Protection of Long-Reach PON Traffic

David K. Hunter
University of Essex
Department of Electronic Systems Engineering
Colchester CO4 3SQ, UK
+44 1206 872416
dkhunter@essex.ac.uk

Tim H. Gilfedder
British Telecom, OP7/2 Antares Building
Adastral Park
Ipswich IP5 3RE, UK
+44 1473 643803
tim.gilfedder@bt.com

ABSTRACT
A resilience strategy is introduced for networks implementing dual homing (dual parenting) of customers, specifically those employing Long-Reach PONs (LR-PONs). Assuming that some mechanism exists to detect network element failures, the discussion concentrates on the protocols which propagate information about the reachability of LR-PONs and those that re-route traffic in the event of a fault. A protocol called FROTH (Fast Recovery for OLTs via Transmission of Hellos) informs each edge router which LR-PONs are available over the whole network. This information is used by another protocol called LATTE (Label and Address Tunneling via Tables at the Edge), to re-route traffic in the event of failure. A target of 50 milliseconds is assumed for end-to-end recovery time, since this has been used for many years with voice services, and is adopted in the absence of other information. Here, FROTH transports signaling traffic over IP – due to the probabilistic nature of packet transmission, it is impossible to place a hard bound on recovery time, but rather the probability is calculated that it will be met.

Modeling studies suggest that in the event of a cable failure or single equipment element failure, re-routed data will almost always leave the transmitting edge router in under 50 milliseconds. For more catastrophic failures (such as router failure or loss of an edge router), recovery might take between 100 – 200 milliseconds. Reachability information for each LR-PON is distributed over each area of the network by IP, and used to activate re-routing of traffic via tunnels or address substitution. However the underlying services need not use IP, and may be of any type (for example private line), making the scheme separate from the service, customer or provider.

Categories and Subject Descriptors
C.2.2 [**Network Protocols**]: Routing Protocols

General Terms
Algorithms, Performance, Reliability.

Keywords
passive optical networks, protection and restoration, GPON, Internet routing protocols.

1. INTRODUCTION
Enhanced customer experience of telecommunications services is founded on the underlying reliability of the network over which these services are provided. At the lowest (physical) layer of the network, connections between the local exchange and inner core nodes are invariably protected via one or more rings or protected chains. These provide sub 50 millisecond protection in the event of equipment or fiber failure and hence allow all customers to experience good overall service availability. Indeed, some customers are served by two independent local exchanges, so even failures in the access network can be alleviated.

It has been reported elsewhere [1] that the deployment of Long Reach PONs could offer significant capital and operational savings, by effectively merging backhaul and access technologies into a single transport solution. However, this approach must meet (or preferably exceed) all the inherent resilience capabilities and options currently available to customers. At the very least, any solution must be able to offer the following:

- Sub 50 millisecond protection for all customers from failures in the backhaul network (e.g. cable dig-ups).
- Fast and automatic restoration (hundreds of milliseconds to a few seconds) of all services in the event of a major edge router failure.
- Complete end-to-end service separation for selected customers.

Being able to re-route traffic from the customer to a secondary edge router over a separate path is only half of the solution to re-establishing services on an end-to-end basis. All traffic originating elsewhere in the network must be made aware of the failure within the LR-PON system, or within the edge router, then it must be re-addressed and re-routed correctly. This process takes a significant time with existing protocols and methodologies. Therefore, this issue is addressed here, permitting significantly faster recovery times for all customers using dual-homed Long Reach PONs.

The ITU-T [2] specifies resilience provisions which are perfectly adequate for Gigabit Passive Optical Networks (GPONs), permitting recovery from broken cables or equipment faults. This work extends GPON's resilience capabilities to the point where dual-homed LR-PONs are possible. Networking configurations are evaluated which permit a customer to access another edge router if the current one fails, thus providing enhanced fault tolerance and superior service to the customer. The solution achieves fast recovery from a range of faults, and does not carry out restoration switching in the physical layer.

To address these issues, this paper introduces two new concepts [3]. Firstly, a re-routing technique called LATTE (Label and Address Tunneling via Tables at the Edge) re-routes traffic to a secondary LR-PON Optical Line Termination (OLT) in a second edge router if the primary LR-PON OLT is not available (either due to link failure within the LR-PON itself, equipment failure of the OLT, or edge router failure due to fire or flood etc.). Also, a simple LR-PON OLT

AccessNets'06, September 4–6, 2006, Athens, Greece.

discovery protocol, optimized for speed, called FROTH (Fast Recovery for OLTs via Transmission of Hellos) collects LR-PON OLT availability information, which is passed on to LATTE.

Figure 1 illustrates the general principle. The core network may be any network implementing IP or any other protocol capable of carrying signaling information. In the diagram, there are six access networks (U-Z), which may be implemented via PONs or LANs. Each customer is dual homed, i.e. connected to two access networks, each of which is interfaced to the core network via a different edge router. Consider how customers A and B communicate. Normally, customer B uses access network W, however, if it becomes unavailable due to a fault, access network X is used instead. FROTH informs all edge routers which access networks are functioning and accessible from the core network. Normally, A and B communicate via the "working" path through edge router 3, however, if W fails, then edge router 1 is informed of this, and re-routes traffic for B over the "protection" path to edge router 4. This re-routing function is carried out by LATTE in this proposal.

Figure 2 shows dual homing of two distinct customer types: those that require complete circuit separation between their premises and the edge router (termed *commercial customers* in this paper) and those that require protection only in the backhaul (herein known as *residential customers*). If link L1 fails, a commercial customer merely re-routes traffic onto the secondary LR-PON marked "Commercial Protection". Before the failure, this customer may have employed load balancing by using both "Working" and "Customer Protection" LR-PONs. The "Commercial Protection" LR-PON leads to OLT(S2), whereas residential customers must now re-route their traffic to the secondary edge router via their secondary LR-PON through OLT(S1).

During normal operation, a primary LR-PON is permanently connected through OLT(P1), with ranging having taken place as normal. However, if L1 fails, the protection OLT that serves residential customers through OLT(S1) must activate and range prior to traffic being restored. Full re-ranging may be unnecessary, as some ranging parameters may already be available, or may have been stored previously. OLT(S1) cannot be fully activated for protection as its ranging and status messages would interfere with the primary LR-PON, although in principle it could monitor upstream traffic from customer-located Optical Network Units (ONUs).

Multiple operators co-located at edge routers can deploy their own LATTE and FROTH protection capabilities independently from one other thus preserving their own network integrity. Similar capabilities would be required at network-network interfaces such as peering points.

2. LABEL AND ADRESS TUNNELING WITH TABLES AT THE EDGE (LATTE)

When FROTH reports a failure to another edge router, IP datagrams may need to be re-routed to the appropriate secondary LR-PON OLT. To do this, LATTE either places them within tunnels (IP-in-IP), or overwrites their IP destination address fields. LATTE may be adapted to re-route MPLS, ATM and other technologies in a similar way. This type of re-routing is possible whenever a packet enters or leaves either the core network, or an "area" of the network.

2.1 Allocation of IP addresses to customers

LATTE requires information about IP addresses, and the availability of LR-PONs throughout the network, which is provided by FROTH. Each LR-PON is assigned several prefixes, or IP address ranges, which may be defined and dimensioned on an individual basis by the service provider, and indicate all permissible IP addresses for either commercial customers or residential customers. This is because these two groups of customers are treated differently in event of a fault.

If required, a customer may define its own block of IP addresses though a prefix (which can define just one IP address if necessary), allowing their existing IP addresses to be retained if required. Another block of the same size is assigned to a secondary prefix for interfaces on the secondary LR-PON. Thus a pair of LR-PONs configured for dual homing may have more than one pair of prefixes associated with them. Each pair would be communicated to LATTE, for use when making re-routing decisions. DHCP can configure such selected customers with static addresses if necessary. Providing the relevant address tables are configured correctly, LATTE can cope with such exceptions.

2.2 Re-routing in the event of failure

Table 1 is a sample routing table generated by FROTH, which is used by LATTE to make IP re-routing decisions. Normally, both the primary and secondary OLTs are available, so there is no re-routing – an IP datagram addressed to the primary LR-PON is indeed routed via that LR-PON's OLT. The first three columns in the table indicate the prefixes associated with the primary and secondary LR-PONs, and the number of IP addresses that exist within these prefixes. The next two columns indicate whether the primary and/or the secondary LR-PONs are working and available. In this case, if the address of an incoming IP datagram falls within the range of 1024 IP addresses defined by the prefix 177.67.40.0/22 (row 1)[1], it is re-routed to the corresponding secondary address, which can be deduced from the primary address. An IP-in-IP tunnel [4] is one way to implement re-routing, where each IP datagram to be re-routed is placed in the payload of another IP datagram.

A source sending datagrams to a commercial customer is unaware that the ultimate destination address may be on the secondary LR-PON. This is illustrated in Figure 2, where the customer's primary IP address is 177.67.40.13, and its secondary IP address is 177.67.44.13. The total number of customers that require full end-to end equipment diversity and circuit routing separation (and hence are protected by the above scheme) will be considerably smaller than the number of residential customers, hence the size of prefix for such commercial customers will be correspondingly smaller than for residential customers.

All traffic (IP, Ethernet, TDM etc.) raises the general underlying issue of many traffic streams from multiple points in the network requiring routes to primary and backup locations. Different platforms could be adapted to take advantage of the approach introduced here, with signaling within FROTH still being carried over IP. For example, another table similar to Table 1 could contain MPLS LSP identifiers, and an MPLS packet could be tunneled within another MPLS LSP, instead of using IP-in-IP tunneling as before. Additionally, re-routing of some Time Division Multiplexed (TDM) technologies, such as private lines, would require an interface between LATTE and the SDH management system, but could avoid bandwidth duplication when providing protection across the inner core network.

[1] The "/22" indicates that the first 22 bits of the IP address represent the network address and the remaining 10 bits correspond to the host id and hence, 2^{10} = 1024 IP addresses are defined by this prefix.

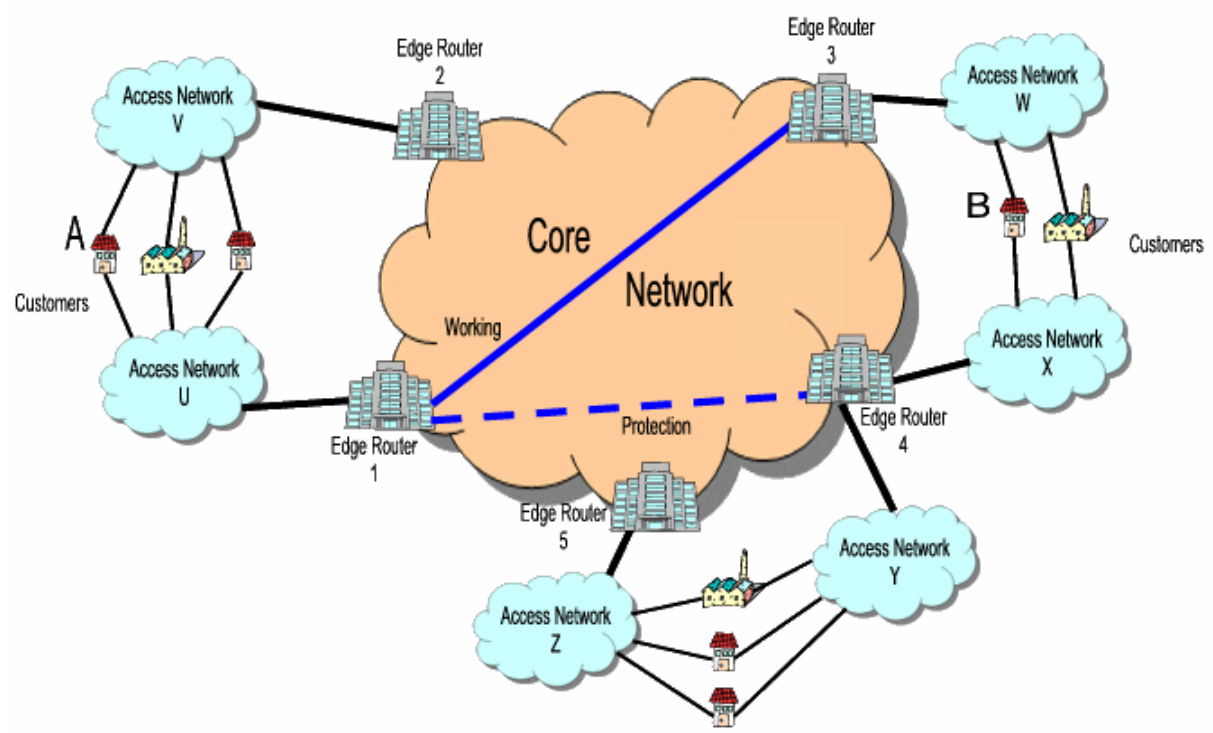

Figure 1. A generic illustration of the dual homing concept studied here. Edge routers 1 to 5 form an interface between the core network and the access networks U-Z. The access networks may be represented by several technologies, with LR-PONs being of special interest. Customers A and B normally communicate via the "working" path.

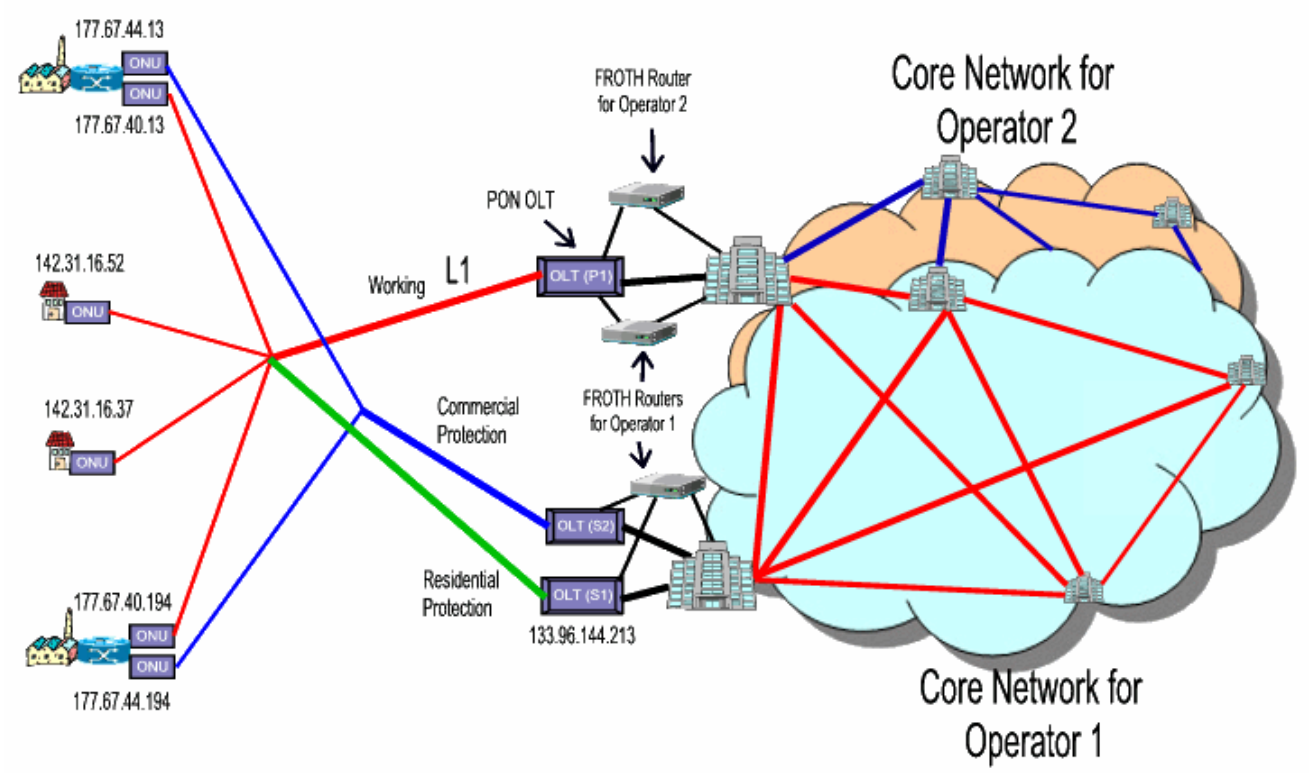

Figure 2. Example network configuration to illustrate the operation of LATTE. All customers are protected from cable failures in the backhaul regime with some (predominantly commercial) customers experiencing full end-to-end protection. The first two shaded rows of Table 1 correspond to the IP addressing scheme shown here. Multiple operators may deploy their own LATTE and FROTH capabilities or implement different protection schemes independently of one other.

Table 1. An example of a routing table supplied to LATTE by FROTH for re-routing IP datagrams. The shaded rows are illustrated in Figure 2.

Primary prefix	Secondary prefix	Maximum number of IP addresses in prefix	Primary working?	Secondary working?	OLT address for residential customers	Manual re-route?
177.67.40.0/22	177.67.44.0/22	1024	no	yes	0.0.0.0	no
142.31.16.0/25	142.31.16.128/25	128	no	yes	133.96.144.213	no
152.52.4.0/24	152.52.5.0/24	256	yes	yes	0.0.0.0	yes
etc.…	etc.…	etc.…	etc.…	etc.…	etc.…	etc.…

Residential customers will not have paired (primary and secondary) IP addresses, hence it is difficult to route datagrams to such a customer via two different OLTs. To ensure that each datagram is re-routed correctly, the tunnel must terminate at the secondary OLT, which then de-capsulates the IP datagram and forwards it to the customer. The secondary OLT is connected to a different edge router from the primary OLT (Figure 2), and, besides being capable of de-capsulating IP-in-IP, it must have its own IP address on its interface to the edge router. In the example of Figure 2, the IP address allocated to the lower OLT is arbitrarily set to 133.96.144.213. Although performing a manual revert upon repair of the main link may have a negative customer impact, the lack of IP-in-IP encapsulation would reduce the amount of traffic in the core network. Furthermore, IP-in-IP has other drawbacks which are discussed shortly.

As noted earlier, the secondary OLT associated with residential customers should re-range prior to receiving the first re-routed data packet from the source. It is instructed to do so either by a notification packet received from the FROTH router adjacent to the fault, or via the usual FROTH timeout mechanism. If the diverted data packets arrive before re-ranging is complete, they must be buffered (or dropped if time to re-range is excessive) by the local FROTH router until the OLT is ready for operation.

The last column of the table is labeled “manual re-route” and may be configured manually by the network administrator. This allows re-routing without any fault to take place if necessary, primarily for network maintenance. LATTE could be implemented on each OLT, but it is probably better located in the multiplexer which concentrates OLT signals into the edge router, permitting economies of scale to be realized.

2.3 Address substitution and IP-in-IP encapsulation

IP-in-IP encapsulation [4] has several disadvantages in this context, despite being a widely accepted technique. It increases the size of each datagram by 20 bytes, implying that the hardware to implement it must do more than merely re-write IP header fields. Given the impressive capabilities of current technology, this is unlikely to be a significant difficulty, especially since LATTE is best located in the multiplexer feeding the edge router, which would implement such functions in any case.

More significantly, for any given value of MTU (Maximum Transmission Unit), IP-in-IP encapsulation reduces the possible payload size by 20 bytes. This may cause problems with those applications using UDP which tend to transmit long packets, as it would cause fragmentation. The effect on TCP is more subtle. When TCP sets up a connection, it determines the MTU via path MTU discovery. If, while a TCP connection is in progress, it is diverted into an IP-in-IP tunnel, the MTU will effectively decrease by 20 bytes without TCP’s knowledge. If TCP now transmits a packet with the size of the old MTU (which is very likely), it will be fragmented by IP when it passes through the tunnel.

Hence fragmentation takes place with both UDP and TCP. Although the resulting degraded efficiency is generally undesirable, it may be tolerable for a short time, since it only arises temporarily after a fault. The additional headers use up transmission resources, while the transmission of additional fragments requires more computation in routers. Also, loss of an individual fragment means that everything in the original datagram must be re-transmitted, with re-assembly then being complex. Incidentally, if a routing protocol such as OSPF reconfigures, a similar effect could occur.

These problems may be avoided by substituting the new IP address into the datagram’s header, without tunneling. However, in some circumstances, a local IP address is specified by a host when setting up a socket, meaning that only datagrams having that destination IP address are acceptable. Such address substitutions would then confuse IP. This is resolved by making the secondary IP address *on a dual-homed customer’s premises* equal to its primary address, thus avoiding such addressing difficulties there. However, the secondary IP addresses shown in Figure 2 are used by the core and its routing protocols, although both customer ports have the same IP address. To permit forwarding through the core, this secondary IP address is replaced by the primary IP address in each datagram emerging from the core, when it passes through LATTE to a LR-PON.

3. FAST RECOVERY FOR OLTS VIA TRANSMISSION OF HELLOS (FROTH)

The FROTH protocol distributes availability information about each cluster of LR-PONs around the network area, where a “cluster” constitutes all the LR-PONs connected to a particular edge router. To avoid any complications with vendor support, and to make the discussion as general as possible, it is assumed that FROTH runs on a separate computer called the FROTH router, which is physically located beside the corresponding edge

router, but is relatively inexpensive. This avoids the need to modify the edge router or its software, or to introduce further complexity elsewhere. In Figure 1, FROTH is implied within each edge router.

Besides distributing LR-PON availability information, FROTH also ensures that LATTE can deduce the secondary IP address of each dual-homed customer from the primary IP address. This information is held in prefix form to reduce table size and facilitate rapid searching.

The implementation of FROTH discussed in this Section uses IP, a layer 3 protocol, for signaling. Two points must be borne in mind:

1. The general concept underlying FROTH makes availability information for all LR-PONs available at all edge routers. Although it is convenient to carry the signaling information for this protocol over UDP/IP (as discussed below), it could in principle be carried by any suitable means. Indeed, there may be benefits to using a circuit-switched TDM technology, because, as is discussed below, best effort IP offers no hard bounds on the time for information to propagate through the network – it is only possible to specify the probability that a given bound will be met.
2. Although transport of signaling over IP is assumed here, this does not imply that the information used by FROTH can only be used to re-route IP. It can be used in principle to re-route almost any protocol, in the link layer, MPLS or the network layer. Indeed, it is conceivable that FROTH could be used with other re-routing schemes, other than LATTE.

FROTH uses a much simpler signaling method than, for example, OSPF or SDH [5], and implements simpler computation at each router. This is because it carries signaling information over the existing core network while avoiding complex processing in intermediate routers along the path, in order to reduce signaling time and hence convergence time. Each FROTH router is informed of local OLT or LR-PON failures via a specially designed and optimized local area network, and forwards this information via IP status packets to all the other FROTH routers in the network. Status packets may also contain any relevant IP addressing information for LR-PONs in the cluster. Two classes of failure are considered – failure of the LR-PON link (or path) or LR-PON OLT, and failure of the edge router or FROTH router.

3.1 Operation of FROTH

For LATTE to function, each FROTH router must have global knowledge, over the area or whole network, about which other LR-PONs are available, and what IP address prefixes are associated with commercial and residential customers in each of them. If another LR-PON OLT is unavailable, it could be due to the failure of, for example, a LR-PON path, its OLT, or the associated edge router or FROTH router. Based upon this information, a FROTH router updates its routing table, which is passed on in modified form to LATTE, so that if the primary destination LR-PON OLT is unavailable, datagrams can be re-routed to the corresponding secondary destination LR-PON.

3.2 Status packets

Each status packet contains several records, as well as overhead information. The latter includes an MD5 (or other) digest of a text password, to address security issues. Each record relates to one OLT, or to an edge router and its associated FROTH router. Each FROTH router sends a status packet to every other FROTH router at fixed intervals via the core network, which reports on one OLT at a time (taken in turn) and also the edge router and FROTH router. Also, a status packet is sent out immediately when an OLT fails or is repaired, and it is hence necessary to inform all other FROTH routers. A FROTH router can only originate a new record in this way (rather than relaying existing information) when changes occur adjacent to its own edge router. Unlike more catastrophic failures, this does not require a timeout, since all other FROTH routers can be informed as quickly as possible. Hence each FROTH router knows exactly which other OLTs are available and which are not, a requirement for LATTE to operate. Each FROTH router forwards incoming records about all other LR-PONs elsewhere in the area or network, providing the associated timestamp indicates that these records supersede its existing data.

When a FROTH router receives a status packet it determines whether the information is more recent than its existing record for each device represented. If so, it updates this record and immediately broadcasts another status message containing the new information to all other FROTH routers. This is crucial to the robustness of the protocol, ensuring that even if a link or router in the core has failed, the information still reaches its destination via another route. Its impact on performance is demonstrated in the modeling work later.

Also, this dramatically reduces the probability that FROTH erroneously detects failures triggered by core packet losses, for example due to protection events or congestion. FROTH routers forward all new updated information they acquire to all other FROTH routers. Hence, regardless of packet loss, availability information about any particular cluster still reaches every other FROTH router, although perhaps by an indirect route, except in pathological cases of network failure. Furthermore, this arrangement has the desirable advantage of expediting propagation of signaling information between FROTH routers, since the receiver need only wait for the first of several similar messages to arrive.

4. PERFORMANCE AND SCALABILITY OF FROTH

4.1 Recovery time

The performance and scalability of FROTH has been evaluated, with respect to both fault recovery time and signaling traffic level. The recovery time is modeled by combining appropriate delay components for each recovery scenario. These components include reporting the fault, updating a FROTH routing table, signaling from FROTH to LATTE, or re-ranging operations on a LR-PON. There is also a variable delay term, modeled via statistical measurements from the Internet, representing the transit time of an IP datagram from one point in the network to another. Existing measurements of TCP Round Trip Times (RTTs) provide useful insights into the statistical distribution of propagation delays in any IP network. Based upon reported measurements

[6], the propagation delay along a path through the network is modeled as a random variable whose tail is Pareto distributed, while the remainder of the probability density function is Gamma distributed.

Four principal scenarios are modeled, each with different timing. In each there can either be failure of a LR-PON, OLT or backhaul connection, or there can be failure of an edge router or FROTH router. Also, there is a choice between residential customers and commercial customers. As noted earlier a residential customer cannot have service restored if the access fiber fails. The end-to-end recovery time is evaluated by the model, being defined as the time from the primary destination LR-PON becoming unavailable, to the first packet of re-routed user data arriving at the secondary destination LR-PON OLT.

Assume that traffic is destined to a customer on one primary LR-PON OLT but the main link fails, hence the traffic must be re-routed to the secondary LR-PON OLT. Table 2 summarizes the results. The first column indicates the type of failure being examined – 'PON' is for cable failure or other failures for which the FROTH router can transmit an appropriate failure message, whereas 'edge' indicates failures for which a timeout is necessary. The second column represents the classification of customer being considered – commercial or residential. The third column indicates the length of time required to pass before a timeout is detected, hence initiating fault recovery. The fourth column indicates the target recovery time for failures, and the remaining columns indicate the probabilities that traffic can be restored to a secondary edge router within such a time given the mean transmission delay of the network (indicated in milliseconds).

Table 2. Summary of results. The last three columns provide the probability that traffic would be restored within the delay target, for mean network path delays of 5, 10 and 15 milliseconds.

Fail Type	Comm/ Res	Time -out	Delay target	5ms	10ms	15ms
PON	comm	N/A	50ms	98.9%	97.5%	96.2%
PON	res	N/A	50ms	98.8%	97.4%	96.1%
edge	both	50ms	100ms	99.8%	99.4%	99.2%
edge	both	50ms	150ms	99.9%	99.7%	99.5%
edge	both	18ms	50ms	99.8%	99.4%	99.2%
no false timeout		18ms	N/A	8 nines	6 nines	6 nines
no false timeout		50ms	N/A	8 nines	6 nines	5 nines

The first and second rows show that between 96% and 99% of traffic could be restored within a target of 50 milliseconds (dependent on size of core network) given a failure for which the FROTH router can transmit an appropriate message. Additional analysis shows that the time taken for traffic to start leaving the far-end correctly addressed to the secondary LR-PON OLT, having received the signal to re-route the traffic, exceeds 50 milliseconds with negligible probability. The possibility of higher delays when re-establishing traffic is largely due to the data traffic traversing the network as opposed to signaling delay. Perhaps the statistical density function of network transmission delay chosen here exacerbates this effect, and a fuller study, considering a variety of further experimental data, would be advisable to understand the sensitivity of this effect. Recovery from edge router failure (rows 3, 4 and 5) is slower, but, as expected, the time between transmissions of status messages may be reduced (within the limits of feasibility), implying a corresponding reduction in timeout interval, and hence improving performance. A timeout interval of 18 milliseconds, for example, is expected to allow sufficient time for service recovery to complete within 50 milliseconds. Finally the last row emphasizes that 'false' timeouts are possible (albeit with low probability). depending on the timeout limit and the mean delay through the network for status messages.

4.2 Traffic level due to status packets

The level of network signaling traffic is important since it influences scalability. Assume that all datagrams have a 20-byte IP header, an 8-byte UDP header, and 22 bytes of status packet overhead. Each record is 6 bytes long. Status packets are transmitted every 15 milliseconds via IP multicast, the number of OLTs per edge router is 200, the number of edge routers is 100, and a core link carries data at 60Gbit/s (6 wavelengths, each at 10Gbit/s).

With these figures, it can be shown that status packets occupy 0.02% of the total available network bandwidth. If they are now transmitted every 5 milliseconds, the signaling traffic approximately triples, and 0.06% of the total available network capacity is now used for signaling. This may be desirable in some scenarios, since it reduces recovery time when an edge router or FROTH router fails.

4.3 Use of areas with multiple operators

It is now shown that FROTH as proposed does not scale to very large networks. As the number of edge routers increases, the amount of signaling traffic becomes impracticably large. To ensure scalability, the network may be divided into areas, although these need not correspond to OSPF areas. LATTE is implemented on each link at the boundaries between areas, or between an area and the international peering points; in fact whenever a signal enters an area. Between areas, carrier grade equipment is required, due to the traffic volume.

Assume the worst case, where the full network of 100 edge routers constitutes one area. The amount of signaling traffic is $O(N^2)$, because each edge router sends information about each other edge router at regular intervals. Multicast routing reduces the overall network capacity devoted to signaling traffic, and enhances scalability. If the network of 100 edge routers were divided into four areas of 25 edge routers, then the amount of signaling traffic would be reduced by a factor of approximately 16. In general, the mean propagation delay through a network is approximately proportional to the mean number of hops in a path, $O(\sqrt{N})$, so while propagation delays through each area would on average be roughly half those through the whole network (with a corresponding decrease in recovery time), the principal improvement would be in reducing signaling traffic.

Areas also permit several telecommunications operators to co-exist, even if they do not all use FROTH, LATTE or LR-PONs. Each area may be owned by a different operator, although no operator need use FROTH and LATTE if it does not wish to. In that case, the operator does not need to be aware of these

protocols, nor does it need to make any concessions or adjustments because the other operators are using them.

In such a multi-operator scenario, LATTE is only necessary when re-routing traffic flowing into an area which uses FROTH, either from another area or from a LR-PON. FROTH only gathers information about its own area – there is no communication between FROTH implementations in different areas. Operators not using FROTH and LATTE are unaware that they are being used elsewhere.

Furthermore, several operators may access one LR-PON, where traffic on the core side of an OLT is split between them. Thus FROTH and LATTE permit flexible configuration of the network to support multiple operators.

5. CONCLUSIONS

If a simplified and low-cost network architecture, consisting of Long Reach PONs linking customers directly to a core of a hundred or so switching and intelligence centers (edge routers) is to be realized, then levels of resilience and protection as good or better than those currently experienced are essential. Conventional techniques for protecting dual-homed traffic onto two separate routers can be slow to re-converge after failures such as cable breaks, thus potentially leading to poor customer satisfaction. Clearly new approaches are required.

This paper has described two new techniques – LATTE and FROTH – that can re-route traffic in the event of a variety of failure scenarios, thus ensuring that traffic originating from elsewhere in the network reaches the correct destination. Calculations have shown that for LR-PON path and OLT failure, it is realistic to expect re-routed data to leave the transmitting edge router well within 50 milliseconds. Transmitting user data to the secondary LR-PON limits the overall speed of recovery. For more catastrophic failures (namely edge router and FROTH router failure), recovery may take 100 – 200 milliseconds, depending on how frequently each FROTH router transmits status messages. Essentially there is a trade-off. If status messages are transmitted more frequently, there is a higher network load due to signaling messages, but recovery from edge router or FROTH router failure is faster. The speed of recovery from LR-PON path or OLT failure is not affected by how frequently status messages are transmitted, since a status message is sent immediately if a LR-PON or OLT fails.

The discussion above assumed that edge router failures are detected by the absence of status messages, invoking a timeout. Alternatively, an edge router failure could be detected by its neighbors through BFD (Bidirectional Forwarding Detection) or similar, which would relay this information via status messages to all other edge routers. This would avoid multicasting of status messages to all edge routers at regular intervals, since "hello" messages would be sent by BFD, but only between pairs of neighboring edge routers. It could expedite failure detection, without the need to multicast repeated status messages, while potentially reducing signaling traffic.

This paper reports early proposals identifying robust and scalable mechanisms to meet the reliability requirements of future communications services. Alternative solutions may well arise during the course of further investigation into new protection mechanisms; nevertheless, this paper underlines the importance that the issue of resilience presents when considering the evolution of future communication networks towards Long Reach PON deployments.

6. ACKNOWLEDGMENTS

British Telecom supported the work of David Hunter under its Short Term Fellowship scheme. The authors acknowledge Rob Booth, Russell Davey, Alan Hill, Alan McGuire, Ben Niven-Jenkins, Albert Rafel and Peter Willis (all of British Telecom) for their constructive and helpful comments during this work. Similarly, Ian Henning and Martin Reed of the University of Essex are thanked for their input.

7. APPENDIX – MODELING OF FROTH

In this Appendix, the mathematical model of recovery time used in this work is outlined briefly. Since packet transmission is a statistical process, recovery time is expressed as a statistical distribution rather than an absolute value. The probability that the recovery time is greater that some target delay, usually 50 milliseconds, may be calculated from the model. Table 2 was produced by developing analytical formulae for tail probabilities of recovery time in various scenarios. These were expressed in terms of the components described below, and were then evaluated numerically via a short program written in C.

7.1 Components of network recovery time

The recovery time is modeled by combining several appropriate delay components, thus modeling several different recovery scenarios. Firstly, there is a variable delay term, modeled via statistical measurements from the Internet, representing the transit time of an IP datagram from one point in the network to another. Secondly, there is a fixed delay term representing, as appropriate, the time to:

- Diagnose a fault on a LR-PON or OLT, and report it to FROTH: assumed to be 1 millisecond throughout
- Update the routing table on a FROTH router: assumed to be 1 millisecond throughout
- Time for FROTH to inform LATTE of a change in the routing table: assumed to be 1 millisecond throughout
- In the case of backhaul protection, the time for a LR-PON to re-range: assumed to be 30 milliseconds throughout

One FROTH router informs another that an OLT or LR-PON has changed status through several different parallel datagram propagation paths. Besides a direct transmission from source FROTH router to destination, transmission also takes place in parallel via other similar intermediate routers (see Section III). The model considers the delay distribution of the first status packet to arrive at the destination via these different routes, since this carries out the necessary notification. In fact, this significantly accelerates the distribution of status information around the network. As inputs to the model, each of these separate paths has its own values of mean propagation delay and fixed delay – as noted above, the latter is assumed to be a constant for each path.

By employing existing published traffic measurements, a model was developed which determines the approximate probability that

the recovery time is greater than a specified value τ_R, which is usually assumed here to be 50 milliseconds. This model is not described in detail here, but can be found in Reference [3].

7.2 Modeling of network propagation delay

Although Internet Round Trip Times (RTTs) superficially appear to be Gamma distributed [7], other studies have shown that in fact the distribution has a heavy tail. The tail probability is critical since the probability that the recovery time is greater than 50 or 100 milliseconds is required. Two independent studies have shown, via rigorous statistical tests, that the tail can be approximated reasonably accurately by a Pareto distribution [8, 9].

Based upon reported measurements [6], the propagation delay along a path through the network is modeled by a random variable with a Pareto distributed tail, while the main body of the pdf is Gamma distributed. The following mathematical definition of the probability density function was derived from published network measurements [6]:

$$f(t) = \begin{cases} 0 & 0 \le t < 7.78 \\ \dfrac{\left(44.44\left(t-7.78\right)\right)^{3.416}}{e^{44.44(t-7.78)}} & 7.78 \le t < 8.13 \\ 0.244/t^{1.637} & 8.13 \le t < 300 \\ 0 & t \ge 300 \end{cases}$$

The mean value of this distribution is 11.05 milliseconds, but it may be scaled to have any desired mean. This distribution models propagation delay through the network, but *t* may have a constant term subtracted from it, to reflect fixed delays due to processing or timeouts, for example. Four scenarios are modeled, each with different timing:

- Failure of LR-PON, OLT, backhaul, or access fiber – recovery time for commercial customer
- Failure of LR-PON, OLT or backhaul – recovery time for residential customer
- Failure of edge router or FROTH router – recovery time for commercial customer
- Failure of edge router or FROTH router – recovery time for residential customer

8. REFERENCES

[1] D. B. Payne, R. P. Davey, "The Future of Fiber Access Systems?", *British Telecom Technology Journal*, vol. 20, no. 4, October 2002, pp104-114.

[2] ITU-T, "Gigabit-capable Passive Optical Networks (GPON)", recommendation G.984.1,2,3,4.

[3] D. Hunter, "Failure Recovery in Networks with Dual Parented Customers", *BT Technical Report*, June 2006.

[4] C. Perkins, "IP Encapsulation within IP", RFC 2003, October 1996.

[5] J.-P. Vasseur, M. Pickavet, P. Demeester, *Network Recovery – Protection and Restoration of Optical, SONET-SDH, IP and MPLS*, Morgan-Kaufmann, 2004.

[6] A. Corlett. D. Pullin, S. Sargood, "Statistics of One-way Internet Packet Delays", *IETF Internet Draft*, draft-corlett-statistics-of-packet-delays-00.

[7] A. Mukherjee, "On the Dynamics and Significance of Low Frequency Components of Internet Load", *University of Pennsylvania*, Technical Report MS-CIS-92-83/DSL-12.

[8] D. Loguinov, H. Radha, "Large-scale Experimental Study of Internet Performance using Video Traffic", *ACM SIGCOMM Computer Communication Review*, vol. 32, no. 1, January 2002, pp7-19.

[9] K. Fujimoto, S. Ata, M. Murata, "Statistical Analysis of Packet Delays in the Internet and its Application to Playout Control for Streaming Applications", *IEICE Transactions on Communications*, vol. E84-B, pp1504-1512, June 2001.