MITIGATING SOFT FAILURES USING NETWORK ANALYTICS AND SDN TO SUPPORT DISTRIBUTED BANDWIDTH-INTENSIVE SCIENTIFIC INSTRUMENTS OVER INTERNATIONAL NETWORKS

Jeronimo Bezerra, Julio Ibarra (Florida International University); David Boertjes, Franco Santillo, Lance Williford (CIENA); Heidi Morgan (University of Southern California); Chip Cox (Vanderbilt University); and Luiz Lopez (University of Sao Paulo).
Email: jbezerra@fiu.edu

Florida International University. 11200 S.W. 8th Street, Miami, FL, USA, 33199.

Abstract: With the consolidation of high-speed networks and worldwide scientific deployments, new experiments are being conducted remotely. The control and data gathering of these bandwidth-intensive mission-critical instruments require a reliable network infrastructure capable of reacting in real-time to soft failures, such as packet loss. To address the mission-critical real-time instruments' Service Level Agreement (SLA), streaming telemetry and data-driven analytics are required. In recent years, the industry has created many open consortiums and specifications, such as OpenConfig and Inband Network Telemetry (INT). As a result, we have new levels of interconnections, interoperation, and disaggregation allowing Software-Defined Networking (SDN) applications to use protocol agnostic, common APIs, Artificial Intelligence and Machine Learning to create reliable and adaptive networks. This paper aims to present the ongoing effort to create an adaptive network infrastructure capable of identifying and isolating soft failures in an automated approach to optimize bandwidth-intensive data transfers. Our approach leverages the most recent solutions offered by the optical and packet layers using SDN and network analytics.

1. INTRODUCTION

Science applications are becoming more data-intensive, generating and moving petabytes of data across wide area networks to data center facilities. These bandwidth-intensive applications are becoming more popular because new experiments are being conducted remotely. The Large Synoptic Survey Telescope (LSST) [1] is an example of a bandwidth-intensive remote-controlled application. Being constructed in Chile, it will move petabytes of data to the U.S. and be remotely monitored in the U.S. As a result, LSST requires a highly available and reliable network infrastructure to support its experiments.

Nowadays, deploying a new network infrastructure to support high bandwidth applications has never been easier. With coherent technologies and highly efficient Digital Signal Processing (DSP), optical Wavelength Dense Modulation (WDM) devices can support terabits per second per fiber. New optical cables are being buried on a daily basis, and new submarine cable systems are being built to connect countries and continents, not just by service providers but also by Content Delivery Networks (CDN). Network users are being offered not just lit services and dark fiber, but alien waves and optical spectrum. For network operators aiming to create a resilient network, initiatives such as TeleGeography's Submarine Cable Map [2], and shared Google Earth (.kmz) files facilitate the selection of different physical paths. To handle network outages, path convergence can leverage functionalities at the optical layer as well as the packet layer in the sub-50

milliseconds time frame. Throughout the world, Research and Education Networks (RENs) and CDNs are leveraging new fibers and coherent technology to support bandwidth-intensive user applications.

However, for these applications, bandwidth and reliability are not enough to guarantee a successful data transfer. Depending on the distance between endpoints, packet loss must not be ignored. For a network with propagation delay up to 5ms, a packet loss rate of 1 out of $1x10^8$ would not impact a data transfer operating at 100Gbps. Using the Mathis equation [3], if the propagation delay were 25ms, the same packet loss would limit a data transfer to 14Gbps. If the propagation delay were 40ms, the data transfer would be constrained to 9Gbps. Mitigating a 1 out of $1x10^8$ packet loss is complex and extraordinarily time-consuming.

Even though it is possible to create an international network that fully operates at the optical layer, because of the number of carriers and technologies involved, RENs and CDNs usually prefer to leverage a combination of solutions using both optical and packet layers and multiple operation techniques, ranging from optical spectrum management and Optical Transport Network (OTN) to MPLS and RSVP for provisioning and path protection. Hard failures, such as fiber cuts, are easily overcome, most of the time in a sub-second time frame.

For network operators supporting bandwidth-intensive applications, soft failures are still a very complex issue. For this paper, soft failures [4] are network issues impacting performance but are not easily detected or, most of the time, issues that do not trigger alarms. A small packet loss rate is considered a soft failure. As soft failures don't necessarily trigger a path convergence operation or an alarm, they can go undetected by Network Management Systems (NMS) for days or even longer.

Time-restricted and real-time bandwidth-intensive data transfer applications cannot achieve their Service Level Agreements (SLAs) when soft failures are present. In these cases, network operators need to instrument the network using multiple passive and active monitoring solutions and active performance measurement tools. Even when a monitoring solution is useful for identifying a soft failure, network operators still need to manually isolate the root cause. It is not unusual for soft failure mitigation activities to take days or even weeks to resolve.

For time-restricted and real time bandwidth-intensive data transfer applications, soft failure mitigation must be driven in an automated approach. This paper aims to present the ongoing effort to create an adaptive network infrastructure capable of identifying and isolating soft failures in an automated approach to optimize bandwidth-intensive data transfers. Our approach leverages the most recent solutions offered by the optical and packet layers using Software-Defined Networking (SDN) and network analytics.

2. THE USE CASE

The Large Synoptic Survey Telescope (LSST) [1] was designed to be remotely monitored from the LSST Headquarters in Arizona, with the telescope located in the Andes mountains of northern Chile, and the data archive site in Illinois. Every 27 seconds, LSST will produce a 13GB data-set that must be transported to the archive site in less than 5 seconds over a wide-area network that extends 8,600 miles. Due to costs and geography, the LSST network was designed to leverage a mix of terrestrial and submarine network infrastructures. The LSST network uses optical spectrum and alien waves over multiple optical networks. An SDN orchestration framework is responsible for the provisioning and monitoring of the network functions.

The LSST network design estimates that packet loss must be less than 0.0001% in order to satisfy the SLA specified. Moreover, a 0.001% packet loss rate over an RTT of 140ms is enough to impact LSST's 5-second transfer window.

3. LIMITATIONS TO MITIGATE SOFT FAILURES IN REAL TIME

Optical layer devices and transponders usually collect telemetry data every few seconds but export reports with minimum, maximum and average results every 15 minutes [5]. This time was defined to enable basic monitoring without overwhelming management and control planes. Also, in many cases, amplifiers and ROADMs are purely analog devices, not being able to provide telemetry information along the path. In case of multi-vendor environments, as some vendors only support proprietary monitoring interfaces, network operators have to use multiple network management solutions in parallel, lacking an end-to-end network visibility.

In the packet layer, setting aside faulty network elements causing packet drops, soft failures are mostly driven by oversubscription and traffic bursts. When one of these issues occurs, network devices need to tail-drop packets due to full buffer utilization or traffic prioritization. The main challenge is understanding if a soft failure was caused by a faulty network element or full buffer utilization. Monitoring a buffer utilization in a sub-second scale is extremely complex using current technology. As network devices need to protect the control plane from overutilization, data plane counters are not updated in the control plane in real time. As a result, it is not unusual to see updates happening after a few seconds [6]. Also, NMSes should avoid polling network devices' counters in a second interval to avoid CPU overutilization. In light of these limitations, detecting the source of a soft failure using current technologies is mostly based on time-series information, which means, not nearly close to real time.

Besides all described per-layer challenges and limitations, NMSes don't usually integrate telemetry data gathering from both optical and packet layers, making it impossible to have multi-layer visibility and event correlation.

4. NEW APPROACHES FOR EXPORTING TELEMETRY DATA FROM OPTICAL AND PACKET LAYERS

With the availability of higher bandwidth for the control and management planes and new monitoring and management solutions relaying on autonomic monitoring [7], and proactive event-driven solutions, the optical layer monitoring is evolving from a device- to a DSP-based approach. Such an approach allows network elements to stream telemetry information to network controllers in a time-interval or even event-driving basis. Telemetry data such as Q-factor, latency, BER, and errors can be exported via a stream telemetry approach every few seconds.

New analytics tools using Machine Learning and Artificial Intelligence can consume events logs, historical telemetry data, topology information, and network operators' inputs to create models that predict failures and change provisioned network services to isolate specific network components and paths. Such information can also be available to be used by external tools using open interfaces, such as RESTful.

Today, many optical layer devices offer stream telemetry over protocols such as gRPC [8]. The mechanism is typically a publish/subscribe model wherein the network element publishes a specific set of performance monitoring points, which are of interest to that specific subscriber, and which is typically the network controller. The

cadence of this method is typically on the order of seconds, which is clearly an improvement from previous implementations.

The main limitation of streaming telemetry is the sheer volume of data that can be produced. Although it may seem small in the context of today's cloud computing, it is not typically carried over the same transport mechanisms as the revenue generating traffic. Rather it is necessary to carry telemetry over a dedicated management plane that is typically low bandwidth, but highly available – one needs to be able to reach network elements even when there is a network outage of the traffic carrying equipment for troubleshooting and recovery.

One evolution of streaming telemetry is the so-called "dial out" mechanism [9], where the network elements reach out to the controller. This can be a method to limit the publication volume of data. It could also mean that one need not wait for the next "tick" of the publication cycle to alert the controller of potentially interesting changes, using an event-driven approach.

In addition to simple performance monitoring, there is a trend toward instrumentation in the DSP based modems. Strong FEC means that there is a direct estimate of the signal-to-noise ratio (SNR) at the receiver and the operating margin available before hitting the FEC threshold. Although this is critical information and is used in many Liquid Spectrum [10] applications, there are other capabilities which are only now being made available through the application of streaming telemetry. The modem DSPs are able to compensate for all linear and some non-linear effects of propagation and as a result, we have a means to estimate the degree to which each of the SNR contributors is affecting the operating margin. For example, a sudden change in the ratio of linear to non-linear noise even at a constant SNR may be an indication of imminent failure of an optical amplifier.

Monitoring changes in optical signal quality helps predict (probabilistically) future outages, including soft failures. Some signal quality changes lead to a poor signal that leads to corruption and even silent packet drops [11]. [12] reports that there is a 50% chance of an outage within an hour of a drop event and a 70% change of an outage within one day. This means Q-factor drop events are strong predictors of future outages (leading to hard and soft failures). As Q-drop events raise the probabilities of outages, high priority network services at the packet layer should be moved away from the impacted link.

In 2017, the network chip manufacturer Barefoot Networks launched the first fully programmable Protocol Independent Switch Architecture (PISA) switch. PISA switches allow network operators to fully program the switch's data plane with no performance penalty to both the forwarding processor and the control plane. Combined with a data plane programming language called Programming Protocol-Independent Packet Processors (P4) [13] created in 2014, addressing the packet layer monitoring limitations is now possible. Using PISA switches and P4's In-band Network Telemetry application [14], network operators can extract any data plane information from the network device without requiring any intervention from the control plane. With a data plane capable of supporting network telemetry at line-rate, any and every packet can be monitored in real time. Monitoring buffer utilization can be done at a millisecond time interval, making it possible to detect bursts that will lead to tail-drops.

5. PROPOSED SOLUTION

The LSST international connectivity orchestration was designed using the

Software-Defined Networking (SDN) paradigm. With SDN, optical and packet layers' network functions are controlled by a centralized orchestration entity with a complete view of the network resources. The LSST SDN framework was tailored to operate a network infrastructure with a mix of dedicated and shared networking resources. Resources from optical and packet layers are integrated into a solution that manages not just the high-priority LSST telescope data transfer flows, but also telescope remote control flows, IT and network services, as well as Internet access.
The LSST SDN framework is composed of two main entities: the PathManager and the MonitorManager. The PathManager is responsible for control plane activities, such as topology discovery, pathfinding, path instantiation, path convergence, and path optimization. The MonitorManager is accountable for gathering, processing, and storing network state information, counters, and logs, as well as triggering alarms to network operators and the PathManager, in case of abnormalities. Figure 1 has a representation of the LSST SDN framework.

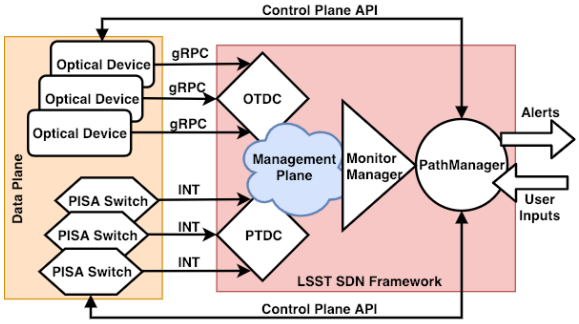

Figure 1 LSST SDN Framework

As part of the optical layer monitoring infrastructure, a new monitoring component was created: the optical telemetry data collector (OTDC). Each OTDC is responsible for collecting telemetry data from optical devices using gRPC, using both time-interval and event-driving approaches. Information such as OSNR, pre- and post-FEC, and event logs are collected every 5 seconds. The OTDC is capable of processing the telemetry data, for instance, computing Q-factor and analyze log events, and, if available, the OTDC can use external analytics tools provided by the optical manufacturers to receive trends and more complex information that could predict network outages. In case any abnormality is detected, the OTDC sends a notification to the PathManager to act upon and the network operator is notified.

As part of the packet layer monitoring infrastructure, due to the sheer amount of data received from PISA switches, a new monitoring component is needed: the packet telemetry data collector (PTDC). Each PTDC is responsible for collecting multiple gigabits per second of telemetry metadata exported by the directly-connected PISA switch. At every configurable interval (currently 150 milliseconds) of telemetry data collected, the PTDC looks for specific user-defined mitigation patterns, such as per-packet instantaneous interface buffer utilization and interface output utilization. In the event the average interface buffer utilization reaches higher than 80% (also a user-defined variable), an event log is triggered and sent to the MonitorManager. If interface buffer utilization is continuously above a pre-defined threshold, network services redistribution and policing activities are performed by the PathManager to avoid soft failures.

Soft failures resulting from packet loss caused by a damaged network element cannot be detected by just exporting metadata and counters from a PISA switch. In these cases, the chosen approach to mitigating packet loss is tagging each packet at the source PISA switch with a sequence number. As each PISA switch in the path exports metadata to the TDC and this data becomes available to the MonitorManager, the MonitorManager can query all PTDCs in the network service path to track where a packet was last seen. Using this approach, the

MonitorManager can isolate packet loss. In case packet loss continues increasing, an event log is generated, and the PathManager is requested to perform a network services optimization – find a path without known soft errors. The network operator is then notified of an issue between two PISA switches.

6. FUTURE WORK

Future work shall include an assessment of the efficacy of the various methods and collecting interval, in particular, to detect if there is a timeframe for monitoring latency that has a more impact on the avoidance of outages. Further work is needed to understand which DSP parameters correlate most strongly with network events to lower the amount of telemetry data export. The same applies for the telemetry data collected from PISA switches.

7. CONCLUSION

Combining multiple sources of real-time event-driven telemetry data obtained from new monitoring technologies at the optical and packet layers with the power provided by Machine Learning and new analytics tools is a valuable approach to anticipate, mitigate, and isolate complex soft failures. With the capabilities provided by SDN using many newly launched open and standard control and management interfaces, network operators can develop and customize tools to focus on domain-specific challenges and to automate most of their troubleshooting and capacity planning activities.

8. ACKNOWLEDGEMENTS

Financial support for LSST comes from the National Science Foundation (NSF) through Cooperative Agreement No. 1258333, the Department of Energy (DOE) Office of Science under Contract No. DE-AC02-76SF00515, and private funding raised by the LSST Corporation. The NSF-funded LSST Project Office for construction was established as an operating center under management of the Association of Universities for Research in Astronomy (AURA). The DOE-funded effort to build the LSST camera is managed by the SLAC National Accelerator Laboratory (SLAC).

The National Science Foundation (NSF) is an independent federal agency created by Congress in 1950 to promote the progress of science. NSF supports basic research and people to create knowledge that transforms the future.

9. REFERENCES

[1] Ivezic, Z., et al, LSST: From Science Drivers To Reference Design And Anticipated Data Products 2019, ApJ, in press, arXiv:0805.2366

[2] Staple, G.C., 1994. TeleGeography 1994: Global Telecommunications, Traffic Statistics & Commentary. TeleGeography.

[3] M. Mathis, J. Semke, J. Mahdavi, and T. Ott, “The macroscopic behavior of the TCP congestion avoidance algorithm,” SIGCOMM Comput. Commun. Rev., vol. 27, no. 3, pp. 67–82, Jul. 1997. [Online]. Available: http://doi.acm.org/10.1145/263932.264023

[4] Vela, A.P., Shariati, B., Ruiz, M., Cugini, F., Castro, A., Lu, H., Proietti, R., Comellas, J., Castoldi, P., Yoo, S.J.B. and Velasco, L., 2018. Soft failure localization during commissioning testing and lightpath operation. Journal of Optical Communications and Networking, 10(1), pp. A27-A36.

[5] Paolucci, F., Sgambelluri, A., Cugini, F. and Castoldi, P., 2018. Network telemetry streaming services in SDN-based disaggregated optical networks. Journal of Lightwave Technology.

[6] Hendriks, L., Schmidt, R.D.O., Sadre, R., Bezerra, J.A. and Pras, A., 2016, April. Assessing the quality of flow measurements from OpenFlow devices. In 8th International Workshop on Traffic Monitoring and Analysis (TMA).

[7] Boitier, F. and Layec, P., 2018, July. Automated Optical Networks with Monitoring and Machine Learning. In 2018 20th International Conference on Transparent Optical Networks (ICTON) (pp. 1-4). IEEE.
[8] A high-performance, open-source universal RPC framework (gRPC), https://grpc.io/er,
[9] An OpenConfig/gRPC interface for dial-out streaming. https://github.com/openconfig/reference/issues/42
[10] D. W. Boertjes, A. Leong, and D. Attard, "Field Trial of Short-Term Capacity Optimization on a Live System," in *Photonics and Fiber Technology 2016 (ACOFT, BGPP, NP)*, OSA Technical Digest (online) (Optical Society of America, 2016), paper AW5C.3.
[11] Tremblay, Y., Fitel Photomatrix, 1998. Circuit and method of testing for silent faults in a bi-directional optical communication system. U.S. Patent 5,781,318.
[12] Ghobadi, M. and Mahajan, R., 2016, November. Optical layer failures in a large backbone. In Proceedings of the 2016 Internet Measurement Conference (pp. 461-467). ACM.
[13] Bosshart, P., Daly, D., Gibb, G., Izzard, M., McKeown, N., Rexford, J., Schlesinger, C., Talayco, D., Vahdat, A., Varghese, G. and Walker, D., 2014. P4: Programming protocol-independent packet processors. ACM SIGCOMM Computer Communication Review, 44(3), pp.87-95.
[14] Kim, C., Sivaraman, A., Katta, N., Bas, A., Dixit, A. and Wobker, L.J., 2015, August. In-band network telemetry via programmable dataplanes. In ACM SIGCOMM.

