Experimental End-to-End Demonstration of Shared N:M Dual Homed Protection in SDN-controlled Long Reach PON and Pan-European Core

Séamas McGettrick, Frank Slyne, Nattapong Kitsuwan,
David B. Payne, and Marco Ruffini, *Senior Member, IEEE*

***Abstract*—Seeking reduction of capital and operational costs on next generation fibre networks is the holy grail of network planning and deployment for all operators world wide. In particular, efficient deployment strategies of next generation fibre access networks is of paramount importance to enable wide scale ubiquitous deployment of high-speed broadband services. Exploiting the large capacity of fibre networks to support heterogeneous services from residential and business users is a promising strategy towards this goal. Long-Reach Passive Optical Networks (LR-PON) is one such strategy which also adopts greater sharing of active and passive components, together with consolidation of central offices, to reduce capital and operational expenditures. However due to its long reach and large split ratio, protection mechanisms become a major consideration when designing an LR-PON, as a single feeder cable cut could disrupt services for several thousand users. In this paper we demonstrate fast restoration of LR-PON services using a dual-homed, shared-OLT protection mechanism. Our end-to-end testbed connects our optical access broadband laboratory operating on custom built LR-PON ONU and OLT FPGA prototypes with a Europe-wide testbed core network (GÉANT). We use an SDN control plane to manage the dual-homed N:M protection switching and traffic reroute in the core, and achieve access protection and end-to-end services restoration times of 40 ms and 80 ms respectively.**

***Index Terms*—Long-Reach Passive Optical Network, GPON, XG-PON, N:M protection, SDN, end-to-end, service restoration, converged architecture.**

I. Introduction

MANY operators worldwide are currently upgrading their access network to fibre network, and there is a strong push on developing efficient network designs able to reduce capital and operational costs for the deployment of the current and next generation fibre access networks, while increasing the revenue they can generate. Firstly, a fiber access network with its high capacity lends itself to carry much more than residential traffic. This could enable the access network to be used in new ways, for example to offer dynamic on demand services to residential and small/medium businesses alike, and to be utilized for mobile X-hauling (i.e., including backhauling, fronthauling and any other means of transporting mobile access data). Such a heterogeneous utilization of the fiber access network will increase its efficiency of use, increasing the revenues it can bring and thus the time to positive cash flow. Secondly, more efficient network architectures can be investigated, aimed at reducing the initial deployment costs as well as its future upgrade and maintenance requirements. One such access architecture is the Long Reach Passive Optical Networks (LR-PON), which introduces larger split ratios and longer reach compared with current PON systems, allowing the bypass of a portion of the metro transmission network [1].

A significant aspect of network design is that of providing protection mechanisms, which are essential to ensure adequate network availability. However, providing redundancy adds costs to the network and so protection mechanisms are mainly limited to the core and metro networks, where the cost of adding redundancy to the network can be shared among many users. Access networks for residential users are typically unprotected, and this is reflected in the terms of Service Level Agreements (SLAs) which generally allow for service restoration times of days or weeks. Access networks for business users can instead operate on much stricter SLAs, although businesses normally apply for private line connections and incur high cost to guarantee protection. While ultra-high availability can still be guaranteed to customer willing to absorb its cost through bespoke solutions, a multi-service access network as the one we envisaged requires offering adequate protection at acceptable price in order to satisfy a wide variety of services and customers.

This is especially true for architectures like LR-PON that provide cost savings by promoting larger customer aggregation and bypass of transmission networks that are normally protected: thus special consideration needs to be taken to ensure appropriate protection strategies. Figure 1 compares the layout of a future LR-PON to that of current PON architectures. Similarly, Table I summarizes the points of failure for the LR-PON and current PON systems (XG-PON in this case) together with an approximation of the number of people affected by such a failure. We can see that since the LR-PON also operates as a metro aggregation point, the likelihood of one individual failure affecting a large number of customers is higher compared to a conventional access network: for this reason protection mechanisms become a requirement in LR-PON. A dual homing architecture is also important as it provides additional resilience through link and node diversity, and it can be implemented as far as the constraint over the

This material is based upon works jointly supported by the Science Foundation Ireland under Grant No. 10/CE/I1853 and European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 318137 (Collaborative project ”DISCUS”).

S. McGettrick, F. Slyne,N. Kituswan, D.B. Payne and M. Ruffini are with CTVR/CONNECT, The Telecommunications Centre, Trinity College Dublin, Ireland e-mail: (mcgettrs@tcd.ie).

Manuscript submitted Oct, 2015

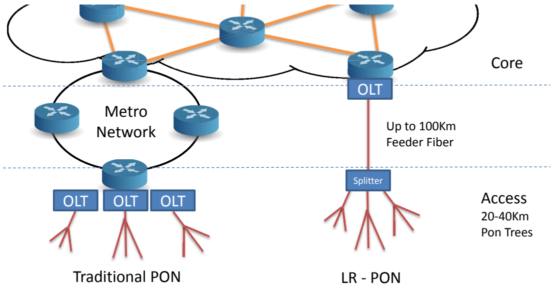


Fig. 1. XG-PON and Long-reach PON network structure compared.

optical reach is satisfied on both primary and backup paths, as for example is shown in the optimisation studies carried out in [2], [3].

TABLE I
FAILURE POINTS OF A PASSIVE OPTICAL NETWORK FOR FUTURE LR-PON IMPLEMENTATIONS AND CURRENT XG-PON IMPLEMENTATIONS

Failure Point	LR-PON		XG-PON	
		Customer Effected		Customer Effected
ONU Failure		1		1
Final Mile	10-20 Km	1-128	10-20Km	1-128
Optical Amplifiers at first splitter	used to extend PON reach	512-1024 (entire PON)	N/A reach achieved using Metro Network	
Feeder Cables	70-100 Km	10s thousands	N/A part of the Metro Network	
OLT Node	Already Protected by core and Metro network			

In this paper we investigate end-to-end service restoration times for networks using next generation Long-Reach Passive Optical Networks and Software Defined Network (SDN) to control both access and core networks. Our PON protocol is implemented on FPGA and is networked to an SDN controller running Openflow as southbound interface. We present and evaluate a dual-homed LR-PON protection mechanism where backup OLTs are shared among PONs in an N:M scheme, and the service restoration is provided over an end-to-end Software Defined Network (SDN) controlled network. Our goal is to demonstrate ultra-fast protection of PON access systems. We argue that fast protection is required in order to satisfy user requirements in a multi-service shared PON environment (especially considering business and mobile backhaul applications). In addition, fast hitless PON protection allows the implementation of protection load balancing schemes, such as those introduced in [4] which allow reducing substantially cost of both IP [2] and PON backup resources [3], by increasing the ability to share protection equipment across the network.

In our previous work [5] we looked at 1+1 protection mechanisms in the LR-PON, as showed in Figure 2 (left). This requires downstream data to be replicated at both the primary and backup OLT, and also requires a dedicated backup OLT to remain active to monitor the optical channel. In addition, if, like for the case shown in the Figure 2 (left), the protection is through dual homing, traffic is also replicated through the core.

In our experiments we showed that the PON protocol together with hardware optical monitoring could re-establish control of all ONUs on the PON in as little as 4 - 5ms. Although this result represents the fastest possible switchover time when working with the LR-PON, this method also increases the cost of the network as extra capacity is needed both in the access and in the core to duplicate the downstream traffic. We then progressed this work in [6], by showing a 1:1 protection scheme that removed the need for duplicated data in the core, see Figure 2 (middle). After failure, our mechanism would restore service in the access while also re-routing traffic through the core. Our experiments showed protection times of 125ms, when re-routing core traffic through a PAN-European core network [7], although we also made propositions that this could be reduced to about 40 ms by optimizing the controller and reducing the network latency to a more realistic level.

In order to fulfil the goal of the DISCUS architecture [8] to provide cost effective end-to-end solutions for ubiquitous optical access networks, we have in this paper, further expanded our LR-PON protection mechanisms towards dual-homed N:1 protection [9], more generally referred to as N:M, as M backup OLTs can be used to protect for N active OLTs. As shown in Figure 2 (right) in the event of a failure the backup OLT is allocated to the failed PON and data is re-routed to the backup node. Furthermore, the backup OLT is located at a different location to the primary OLT (i.e., dual-homed to a different Metro-Core node) and utilizes a backup feeder fiber path. Thus, unlike co-located backup OLTs which only protect the network from OLT board failure, this scheme protects against fiber dig-ups, OLT failure and metro node failure. It should be noticed that in the N:M case an optical switch is required in order to allow any of the M backup OLTs to protect any of the active OLTs, with $M \leq N$. Differently from the 1:1 case, the backup OLTs are not statically linked to a specific PON, but can be dynamically redirected, through the optical switch, to whichever PON requires protection. It should also be noticed that since in the N:M case the backup OLT is not directly connected to the PON, the fail detect operation cannot be operated by the backup OLT, as in the 1+1 [5] and 1:1 [6] cases, but it needs to be activated by the primary OLT, as described in section II.B.

To test the scheme we have partially implemented the OLT and ONU units for LR-PON on Xilinx VC709 FPGA boards. We connect our PON system to the GÉANT OpenFlow testbed [7] to act as the core network for our experiments. We use this setup to measure the end-to-end protection time after a failure occurs. In 2008 a similar experiment was carried out in BT using commercial GPON hardware and the restoration time was found to be in the order of 30s [10]. Additional experiments results were published in 2013 [11], showing an automated protection mechanism based on VLANs that reduced average access protection times to 4.5 s (with maximum values of 9.5 s). Other work carried out by Tsutsumi et al. [12] showed that **shared home** N:1 protection could be carried out in as little as 30ms and that it was possible in certain scenarios to carry out a protection event without losing any data.

As we have implemented the OLT and ONU on FPGA

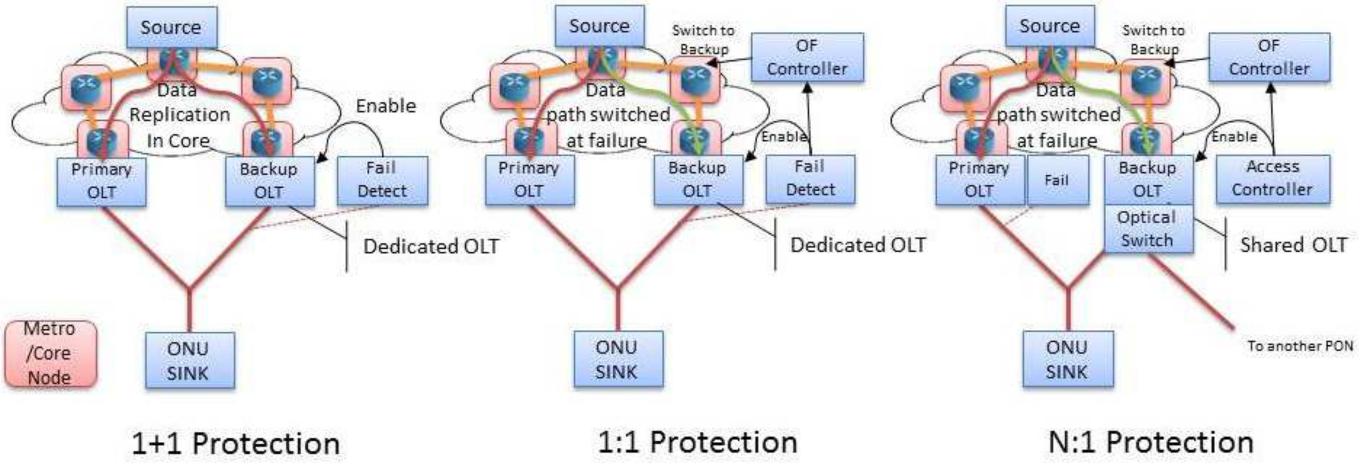


Fig. 2. Comparison of 1+1, 1:1 and N:M Protection schemes.

and are using an OpenFlow core network we have full access/control to all elements in our system. Furthermore, The XG-PON standard on which our LR-PON protocol is based - has greatly reduced the time required to register and range the PON when compared to the GPON standard [13].

We previously published an initial set of results for end-to-end dual homed protection in [9] where a protection time of approximately 155ms was achieved. In this paper we extend this work by:

- developing a brand new SDN controller to reduce the variance in message passing to the openflow controller thus halving the protection time to 80ms.
- developing a system to measure the times of various events in the system so that we can understand how long each step in the protection of the LRPON takes.
- giving details of FPGA hardware, failure detection mechanism, experimental setup and switchover timing mechanism.

The new protocol together with the accessibility and flexibility of our test bed means that we can show that end-to-end protection can be achieved in 80ms across a Pan-European core network. We can also use our system to give a detailed breakdown of where this time is lost in the protection switch over. From this analysis we can also see that much of the protection time is due to network latency, which is understandable considering the large span of the core network we have used for experiments. We envisage protection times will be below the 50 ms for typical use cases of European national networks.

II. Details of the PON protection system

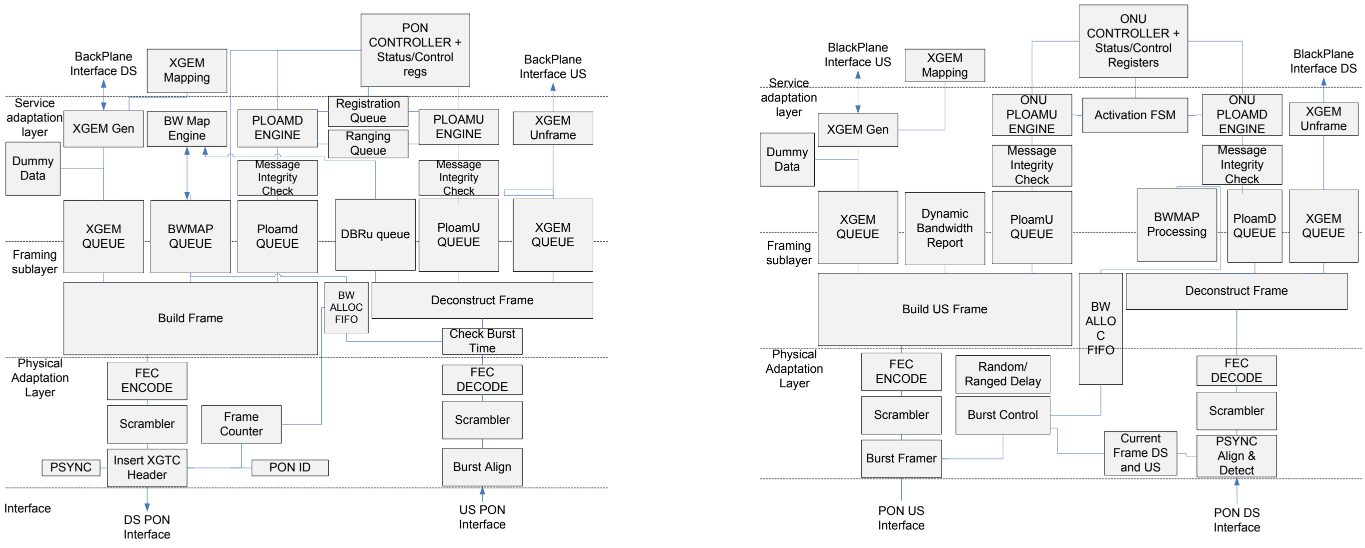
A. Optical network implementation and protocol

The LR-PON Protocol hardware is implemented over Xilinx VC709 boards [14]. The PON hardware is a partial implementation of the XG-PON standard, modified where necessary to allow for the longer reach and higher split ratios needed by the DISCUS architecture. Figure 3a and 3b show the hardware implemented on the OLT and ONU FPGA respectively. The OLT and ONU hardware is implemented in five layers. These are the core interface layer, service adaptation layer, framing sublayer, physical adaptation layer and access interface layer.

The core interface layer is the PONs backplane connection to the core network. It contains a 10G Ethernet physical layer and MAC. This allows the PON hardware units to be plugged into any 10G capable network element. In this experiment it is connected to a 48-port 10G Openflow switch which is used to direct traffic from the core network on the PON. The North bound interface layer also contains a microblaze soft processor which acts as a controller for the OLT and ONU. The Microblaze is connected to a Universal Asynchronous Receiver/Transmitter (UART) interface which is used as a management link to the PON hardware. Through this link various aspects of the PON can be controlled. These include resetting the hardware, viewing hardware status, triggering hardware failure (for experimentation), loading bandwidth map and XGEM mappings (each of which is explained below).

The access interface is implemented as a raw data interface. This means that the data from the PON protocol is sent directly to the SFP+ laser without any additional encoding or framing. This raw interface gives the protocol hardware complete control over what is being sent on the PON. The core and access interfaces are both specific to the prototype board/FPGA being used. The remaining three layers constitute the LR-PON protocol and are based on protocol setup described in the ITU standard for XG-PON [15]. These layers would not require further changes if the system was moved to a different platform.

The physical adaptation layer of the PON hardware is responsible for ensuring the PON data is sent and received correctly between the OLT and ONU. In the downstream direction it creates the GTC frame structure and in the upstream direction it controls the burst frames structure. This layer places a synchronization word at the beginning of each frame to ensure data alignment on the PON. In the downstream direction it also adds a PON ID tag and a frame counter which allows the ONUs to identify the connected OLT and ensures the ONUs are synchronized downstream respectively. In the upstream direction it ensures that data bursts from the



(a) OLT (b) ONU

Fig. 3. Block Diagram of hardware implemented on FPGA.

ONU meet the requirements of the PONs burst profile message received at activation. The PON data is also scrambled to aid clock recovery and FEC is applied to the data to ensure correct transmission. In this implementation we have not included FEC as it is not necessary in our current setup. For simplicity we have also omitted the HEC signing of the Frame counter and PON ID. This would have no bearing on protection timing of the PON. In the upstream direction the physical adaptation layer is also responsible for controlling the burst nature of the ONU.

The framing sublayer is responsible for muxing and demuxing the various components of the data frames, these are the bandwidth map, the Physical Layer Operations, Administration and Maintenance (PLOAM) messages and the XG-PON Encapsulation Method (XGEM) data in both the upstream and downstream directions. The bandwidth map contains precise instructions as to when each ONU can send data bursts upstream. This ensures that only one ONU uses the upstream channel at any time thus avoiding contention issues. PLOAM messages are management messages sent between the OLT and ONU to pass vital operational information about the PON. The XGEM data is network data coming from the core interfaces (Core network for OLT and User network for ONU). This is the information to be transported on the PON.

The Service adaptation layer is responsible for translating data from the backend network to data structures used on the PON or visa-versa. Data frames from the core/user network need to be translated and encapsulated into PON data structures called XGEM frames. To do this the Ethernet frame headers and preambles are removed. All VLAN and MPLS tags are removed. The data in these tags together with the destination address is used to address the XGEM frame to a specific ONU on the PON. The PLOAM engine which is responsible for creating and interpreting the PLOAM management messages resides in this layer. In this implementation the OLT can send burst profile, ONU registration and ONU ranging PLOAMs to the ONU. The ONU can respond with Serial number and Acknowledgement PLOAMs. These PLOAMs are all required for ONU discovery on the PON, as described below. The DBA engine responsible for creating and interpreting the Bandwidth Map also resides in this layer.

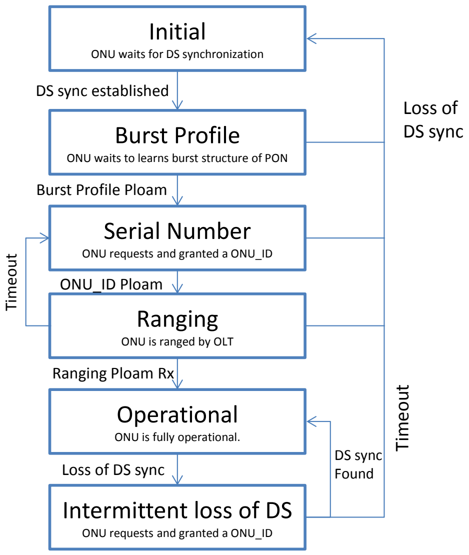


Fig. 4. ONU activation state machine (simplified) [15].

1) PON activation: In order to test protection switching in the PON it is important that the ONU implements the full activation state machine; see Figure 4, described in the XG-PON standard [15]. To get to the operational state the ONU must first synchronize to the downstream data. The ONU must then wait until it receives a burst profile PLOAM which is periodically sent by the OLT. This PLOAM message contains details of how the OLT expects upstream transmissions to be formatted on the PON. Once the burst profile has been received the ONU must wait for the OLT to initiate an activation cycle. Since the OLT does not know when a new ONU might come online, it periodically creates a break in upstream transmission and invites new ONUs to send their serial numbers in a PLOAM message. It does this by sending a broadcast bandwidth map entry. All new unregistered ONUs reply to this message and so collisions can occur. All ONUs wait a random time before bursting serial number to reduce probability of collisions. The OLT responds to all successful serial number PLOAMs by sending directed registration PLOAMs that register the ONU on the PON and give it an ONU-ID. The OLT then ranges the ONU to ensure that it is synchronized with all other ONUs on the PON. The ONU is now in the operation state and is ready to receive traffic.

In the event of a failure of the OLT or the fiber on the PON the ONU does not return to startup straight away. Instead it enters a temporary loss of synchronization state when it waits for 100ms or until downstream synchronization is restored. It is during this time that the backup OLT can take control of the PON without needing to register and range the entire ONUs again.

2) Limitations of the physical layer: This implementation uses a simplified physical layer and not an LR-PON physical layer. The OLT is directly connected to the ONU via an optical patch cable. The Secondary OLT is connected to an optical switch and then to the ONU via a second optical port. The PON first stage splitter is therefore implemented in the FPGA logic of the ONU instead of being a standalone unit. We do not use optical amplifiers which would require a management unit at the first stage splitter to also detect the failure and switch the amplifiers accordingly. However, LR-PON transmission latency was emulated in the FPGA hardware. We have future plans to incorporate the full LR-PON physical layer into the testbed. However since the protocol and FPGA implementation model the full physical layer correctly we believe the results of the protection experiments obtained in this setup are in line with what might be achieved in the full PON system. Furthermore, our experiments are being run at a 1ms precision and so even significant changes in PON round trip time will not affect the results by more than 1-2ms.

B. Failure detection

In our test scenario we simulate a cut in the feeder fiber between the primary OLT and first stage splitter on the PON. This stops all upstream and downstream data on the Primary link. A hardware unit in the primary OLT FPGA monitors the upstream data path. Even in cases were the two directions of communication are operated over separate fibers, a cut in the downstream fiber will prevent all ONUs from receiving messages form the OLT: in this case all ONUs will automatically stop transmitting, so that the OLT will not receive any upstream data. This upstream silence activates a timer. If this timer expires an alarm is raised to initiate a protection switchover. The duration of this timer must take into account all normal silences on the PON, namely quiet windows and normal roundtrip time, in order to make sure that an alarm is only raised when a failure occurs. Thus on an LR-PON of 125 Km, like the one proposed by the DISCUS project, failure detection could take approximately 2.5ms in worst case conditions (i.e. 1.25ms for round-trip fiber delay and 1.25 ms for quiet window). It should be noted that this worst case condition is the same regardless of the number of ONUs on the PON provided there is one or more active units.

Since our aim is to investigate end-to-end network protection times, we need to also involve the control plane for the access and core nodes. Thus, when a failure is detected, the hardware detection unit alerts the OLT controller which sends an in-band upstream alarm to the Openflow access network controller. An upstream alarm is required, because the Openflow bridge does not physically terminate the connection between the ONU and the OLT, which means that it is not possible for Openflow path switching rules (such as Group based port protection) to be invoked due to the fiber cut.

C. Development of a new low-latency message passing mechanism for Open Flow controller

Our test scenario uses two independent sets of Openflow controllers, the core controller and the access controllers. As described in the next section, in our tests the access data plane and controller [16] were located in our Optical Network Architecture (ONA) lab in Dublin, Trinity College University, while the core transmission network was overlaid as a virtual network operating over the Pan-European GÉANT network. The SDN core controller was also located on one of the GÉANT nodes and implemented in POX. The test scenario initiates when the OLT issues an alarm to the access Openflow controller following a failure detection. The access controller enables a data route through the backup Openflow bridge, activates the backup OLT and tunes the optical switch to route data from the backup OLT to the failed PON backup fiber. In parallel it communicates with the core controller to activate the pre-calculated protection route in the core, which connects the remote server with the MC node where the backup OLT is located.

In our previous experiments [9], we treated PON protection events as generic openflow events which were passed between the access and core controller using openflows default communication channels. This meant that processing of these messages was left to the discretion of the openflow controllers. In our experiments we thus noticed significant variation in the time for the protection path to be fully activated, ranging from 79ms up to 133ms. On analysis, this high level of variation was caused by two factors. Firstly, there was variation in the time between the receipt of a PON failure alarm by the access Openflow controller and the subsequent action taken by both the access and core controllers. Secondly, there was significant variation in the relaying and tunnelling of signals between the ONA and GÉANT POX controllers. The issue is that most controllers are designed with functionality abstraction in mind, using class/objects hierarchies in order to increase developer productivity [17]. The drawback is in terms of performance, as interpreted languages such as python do not perform as well as compiled languages [17]. To rectify this issue we needed to have more control over when and how the SDN controller dealt with protection events. We thus developed a new SDN event plane based on fast, low latency distributed message queuing architecture in order to reduce the latency and variation in both the controllers and the tunnels between the controllers. Our development optimised the real-time event-handling capability of the standard POX controller and extended the functionality across multiple controllers. We retained however the basic shell of the POX controller for the purposes of a standard Interface to the south-bound Openflow devices in order to leverage our existing code-base.

We elected to use the open source ZeroMQ library which can handle up to 2.8 million messages per second and can open a TCP socket and process data within 28.45 microseconds. The event plane allows all major elements in the test bed such as the Openflow Controllers and the PON components to publish events using a common message format, as well as to subscribe to system wide broadcast events. We implemented four sets of messages which have a common format for control and co-ordination of events within the test bed. The message format is composed of a major category (called a ZeroMQ topic); a global timestamp which is synchronized to Dublin time; a minor category which is used for a command or message payload. The four messages types are as follows:

a) Unsolicited events of large significance such as the failure of major nodes and links: an example of a primary PON failure event is: 'NetEvent, 1422736912.30, OLT_P_Failure'.
b) Reactive control messages, which are generated in response to unsolicited events, for example, messages that trigger takeover of service by a standby piece of equipment or Service. An example of an Optical Switch control message is: 'GlimEvent, 1422737623.07, upSdownP'.
c) Proactive configuration of elements or sub-systems within the testbed. e.g. 'SysControl, 1422737623.07, Restart" is the sub system event to reboot the system.
d) The control of or alerting within test routines. These messages serve to co-ordinate the actions of a number of agents involved in a test cycle which are located across the wide area testbed.

1) Control Plane based on distributed message queue: Figure 5 shows the logical configuration of the Test bed control plane based on a distributed ZeroMQ Message Queue. Any elements throughout the test bed can either publish or subscribe to topics on the event plane, through the use of the lightweight ZeroMQ API. This API is available for scripting and programming languages such as Perl, Python, Java and C. A key feature of ZeroMQ is that the Message Queue is logically centralized; there is no physical hub through which all messages flow. This removes both single points of failure and performance bottlenecks.

The TCD_ONA Openflow Controller intercepts the downstream failure alarm in the primary PON. As well as triggering Openflow switching rules in the Metro Core network, the TCD_ONA controller also publishes Message Queue NetEvent and GlimEvent messages. The NetEvent broadcasts to all subscribed components about the PON failure. One such subscribe component is the GÉANT Openflow Controller which executes secondary routing in the network core. The Message Queue was extended between the TCD_ONA controller and GÉANT Openflow Controller using an SSH tunnel. The GlimEvent triggers Optical Switch path selection or protection Path in the Access Optical Switch. The Optical Switch subsystem was developed to provide a concurrent Message Queue interface to the TL1 interface of the Glimmerglass Optical Switch. Table II shows the array of test bed components on the Message Queue and the type of messages which they publish or to which they subscribe.

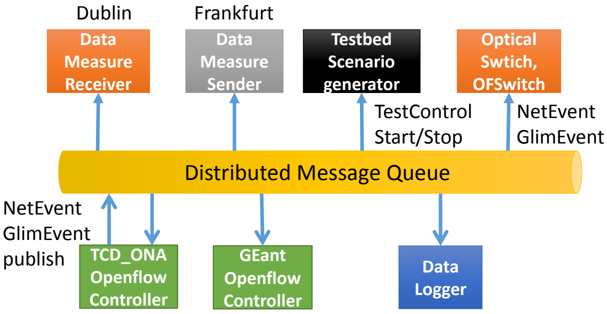


Fig. 5. Logical testbed control plane configuration based on distributed ZeroMQ Message Queue.

III. Experimental Setup

We tested our end-to-end protection service with dual-homed, N:M backup OLT sharing by combining the optical architecture testbed in Trinity College Dublin and the GÉANT pan-European research network, as shown in Figure 6. The testbeds are connected through two dedicated GE links. The experiment replicates both the metro-access and core networks of a high-speed fixed line telecommunications network. The end-points replicate a data center generating traffic (located in Frankfurt) and a reception or termination point located on a PON ONU in Dublin. The Metro-Access portion of the network is created in the Optical Network Architecture lab in Trinity College Dublin. The Core network is replicated using the GÉANT Openflow testbed facility, which spans continental Europe.

The GÉANT OpenFlow facility is a test-bed environment deployed on top of the GÉANT production network. The facility is built on network resources such as software-based OVSwitch OpenFlow switches and interconnecting network links. It is collocated with five of the Points-of-Presence in Vienna (AT), Frankfurt (DE), London (UK), Amsterdam (NL) and Zagreb (HR). The OFELIA Control Framework (OCF) is used by the GÉANT OpenFlow facility to manage requests for slice submission, instantiation, and decommissioning. OCF is a set of software tools for testbed management, which controls the experimentation life cycle such as resource reservation, instantiation, configuration, monitoring.

A server co-located at the DE node acts as the source for data in this experiment. The primary data path in the core is between nodes DE, AT and NL (shown in green in Figure 6).

TABLE II
ASSOCIATION OF MESSAGE QUEUE TYPES AND TESTBED COMPONENTS

Subsystem	Location	Topic	Direction	Purpose
TCD Openflow controller	TCD_ONA Controller	NetEvent	Publish	Interpret failure signal from PON and broadcase event on Message Queue.
TCD Openflow controller	TCD_ONA Controller	GlimEvent	Publish	Trigger Primary or Secondary Path in Optical switch.
Testbed Controller	TCD_ONA Controller	TestControl	Publish	Broadcast signals for stop, start, restart of test cycles.
GÉANT Controller	GÉANT Openflow Controller (NL)	NetEvent	Subscribe	Execute secondary routing in network core.
Data Logger	TCD_ONA Controller	Test Control	Subscribe	Aggregate and format results of test events throughout Testbed.
Data Traffic Gap Measurement Sender	GÉANT Data Centre (DE)	TestControl	Subscribe	Ascertain when test scenario has started or stopped.
Data Traffic Gap Measurement Receiver	TCD_ONA Testbed	TestControl	Subscribe	Ascertain when test scenario has started or stopped.
Optical Switch (Glimmerglass system 100)	TCD_ONA Testbed	GlimEvent	Subscribe	Execute Optical switch path selection.

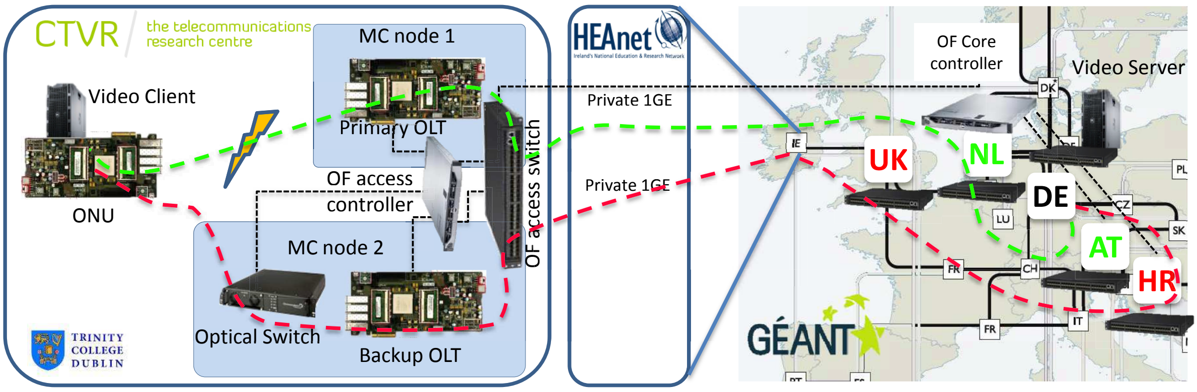


Fig. 6. Logical view of combined LR-PON access and SDN Core network.

The backup path follows the route: DE, HR and UK (shown in red in Figure 6). We implement two paths to emulate dual-homed PON network where the primary and the backup OLT are in different locations. Data on the primary data path is routed to the Primary OLT and data on the backup link is routed to the backup OLT. The NL node also hosts the Openflow core network controller which can be used to control which of the two paths data takes to our access network.

The TCD ONA is setup as two Metro/Core nodes together with an LR-PON access network. Although we have used one physical switch and server, they are both virtualized to represent independent MC node switches and controllers. The Metro/Access network comprises of a pronto 3780 switch, running release 2.2 (Openflow v1.3 compatible firmware), three Xilinx vc709 development boards [14], acting as primary OLT, backup OLT and ONU, a glimmerglass MEMs-based optical fiber switch, A Dell T620 with 10G SFP+ cards acting as client machine attached to the ONU and a separate Dell R320 Server acting as access Openflow controller. The Pronto switch is configured as multiple virtual bridges to act as standalone bridges each with a separate Openflow controller. These are connected to a gateway machine on the core side of the network and one of the Primary or backup OLTs on the access side.

The two testbeds, TCD and GÉANT are connected via two dedicated 1GE links to UK and NL respectively. In our previous experiment in 1:1 protection [6] we utilized tunneling over the internet for these links. This added a variance to our results that was very hard to measure and account for in our results. Although the dedicated 1Gb links are well below the 10Gb capacity of the GÉANT and TCD testbeds they do offer a stable link that enables us to carry out the protection experiments to the desired precision. The Glimmerglass optical switch is connected between the backup-OLT and the ONU so that the backup OLT can switch between a number of different PONs allowing us to test the N:M protection timing. In order to extend Message Queue from TCD testbed to GÉANT, we implemented tunnels to NL and DE.

A. Testing Procedure

The TCD ONA testbed has been designed to ensure all tests are easily reproducible. All testbed components are completely programmable using a Python (v2.7) based object Framework. This allows us to centrally control all testbed components so that test scenarios can be set up quickly and consistently. Likewise all test components log information to a central repository, with clear and consistent messages and time stamps. This allows us to run test scenarios repeatedly allowing for statistical analysis of means and deviations of measurements. In this scenario, a Python script is initially used to check the status of all relevant components. It uploads a bootstrap configuration to an Openflow Switch, sets up the Openflow controllers, sets up the event logger, configures the optical switch and uploads the FPGA images. The OpenVSwitch (OVS) is restarted and connected to the controller. The switches and ports are defined and associated with the OVS instance. For each switch, the required flows are configured. The testbed controller then links to the OLT and ONU microprocessors to ensure the PON is operational. Finally, the data service on the DE node of the GÉANT network is enabled and end-to-end operation of the system is confirmed.

Once the experiment is started, data start flowing over the primary link from node DE through AT and NL to our testbed and through the PON to the data sink connected to the ONU. After some time a trigger signal is sent to the primary OLT FPGA which resets the optical channel to simulate a fiber dig up. At this stage the protocol hardware is unaware of the break and packets are lost. Once the fiber breaks the OLT will no longer receive upstream data. This in turn starts the failure detection timeout timer. This timeout timer ensure that US silence is not just the result of normal PON operation (i.e. registration quiet window or RTT). When the timeout expires the primary OLT issues an alarm which is sent in band upstream to the Openflow controller. The Openflow access network controller notifies the optical switch to connect the backup OLT path to the failed PON and finally the management controller notifies the backup OLT to take control of the PON (following the procedure described in section 2.1.1). The ONU meanwhile has entered the Loss of Downstream Sync state and will remain there for 100ms or until the backup OLT begins to send synchronization words downstream. If the backup OLT does not take over before the 100ms time out the entire PON will have to be reactivated, and re-ranged to resume transmitting data. Once the backup OLT has taken control of the PON the PON is ready to start receiving data again.

In parallel with the backup OLT taking control of the PON, the Openflow access network controller passes a message to the core network Openflow controller. This causes the core network to redirect data from the primary path to the backup path to emulate more closely the dual homed nature of the Long-Reach PON. When the service resumes data will flow over the backup-path from DE to HR, UK through our backup OLT to the ONU. Since each of the packets being sent on the PON has a sequence number the ONU can easily work out how many packets were dropped during the switchover. Once this number has been calculated the scenario script is ready to restart a new iteration of the experiment.

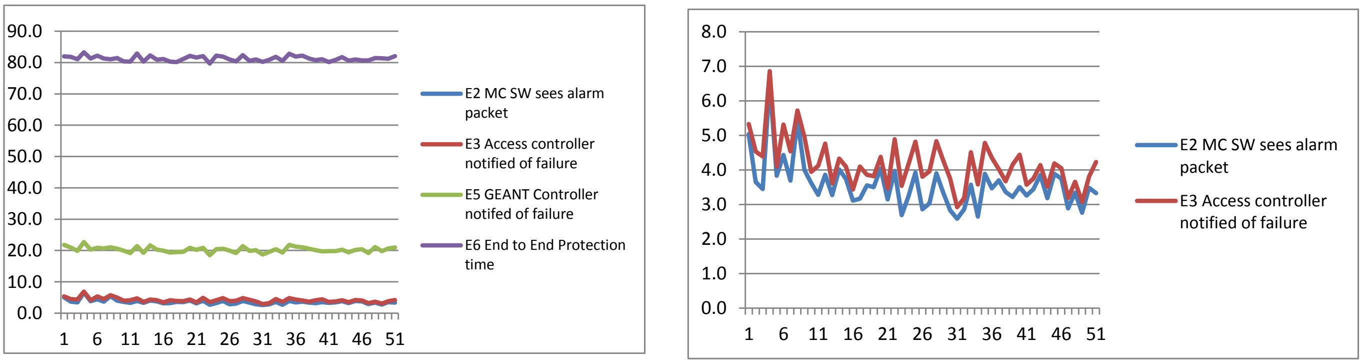
B. Results

In our previous work [9] using our previous SDN controller we measured an average end-to-end protection time of approximately 155 ms. We saw large deviations in the network reconfiguration time from 79 ms to 133ms. This was caused by an uncertainty in processing the failure alarm at the access controller and the subsequent actions being taken by the access and core controllers. Figure 7 shows the new SDN controllers end-to-end N:M dual homed protection time over 50 experimental iterations, using the initial break in the fiber as a reference point.

Two things are immediately clear from the results in Figure 7. Firstly, the new message passing system in the SDN controllers have removed the variance associated with the previous results. Secondly, the average restoration time as almost halved from 155 ms to 80 ms. Figure 7 also shows the timing of various events that occur during the protection switch for all 50 experimental iterations. Since the trigger failure event was issued to the FPGA board over a UART interface it was not possible to read an absolute time value from the FPGA boards for when the break in the primary fiber occurred. However, we were able to work back from the restoration point of traffic by subtracting the outage duration within each cycle. On average, the alert that identifies loss of the primary PON (the E2 event in the figure) occurs 3.5ms after the break. The Openflow controller within the TCD ONA testbed sees the alert 0.59ms (E3) after this and publishes a NetEvent failure alert as well as a GlimEvent message. The NetEvent alerts the GÉANT controller to invoke the alternate path through the core. The GlimEvent message invokes the secondary path in the optical switch. Within this experiment, the GÉANT controller sees the NetEvent message 20.3 ms after the initial failure (E5). The average switching time of Glimmerglass optical switch was measured at 23ms. Overall, Restoration time of the data traffic is measured as 81.29ms. The analysis of the path restoration time can be split into three distinct phases. These are the time it takes to detect a failure on the network, the access network recovery time and the core network recovery time. In Figure 8 we show how these times stack up to give the 81ms protection figure. As discussed previously the hardware monitoring at the OLT can detect a failure in the network in about 2.5 ms. We can see from the results in Figure 7 that a further 1ms is taken for the alarm packet to be created and sent to the Metro core node switch.

We found the access recovery time to be approximately 30 ms, which can be further broken down into time required to tune the optical switch (23 ms) and time needed by the protocol to re-establish downstream synchronization (2-3 ms). From our previous work [13] we know that some time may be needed to re-range the ONUs in addition to the synchronization time (between 2 and 4 ms), however in this work we assume that ranging to the backup OLT can be done during normal operation of the PON. The remaining time is needed by the local Openflow controller to communicate with the optical switch to connect the backup OLT.

The core network recovery time happens in parallel to the access network recovery time. The core controller sees the failure event approximately 20 ms after it happens. It then begins to reconfigure the network to reroute data to the backup path. Since we know the latency of this path is approximately 50 ms we can calculate that core network switch over takes approximately 10 ms. Thus we believe that



(a)

(b)

Fig. 7. Switchover time (ms) for 50 iterations of N:M protection experiment. (a) shows all timing results and (b) shows results zoomed into E2 and E3 to show SDN controller responsiveness to failure in higher resolution.

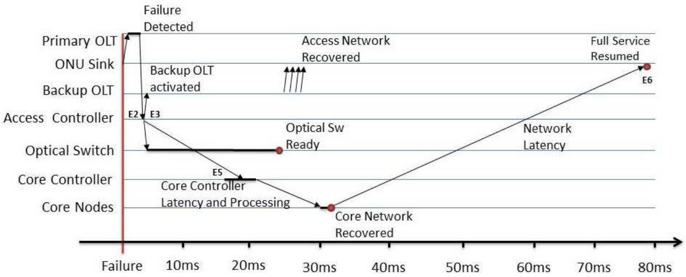


Fig. 8. Breakdown of approximate N:M protection events and timing.

a similar experiment using this new controller run through a core network of a typical European country would lead to service restoration times below 50 ms.

IV. Lessons Learned

As with any complex implementation and testbed setup this build ran into a number of obstacles during its development. We summarize a number of the bigger issues faced here so that future readers who attempt to create a similar setups might benefit from our experience. Developing a new high speed protocol on off-the-shelf FPGA development boards proved difficult. Initially this work was targeted at the NetFPGA 10G board, however after many failed attempts trying to bypass the 10 GE physical layer chipset on the NetFPGA we had to migrate the design to the VC709 boards. The VC709 boards allowed us to have full control over the optical channel, however we had to split our design into a number of clock domains to correctly transmit and receive data. The 10G I/O GTH pins on the FPGA have also limited parts of design as they have limited use in burst mode scenarios and can be slow to recover clock synchronization after a quiet period on the channel.

From a control plane perspective, in our earlier results we saw large variations in the timing measurements across the test bed. These variations were caused by latency within the processing of both the Core and Metro Openflow Controllers, as well as transport latency across the overall testbed. We overcame these latency issues by implementing a low-latency high-volume distributed Message Queue to which all major components could interface. The Core and Metro Openflow Controllers were enhanced to interface directly with the Message Queue.

V. Conclusions

In this paper we set-out to demonstrate N:M dual homed protection of an end-to-end testbed interconnecting an LR-PON access system with a Pan-European core network. We developed an SDN controller for our metro-core node capable of managing the feeder fibre protection mechanism and interacting with the GÉANT network controller to re-route data in the core. The implementation of mechanism for fast protection in the FPGA-based OLT and the development of a fast message-passing mechanism between controllers allowed us to show service restoration times of the order of 80 ms, even through a very long core data latency of 50ms. Our system was subsequently demonstrated over a full Long-Reach PON physical layer operating over multiple wavelengths, and with an extended set of Openflow features allowing dynamic control of wavelength and bandwidth assignment, flow setup and user registration [18]. Finally, while our implementation has focused on the Long-Reach PON, as the increased latency provides a worst-case scenario for our test, the SDN protection system used in the paper could be deployed along side any PON system.

Acknowledgment

The authors would like to thank HEAnet and GÉANT for providing dedicated access to the GÉANT openflow testbed facilities.

References

[1] M. Ruffini *et al.*, “Discus: End-to-end network design for ubiquitous high speed broadband services,” in *Transparent Optical Networks (ICTON), 2013 15th International Conference on*, June 2013, pp. 1–5.

[2] ——, “Deployment strategies for protected long-reach pon,” *Optical Communications and Networking, IEEE/OSA Journal of*, vol. 4, no. 2, pp. 118–129, February 2012.

[3] A. Nag *et al.*, “N:1 protection design for minimising olts in resilient dual-homed long-reach passive optical network,” in *Optical Fiber Communications Conference and Exhibition (OFC), 2014*, March 2014, pp. 1–3.

[4] M. Ruffini *et al.*, "Protection strategies for long-reach pon," in *Optical Communication (ECOC), 2010 36th European Conference and Exhibition on*, Sept 2010, pp. 1–3.

[5] S. McGettrick *et al.*, "Ultra-fast 1+1 protection in 10 gb/s symmetric long reach pon," in *Optical Communication (ECOC 2013), 39th European Conference and Exhibition on*, Sept 2013, pp. 1–3.

[6] F. Slyne *et al.*, "Design and experimental test of 1:1 end-to-end protection for lr-pon using an sdn multi-tier control plane," in *Optical Communication (ECOC), 2014 European Conference on*, Sept 2014, pp. 1–3.

[7] Geant research network. [Online]. Available: www.geant.net

[8] M. Ruffini *et al.*, "Discus: an end-to-end solution for ubiquitous broadband optical access," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. S24–S32, February 2014.

[9] S. McGettrick *et al.*, "Experimental end-to-end demonstration of shared n:1 dual homed protection in long reach pon and sdn-controlled core," in *Optical Fiber Communications Conference and Exhibition (OFC), 2015*, March 2015, pp. 1–3.

[10] J. Kang *et al.*, "Restoration of ethernet services over a dual-homed gpon system - operator requirements and practical demonstration," in *Optical Fiber communication/National Fiber Optic Engineers Conference, 2008. OFC/NFOEC 2008. Conference on*, Feb 2008, pp. 1–3.

[11] A. Rafel *et al.*, "Automatic restoration over a type b dual parented pon using vlan switching," in *Optical Communication (ECOC 2013), 39th European Conference and Exhibition on*, Sept 2013, pp. 1–3.

[12] T. Tsutsumi *et al.*, "Long-reach and high-splitting-ratio 10g-epon system with n:1 osu protection," in *Optical Communication (ECOC), 2014 European Conference on*, Sept 2014, pp. 1–3.

[13] S. McGettrick *et al.*, "Improving hardware protection switching in 10gb/s symmetric long reach pons," in *Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC), 2013*, March 2013, pp. 1–3.

[14] Xilinx virtex-7 fpga vc709 connectivity kit. [Online]. Available: http://www.xilinx.com/products/boards-and-kits/dk-v7-vc709-g.html

[15] "Xg-pon ten - gigabit - capable passive optical networks," *ITU-T Std. G 987*, vol. June, 2012.

[16] M. Ruffini *et al.*, "Software defined networking for next generation converged metro-access networks," *Optical Fiber Technology*, vol. 26, no. A, pp. 31–41, December 2015.

[17] D. Erickson, "The beacon openflow controller," in *Proceedings of the Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking*, ser. HotSDN '13. New York, NY, USA: ACM, 2013, pp. 13–18. [Online]. Available: http://doi.acm.org/10.1145/2491185.2491189

[18] G. Talli *et al.*, "Demonstration of sdn enabled dynamically reconfigurable high capacity optical access for converged services," in *Optical Fiber Communication Conference Postdeadline Papers*. Optical Society of America, 2016, p. Th5B.1. [Online]. Available: http://www.osapublishing.org/abstract.cfm?URI=OFC-2016-Th5B.1