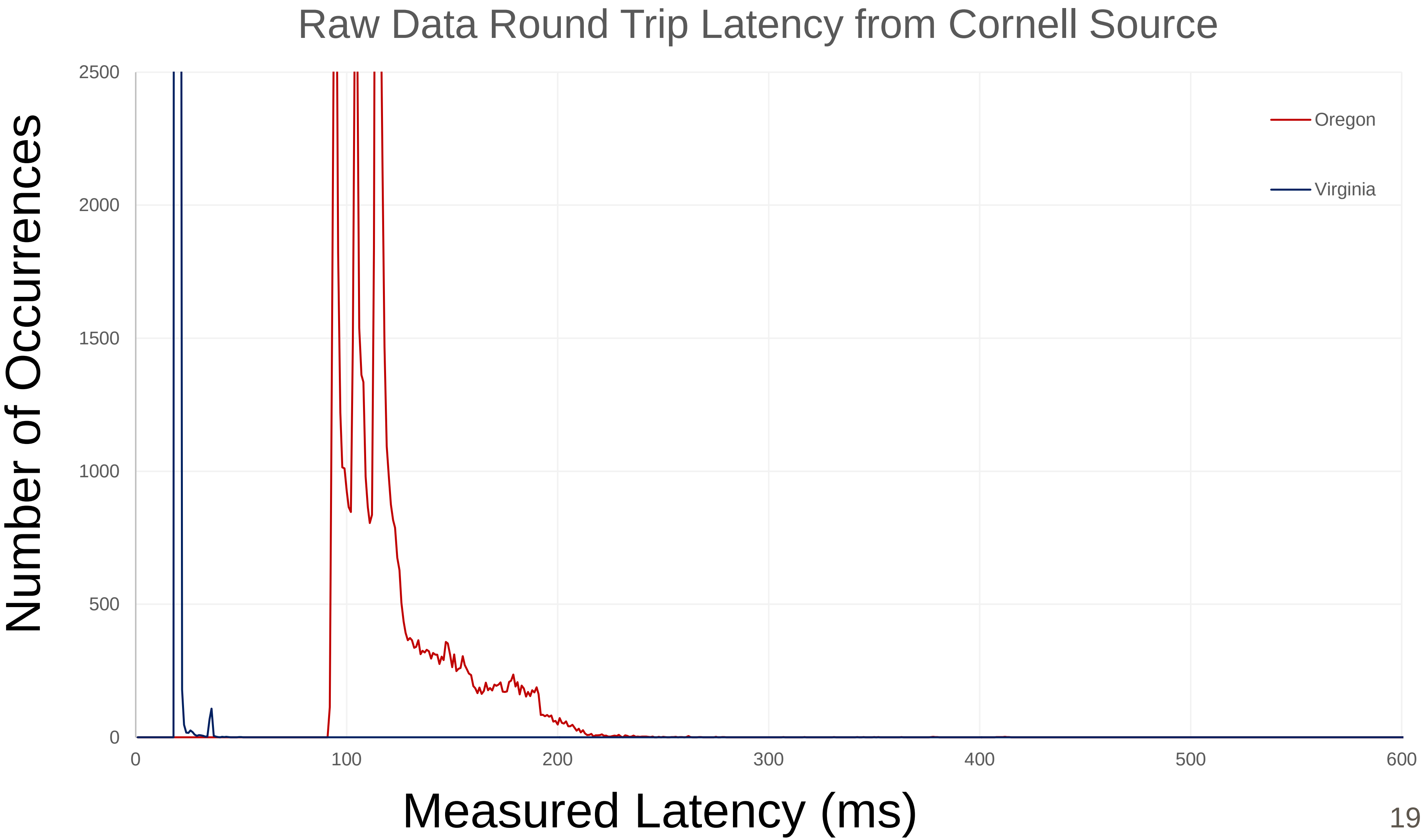
Histogram: Round Trip Latencies

Graphs: Number of times a particular latency occurs



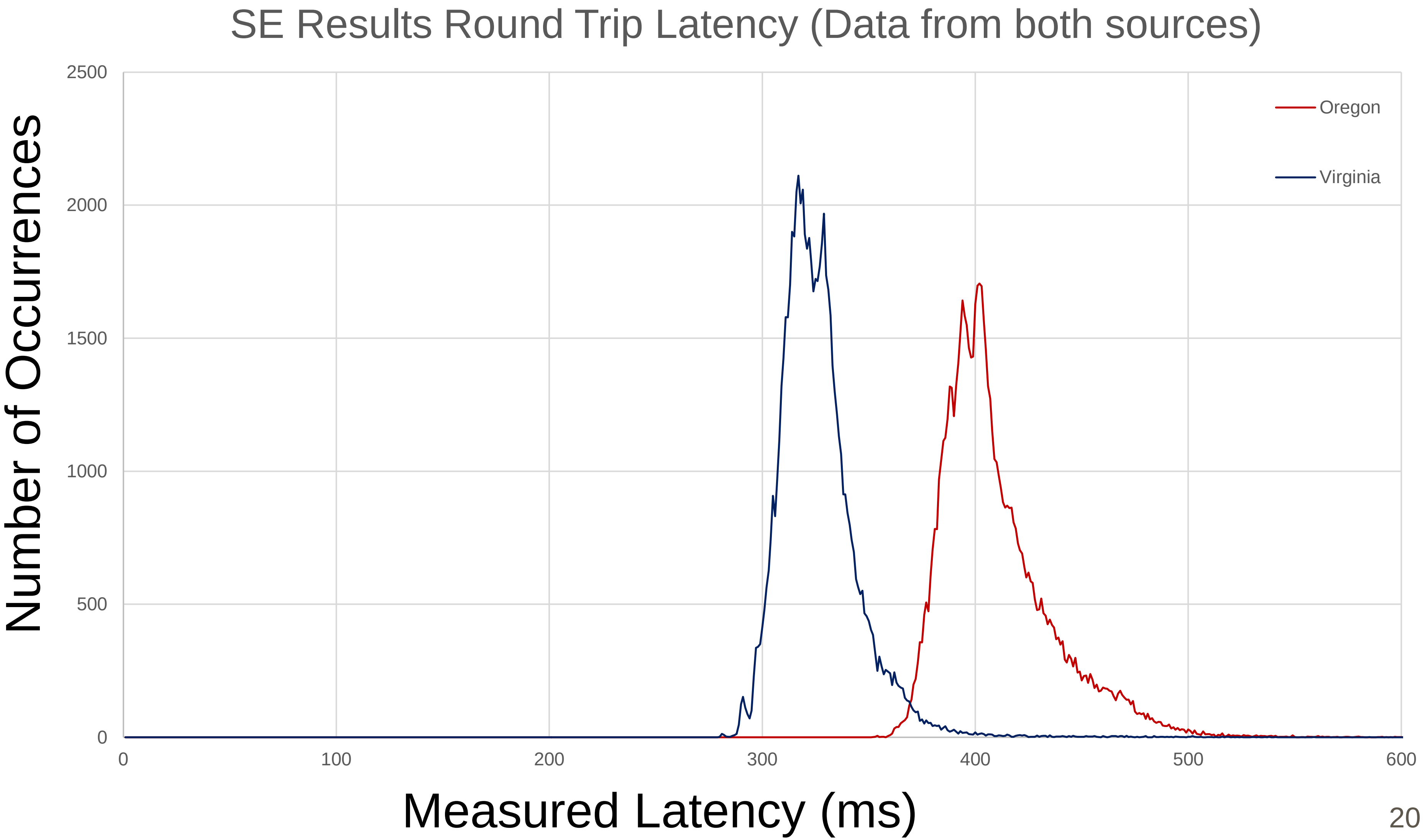
Histogram: Round Trip Latencies

Graphs: Number of times a particular latency occurs



Histogram: Round Trip Latencies

Graphs: Number of times a particular latency occurs



Latencies (milliseconds)

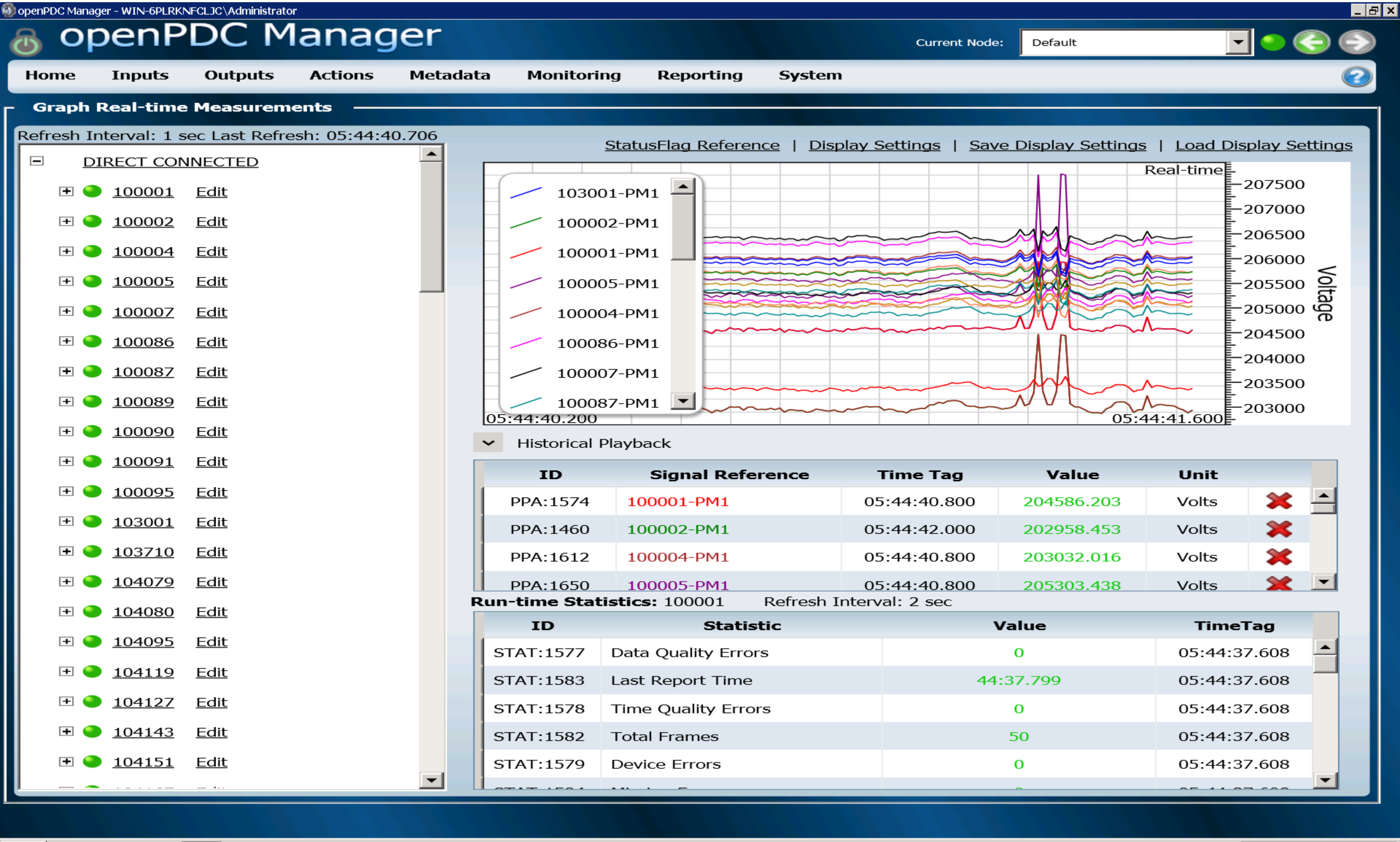
	Virginia	Virginia-Internal	Oregon	Oregon-Internal
ISONE Raw-Low Min	20		88	
ISONE Raw-Low 1st Percentile	22		89	
ISONE Raw-Low Average	**25**		**102**	
ISONE Raw-Low 99th Percentile	58		152	
ISONE Raw-Low Max	611		696	
ISONE Raw-High Min	22		90	
ISONE Raw-High 1st Percentile	25		99	
ISONE Raw-High Average	**46**		**127**	
ISONE Raw-High 99th Percentile	82		179	
ISONE Raw-High Max	612		697	
Cornell Raw-Low Min	17		90	
Cornell Raw-Low 1st Percentile	17		91	
Cornell Raw-Low Average	**18**		**115**	
Cornell Raw-Low 99th Percentile	20		191	
Cornell Raw-Low Max	49		407	
Cornell Raw-High Min	18		91	
Cornell Raw-High 1st Percentile	18		92	
Cornell Raw-High Average	**19**		**120**	
Cornell Raw-High 99th Percentile	20		199	
Cornell Raw-High Max	49		413	
SE Results Min	279	242	351	240
SE Results 1st Percentile	294	267	370	273
SE Results Average	**325**	**300**	**409**	**317**
SE Results 99th Percentile	384	348	490	393
SE Results Max	911	469	962	642

Latencies (milliseconds)

	Virginia	Virginia-Internal	Oregon	Oregon-Internal
ISONE Raw-Low Min	20		88	
ISONE Raw-Low 1st Percentile	22		89	
ISONE Raw-Low Average	**25**		**102**	
ISONE Raw-Low 99th Percentile	58		152	
ISONE Raw-Low Max	611		696	
ISONE Raw-High Min	22		90	
ISONE Raw-High 1st Percentile	25		99	
ISONE Raw-High Average	**46**		**127**	
ISONE Raw-High 99th Percentile	82		179	
ISONE Raw-High Max	612		697	
Cornell Raw-Low Min	17		90	
Cornell Raw-Low 1st Percentile	17		91	
Cornell Raw-Low Average	**18**		**115**	
Cornell Raw-Low 99th Percentile	20		191	
Cornell Raw-Low Max	49		407	
Cornell Raw-High Min	18		91	
Cornell Raw-High 1st Percentile	18		92	
Cornell Raw-High Average	**19**		**120**	
Cornell Raw-High 99th Percentile	20		199	
Cornell Raw-High Max	49		413	
SE Results Min	279	242	351	240
SE Results 1st Percentile	294	267	370	273
SE Results Average	**325**	**300**	**409**	**317**
SE Results 99th Percentile	384	348	490	393
SE Results Max	911	469	962	642

L3 Raw (Virginia, Raw rows)
L3 Raw (Oregon, Raw rows)
L3 SE (Virginia, SE rows)
L2 (Virginia-Internal, SE rows)
L3 SE (Oregon, SE rows)
L2 (Oregon-Internal, SE rows)

OpenPDC Manager (Visualizer) Displaying SE Results



Cyber-security Performance Cost

- EC2 Latency
  - Average = 245ms
  - 1st Percentile = 211ms
  - 99th Percentile = 255ms
- VPC Latency
  - Average = 261ms
  - 1st Percentile = 228ms
  - 99th Percentile = 270ms
- Delta is approximately +15ms
- These numbers (L1 latencies) do not include SE compute time (75ms-100ms)
- Adding SSH tunnels added less than 2ms to RTT

Main Findings

- **Cost:** As configured for testing
  - 13 AWS instances total per datacenter (Vizualizer, CloudRelay, CloudMakeLeader, StateEstimator, 3xRawArchiver, 4xSEArchiver, 2xForwader)
  - $2.47/hr to run per datacenter
- **Latency:** Round-trip time including LSE solution on an eastern data center was 300ms; on the western data center was 500ms
- **Consistency:** Returned raw data and LSE results from the two data centers were identical
- **Security Effect on Latency:** Cost of AES256 encryption at noise level; cost of SSH 2ms; data loss & delays were not observed and did not affect latency
- **Fault Tolerance:** Loss of one data center did not impact results from other data center. Restart of lost data center took 175sec

Additional Platform Features

- Distribute real-time data streams to multiple applications in the cloud
- Freeze-Frame File System (FFFS)
  - Distributed, time-consistent snapshots of stored data
  - Tamper-proof data
- CloudMake
  - Declarative specification of GridCloud components, their interconnection and the cloud resources that they use
  - Automated instantiation, monitoring, and repair of GridCloud components when instances or communication fail

For Further Investigation

- Flexibility to incorporate multiple entities (actual sharing)
  - Naming and configuration for sharing
  - Cyber-security
- Recording time-synchronized system topology along with PMU data (FFFS should help)
- New project starting with NYPA to investigate these and other issues

Questions?

Carl Hauser

(hauser@eecs.wsu.edu)