Light-Trail Testbed for Metro Optical Networks

Nathan Vanderhorn, Srivatsan Balasubramanian, Mani Mina, Robert J. Weber, Arun K. Somani
Department of Electrical and Computer Engineering
Iowa State University, Ames, IA 50011
E-mail: {nvander,vatsan, mmina,weber,arun}@iastate.edu

Abstract—Telecommunication networks have rapidly added staggering amounts of capacity to their long haul networks at low costs per bit using DWDM technologies. Concurrently, there has been a wave of new access technologies that are driving customers to demand high-speed, robust and customized data services. These dynamics have led to what is called the "metro gap" - the inability to leverage the backbone capacity to create and distribute revenue generating services. This paper presents work1 in progress at Iowa State Universities’ High Speed Systems Engineering laboratory to address the metro gap problem. As an initial step towards solving this problem, we demonstrate a streaming media application implemented utilizing Field Programmable Gate Arrays (FPGAs) on a 3 Gbps optical fiber network employing light-trail technology [1]. The testbed and application presented within illustrates a cost-effective platform and outlines high-speed system level design challenges and solutions. This complete solution enables high-bandwidth services to move closer to the user premises by combining commercially available network components and emerging network technologies.

I. INTRODUCTION

The telecommunications industry has been witnessing an exponential growth of network traffic in the past few years. The aggregate bandwidth requirement of the Internet is expected to be well over 5000 petabits/day by 2007. While voice traffic growth has been slow for many decades, there has been a surge in the growth of data traffic. Data is expected to be over 75% of the total network traffic seen in the Internet. This brings to highlight some interesting and challenging issues related to self-similar nature of Internet traffic, asymmetry in IP flows and server bound congestion [2]. With the continuing proliferation of bandwidth-intensive multimedia applications and widespread availability of broadband access technologies, this paradigm shift in capacity demands are having profound impacts on today’s network design and deployment.

Telecommunication networks can be roughly organized into a three-tiered hierarchy: access, metro and long haul [3]. The access networks provide the subscriber interface to the communication network. It hosts a broad range of protocols/technologies and supports a wide variety of application devices. A discussion on the plethora of access solutions pervading the market is beyond the scope of this paper. On the other end of the hierarchy is the long haul, which provides large tributary connectivity between regional and metro domains. There has been an unanimous agreement among backbone service providers that DWDM offers the best cost-capacity trade-off and hence is the technology of choice for long haul networks. Interfacing the access with the long haul is the metro. The metro segment provides high speed media and application devices required to interconnect the access networks to the core. The emerging trends in traffic have significantly altered the domains bordering the metro and have made service providers seriously rethink the current technologies that are in place. Evaluating various alternatives and providing new solutions with good price-performance characteristics for the metro space is the theme of this paper.

Towards this end, this paper is organized as follows. In section II, we describe the traditional architectures that were designed with primary focus on voice. Section III looks at some of the emerging trends and network requirements that will clearly bring out the reasons why the conventional architectures are ill-suited to cater to the needs of the evolving demands. Next, section IV discusses some of the proposed solutions based on Next Generation SONET (NGS), Next Generation Ethernet (NGE) and WDM technologies and analyzes their capabilities and limitations. We introduce an architecture called light-trails in section V which when deployed with WDM can lend itself naturally to the service provisioning requirements in the metro space. We illustrate the light-trail ring and the mesh switch architectures and explain how it compares with the traditional circuit/burst/packet switched WDM architectures. As a proof of concept, we demonstrate a streaming media application implemented on a 3 Gbps optical fiber network employing light-trail technology in section VI. Section VII describes some possible future directions and section VIII presents our conclusions.

To the best of our knowledge, we are not aware of other existing light-trail test beds. Our idea is synchronous with the general observed trend of optical technologies gradually propagating from the core towards the metro and eventually to the access. The detailed discussions that follow bring a practical perspective to the testbed that we describe later.

II. TRADITIONAL ARCHITECTURES

Currently, metro networks are based upon SONET/SDH ring architectures and are organized into a two-level hierarchy: metro edge and metro core. The metro edge refers to the space between subscriber access and central office location. Metro edge rings span about 10 to 40 kms, operate at OC-3/STM-1 or OC-12/STM-4 rates and employ Add Drop Multiplexers (ADMs) that connect to digital loop carrier setups, enterprise networks, telephone public branch exchanges etc.

1The reported research is funded in part by NSF grant ANI-0087746, ANI-0434872 and ANI-0306007
Most edge traffic is usually outbound from the local ring and hence exhibit strongly hubbed traffic patterns [3] with central office as the hub. This makes edge networks well suited for Unidirectional Path Switched Ring (UPSR) architectures.

The metro core refers to the rings that interconnect major central office hub locations and that feed into long haul networks. Metro core rings span about 40 to 80 kms, operate at OC-48/STM-16 or OC-192/STM-64 rates and perform a higher level of aggregation than the corresponding edge rings. The traffic demands in metro core are much more meshed and improved bandwidth efficiency is obtained through Bidirectional Line Switched Ring (BLSR) architectures. Digital cross-connects that can switch in both space and time are used to interconnect rings and to provide fine granular bandwidth management. The traditional ring architectures performed well when the dominant traffic was voice. However, there have been some emerging trends (discussed below) in design and deployment of optical networks that bring to the forefront the inherent deficiencies in existing architectures.

III. EMERGING TRENDS

A. Growing demands

The tremendous growth in internet traffic volumes is fueled by content-rich applications like packetized voice, internet gaming, video on demand, and streaming multimedia. New services that are offered include interconnecting and consolidating data centers and transparent extension of the LAN across the MAN. There is a trend towards supporting Storage Area Network (SAN) architecture, real-time transactions backup, high-speed disaster recovery, grid computing and the more futuristic optical virtual private networks. Concurrently, there remains a very healthy demand for legacy voice and leased-line services, arising from a huge, entrenched base. It is important to note that the bursty nature of data traffic requires that network design be different from the conventional telephony design. For instance, the edge buffering capabilities have to be increased significantly to account for the self-similar nature. Also, it is observed that IP flows are asymmetric which is attributed to the pattern of big server farms sending out large data in return for small requests. Current SONET based networks are bi-directional and hence half the resources are idle leading to lopsided network utilization.

B. Advancing access technologies

Many technologies are emerging in the access domain including cable, DSL, high speed wireless, wavelength leasing and wavelength on demand. Improved access technologies make possible wide spread use of bandwidth-intensive applications which in turn create the need for more efficient access networks thereby entering a positive regenerative cycle. Thus, there is need for a scalable, robust and easy to manage network architecture that can support multiple access technologies and provide intelligent handling of broadband user data flows.

C. Increasing need for transparency

Transparency is one of the key requirements of a future-proof network. An all-optical network is transparent to bit rates, modulation formats and protocols and can upgrade to higher bandwidths without resorting to "forklift upgrades" that require massive overhaul of existing infrastructures. This enables metro operators to scale their networks to meet customer requirements and enhance their service velocity. Elimination of electronics in the intermediate nodes lowers costs and power consumption. It also simplifies operations, since there is no need to manage disparate network elements. It offers support for legacy services and gives operators the ability to bundle services with different optical quality-of-service and service-level agreements. This feature allows the service providers to tailor service offerings to meet the needs of specific customers.

D. Migrating to mesh

Traditional telecommunication networks were configured as rings since they guarantee recovery and lead to predictable restoration paths thereby simplifying management. Fiber usage can be low in ring solutions because of the requirement for protection fibers on each ring. A mesh physical topology is more efficient when the demand pattern is also meshed. Besides, network designs rarely resemble rings since fibers can be routed only along rights-of-way which may not facilitate a ring topology. Building rings on top of meshed fibers results in a logical overlay which is harder to design and maintain. Mesh networks allow a topology similar to fiber routing. Also, the benefits in flexibility and efficiency of mesh networks are potentially great. Protection can be based on shared paths, thereby requiring fewer fibers for the same amount of traffic and can lead to efficient wavelength utilization. However, mesh networks require a high degree of intelligence to perform the functions of protection and bandwidth management, including fiber and wavelength switching.

E. Improving reconfigurability

Conventional networks are circuit switched and are interconnected by leased lines with long holding times. However, there is an increasing need for reconfigurability in optical networks that allow bandwidth creation in real time between end users to accommodate dynamically changing traffic demands. The routers and switches should acquire the ability to set up circuits of wavelength or sub-wavelength granularity across optical backbones within seconds. Such provisioning will allow customers to buy high bandwidth for short-term use, such as a high-definition video transmission that a television network might need.

IV. METRO SOLUTIONS

In light of above trends, SONET based metro networks are facing some serious limitations [3], [4]. We describe some of the challenges faced by the conventional networks and then critically assess some of the promising solutions that are available to meet these specific requirements.

A. Challenges

With growing demand, capacity exhaust problems gain significance. Capacity upgrade in SONET is possible either
through deploying new rings or through increasing TDM rates. The former requires new fiber routes while the latter necessitates equipment upgrades on all ring nodes both of which are expensive and time consuming. In SONET, each transport path has a fixed bandwidth defined over a rigid rate hierarchy. This precludes the possibility of supporting a multitude of client data applications resulting in large bandwidth inefficient mappings. Besides, the burstiness in traffic cannot be handled well since re-provisioning requires careful capacity planning and takes a long time. The network is not transparent, supports only constant bit rates and provides very little room for service differentiation. So, a need for a transparent, cost-effective architecture that can respond to dynamic traffic needs and allow for service differentiation while still offering support for legacy services is being increasingly realized.

B. Metro core solutions

The requirements for metro core are different from that of the metro edge. In the metro core, the emphasis is on scalable bandwidth provisioning. With maturing optical technologies, ring- or mesh-based wavelength routed DWDM networks is an ideal fit here since it offers rapid provisioning, service transparency and low network costs (since they are amortized over a large user base). However, in the metro edge, the focus is on protocol heterogeneity, heavily sub-wavelength traffic and a price-sensitive limited user base. Hence the metro edge is seeing more diverse possibilities, ranging from improved SONET/SDH and Ethernet offerings to optics-based paradigms. We discuss each of them in turn below.

C. Metro Edge solutions

1) Next Generation SONET: Recently, new techniques for bettering transport over fiber have been added to NGS [5] while still retaining its original protection and performance monitoring features. This includes the Generic Framing procedure (GFP), Link Capacity Adjustment Scheme (LCAS) and Virtual Concatenation (VC) mechanisms. VC allows for concatenation of several payloads to provide flexible bandwidth and to minimize mismatch in data and port rates. GFP provides a simple framing technique [6] to multiplex multiple client protocols and LCAS specifies a control mechanism to dynamically adjust the number of tributaries assigned to a connection. Collectively, these features are the building blocks of the new data-aware NGS transport networks.

Despite the above enhancements, NGS is still an approach that attempts to bridge the packet and circuit switching paradigms, both of which differ fundamentally in their philosophies. NGS systems process the signals electronically in all the intermediate nodes thereby precluding transparency, reducing scalability and leading to increased equipment costs. Besides, NGS also has some framing requirements like STRATUM timing [7] and pointer processing which can become expensive at high data rates like 40 Gbps.

2) Next Generation Ethernet: The features that are exclusive to SONET is its efficient support for survivability and performance monitoring. Ethernet services, on the other hand, are easily upgradeable and has the advantages of familiarity, simplicity and low cost. While Ethernet does not offer TDM-level guarantees for bandwidth and delay, SONET does not offer efficient data mappings. NGE is a ring based cost-effective and fault tolerant data transport solution that combines statistical multiplexing along with a fairness based access scheme called Resilient Packet Rings (RPR) [8].

However, there are some problems associated with the packet scheduling and rate adaptation approach followed by RPR. The scheduling stream gives priority to transit traffic over local traffic and hence delay seen by a node is dependent on upstream traffic patterns. In addition, if the bandwidth requirement of a newly arriving traffic is lowest among the contending traffic flow, this causes all the upstream nodes to throttle their rate to this lowest rate, creating large oscillations in bandwidth allocation. Such a reactive approach in the presence of bursty traffic may result in large settling times for the oscillations. In general, packet rings have been designed based on enterprise requirements and consequently there is less support for TDM traffic. Since RPR terminates traffic on every node like NGS, their capacity scalability and cost-effectiveness is also questionable.

3) Course Wavelength Division Multiplexing (CWDM): WDM is the sole technology that can support TDM, data, SAN, cable video etc. independent of bit rates and protocol formats. Although the other alternate solutions presented above may delay the deployment of ring based WDM systems, it appears to be the most compelling solution in the long run - one that combines scalability and transparency with simplicity and cost-effectiveness. Since the traffic volumes in the metro edge may not be excessively heavy as in the core, CWDM can be deployed. CWDM does not place stringent requirements on optical equipment, thereby leading to significant cost savings [4]. CWDM will allow operators to expand service offerings, support legacy services and prepare for future traffic growth.

Having cited the advantages of deploying WDM on both the metro edge and core networks, we have roughly four architectural choices: circuit, packet, burst or trail switched paradigms. We have developed light-trails as a WDM solution to address IP-centric data communication at the optical layer and we argue why it is a viable candidate for the metro networks in the forthcoming sections.

V. Light-Trails

Traditional circuit switched WDM networks [9] are provisioned for peak rate traffic due to lack of buffering capabilities in the optical domain and hence may be severely underutilized. Network utilization can be improved by equipping nodes with electronic grooming (e-grooming) capabilities that allow efficient packing of low rate streams onto high rate channels. However, grooming brings along with it concerns related to complexity, scalability, delay and transparency. Traffic engineering and statistical multiplexing gains are achievable in optical packet switched networks [10] but high speed optical switches, scalable packet parsing mechanisms and fast and large random access units have not been realized for large scale commercial deployment. Burst switching [11] provides a hybrid approach between circuit and packet switched
paradigms, but the requirement of low switch reconfiguration times as compared with the burst duration leads to significant challenges in optical switch design.

A. System Architecture

As a solution to providing high resource utilization and sub-wavelength support, we discuss light-trail technology [1], [12], [14]. A light-trail is similar to lightpath in that, it requires the establishment of a unidirectional optical circuit between the source and destination. The key difference is that some intermediate nodes can also receive and transmit data on the same channel in a time multiplexed manner.

Figure 1(a) shows a four node uni-directional light-trail in a ring network, which is a small variant of the system suggested in [1]. At every node, the signal passes through a light-trail access unit (LAU) that consists of a splitter, a shutter, a combiner and optionally an amplifier that enables drop-and-continue functionality. A simple medium access control protocol (MAC), discussed in [1],[12]-[17], may coordinate communication among nodes in the trail. A signal sourced by a node traverses all nodes downstream to it on the trail. At the splitter, a part of the incoming signal power is tapped by the receiver for local processing and the rest of the optical signal is amplified and passed to the shutter. The shutter is configured to either block or let the wavelength pass through. If the current node is the last or the first node on the trail, the shutter is configured to block this wavelength. For all intermediate nodes on the trail, the shutter lets the signal pass through. If the signal is not blocked by the shutter, it traverses the combiner before exiting the node. The combiner enables the intermediate nodes to insert its data on the trail based on the MAC protocol.

At first glance, it may not be readily apparent why the signals are split by a coupler and a part of it is sent to the next node while it could have been locally terminated and retransmitted to the next node in steady. While local termination cleans up the signal, it may lead to increased resource consumption. For instance, consider a trail with only one source and multiple destinations. If couplers were used, each of the destination nodes need not have a transmitter else each destination needs to be equipped with a transmitter that has the sole purpose of relaying traffic.

The simple MAC protocol proposed in [1] allows an upstream node to send an out-of-band control packet to indicate its desire to transmit and send its data packet after a guard band gap. The guard band is set so as to preempt any possible ongoing downstream communications. For a good description of the various MAC protocols and issues related to fairness in shared network mediums, readers may refer to [16], [17].

This bandwidth on demand mechanism helps the network handle bursty and highly variable traffic in a more efficient way as opposed to conventional circuit switched networks. The key point to note in the architecture is that the optical shutters are not switched on a per packet basis but configured only on a longer time scale as opposed to burst or packet switched networks. This prevents light-trails from being constrained by optical switching technologies since switches with large port count, low cross-talk, nano-second switching times are not commercially feasible yet. Despite the absence of dynamic switching, by sharing the medium statistically, by expanding trails to meet new demands and by tearing down unused trails in a distributed manner as mentioned in [1], light-trails are able to provide the granularity required for data-centric communication.

B. Resource Requirements

Light-trails share resources at the optical layer (called o-grooming), leading to significant wavelength and equipment savings. For example, consider a three node network (N1, N2, N3), whose physical topology resembles a simple path as shown in Figure 2. Suppose, wavelength capacity is three units, each node is equipped with a transmitter/receiver unit and traffic demand is as shown in the figure. A single light-trail can be set up passing through all the three nodes, and the wavelength can be shared by all the connections on a need basis. This traffic requires a transmitter each on N1 and N2 and a receiver each on N2 and N3. However, the same pattern cannot be carried using lightpaths (non e-groomed) even if more than one wavelength is available on each link. This is due to the fact that two connections cannot be sourced by N1 since it has only one transmitter. We designed heuristics for quantifying and comparing the wavelength and transceiver
requirements of light-trail and lightpath networks in the presence of dynamic traffic. The heuristics for static single hop trails are provided in [13] and for dynamic multi hop trails are provided in [15]. The conclusion of both the work is that light-trails can do better than non-groomed lightpaths. Electronically groomed lightpaths and light-trails may result in some additional equipment savings [15] but comes at the price of losing transparency. Light-trails, by sharing the wavelength in the optical domain achieves efficient wavelength utilization while retaining the transparency property which is key to developing a scalable and cost-effective network.

C. Mesh Switch Architecture

Figure 1(b) shows a wavelength plane switch in a mesh network [13], [15]. As in the case of the ring, every signal goes through one LAU unit. It is likely that the local node may not be active on all the trails and hence such trails are not required to be received using a local receiver. It is only for the trails on which the node is active, either through transmission or reception or both, a transmitter and/or receiver is allocated as required. This is made possible using cross-connects and tunable transceivers and leads to transceiver savings.

D. Light-Trail Metro Solutions

Having discussed the requirement of WDM in metro networks and the rationale behind trail switching in WDM networks, we see how light-trails can be designed for metro networks. We propose a CWDM light-trail ring architecture for metro edge networks. Alternate architectures like point-to-point and bus topologies are also possible. A variety of devices like GigE routers, ESCON main frames, Fiber Channel based SAN switches, ATM and telephony switches can connect to the subscriber access points on the CWDM ring/bus. Since the edge networks have hubbed traffic patterns and heavily fractional traffic as discussed before, we consider two unidirectional trails being set up as shown in Figure 3. The downstream trail is used for the hub (central office) to transmit data to all the other nodes (access points) on the ring and the upstream trail is used for the access points to transmit data to the hub. While the downstream trail has only one source, the upstream trail has multiple sources and hence needs a medium access control for upstream communication.

The demands in metro core, however, are more meshed and voluminous and hence we propose either a ring or a mesh DWDM architecture. Figure 3 shows an example metro core network configured in the form of a mesh. Nodes in the core are connected to the long-haul but is not shown in the figure. If node N1 in Figure 3 has data to be sent to node N2, the data is first sent on the upstream trail to the central office CO1 and then routed via the metro core which then reaches node N2 via the downstream trail originated by CO2.

E. Light-Trail Feasibility Check

The drop-and-continue functionality allows a wavelength to be shared by multiple nodes in time, but also leads to power budget constraints and signal-to-noise ratio (SNR) impairments. The transmission quality is measured by the received SNR which is defined as the ratio of the signal power to noise power at the decision point. The system needs optical amplification to compensate for the fiber attenuation, splitting and insertion losses due to the optical components. The effect of amplified spontaneous emission (ASE) that is introduced by Erbium Doped Fiber Amplifiers (EDFAs) on the Optical SNR (OSNR) needs to be assessed. It is important to investigate if the proposed architecture can still meet the quality of service requirements without the requirement of regeneration.

Another important parameter is receiver sensitivity, which is defined as the minimum optical power required at the detector for a specified bit error rate (BER). Due to fact that signals being sourced by nodes in the network are separated by large geographical distances, it is possible that signals traversing a link have widely differing power levels. Signals of high optical power can saturate the EDFA gain, limiting the available gain for other lower power channels or may lead to other undesirable side effects like cross modulation. This may have severe detrimental impact on signal quality, and thus, gain equalization of different wavelength channels is required. More specifically, all channels incident on an amplifier should have approximately the same signal strength to avoid the above mentioned near-far effects. Input launched power per wavelength is an important design parameter since it decides the number of wavelengths that can be launched into the fiber without saturating the amplifier or without entering the non-linear region of the fiber. The problem becomes even more
complex in the presence of dynamic traffic where connections can originate from any source and terminate at any destination passing through any set of intermediate nodes. We propose a simple analytical model that can identify the network operation point in terms of launched power levels and configuration and placement of EDFAs subject to all the constraints mentioned above.

In this section, we discuss how light-trail networks need to be designed so as to meet the gain equalization and OSNR requirements. We first reason out why signal regeneration may not be required in such a metro setting. An amplifier is assumed to be present on the incoming line section of every node. For simplicity, channel independent gain is assumed for all EDFAs. The amplifier is a single-stage, constant gain EDFA regulated by a Variable Optical Attenuator (VOA) at its input with a maximum gain per channel of 30 dB and a maximum output power per channel of 0 dBm. Without loss of generality, we assume all the links to be of equal sizes (10 kms) and with just a slight modification, the analysis can model links of unequal sizes as well. A light-trail system of size $m$, $N_1, \ldots, N_m$ can be modeled as a series of gain and loss elements as shown in Figure 5 (c) based on a variant of the approach suggested in [20]. Consider a connection that traverses from $N_1$ to the last node $N_m$. This connection suffers the maximum loss and we need to check if it meets the OSNR requirements for a specified trail length. The internal structure of nodes $N_1$ and $N_2$ are shown in Figure 5 (a). Figure 5(b) shows all links incident on nodes $N_1$ and $N_2$ apart from the links traversed by the trail (shown in red). We define a few variables:

- $P_i^S$ - signal power transmitted by a specific source on node 1.
- $P_i^S$, $i > 1$ - signal power as it enters node i (just after EDFA)
- $P_i^\alpha$, $i > 1$ - signal power at a specific receiver on node i
- $P_i^\gamma$ - ASE noise power as it enters node i (just after EDFA)
- $g_i$, $i > 1$ - gain of the $i^{th}$ EDFA block
- $\alpha_i$ - losses suffered by the signal from the point it leaves the source (after transmitter) to the point it enters node 2 (before EDFA at node 2).

Refer Figure 5(a)

- $\alpha_i$, $i > 1$ - losses suffered by the signal from the point it enters node i (after EDFA at node i) to the point it enters node i+1 (before EDFA at node i+1). Refer Figure 5(a)
- $n_f$ - noise figure of the amplifier
- $L(Device)$ - Loss in dB due to an optical device
- $G(EDFA)$ - Gain in dB due to an EDFA

At the output of the $i^{th}$ amplifier, the signal power is

$$P_i^S = P_{i-1}^S \alpha_{i-1} g_i, \ \forall i > 1$$

At the the $m^{th}$ node input (or at the $m^{th}$ amplifier output), the signal power can be calculated to be

$$P_m^S = P_1^S \prod_{i=1}^{m-1} \alpha_i g_{i+1}$$

The total ASE power out of the $i^{th}$ amplifier consists of the amplified accumulated noise and the locally generated ASE noise and is approximated as given in [?] to be

$$P_i^N = P_{i-1}^N \alpha_{i-1} g_i + g_i h \nu n_f B_0 \ \forall i > 1$$

where $B_0$ is the optical bandwidth, $n_f$ is the noise figure (usually, 6 dB) and $h \nu$ is the photon energy at 1550 nm. $B_0$ is related to the bit rate (B) and equals 10*B to guarantee negligible penalty due to output filtering. The ASE power at the $m^{th}$ node is computed as

$$P_m^N = g_m h \nu n_f B_0 [1 + \sum_{j=2}^{m-1} \prod_{i=j}^{m-1} \alpha_i g_i]$$

After some manipulations, OSNR can now simplified to

$$OSNR = \frac{P_i^S}{h \nu n_f B_0 T} \text{ where}$$

$$T = \frac{\sum_{j=2}^{m-1} \prod_{i=j}^{m-1} \alpha_i g_i + 1}{\prod_{i=1}^{m-1} \alpha_i g_{i+1}}$$

Now, suppose the gain blocks are configured such that $g_i$ compensates for $\alpha_{i-1}$, then,

$$\alpha_{i-1} = \frac{1}{g_i} \ \forall i: i \neq 1, 2$$

Equation reduces to

$$T(dB) = -\alpha_1 (dB) + 10 \log_{10} (m-1)$$

Typical insertion losses are as follows: multiplexer (6 dB), OXC (3 dB), splitter (3 dB), shutter (2 dB), combiner (3 dB), demultiplexer (6 dB) and fiber (2 dB @ 0.2 dB/km for 10 kms). As the signal leaves the source, it passes through the combiner, multiplexer and a span of length 10 kms before it reaches the amplifier of the second node. Based on the definition of $\alpha_1$ given above, $\alpha_1 (dB) = -11$ dB which allows $T$ to be calculated for $m = 20$ as $T(dB) = 23.78 dB$

The OSNR can now be computed as,

$$OSNR = P_i^S (dB) - 10 \log_{10} (h \nu B_0)$$

$$-10 \log_{10} (n_f) - T(dB) = P_i^S (dBm) + 49.14 dB$$

We first calculate the power level of the launched signal and then estimate the OSNR. From the point a signal enters a node i, $i > 1$ (after EDFA), to the point it enters the next node (before EDFA), it encounters the optical components in the following order - Demux (6dB), OXC (2dB), splitter (3dB), shutter (2dB), combiner (3dB), Mux (6dB), and fiber span of length 10 kms (2dB). The loss $\alpha_i$, $i > 1$

$$\alpha_i (dB) = L(Demux) + L(OXC) + L(Splitter)$$
+L(Shutter) + L(Combiner) + L(Mux) + L(Fiber)

and the value is computed to be \( \alpha_i(dB) = -24 dB \) \( \forall i \neq 1 \). Since \( g_i = 1/\alpha_i \), \( g_i(dB) = 24 dB \) for every EDFA block. The power level at the receiver on the \( i^{th} \) node is computed as follows. Since each amplifier at node \( i \) compensates for component losses at node \( i-1 \), and attenuation on the \((i-1,i)\) span, the power available just after the amplifier on node \( i \) is the same as the power available just after the amplifier on node \( i-1 \). So, \( P_i^g = P_{i-1}^g = P_2^g \). Upon entering node \( i \), the signal has to go through a Demux, OXC and a splitter before being detected by the local receiver.

\[
P_i^r(dBm) = P_i^g - L(demux) - L(OXC) - L(splitter) = P_i^g - 11 dBm
\]

Also, the signal powers \( P_2^g \) and \( P_1^g \) are related as:

\[
P_2^g(dBm) = P_1^g + L(Combiner) + L(Mux)
\]

\[
+ L(Fiber) + G(EDFA)
\]

Set \( P_1^g \) as follows

\[
P_1^g(dBm) = L(demux) + L(OXC)
\]

\[
+ L(splitter) + L(shutter)
\]

After plugging in the values, \( P_1^g(dBm) = -13 dBm \) and hence \( P_2^g = 0 dBm \). Since, \( P_1^r = P_2^g \), \( P_1^r = -11 dBm \).

The typical sensitivity of a PIN photodiode at 10 Gbps is about -17 dBm and hence the received power is within the sensitivity limits of a PIN photodiode. Also, the OSNR of the signal at the \( m^{th} \) node is given by,

\[
OSNR = P_1^g(dBm) + 49.14 dB = 36.14 dB
\]

which is well above the conservative minimum required OSNR of 26.3 dB estimate in [21], that has been calculated after accounting for safety margins and linear and non-linear impairment penalties for 10 Gbps systems operating at BER of \( 10^{-12} \).

Consider a signal \( S_1 \) sourced by the first node and a signal \( S_2 \) sourced by the second node as shown in Figure 5 (a). The signals, of different wavelengths, are multiplexed onto the output link \( H \) of node 2. The signal power of \( S_1 \) as it enters node 2 (after EDFA) was calculated above to be 0 dBm. \( S_1 \) passes through the demultiplexer, OXC, splitter, shutter and combiner before reaching the multiplexer on this node. Just before it reaches the multiplexer, \( S_1 \) has a signal strength of -16 dBm.

The signal \( S_2 \), when sourced at node 2, measures -13 dBm (as decided above), and after encountering combiner losses, it measures -16 dBm. At the input of any EDFA, the power level of any signal is -16 dBm. For this specific example, if all the signals are sourced at -13 dBm and the EDFAs are set to 24 dB constant gain value, all receivers can detect at -11 dBm, near far effects can be avoided and gain equalization can be achieved. The EDFA gain is determined by equation (1) and is different for different link lengths. The exact link loss information can be conveyed real time to the amplifier section through an optical supervisory channel. The analysis presented here assumes that EDFAs have a frequency-independent gain, but, the gain spectrum of the amplifiers are not completely flat and hence gives rise to a small dynamic range in the received signal power which has to be accounted for. The conclusion of this section is that, by using slightly higher gain EDFAs, light-trail design can be made feasible without requirement for regenerators in the metro network.

VI. LIGHT-TRAIL TEST BED

The following sections outline our progress and ongoing work in building a testbed to prototype, to evaluate advanced
network technologies and to demonstrate a light-trail solution for high speed metro networking services. The following issues are addressed below: system components specification, high-speed source/destination synchronization, design of optical on/off shutters and description of testbed operation.

A. System Description

Our goal is to set up a uni-directional light-trail on a single wavelength metro bus that supports a multimedia streaming application. The bus that we implement consists of a four node network \((N_1, N_2, N_3, N_4)\) as shown in Figure 4. The testbed features a sender client station (connected to \(N_1\)), three receiver client stations (connected to \(N_2, N_3\) and \(N_4\) respectively) and the optical bus. In actual deployment, node \(N_1\) may be connected to the metro core as shown in the figure but we do not incorporate this feature in our testbed. A light-trail is statically established with \(N_1\) as the converter node and \(N_4\) as the sink node. The trail undergoes a tap-and-continue functionality on the intermediate nodes \(N_2\) and \(N_3\). As mentioned above, the edge networks are hubbed and hence node \(N_1\) transmits information downstream to all other nodes and depicts a downstream light-trail solution.

A multimedia application is run on the sender client station which needs to be streamed to all the receiver client stations across the optical bus. Towards this end, the test bed implements two primary functions; the first is to provide an interface from the sender and receiver client terminals to the optical backbone and the second is to enable light trail nodes to communicate over the optical channel. The testbed operation proceeds as follows. The sender client station first streams the multimedia content to the optical node \(N_1\). Node \(N_1\) then buffers the stream, encapsulates it in a light-trail frame, and broadcasts it to all the other optical nodes in the trail. A downstream optical node first checks if the data is destined for it and if so proceeds to buffer the data. The information is then transmitted to a receiver client station which in turn plays the multimedia stream. The following sections describe the system components required to realize this testbed.

B. System Components

Each node in the ring is equipped with one LAU to access the channel. The optical transmitter is a continuous wave Small Form Factor Pluggable (SFP) Fabry Perot laser operating at 1310 nm. The generated optical signal is internally modulated at 3 Gbps through an ON/OFF shift keying scheme using a voltage controlled differential serial pair. The receiver is a broadband Silicon Positive Intrinsic Negative (PIN) photodetector with an optical power sensitivity of -17 dBm. The fiber is a multimode fiber with a graded refractive index profile.

Each node is equipped with a splitter and a combiner which are low insertion loss multimode couplers. A number of devices could be used as the optical shutter. One such example is the magneto-optic switch based on Faraday effect designed in [18]. This shutter has an insertion loss of 4.8 dB, extinction ratio of 20 dB with rise/fall time of about 2 \(\mu s\). The insertion loss, however, can be improved greatly by using an index matching epoxy coating. The extinction ratio is reasonable for the specified switching speed. It is to be noted again that the shutter is not configured on a per packet basis and hence high-speed operation is not a stringent requirement.

The four light-trail nodes are implemented on two Xilinx Virtex II Pro field programmable gate array (FPGA) development boards. The FPGA device contains high-speed RocketIO serial deserial (SERDES) multigigabit transceivers (MGTs), which provide the differential signaled serial data stream to the laser modules. Each of the four stations of the testbed assembly maintains block RAM (BRAM) modules for storing transmit and receive data. The BRAM modules of all stations are also connected to the embedded PPC 405 microcontroller which facilitates loading and retrieving of the contents for use with the sender and receiver client interfaces.

We do not use multiplexers/demultiplexers on the ring so as to reduce system costs. Since our solution is scalable, the testbed can be extended to multiwavelength systems and even to mesh (metro core) networks by incorporating the required WDM equipment, which we intend to pursue as future work.

C. FPGA Components

In this section, we describe the components of the FPGA which will more clearly explain the testbed operation. Detailed specification sheets corresponding to the FPGA used can be found in [19].

1) Ethernet MAC: To support Ethernet functionality, Xilinx has provided the Ethernet MAC (EMAC) soft intellectual property core. The EMAC core provides Ethernet communication capability to the PPC 405 microcontroller. The EMAC hardware core is fully compatible with the IEEE 802.3 Media Independent Interface and is addressable from the PPC microcontroller via the On-chip Peripheral Bus (OPB). Using an OPB bus speed of 66Mhz the Ethernet MAC operates at 100Mbps. Data transmission to the EMAC is passed to the higher TCP/IP layer through the use of Light Weight IP.

2) Light Weight IP (LWIP): To interface the EMAC core with the PPC microcontroller and to provide TCP/IP capability, the Light Weight IP Application Programming Interface (API) library is used. The LWIP functions provide an interface to the hardware Ethernet media access controller. LWIP is an open source implementation of the TCP/IP protocol developed with the intention to reduce resource usage while still having full scale TCP capabilities. Our design utilizes the raw API with callback mode. Asynchronous network events (data received, connection established etc.) are communicated to the application through interrupt callback functions. These callback functions are registered during the initialization of the TCP connection using the raw API functions.

3) SDRAM File System: As mentioned earlier, the light-trail nodes maintain memory data buffers to compensate for the transmission speed discrepancy between the client interfaces and the light-trail optical transmission medium. To achieve the required buffering capability, on-board peripheral SDRAM modules are connected to the PPC Processor Local Bus (PLB) which provides memory control. Data is stored in the SDRAM through the use of the Memory File System (MFS) library features available with the use of the PPC.
provides an interface to the SDRAM memory which is directly addressable from the microcontroller. Files created in the MFS can be dynamically created and accessed through the use of one or more file handles which are essentially address pointers into the associated file. Our design utilizes two independent file handles to provide access to a single memory file. The use of two file handles allows the SDRAM MFS to act as FIFO data queue with each file being written by one handle and read through the other; careful tests are performed to ensure that the read handle does not overrun the write handle.

4) MultiGigabit Transceivers: Light-trail communications are enabled through the use of ROCKET I/O Multigigabit transceiver (MGT) modules. These modules provide the differential serial data stream used to modulate the SFP transceivers and receive the incoming serial stream from the optical receivers. The MGT transceiver modules instantiated in the testbed design use a low jitter Low Voltage Differential Signal clock source operating at 150Mhz to provide the 3Gbps data stream. Details of MGT operation and characterization can be found in [19].

Due to the complexity of standalone MGT operation, Xilinx has provided the Aurora protocol which provides a logical link interface between MGT instantiations. The Aurora protocol is a scalable, lightweight, link layer protocol for moving data across one or more point-to-point serial lanes and takes care of configuring and operating the high-speed serial link. The Aurora protocol drastically simplifies the user interface to the complex control structure of the MGT transceivers. In addition to providing frame encapsulation features the protocol enables high-level clock synchronization and clock recovery functionality.

For use in our testbed operation, each Aurora module is configured as a single lane, simplex channel. In single lane operation, the term lane, as described below, is synonymous with channel. For light-trail operation, each Aurora module includes one MGT for each high-speed serial lane. The MGTs, in turn, are driven by their own lane logic module which handles operations like lane initialization, error detection, 8/10B symbol generation and decoding. Upon channel reset, the Aurora module begins by sending clock correction sequences until notified of "lane up" and "channel up" by the destination lane logic module.

To send data through the high-speed channel, a TX interface module provides the data to be transmitted to the Aurora user interface. As mentioned before, the Aurora protocol provides framing capabilities for data transmission. This is accomplished using a TX start of frame (TX_SOF) flag sent at the beginning of frame transmission which is used to signal the intended receiver(s) of the impending frame. Upon reception of the TX_SOF flag, the receiver asserts the RX_SOF flag which signals the RX interface module to begin receiving data into the BRAM. A similar TX_EOF flag is sent to indicate the End of Frame which in turn signals the receiver that the current frame has completed. The size of the light-trail frames are chosen to be 16 KB primarily due to the limit of BRAM resources on the FPGA development board. Because each board utilizes BRAM to accommodate two light-trail transmitters and two receivers in addition to packet FIFOs for the EMAC, using the next larger size of 32KB would exhaust the BRAM resources.

D. Testbed Operation

Winamp’s SHOUTCast streaming media application was chosen to provide the multimedia stream from the sender client. This streaming application is achieved with support from a combination of Light Weight IP, Ethernet MAC, SDRAM Memory File System and the Aurora protocol. The following sections describe the operation of the light-trail testbed. Figure 6 shows the block diagram of the various components of nodes $N_1$ and $N_2$ involved in the testbed.
Our actual implementation of the light-trail network is slightly more complicated than what is shown in the figure since some of the resources were shared and the set up was optimized to use the minimum number of boards. However, for the sake of clarity in illustration, we assume here that each board corresponds to one light-trail node.

1) Sender Client Interface: Node $N_1$ of the light-trail runs a host program that offers the service of accepting datagrams from the sender client. At startup the server program calls `connection_init()` to create a new TCP socket connection, bind the IP address with a well known port and listen for an impending connection from the sender client. When the sender client is ready to stream the multimedia content, it initiates a TCP three way handshake on the listening port which triggers the `senderclient::accept()` callback function on the light-trail node. Node $N_1$ acknowledges the three way handshake and registers the `senderclient::receive()` callback function which processes subsequent packets arriving at the network interface. In addition, a file (EthDataFile) is allocated in the SDRAM MFS to buffer the streaming payload and a new file handle (EthWriteHandle) is associated with the file. Each socket connection is allocated 2MB of buffer space and maintains a byte counter to indicate the location of the file handle. A buffer size of 2MB is designed to be a convenient size to interact with the PPC memory controller and also hold a significant amount of streaming data. When the byte counter indicates that 2 MB of data has been received the file handle is set back to the beginning of the file and the buffer is overwritten. Each socket connection is independent of all other connections and remains intact until it is terminated by the sender client.

2) Light-trail Transmission: The MGT of node $N_1$ transmits 16KB packets using the Aurora protocol. The data for $N'_1$’s MGT is supplied to the BRAM from the EthDataFile. The PPC uses a separate file handle, MgtReadHandle, to mark the location in the data file that needs to be currently transmitted. MgtReadHandle initially points to the beginning of the EthDataFile. Prior to light-trail transmission, a block of 16KB data is copied from the MFS into the dual ported transmission BRAM using `mfs_file_read()`. Subsequent to filling the BRAM, the PPC signals the Aurora protocol to proceed with clock synchronization. The packet is then encapsulated in a frame, addressed to the appropriate destination node and then transmitted. A special address is reserved for broadcasting to all nodes in the trail. In the current experiment, we use the broadcast mode since the receiver clients attached to all the light-trail nodes are interested in the stream. The data to be transmitted is accessed by the MGT through a hardware port of the BRAM.

As the signals traverse down the fiber to the next node, it is split by a drop coupler and a part of it is diverted to the local MGT while the rest of it is sent through the add coupler to reach nodes further downstream. Upon detection of light-trail activity, the local MGT is synchronized with the current transmission and decides whether or not to process the incoming data based upon the 32-bit Start Of Frame (SOF) label. Each light-trail node is assigned a unique software controlled 32-bit SOF label that is register accessible from the MGT hardware receiver modules. Although currently arbitrarily chosen, this light-trail hardware address could be similar to the IP address used to access the node via TCP/IP communication. The SOF label in conjunction with the SOF flag generated by the Aurora protocol (as discussed earlier) provides node addressability.

If the hardware on (say) node $N_2$ detects that it is the intended destination, the impending data is latched into the receive BRAM and a notification signal is sent to the PPC to
indicate that a packet has been received. Following notification of packet reception, the PPC transfers the associated BRAM data into an MgtDataFile at the location of MgtWriteHandle using mfs_file_write(). This handle is created and initialized to the start of MgtDataFile upon reception of the first packet in the communication stream and is set to point to the location where subsequent light-trail packets are placed. The MGTdataFile is associated with the current connection and has characteristics similar to the EthDataFile in that 2MB of space is allocated, and when full the buffer is overwritten.

3) Receiver Client Interface: The receiver client initiates a session with the server program running on the light-trail node (say $N_2$) it is connected to. Similar to the sender connection initiation, the receiver begins communication with a TCP three way handshake on the hosts listening receiver port. Upon reception of the TCP SYN packet from the receiver client, the receiverclient accept() callback function is triggered which completes the three way handshake, associates a new file handle, EthReadHandle (initialized to the start of MgtDataFile), with the connection and registers receiverclient_send() which handles all succeeding data communication. After connection setup has completed, packets are sent from the MgtDataFile at the location of EthReadHandle to the receiver client through receiverclient_send(). A byte counter monitors the location of the EthReadHandle and ensures that it does not overrun the MgtWriteHandle. The datagrams sent to the receiver client completes the streaming media connection over light-trails.

4) MAC protocol: In this hubbed architecture, since node $N_1$ is the only node that transmits, no precautions are taken to avoid collisions and thus node $N_1$ is not restricted from transmitting at any time. Thus, when 16KB of information is collected at the sender client interface it is immediately sent over the light-trail. However, as suggested earlier, a second trail would have to be set up to carry the traffic from all the access nodes to the hub. Since, this circuit will be shared by multiple nodes, a MAC protocol is required and can be implemented as follows.

Various light-trail MAC protocols have been developed and studied in [1], [12], [17], however due to their relative complexity of implementation a more simple approach is taken in the testbed to demonstrate feasibility. To demonstrate upstream operation, a simple MAC is implemented to provide bandwidth sharing between nodes $N_1$ and $N_2$. A connection is set up between nodes $N_1$ to $N_3$ and nodes $N_2$ to $N_4$. In this dual connection case, an out-of-band signal is connected between nodes $N_1$ and $N_2$ to avoid simultaneous transmission. The fact that both nodes $N_1$ and $N_2$ are implemented on the same development board makes this communication relatively simple, however, future work on the test bed will make this communication available between boards. In addition to protecting against simultaneous transmissions, a single pin is connected from the FPGA to the laser TX enable pins which is used to silence node $N_2$’s transmitter while node $N_1$ is sending and visa-versa. The only other modification needed to convert the downstream light-trail into the upstream solution is to address the destination nodes appropriately.

5) Experimental Results: The oscilloscope shown in Figure 7 illustrates a healthy eye pattern at the fourth node in our 4 node testbed. The test pattern was produced with a 3Gbps pseudo random bit sequence sent from node 1 to all 4 nodes of the light-trail. TCP dump traces of the SHOUTcast streaming media application were obtained through ethereal network protocol analyzer. In this demonstration, SHOUTcast listener clients were connected to each of the three downstream receiving nodes of the four node testbed. The traces illustrate that the streaming data, sent to node 1 from the SHOUTcast sender client, was successfully received and relayed by all downstream light-trail nodes to their respective clients. The traces are shown in Figure 8 for The sender client computer IP address is 55.248 and the receiver client IP address is 55.235. The data is streamed to node N1 of the light trail which is on board 1 at address 55.197. The receive data is streamed from the board with the same address because light trail node N2 is actually on the same board.

VII. Future Work

The testbed depicted in this writing illustrates the feasibility of using light-trail technology for metro edge networks. As mentioned, to provide multiple access for light-trail nodes, downstream stations must be silent in the presence of upstream activity. Our current solution uses the TX enable/disable switch of the laser modules to provide this functionality. It is noted however, that this large switching time of approximately 300μs is undesirable considering the high transmission speed of the light-trail network. Empirical results suggest that a downstream laser can remain in the unmodulated continuous wave ON state and not interrupt upstream communications. Thus, as a remedy to this problem we are currently developing a solution that uses an RF single pole single throw switch that can be placed between the serial output of the FPGA and the laser modulation inputs. The switch has been shown to provide an average switching time of 5ns. This method can drastically reduce the transmitter activation time compared to using the TX enable/disable pins. However, it has not yet been determined if this solution will produce the desired functionality when used in the testbed.

In addition to the aforementioned paragraph, we would like to improve the current testbed into a multiple wavelength and multiple light-trail solution. Future work will also implement fairness into the medium access control and investigate power budget issues at greater length. Due to the flexibility that light-trails provide with respect to the dynamic provisioning of sub-wavelength connections in the network, we believe light-trails are well suited for grid networking applications. To this end we are looking into design issues related to light-trail network topologies, architectures and protocols, to enable light-trail communications within our recently acquired 16 workstation FPGA cluster. This solution will allow us to explore and exploit the benefits of parallel processing. Early research into this development suggests that we must consider many practical implementation issues such as measurement and characterization of devices. Investigation into the interoperability and design specifications of light-trail components
such as, couplers, connectors, shutters, laser on/off switches and fiber types must also be considered.

VIII. Conclusion

The work presented in this paper is focused on providing solutions to the metro edge and metro core networks. In this paper, we described a few of the more prominent evolving trends and the inherent limitations in current network deployments and evaluated various solutions in terms of cost and performance. We have proposed an architecture based on CWDM ring/bus topologies for the metro edge networks utilizing light-trail technology and illustrated a DWDM ring/mesh topology for metro core networks. We also substantiated why light-trails may prove to be a good fit for the bursty and sub-wavelength demands imposed by the metro market.

In addition, a 3 Gbps, four node, light-trail testbed is discussed with a description of operation using the SHOUTcast streaming media application. We believe that our testbed has improved our understanding of the light-trail paradigm and we would like to develop on our current understanding of light-trail system level design to lend credence to our theory on design and deployment of light-trails in metro networks.

IX. Acknowledgements

We would like to thank Rashmi Bahuguna and Jin-Wei Tioh for their contributions on information pertaining to the optical shutter discussed in section V and the electrical switch mentioned in the future work section. Rashmi and Jin are currently working to improve the performance and operation of these devices for use in the complete light-trail solution.

REFERENCES


