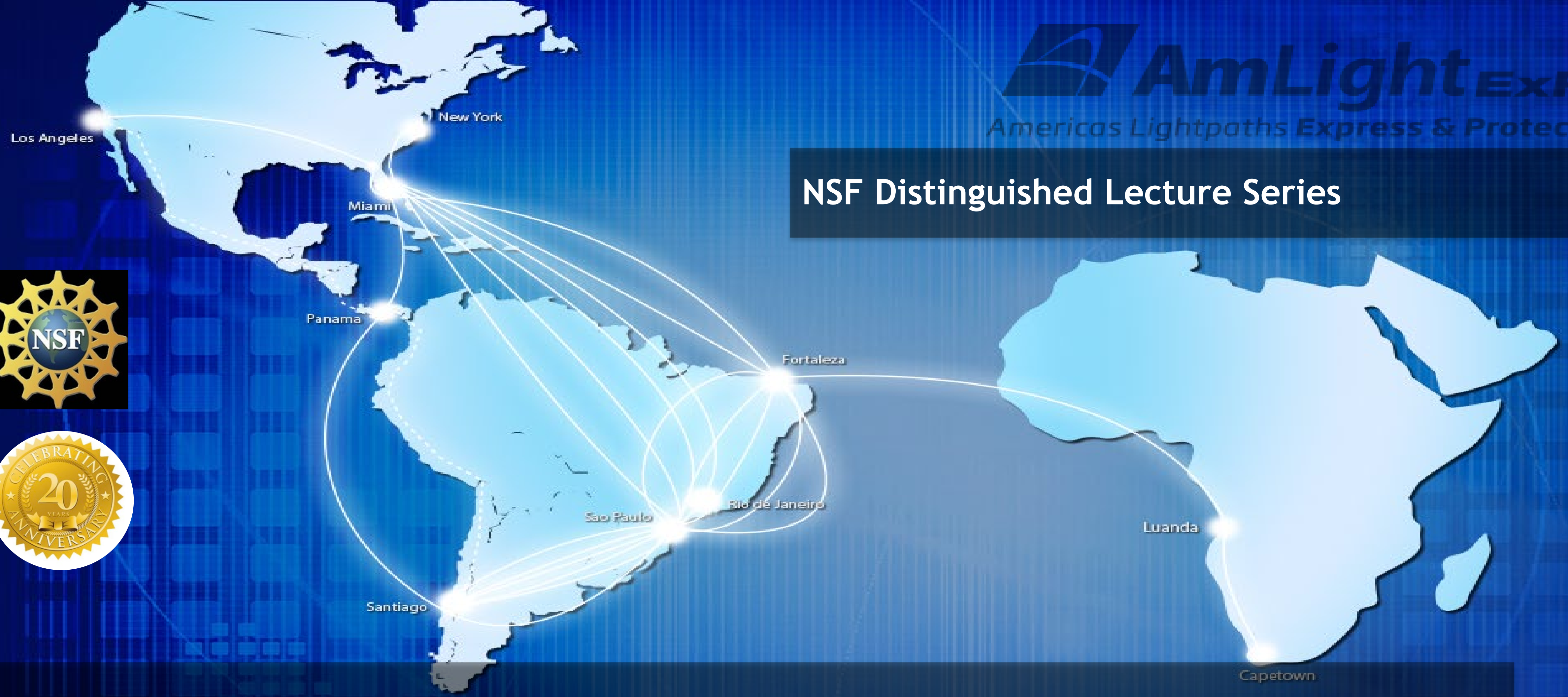
NSF Distinguished Lecture Series



AmLight Express and Protect (AmLight-ExP):

An international production network and platform for network innovation, supporting research and education

January 13, 2022

Julio Ibarra

Principal Investigator

Outline

- Introduction: About AmLight
- International Production Research & Education Network
- Platform for network innovation
- Supporting Science

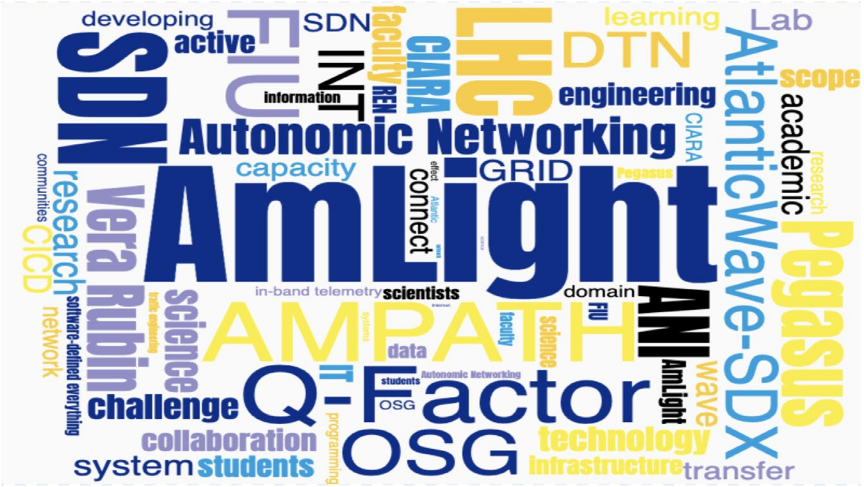
Center for Internet Augmented Research and Assessment (CIARA)

- CIARA supports and conducts research and education through the application of advanced Cyberinfrastructure

- Bridges the technology gaps between researchers and IT practitioners
  - Division of IT
  - College of Engineering and Computing

- Invigorates scholarship for undergraduate and graduate students

- CIARA aligns with FIU's goals as a public research university, contributing to its research, scholarship, and technology development by
  - Advancing international research and education network-dependent collaborations



About AmLight

- Established in 2010 under IRNC award, OAC-0963053
  - Consists of a 20-year buildout, that includes
    - Connections to the R&E networks in Latin America
    - The AMPATH International Exchange Point in 2000
    - Accomplishments of the WHREN-LILA project, IRNC award OAC-0441095
- One of the first to use optical spectrum, combined with leased bandwidth capacity on its backbone
  - Established long-term leases until 2032
- One of the first to deploy and operate its production network with Software-Defined Networking (SDN), since 2014
  - Enabled dynamic service provisioning
  - Significantly increased operations efficiency
- Established the South American Astronomy Coordination Committee (SAACC)
  - SAACC provides a venue for the exchange of information and coordination between the U.S. astronomy projects in Chile and the AmLight network operators
  - 2021 SAACC meeting report https://www.amlight.net/?p=4467

Key Factors for Success

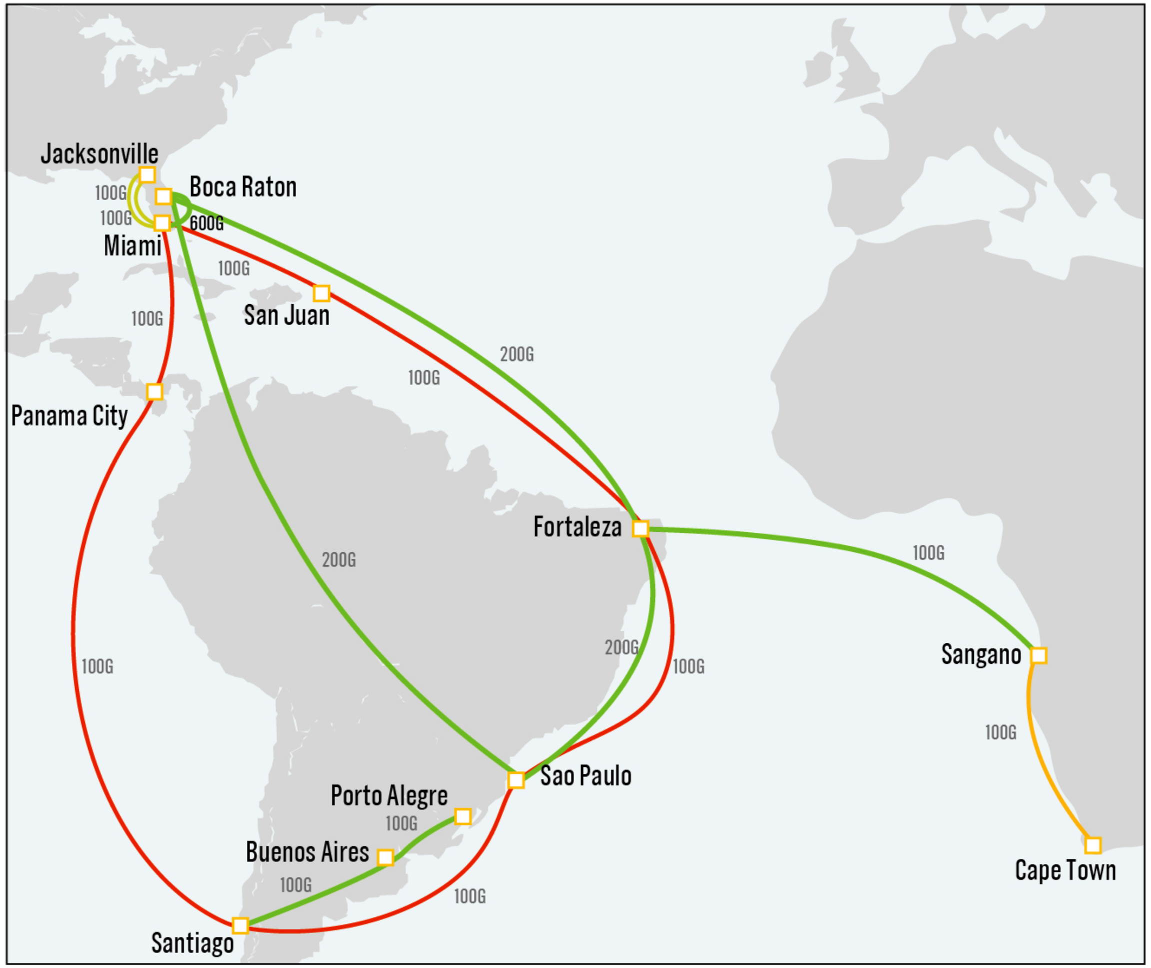
- Support from NSF, OAC, and the IRNC program
- Support from FIU
- Partnerships with R&E networks in the U.S., Latin America, Caribbean and Africa, built upon
  - Layers of trust and openness by sharing
    - Operations resources
      - Network bandwidth, colocation facilities, network and compute resources
    - Human resources
      - Collaboration and cooperation among some of the most talented network engineers in the global R&E networking community

Outline

- Introduction
- International Production Research & Education Network
- Platform for network innovation
- Supporting Science

AmLight Express and Protect (AmLight-ExP) Network, NSF OAC-2029283

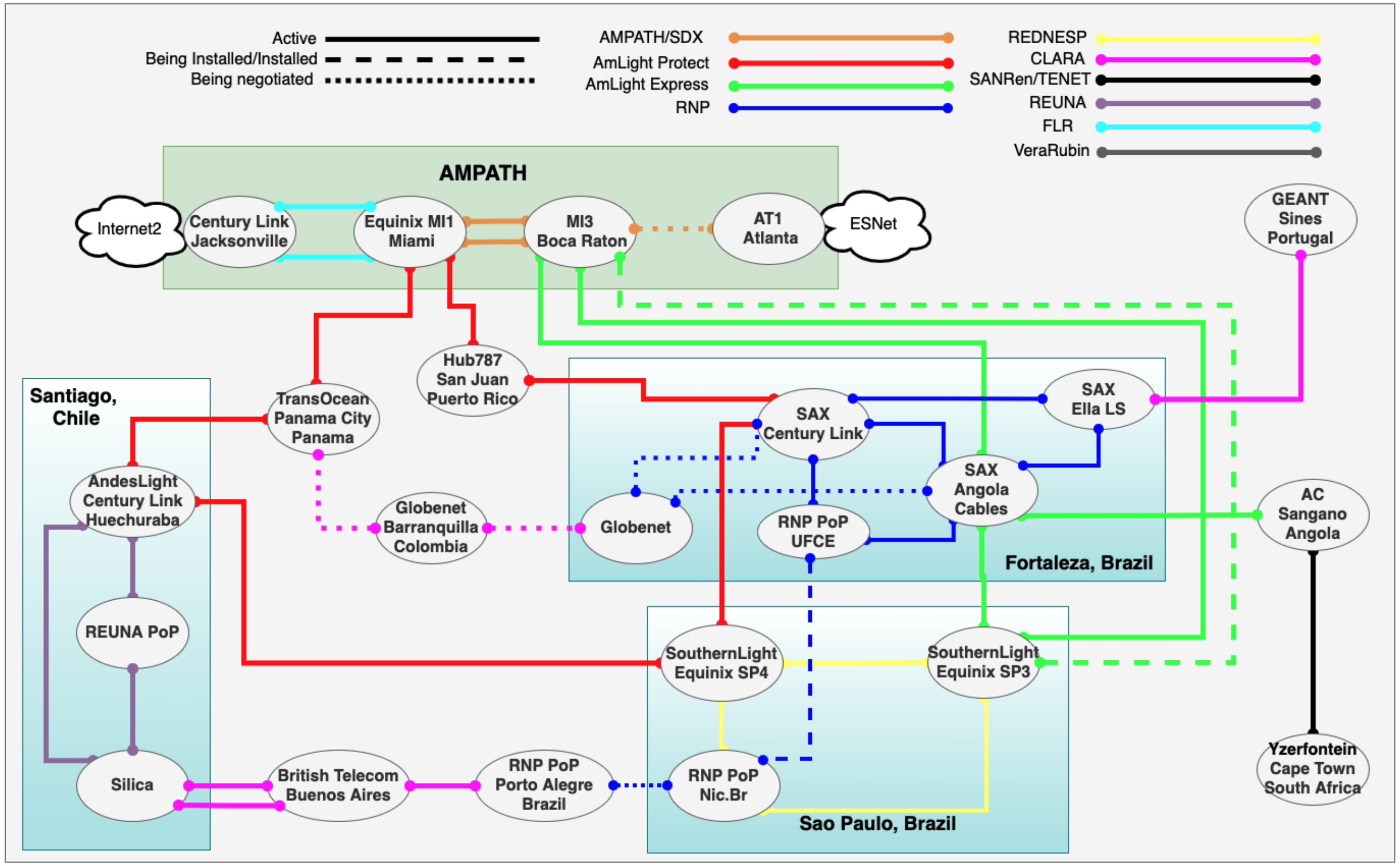
- AmLight Express network (green), Spectrum:
  - 200G Boca Raton to Sao Paulo
  - 200G Boca Raton to Fortaleza
  - 200G Sao Paulo to Fortaleza
  - 100G Boca Raton to Cape Town
  - 100G Santiago to Porto Alegre
- 100G AmLight Protect ring (solid red), Leased capacity:
  - Miami-Fortaleza, Fortaleza-Sao Paulo, Sao Paulo-Santiago, Santiago-Panama, Panama-San Juan, and San Juan-Miami
- 600Gbps of upstream aggregate capacity
- Open Exchange Points: Miami, Fortaleza, Sao Paulo, Santiago, Cape Town



AmLight ExP
Americas Lightpaths Express & Protect

Americas-Africa Lightpaths Express and Protect (AmLight-ExP)

Increasing capacity and adding network paths to increase resiliency



AmLight ExP
Americas Lightpaths Express & Protect
(NSF Award # OAC-2029283)

Outline

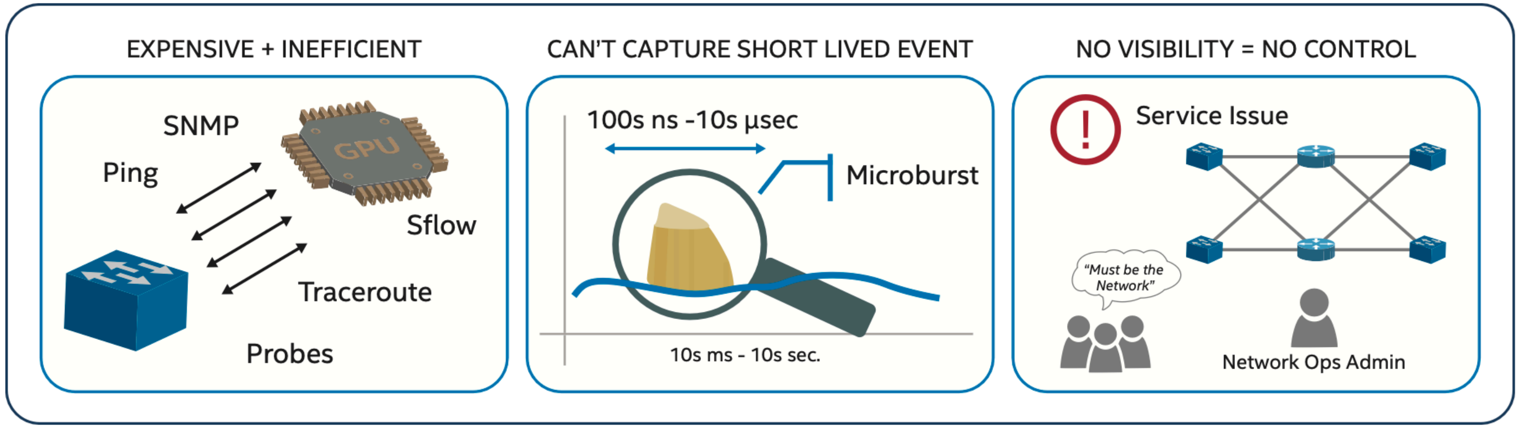
- Introduction
- International Production Research & Education Network
  - In-Band Network Telemetry
- Platform for network innovation
- Supporting Science

Challenge

- Isolating and detecting faults of data transfers in long-haul networks with high latency, such as AmLight, is complex and time consuming

- Detecting what events cause performance degradation often result in questions that have incomplete answers
  - Where is there packet loss and why?
  - Which path did this packet take?
  - How long did this packet queue at each switch?

Challenge: Network Monitoring Pain Points



- Common network monitoring tools fail to detect network transient events
- Network transient events are short-term and sporadic degradations in network performance
  - They are caused by conditions that can lead to failures over time (e.g. attenuation on an optical channel)
  - They often go undetected, such as microbursts
  - They can have a high impact (packet loss) in long-haul networks with high latency, such as AmLight

AmLight ExP
Americas Lightpaths Express & Protect

In-band Network Telemetry (INT)

Creating new methods to see deeper into the phenomena

Adapted from Robertson, D. (2003), and Arthur, W. B. (2009)

In-band Network Telemetry (INT)

- INT records network telemetry information in the packet, while the packet traverses a path between two points in the network
- Telemetry reports are exported directly from the Data Plane, with no impact to the Control Plane
  - *INT tracks/monitors/evaluates EVERY single packet at line rate and in real time*
- Examples of network telemetry information collected
  - Timestamp, ingress port, egress port, queue buffer utilization, sequence #, and many others
- INT enables unprecedented visibility into network states
  - detecting throughput issues due to bottlenecks, failures, or configuration errors

How does In-band Network Telemetry (INT) work?

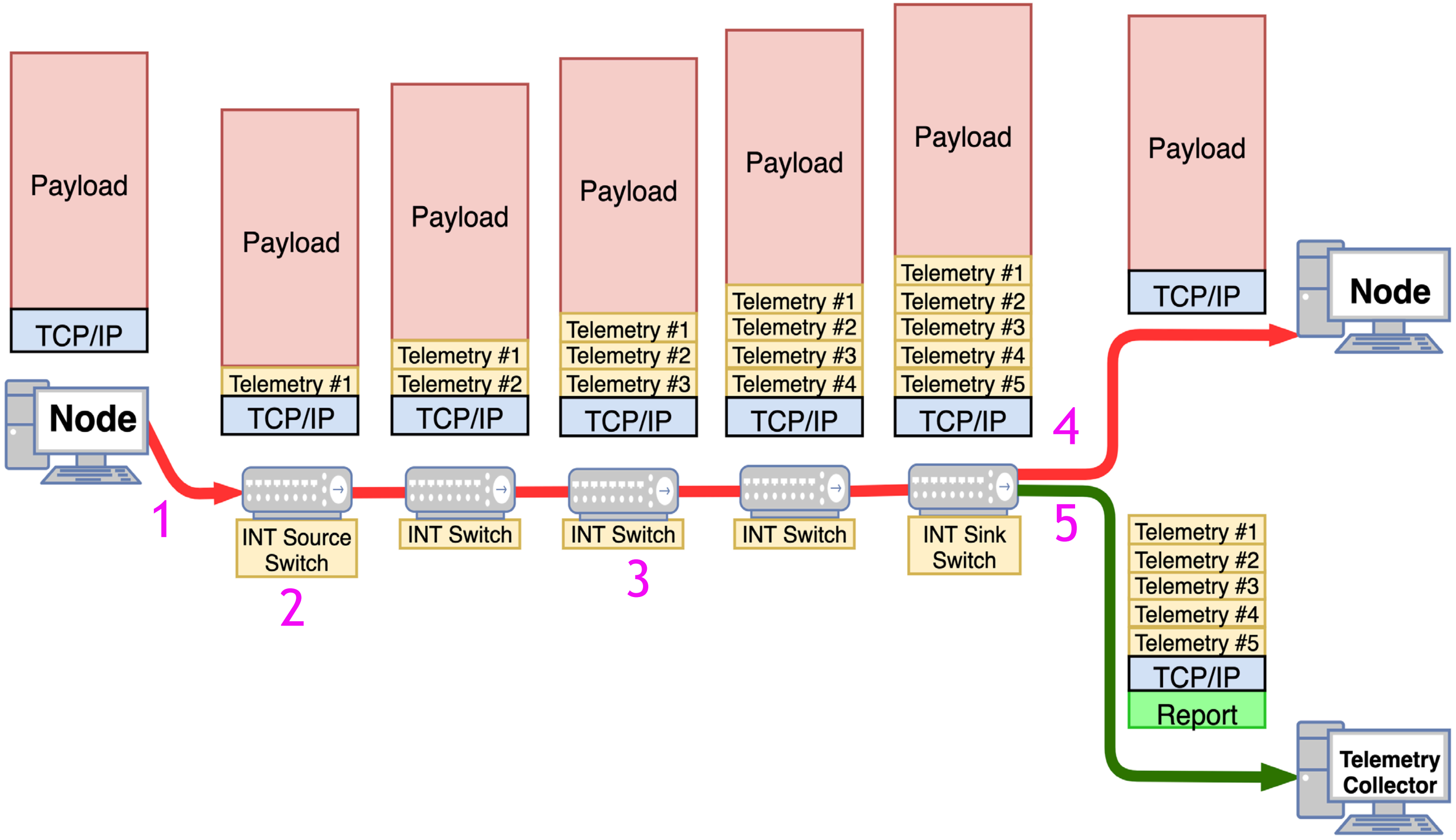
1 – User sends a TCP or UDP packet unaware of INT

2 – First switch (INT Source Switch) pushes an INT header + metadata

3 – Every INT switch pushes its metadata. Non-INT switches just ignore INT content

4 – Last switch (INT Sink Switch) extracts the telemetry, then forwards original packet to the destination node

5 – Last switch (INT Sink Switch) forwards each telemetry report to the Telemetry Collector



INT metadata and telemetry reports

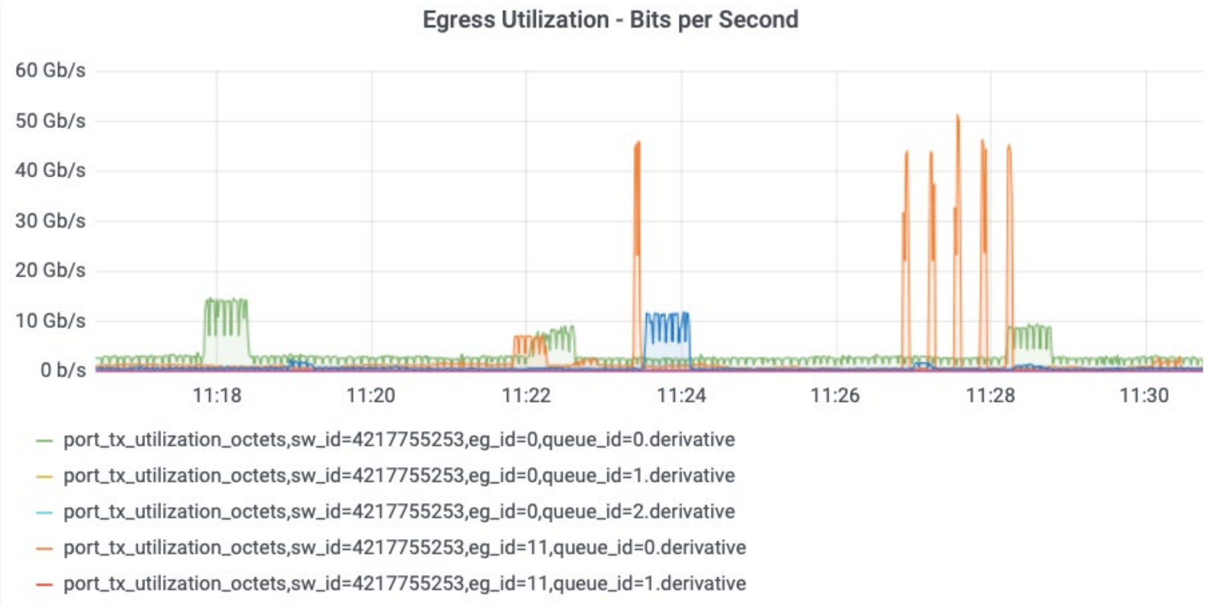
- AmLight INT switches collect the following metadata:
  - Per switch:
    - Switch ID
    - Ingress port
    - Egress port
    - Ingress timestamp
    - Egress timestamp
    - Egress queue ID
    - Egress queue occupancy
  - Per telemetry report:
    - Report timestamp
    - Report sequence number
    - Original TCP/IP headers

Out Time: 123144143 ns	
In Time: 123132143 ns	
Queue: 2	Occ: 15MB
Hop Delay: 12 us	
In: Port 1	Out: Port 2
Switch: 1	
Out Time: 124145243 ns	
In Time: 124144143 ns	
Queue: 0	Occ: 10KB
Hop Delay: 1.1 us	
In: Port 1	Out: Port 4
Switch: 2	
Out Time: 125146343 ns	
In Time: 125145243 ns	
Queue: 0	Occ: 10KB
Hop Delay: 1.1 us	
In: Port 31	Out: Port 28
Switch: 3	
Out Time: 126147443 ns	
In Time: 126146343 ns	
Queue: 0	Occ: 10KB
Hop Delay: 1.1 us	
In: Port 12	Out: Port 13
Switch: 4	
Out Time: 127187443 ns	
In Time: 127147443 ns	
Queue: 0	Occ: 21MB
Hop Delay: 40 us	
In: Port 1	Out: Port 7
Switch: 5	

What INT metadata is being used and how? [1]

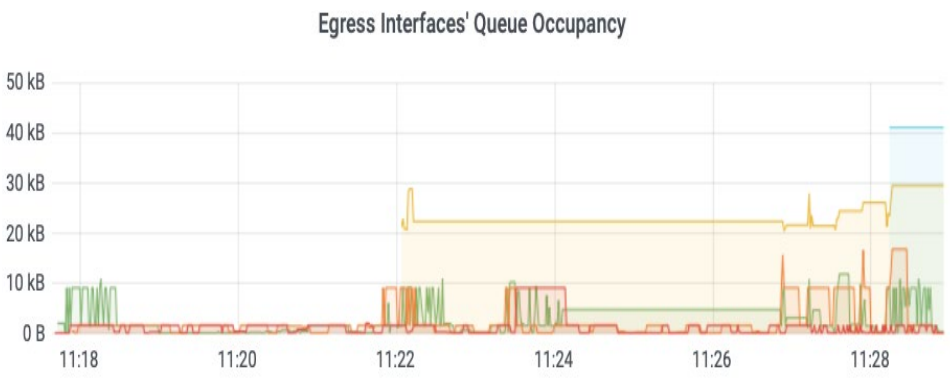
- Instantaneous Ingress and Egress Interface utilization

  - Telemetry Collector monitors and reports egress interface utilization every 100ms
    - Useful for detecting microbursts
    - 100ms can be tuned down if needed
    - Bandwidth monitored per interface & queue

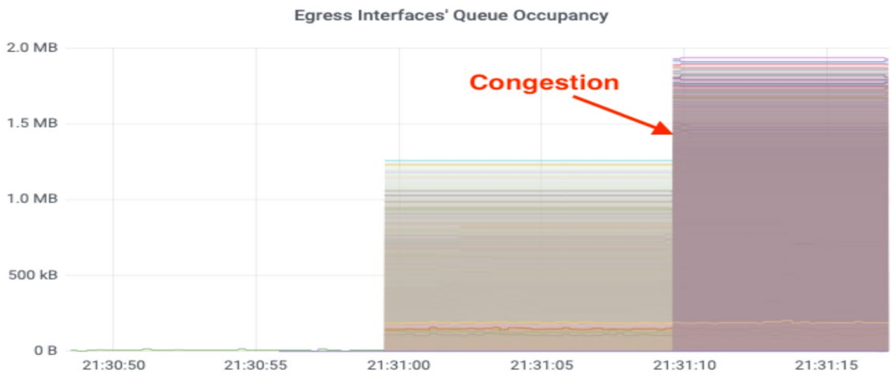


What INT metadata is being used and how? [2]

- Instantaneous Egress Interface Queue utilization (or buffer)
  - Monitoring every queue of every interface of every switch
    - Useful for evaluating QoS policies
    - Useful for detecting sources of packet drops



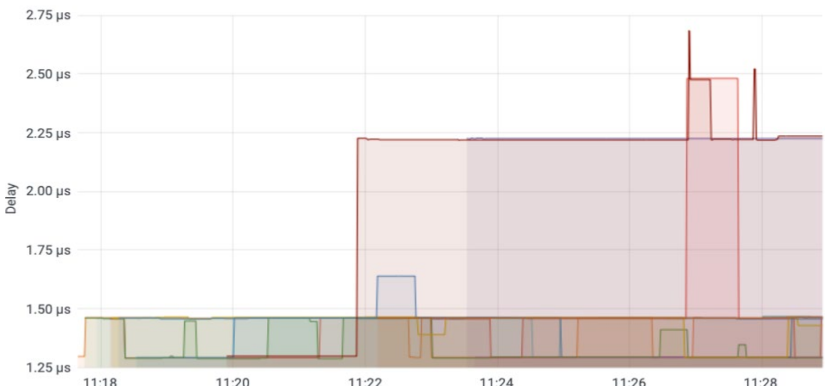
“Normal” Buffer Utilization



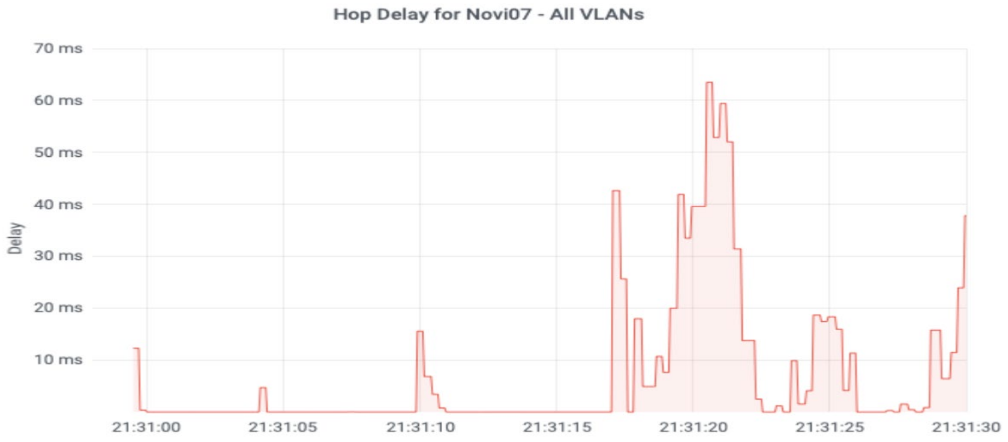
Under-Congestion Buffers

What INT metadata is being used and how? [3]

- Sources of jitter
  - Monitoring per-hop per-packet forwarding delay:
    - Useful for evaluating sources of jitter along the path
    - Useful for mitigating QoS policy issues (under provisioned buffers)
    - Useful for mitigating traffic engineering issues (under and over provisioned links)



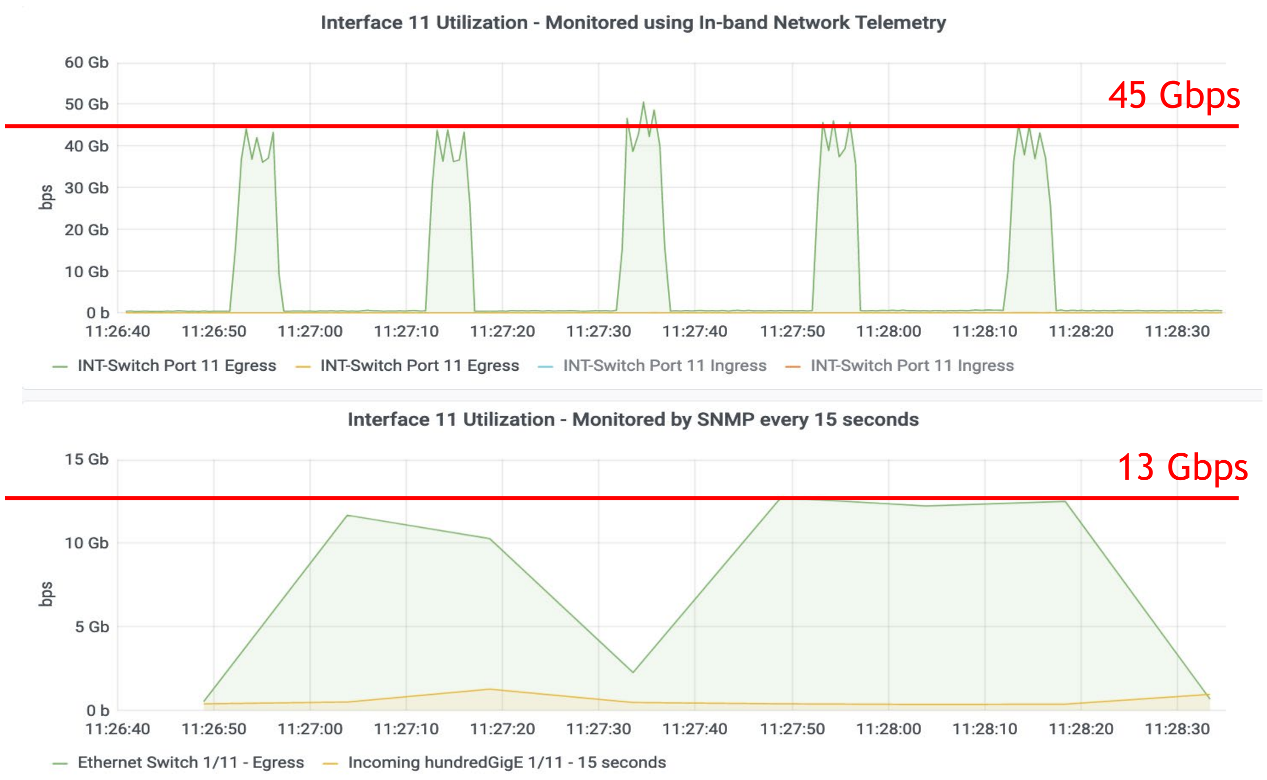
"Normal" Buffer Utilization



Under-Congestion Buffers

Use Case: Observing microbursts

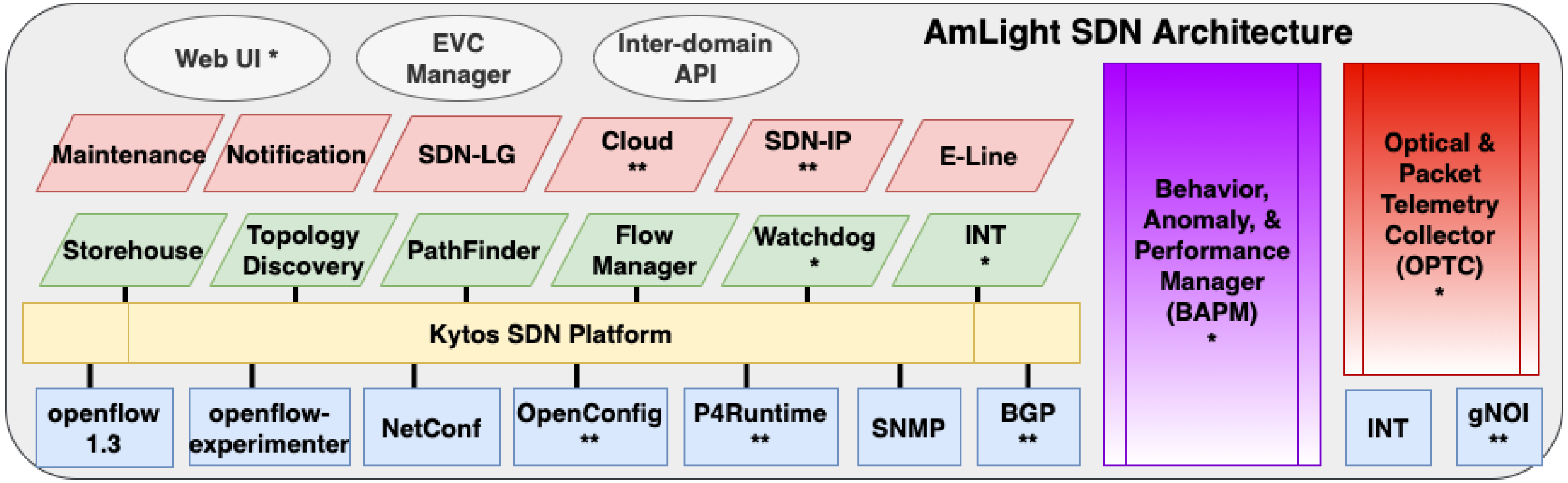
- 5 data transfers/bursts of 40-50Gbps for 5 seconds.
- Top: INT switch, INT metadata exported in real time, per packet
- Bottom: Ethernet switch, SNMP Get running as fast as supported by the switch: 15 seconds
- *By leveraging legacy technologies, such as SNMP, troubleshooting microbursts – malicious or not – is a complex activity that won't be enough to characterize the microburst and determine its impact.*



Outline

- Introduction
- International Production Research & Education Network
- Platform for network innovation
- Supporting Science

AmLight SDN Architecture

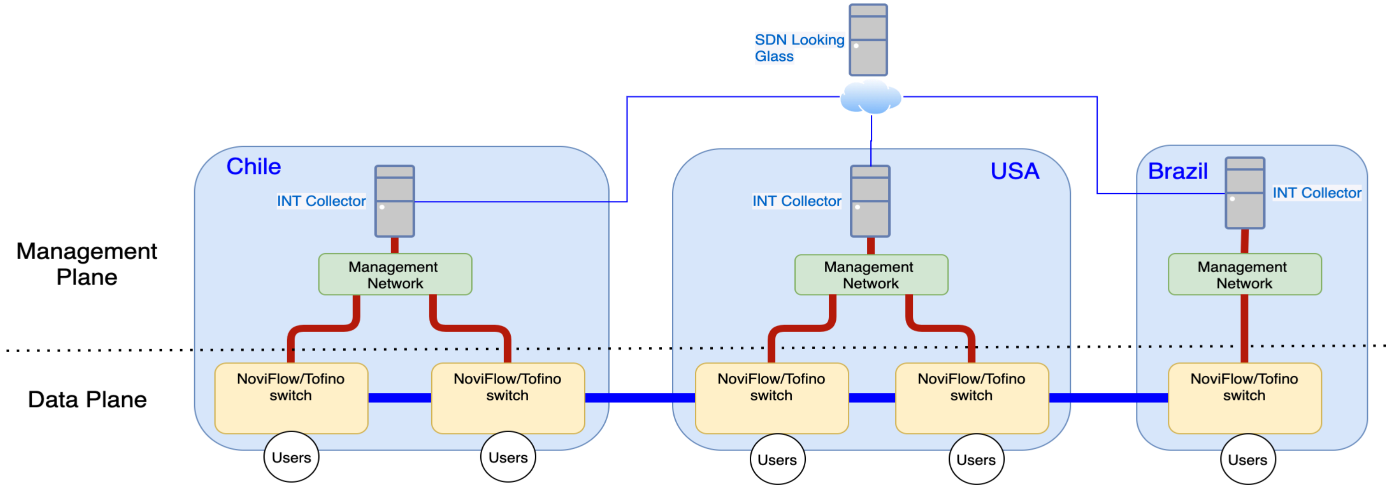


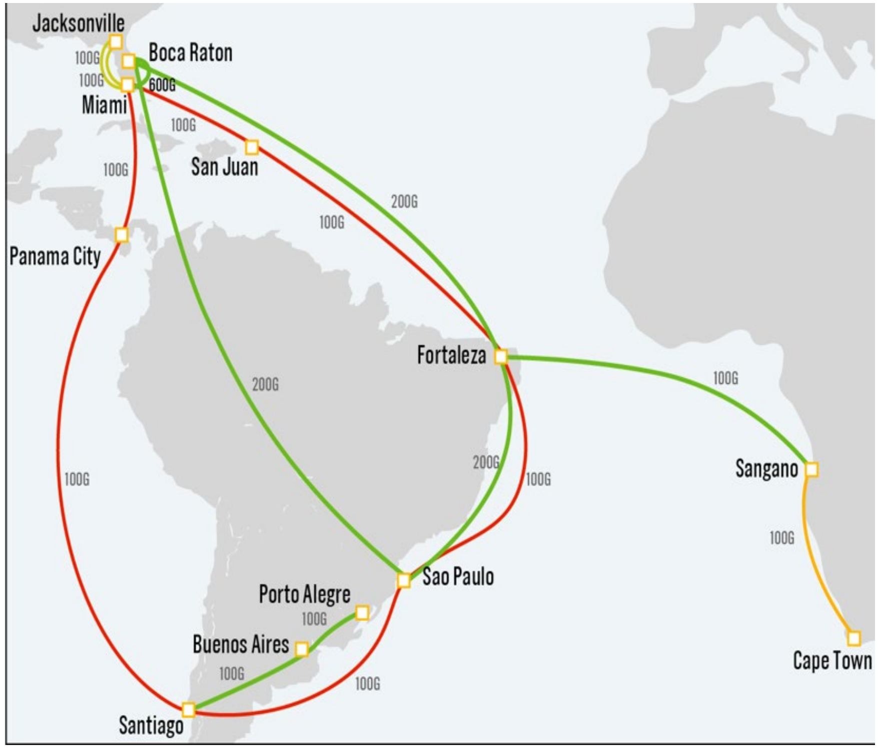
- Blue boxes: Southbound interfaces
- Yellow box: Kytos SDN platform – the core of the architecture
- Green boxes: Kytos' micro applications
- Pink boxes: Business applications
- Ellipses: Applications or interfaces for users to make service requests
- Optical & Packet Telemetry Collector (OPTC)
- Behavior, Anomaly, & Performance Manager (BAPM)

AmLight ExP
Americas Lightpaths Express & Protect

Deployment on AmLight

- Each AmLight site is being instrumented with
  - INT switches, replacing the current data plane
  - A Telemetry Collector to parse Mpps of telemetry reports
  - InfluxDB & Grafana combo to store and display reports
- Goal is for AmLight to be fully INT-capable by Q2/2022





Outline

- Introduction
- International Production Research & Education Network
- Platform for network innovation
  - AmLight SDN Architecture
  - Autonomic Networking
- Supporting Science

Autonomic Networking (Review)

- Autonomic systems were first described in 2001 (Kephart and Chess, 2003)

- Documented in IETF RFC 7575 and other RFCs

- The fundamental goal is self-management, comprised of several self-* properties
  - Reduces dependencies on human administrators or centralized management systems
  - Adapts to a changing environment

- Closed-loop control
  - Mechanism of self-management functions that include Collect, Analyze, Decide, and Act processes
  - AmLight refers to this closed loop control mechanism as Closed-Loop Orchestration

Closed Loop Orchestration

	Automatic	Automation	Closed-Loop Orchestration	Autonomic
Description	User runs a script to change a service or configuration	User runs a “playbook” to change multiple services and to configure multiple nodes at the same time	Orchestrator changes multiple services and node configurations. Nodes export new status and counters. Orchestrator monitors and reacts to the new state, then performs (or not) changes in a closed loop.	Application discovers assets. Configures devices from scratch based on policies and intents. Minimal to no user interaction. Resolution of conflicts defined by administrators
User Input	Scripts, inputs, topology, destination	Scripts, inputs, inventory	Scripts, inputs, inventories, policies/conditions/triggers	Policies and intents

Goal

More Human Interaction

Less Human Interaction

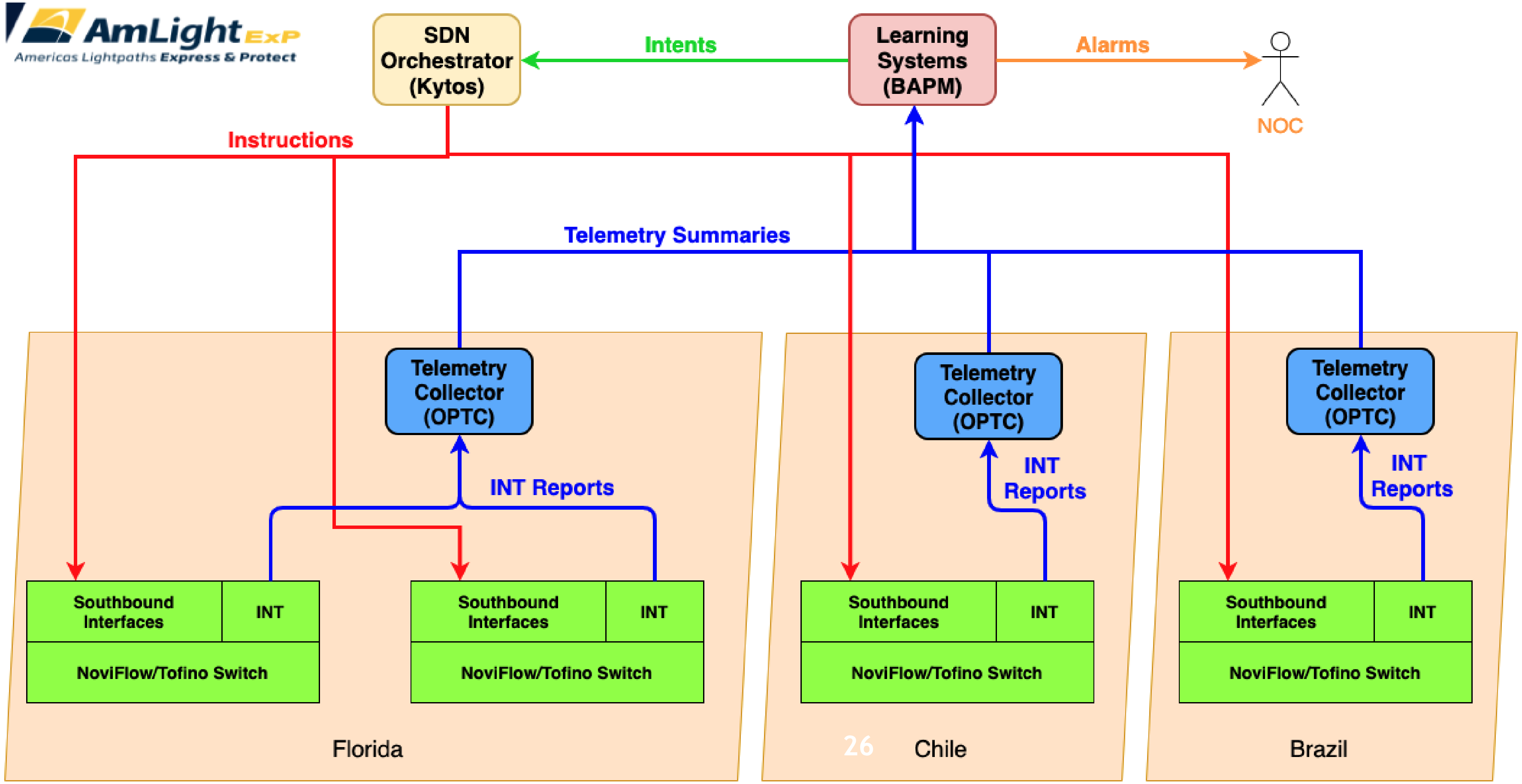
Use Case: Self-optimizing AmLight

Closed-loop network orchestration by

- Processing telemetry reports from the packet and optical layers
- Combined with learning algorithms

Roadmap: Self-Optimizing the network:

- Year 2: < 5 seconds
- Year 3: < 2 seconds
- Year 4: < 1 second
- Year 5: < 500 ms

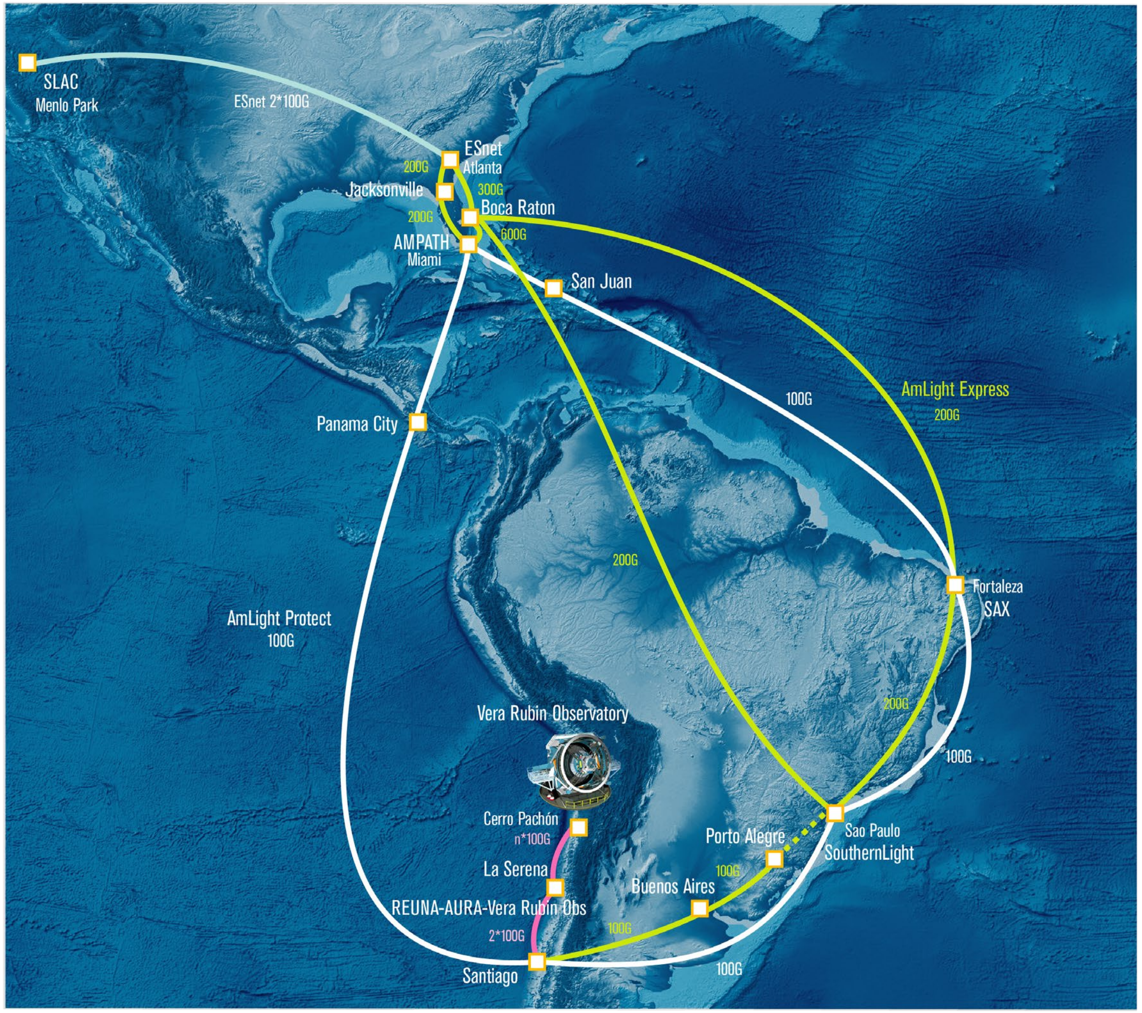


Outline

- Introduction
- International Production Research & Education Network
- Platform for network innovation
- Supporting Science

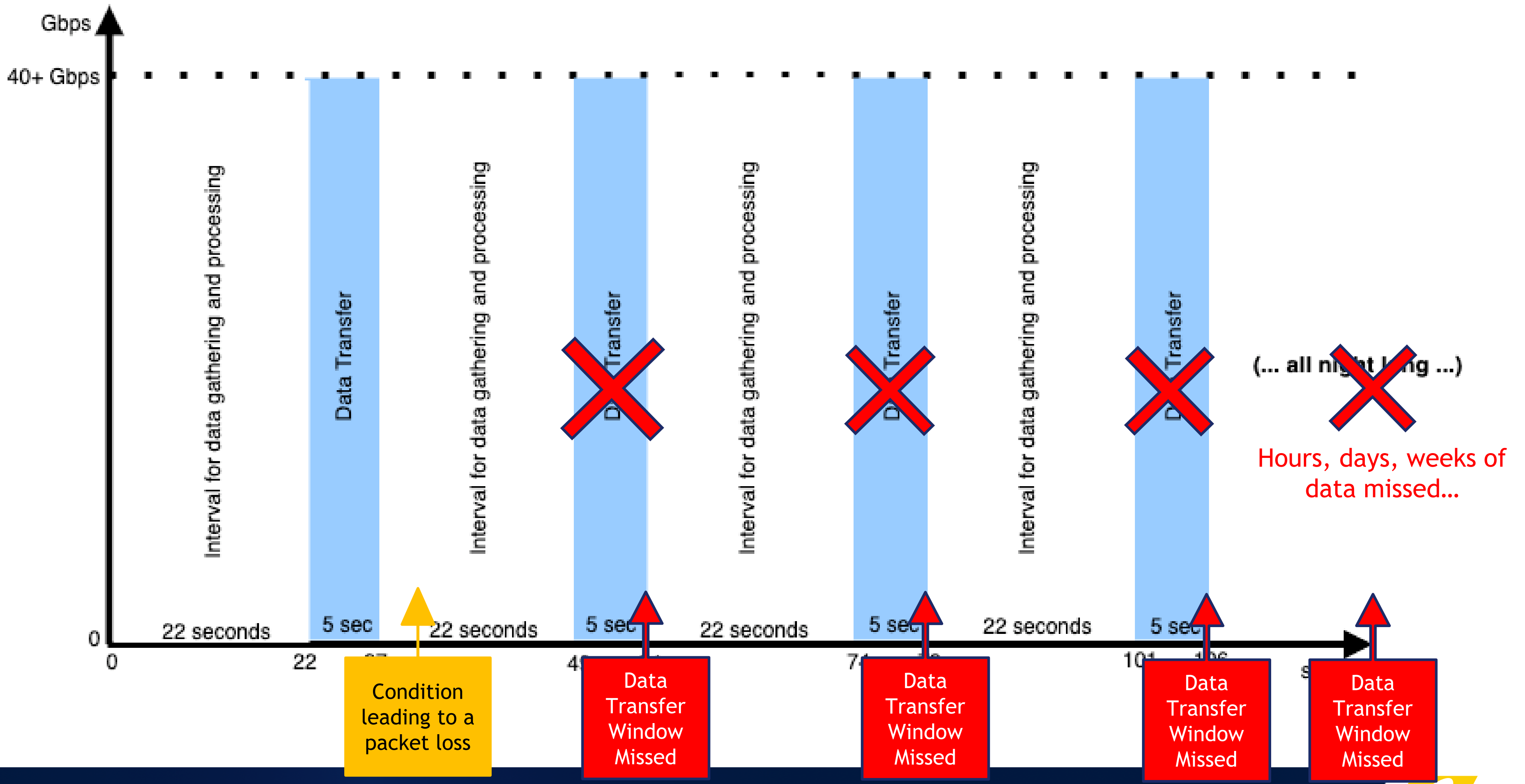
Use Case: Vera Rubin Observatory operation

- Vera Rubin is a large-aperture, wide-field, ground-based optical telescope under construction in northern Chile
- The 8.4 meter telescope will take a picture of the southern sky every 27 seconds, and produce a 13 Gigabyte data set
- Each data set must be transferred to the U.S. Data Facility at SLAC, in Menlo Park, CA, within 5 seconds, inside the 27 second transfer window
- Challenges
  - High propagation delay in the end-to-end path
  - RTT from the Base Station to the USDF is approximately 180+ ms
  - 0.001% of packet loss will compromise the Rubin Observatory application



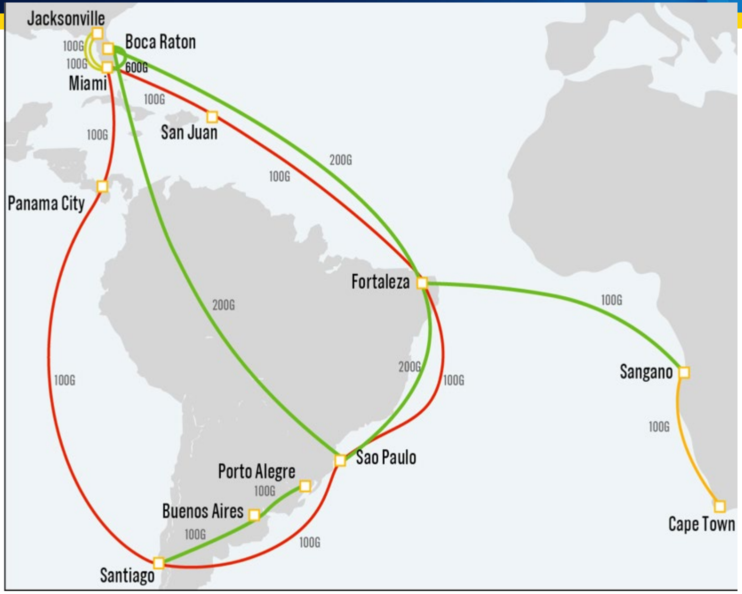
AmLight ExP
Americas Lightpaths Express & Protect

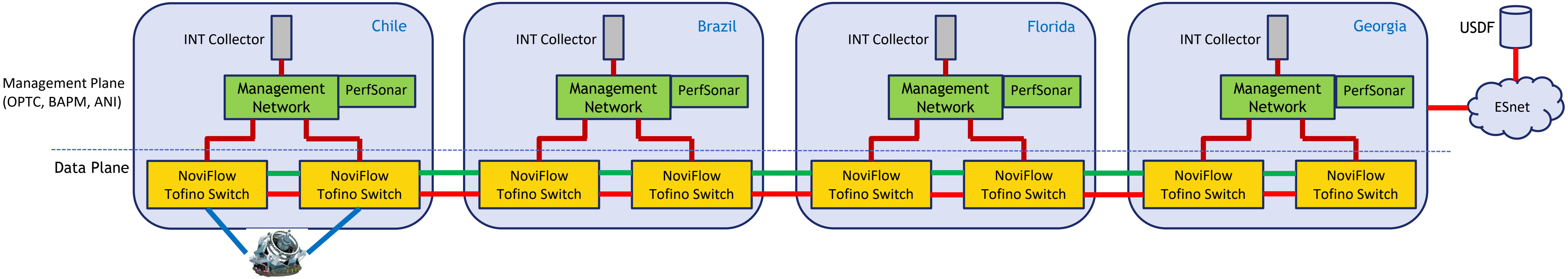
Use Case: Vera Rubin Observatory workflow



Instrumented for SLA-grade network resilience

- AmLight is Instrumented for SLA-grade network resilience to support Vera Rubin
  - Express and Protect paths are instrumented with INT and PerfSonar
- AmLight's Management Plane
  - Processing telemetry report
  - Isolating and detecting traffic anomalies
  - Validating performance thresholds
  - Computing risk profiles of optical and IP layer metrics in a closed loop
  - Reacting to packet loss and packet performance in real-time
- AmLight's metric for success is to not miss a data transfer window

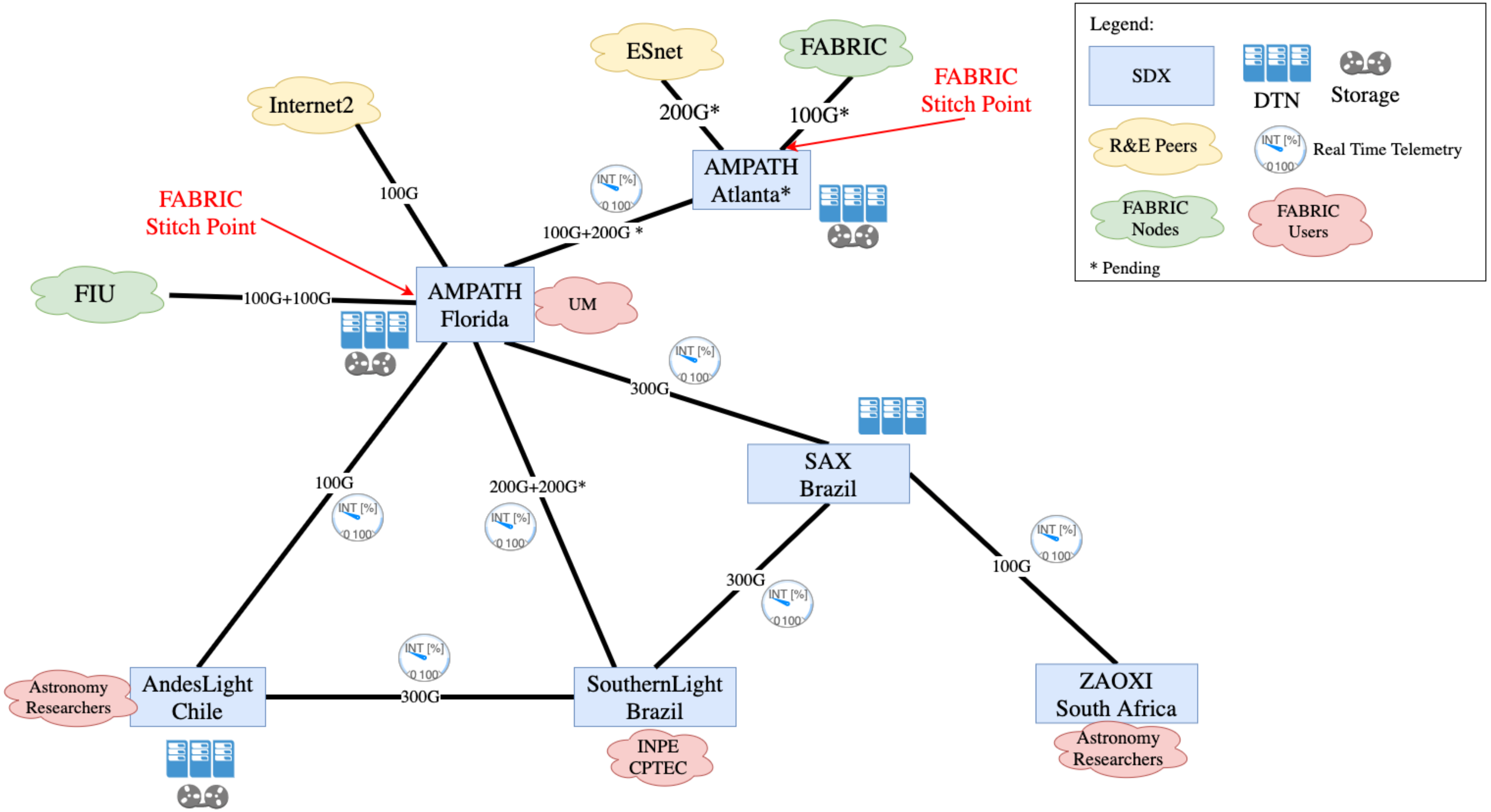




AmLight supports FABRIC

- A dedicated 100G optical path between FIU FABRIC node and Atlanta Core node
- Multiple 100G stitching points:
  - Atlanta (ESnet), Internet2, and AMPATH/Miami (FABRIC node at FIU)
- Up to 50Gbps available over AmLight links during experiments to support reproducibility
- Experiments will have access to per-packet telemetry in real-time

AmLight supports FABRIC [2]



AmLight ExP
Americas Lightpaths Express & Protect

Other science communities supported on AmLight

- Large Hadron Collider Open Network Environment (LHCONE)
- Open Science Grid (OSG)
- Partnership to Advance Throughput Computing (PATh)
- Event Horizon Telescope (EHT)
- Ground-based telescopes in Chile and South Africa

AmLight Team

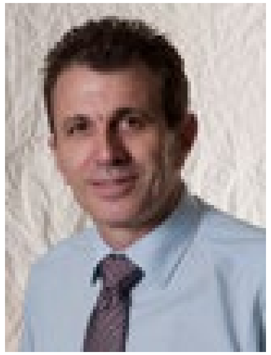
Julio Ibarra
FIU

Jeronimo Bezerra
FIU

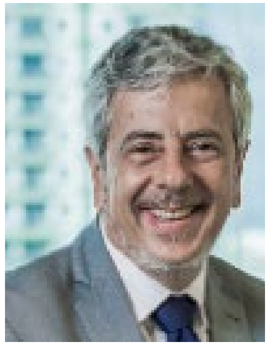
Vasilka Chergarova
FIU

Heidi Morgan
USC-ISI

Chip Cox
Vanderbilt University



Luis Lopez
FIU, USP



Eduardo Grizendi
RNP

Me in one slide

- When, how, and why did you decide to go to pursue a research career?
  - Encouragement from a VP at FIU, and a family member
  - Inspiration from colleagues and team members
  - Motivation from my PhD professor
- Experience was transformational

References

- J. Ibarra et al., "Benefits brought by the use of OpenFlow/SDN on the AmLight intercontinental research and education network," 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM), 2015, pp. 942-947, doi: 10.1109/INM.2015.7140415
- J. Bezerra, “Deploying per-packet telemetry in a long-haul network: the AmLight use case”, INDIS Workshop, 2021
- J. Bezerra, et. al., “In-band Network Telemetry @ AmLight: Our Solution”, Super Computing 2021
- Ghobadi, Monia, and Ratul Mahajan. "Optical layer failures in a large backbone." Proceedings of the 2016 Internet Measurement Conference. 2016.
- “In-band Network Telemetry Detects Network Performance Issues”, Intel White Paper, 2020.
- In-band Network Telemetry (INT) Dataplane Specification, Version 2.1, 2020.
- “Taking the AmLight network to the next level”, White Paper, FIU and NoviFlow, 2020.
- Jeyakumar, Vimalkumar, et al. "Millions of little minions: Using packets for low latency network programming and visibility." ACM SIGCOMM Computer Communication Review 44.4 (2014): 3-14.
- B. Leal, “Using Kytos SDN platform to enhance international big data transfers”, CHEP2018, Sofia, Bulgaria.
- “Software-Defined Networking (SDN): Layers and Architecture Terminology”, RFC 7426, January 2015.
- “Autonomic Networking: Definitions and Design Goals”, RFC 7575, June 2015.
- Kephart, Jeffrey O., and David M. Chess. "The vision of autonomic computing." Computer 36.1 (2003): 41-50.

THANK YOU