Multi-Point Ethernet over Next-Generation SONET/SDH

C. Xie¹, N. Ghani¹, Q. Liu¹, W. Shu¹, A. Gumaste², Y. Qiao³, M. Wu³
¹University of New Mexico, ²IIT Bombay, ³Shanghai Jiao Tong University

Abstract: Advances in SONET/SDH technologies have introduced novel features for improved services mapping and provisioning, enabling many new avenues for new Carrier Ethernet support. However Ethernet-over-SONET studies have mostly focused on provisioning point-to-point Ethernet private line offerings. This paper considers the more challenging case of provisioning multi-point-to-multi-point Ethernet LAN services over advanced SONET/SDH networks and presents novel strategies based upon connection group overlays. Detailed simulation results are also presented along with directions for future work.

Keywords: Next-generation SONET/SDH, Ethernet-over-SONET, Ethernet LAN-over-SONET

I. INTRODUCTION

The last decade has seen many advances in ubiquitous time-division multiplexing (TDM) SONET/SDH technologies. Broadly termed as “next-generation SONET/SDH” (NGS), these new standards include the generic framing procedure (GFP, ITU-T G.7041), virtual concatenation (VCAT, ITU-T G.707), and link capacity adjustment scheme (LCAS, ITU-T G.7042). A key feature here is inverse-multiplexing [1] which enables demand resolution over multiple sub-connections. Also, another key NGS provision is its interoperability with the legacy SONET/SDH base as well as new ITU-T optical transport network (OTN) standards. Hence carriers need only deploy NGS selectively on a need basis at client access points [2].

As the focus shifts towards service provisioning, particularly Carrier Ethernet [3], the Metro Ethernet Forum has standardized a variety of service models for Ethernet private line (EPL) and Ethernet-LAN (E-LAN). EPL offers point-to-point (pt-2-pt) data connectivity between user sites and the additional Ethernet virtual private line (EVPL) variant allows multiple users to share a single user network interface (UNI). Meanwhile E-LAN service defines further multi-point-to-multi-point (mp-2-mp) connectivity between data sites, providing a LAN “extension” service over larger geographic ranges.

In light of the above, many carriers are interested in provisioning Ethernet services over SONET/SDH networks. Now numerous Ethernet-over-SONET (EoS) studies have emerged recently, mostly focusing on multi-path routing and survivability schemes for inverse multiplexing NGS networks [6]-[10],[14]. However these efforts have only studied pt-2-pt connections, i.e., EPL services, and commensurate mp-2-mp Ethernet LAN services have only now started to received attention [16]. Clearly these latter service types represent much more lucrative “value-add” carrier offerings and require further investigation and design.

II. BACKGROUND

Earlier studies on inverse multiplexing in [4] and [5] proposed basic schemes to resolve incoming requests at the STS-1 level. Modified shortest-path heuristics were then proposed for multi-path routing in SONET-DWDM networks. Meanwhile [6] also tabulated several iterative graph heuristic algorithms for multi-path routing in VCAT settings, i.e., shortest-path first (SPF), widest-path first (WPF), and max-flow (MF). The findings indicate that SPF performs best for higher load scenarios. Overall results show that inverse multiplexing yields good blocking reduction, and hence further studies have focused on related survivability provisioning. For example, [7] proposed several low-overhead protection for EoS (PESO) schemes to increase VCG resiliency, i.e., PESO α, β, γ schemes. These schemes try to achieve sufficient multi-path diversity between VCG members to ensure adequate immunity to single-link failures. Also, [8] has tabled a similar strategy for achieving degraded-service-aware bandwidth provisioning across multiple VCG paths. Multi-path load distribution is also done to minimize the maximum incremental link utilization using integer linear programming (ILP). The overall results show good improvements in blocking and load balancing.

Other schemes have also studied sub-connection protection/restoration strategies. For example, [9] proposed two inverse multiplexing shared protection schemes. The first approach, protecting individual VCG member (PIVM), allows backup capacity sharing between link-disjoint VCG members, whereas the second approach, provisioning fast restorable VCG (PREV), only allows sharing between link-disjoint VCG members with the same source-destination. Expectedly, the efficient PIVM scheme gives a slower recovery performance whereas PREV gives faster recovery. Finally [10] developed a new effective multi-path bandwidth metric that accounts for both link bandwidth and availability. Two multi-path routing heuristics are then tabled, with

This paper studies EoS LAN services and presents a more indepth treatment from the authors’ initial work in [16]. The paper is organized as follows. First Section II overviews existing research on traffic engineering and survivability in NGS as well as recent studies on overlay provisioning. Subsequently, Section III introduces the mp-2-mp EoS problem and defines several new topology overlay strategies, including mesh, ring, and star/hub. Finally, the proposed schemes are analyzed in detail in Section IV using simulation and overall conclusions and future directions are presented in Section V.
both showing significant gains over single path provisioning. Furthermore, researchers have also studied overlay topology connection group provision. Most of this work falls under the context of overlay network design to build specialized topologies over physical substrates to support applications, e.g., resilient overlay network (RON) [11], service overlay network (SON) [12], virtual network (VN) [13], etc. Although the above contributions embody many service innovations, additional avenues exist for developing new mp-2-mp EoS services. In particular, no studies have looked at “on-line” overlay provisioning of connection groups for dynamic Ethernet LAN requests in the context of NGS inverse multiplexing. It is envisioned here that ability to split a demand into multiple sub-connections will yield much-improved multi-tiered resiliency and efficiency.

III. ETHERNET LAN OVER SONET/SDH

The key objective in delivering Ethernet LAN services over SONET/SDH is to provision reliable mp-2-mp connectivity across dispersed metro/wide-area sites. However, given the inherent lack of timeslot multi-casting features at the SONET/SDH level, this requires coordinated setup of multiple pt-2-pt SONET/SDH connections to emulate mp-2-mp connectivity, i.e., connection groups, topology overlays. Moreover, these overlays will be further impacted by the availability of “higher-layer” Layer 2 connectivity. Now in [16], the authors have proposed two “on-demand” EoS LAN schemes using mesh and star topology overlays. This effort presents a more extensive treatment of this initial work. First, the necessary notation is introduced.

Consider a physical SONET/SDH network topology comprising of \( N \) nodes and \( M \) links. This network is modeled as a graph \( G(V,L) \), where \( V \) is the set of network nodes and \( L \) the set of links, i.e., \( V=\{v_1,v_2,\ldots,v_n\} \) and \( L=\{l_{12},l_{13},\ldots,l_{ij}\} \). Here, link \( l_{ij} \) is the link between nodes \( i \) and \( j \) of total capacity \( c_{ij} \) and \( c_{ij} \) is the available capacity of \( l_{ij} \). Necessarily, if there is a link from \( i \) to \( j \), then there is also a reverse-direction link from \( j \) to \( i \) since SONET/SDH links are bi-directional. Furthermore, consider the \( i \)-th Ethernet LAN request for Layer 2 connectivity among a subset of network nodes, \( v_i \subseteq V \). This request is used to derive a connection group comprising of a set of \( n_i \) bi-directional pt-2-pt connections, \( \{s_i, d_i\} \), where the set \( s_i=\{s_{i1},s_{i2},\ldots\} \subseteq v_i \) and \( d_i=\{d_{i1},d_{i2},\ldots\} \subseteq v_i \) represent the end-point nodes and the individual connections are \( s_{i1} \rightarrow d_{i1}, s_{i2} \rightarrow d_{i2}, \ldots \). Assuming that the LAN must support a throughput of \( x_i \) STS-1 units, each individual connection must be of size \( x_i \) STS-1 units as well. Now the exact composition of the connection group will depend upon the overlay topology chosen. However, regardless of the particular overlay solution, underlying pt-2-pt connections can still be routed using inverse-multiplexed sub-connections. Namely each individual connection can also be “split” into multiple sub-connections, i.e., up to \( K \) in total. Hence the overall connection group request is denoted as \( (n_i, \{s_i, d_i\}, x_i, K) \). Now consider overlay design and protection schemes.

A. Connection Group (LAN Overlay) Selection

The first step in EoS provisioning is to determine the LAN connection (overlay) group, i.e., to identify the pt-2-pt interconnections between dispersed participating nodes. Clearly the type of overlay chosen will very much depend upon the Ethernet switching functionalities in the core (or client) nodes. For example with no Ethernet switching capability at any nodes, operators must implement TDM connections between every possible node to deliver “LAN-mesh” connectivity, i.e., mesh overlay, Figure 1. Alternatively if one or more core (or client) nodes have Ethernet switching capabilities, star overlay can be implemented, i.e., star overlay, Figure 1.

Mesh Overlay: The mesh overlay is a straight-forward approach as it implements direct connections between all LAN nodes. The scheme simply loops between all source-destination pairs to generate bi-directional connection. Expectedly, the mesh overlay approach is very resource-intensive, generating \( O(|V|^2) \) connections, which may adversely impact LAN blocking.

Star Overlay: The star overlay scheme is designed to be more resource-efficient and uses a designated hub site to provide connectivity to all other nodes. Specifically, it is assumed that this site has Ethernet (i.e., Layer 2) switching capabilities, allowing it to extract incoming Ethernet frames and forward them to destination sites. This Layer 2 switching capability can be provided either
This basic scheme chooses hub selection from the subset of LAN sites with Ethernet switching capabilities. Namely, the design goal is to choose a hub that minimizes overall resource utilization and helps lower blocking for future LAN requests. Indeed this is a key requirement given the higher dimensionality of LAN setup requests as opposed to pt-2-pt EVC requests. Hence three selection schemes are presented:

**Random Hub Selection (RS):** This basic scheme randomly selects a hub site from the subset of LAN connection group nodes which have Ethernet switching.

**Minimum Average Hops (MAH):** This scheme chooses the hub site with the minimum average hop count to all other LAN group end-points. The overall goal here is to minimize resource utilization by the group of LAN connections as follows:

$$h_i = \{v_i\}_{j'}$$

where $j' = \min \left( \sum_{k\in S_{ij}} hop(v_i, j, v_i)_k \right)$ for all $v_i \in V$.

**Minimum Average Cost (MAC):** This scheme chooses the hub-site with the minimum average cost to all other LAN group end-points, where the cost is defined as a per-link distance that is inversely-proportional to the available capacity on link $k$.

$$c_k = \frac{1}{c_k + \epsilon}$$

where $\epsilon$ is a small quantity chosen to avoid floating-point divide errors. Hence the hub selection is:

$$h_i = \{v_i\}_{j'}$$

where $j' = \min \left( \sum_{k\in S_{ij}} cost(v_i, j, v_i)_k \right)$ for all $v_i \in V$.

Unlike the hop-based (MAH) scheme, this scheme uses dynamic link-resource state to choose the hub site. As such, it can avoid “congested” hub sites with heavily loaded links. However, computational complexities here are significantly higher, as $O(2^{|V|} \log V) = O(|V|^2 \log |V|)$ shortest path computations have to be done per request, i.e., $O(|V|^2) = O(|V|^2 \log |V|)$ complexity.

### B. Multi-Tiered LAN Group Provisioning

Carrier LAN service offerings will typically demand flexible, multi-tiered support. For example most “regular” LAN offerings will suffice with no/partial protection against single fault events. Alternatively a subset of offerings may demand very stringent 100% recovery, e.g., financial services. To address these requirements a tiered LAN protection framework is proposed here, extending upon the tiered (connection) protection scheme of [14]. The overall goal is to guarantee a minimum (pre-provisioned) LAN throughput in the event of a single link failure. Hence the solution defines a user-specified fractional protection factor, $\rho$ ($0 \leq \rho \leq 1$), which represents the desired level of pre-provisioned protection for the tiered LAN group. Namely a minimum level of $\rho x$ STS-1 units of dedicated protection capacity must be provisioned for all LAN group connections. By protecting each connection in this manner, aggregate LAN throughput can be guaranteed in the event of a single link failure.

This overall tiered concept is illustrated in Figure 2 which shows the $i$-th LAN request for 12 STS-1 units being provisioned between three nodes, $v_{i1}$, $v_{i2}$, and $v_{i9}$. The request is mapped using a star overlay with an inverse multiplexing factor of $K=3$ and of $\rho=0.5$ (50%) and $v_{ij}$ is the hub. A connection group containing two bi-directional SONET/SDH connections, $v_{i1}-v_{i3}$ and $v_{i4}-v_{i9}$ is formed. In turn, each of these connections is split into three diversely-routed sub-connections of 4 STS-1 units each, e.g., routes $w_{ij1}$, $w_{ij2}$, and $w_{ij3}$ for LAN group connection $v_{ij1}-v_{ij9}$, Figure 2. Meanwhile the protection threshold $\rho$ mandates that each individual connection have at least $(0.5)12=6$ STS-1 units of protection. This objective is achieved using dedicated protection on a per-sub-connection basis, akin to [14]. Namely dedicated link-disjoint protection paths are computed for a minimal subset of working sub-connections until the desired threshold is achieved. Hence two protection sub-connections must be setup for each connection in Figure 2, i.e., 4 STS-1 units along routes $p_{ij1}$ and $p_{ij2}$ to protect the working routes $w_{ij1}$ and $w_{ij2}$ for the LAN connection $v_{ij1}-v_{ij2}$. Note that protection over-provisioning can occur for smaller values of $K$ [14]. For example in Figure 2 with $K=3$, a total of 8 STS-1 of protection capacity is reserved for each connection even though the threshold is 6 STS-1, i.e., 33% over-provisioning. Nevertheless this inefficiency can be resolved by appropriately setting the inverse multiplexing factor or “right-sizing” protection sub-connections. However the above approach simplifies protection switchovers during link failures as it ensures equal-sized working and protection VCG members.
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2009 proceedings

Now consider the working LAN group provisioning algorithm in Figure 3. This scheme first makes a temporary copy of the network graph, \( G'(V,L) \), and then iterates to setup working routes for all \( n_i \) connections in the LAN using inverse multiplexing. Namely each connection is resolved into sub-connections using an “even” distribution approach. Next, a modified successive path computation scheme leveraging \( K \)-shortest path computation is used to iteratively compute the individual sub-connection path (route) vectors, \( w_{i,j,k}^{v3} \) for each requested sub-connection. Here if a sub-connection is successfully routed, its capacity is pruned along all route links in \( G'(V,L) \). Furthermore, the algorithm only proceeds to the next group connection if the current connection is successfully routed otherwise the request is dropped. Note that the scheme places no restrictions upon the level of link overlap between the individual sub-connections.

![Logical View](image)

**Physical Star Overlay**

![Physical Star Overlay](image)

**K=3 inverse multiplexed working sub-connections. (capacity \( x_{ijk} = 1 \times \text{STS-1} \))

Figure 2: Tiered LAN protection scheme (only connection routes shown)

Given \( i \)-th LAN group request \( (n_i, (s_i, d_i), x_i, K) \) and temporary graph \( G'(V,L) \) from working group routing stage (containing leftover capacity)

- Prune capacity on all links in \( G'(V,L) \)
- Prune all routed protection sub-connection routes
- Store protection sub-connection route in vector \( g_{ijk} \)
- Initialize protection capacity to 0
- if (route found)
  - Initialize temporary copy graph \( G''(V,L) = G'(V,L) \)
  - Compute K-shortest path sub-connection routes between \( s_i \) and \( d_i \)
  - Route and route all sub-connections on \( G''(V,L) \)
  - while (all sub-connections routed & no sub-connection failed)
    - Extract \( j \)-th group connection \((s_j, d_j, x_j, K)\)
    - Resolve \((s_j, d_j, x_j, K)\) into \( K \) sub-connections of size \( x_{jk} \)
    - Compute \( K \)-shortest path sub-connection routes between \( s_j \) and \( d_j \)
    - Route all sub-connections on \( G''(V,L) \)
    - if \((j\)-th connection routed\)
      - \( j = j+1 \)
    - else
      - if (all \( n_i \) group connections routed)
        - Drop LAN setup request
      - else
        - Continue to protection stage

Figure 4: Pseudo-code for protection LAN computation

Pending successful provisioning of all working group connections, the final step implements tiered (partial) protection as shown in Figure 4. This final step utilizes the “left-over” capacity in \( G(V,L) \) and implements similar steps as in computing working paths for protection path computation. Here a running count of the aggregate protection capacity, \( p_{ijk} \), is maintained and this value is checked against the desired minimum protection threshold of \( p_{ijk} \) STS-1 units after each successful protection setup. If this threshold is crossed, the LAN connection is adequately protected and the sub-connection protection paths, \( p_{ijk} \), are stored. Otherwise, the tiered LAN request is dropped.

In terms of computational complexity, optimized \( K \)-shortest path implementations can be of \( O(K|V|\log|V|) \)
complexity, where \(|V|\) is the number of nodes, i.e., since the maximum number of LAN group connections is bounded by \(O(n) = O(|V|^2)\). By extension, the added protection provisioning phase will have a total complexity of \(O(K|V|^2 \log |V|)\).

The above protection strategies can also be coupled with optional post-fault restoration. Namely, all non-protected VCG members (e.g., \(w_{ij}\), Figure 2) traversing a failed link can be re-routed after failure notification. The overall goal here is to achieve full recovery in single fault even for partially protected LAN requests. This will essentially enable carriers to achieve improved service recovery for lower-priced offerings as well.

### IV. PERFORMANCE ANALYSIS

Multi-point EoS performance is studied using discrete event simulator OPNET Modeler™. All tests are done using the NSFNET topology comprising of 16 nodes and 25 links (node degree 3.125). It is assumed that network elements are generic NGS-capable broadband digital cross-connects (BBDCS) nodes with STS-1 (50 Mb/s) switching granularity, OC-48 (2.544 Gbps) links. All LAN requests follow random exponentially-distributed holding and inter-arrival time distributions, with means \(\mu\) and \(\lambda\), respectively. Herein, a scaled mean holding time of \(\mu=600\) seconds is used, with the mean inter-arrival times adjusted according to desired load. Additionally, LAN bandwidth sizes are varied from 200-1000 Mb/s in 200 Mbps increments (4 STS-1) to model fractional Ethernet demands. Finally, all runs are averaged over 500,000 randomly-generated LAN requests. Modified Erlang metrics are also defined to account for varying connection groups of differing LAN requests, i.e.,

\[
\text{Load(mesh overlay)} = \sum_{x=1}^{n-1} \frac{n(n-1)}{2} \cdot \frac{\mu}{\lambda} \quad \text{Eqn}(4)
\]

\[
\text{Load(star overlay)} = \sum_{x=1}^{n-1} (n-l) \cdot \frac{\mu}{\lambda} \quad \text{Eqn}(5)
\]

where the LAN groups range in size from \(x_1=3\) to \(x_5=5\) nodes and the \(1/\lambda\) represents the mean inter-arrival rate (i.e., inverse of inter-arrival time).

Initial tests (Figures 5a and 5b) plot the carried load for non-protected LAN scenarios (i.e., \(\rho=0\)) for increased inverse multiplexing factors given a nominal request blocking rate of 2%. These scenarios are very relevant to carriers as they indicate the true “load-carrying” capacity of the network at a typical operating point. As expected, the star overlays yield the highest carried load owing to smaller connection group sizes. Furthermore the load-balancing (minimum distance) routing approach also yields much higher carried loads for mesh overlays, about 50-70% higher (Figure 5a). Commensurate carried load gains with star overlays (and load-balancing) are lower, however, averaging about 7-12%. Another key finding is that “intelligent” hub selection strategies (star overlays) yield the best overall gains. For example the MAH scheme gives the highest carried load, averaging 17% higher than the minimum distance MAC scheme and over 40% higher than random hub selection. These results also show that the sizeable improvement with increased levels of inverse multiplexing, i.e., about 30-50% (25%) higher carried load for mesh (star) overlays.

![Figure 5: Carried load for 2% LAN blocking, \(\rho=0\): a) mesh, b) star](image)

Figure 5: Carried load for 2% LAN blocking, \(\rho=0\): a) mesh, b) star

Next, LAN blocking is tested for the more effective star overlays and load balancing routing. Namely the results for varying protection thresholds (\(\rho=0, 0.25, 0.5\)) and inverse multiplexing factors (\(K=1, 2, 4\)) are shown, Figure 6a. Here it is seen that that increased protection factors significantly increase request blocking. In addition it is seen that increased inverse multiplexing (demand splitting) yields notable gains for equivalent protection factors, e.g., \(K=4, \rho=0.5\) gives about 10-30% lower blocking than \(K=2, \rho=0.5\) (Figure 6a). Furthermore a detailed look at the individual blocking rates for different LAN sizes (\(K=4, \rho=0.25\)) shows that larger 5 node LAN groups experience almost an order magnitude higher blocking than smaller 3 node LAN groups (Figure 6b). This discrepancy is clearly due to the difficulties in routing a larger number of working and diverse protection connections at a hub site, i.e., limited by topological node degree. However, the hop-based MAH hub selection scheme still gives the lowest blocking for all LAN sizes (at least 40% lower at most loads) and even achieves under 1% blocking for 5 node LAN requests at low loads. This blocking level is sufficient for operational settings and cannot be achieved with the other star and mesh overlays.

![Figure 6: Star blocking: a) MAH with varying \(K, \rho\), b) \(K=4, \rho=0.25\)](image)

Figure 6: Star blocking: a) MAH with varying \(K, \rho\), b) \(K=4, \rho=0.25\)
Finally, tiered LAN protection is tested in conjunction with post-fault restoration of non-protected working VCAT sub-connections on a failed link. The goal here is to achieve full recovery for partially protected (lower tier) LAN services. All link failures have exponentially-distributed inter-arrival and mean-time-to-repair (MTTR) values with mean 600 sec. Note that the MTTR values represent truck roll repairs and are typically much larger than protection switchovers. Restoration performance is gauged by measuring recovery rates for carried loads up to 20% blocking, i.e., high loading. Namely, the restoration rate is defined as the percentage of LAN groups which recover all failed non-protected connections, i.e., regain full 100% throughput.

The LAN restoration for star overlays with varying protection thresholds and inverse multiplexing factors is also tested (MAH hub selection and load balancing routing, Figure 7). Namely, $K=1$, $\rho=1$ corresponds to regular non-inverse multiplexed protection (e.g., 1+1, 1:1), whereas $\rho=0.25$ and 0.5 correspond to partial protection. The results here reveal some important findings. First, full protection provides 100% recovery, albeit the carried load in this case is much lower. For partial protection, it is observed that increased levels of resource over-provisioning (protection) translate into lower levels of post-fault restoration. This is due to the fact that higher protection thresholds imply less leftover bandwidth for restoration. For example, $K=4$, $\rho=0.25$ gives higher restoration rates than $K=2$, $\rho=0.5$. Meanwhile the maximum carried load (up to 20% LAN request blocking) here is also higher by almost two times. These findings indicate that post-fault sub-connection restoration is highly beneficial and will allow carriers to offer full (100%) LAN recovery for over 95% of single link failures.

V. CONCLUSIONS

This paper addresses the more challenging problem of multi-point to multi-point Ethernet LAN services provisioning. Specifically, novel schemes are proposed for implementing mesh and star/hub overlays to support multi-point extension over metro-wide-area TDM networks. Detailed findings show much-improved blocking performance with star overlay strategies, especially when coupled with intelligent hub selection and load balancing routing. Future efforts will look at more capable shared protection strategies as well as the design of more efficient overlay topologies.

REFERENCES