University of New Mexico
UNM Digital Repository

Electrical and Computer Engineering ETDs | Engineering ETDs

7-8-2009

End-to-end provisioning in multi-domain/multi-layer networks

Tannous Frangieh

Follow this and additional works at: https://digitalrepository.unm.edu/ece_etds

Recommended Citation

Frangieh, Tannous. "End-to-end provisioning in multi-domain/multi-layer networks." (2009). https://digitalrepository.unm.edu/ece_etds/90

This Thesis is brought to you for free and open access by the Engineering ETDs at UNM Digital Repository. It has been accepted for inclusion in Electrical and Computer Engineering ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.

Tannous Frangieh
Candidate

ECE
Department

This thesis is approved, and it is acceptable in quality and form for publication on microfilm:

Approved by the Thesis Committee:

Nasir Ghani , Chairperson

Chaouki Abdallah

Wei Wennie Shu

Accepted:

Dean, Graduate School

Date

End-to-End Provisioning in Multi-Domain/Multi-Layer Networks

by

Tannous B. Frangieh

B.E., Computer Engineering, Lebanese American University, 2005

THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science
Computer Engineering

The University of New Mexico

Albuquerque, New Mexico

May, 2009

©2009, Tannous B. Frangieh

Acknowledgments

I would like to acknowledge and extend my heartfelt gratitude to the following persons who have made the completion of this work possible:

My advisor, Professor Nasir Ghani, for his help, dedication and advisement in my research,

Professor Chaouki Abdallah, Chair, Electrical and Computer Engineering Department for his support and guidance,

Professor Wei Wennie Shu, Director of Graduate Program, Electrical and Computer Engineering Department for being a great teacher and a source of knowledge,

Dr. Qing Liu, for his help in my research,

All friends, in particular Mr. Joud Khoury and Mr. Rani Abou Ghaida for helping me settle down in Albuquerque, and Dr. Zahi Nakad and Mr. Wadih Barakat, for their help and endless support,

And most especially my parents and family, for their endless love and support.

Dedication

To my father, mother, sister and two brothers...

End-to-End Provisioning in Multi-Domain/Multi-Layer Networks

by

Tannous B. Frangieh

ABSTRACT OF THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science
Computer Engineering

The University of New Mexico

Albuquerque, New Mexico

May, 2009

End-to-End Provisioning in Multi-Domain/Multi-Layer Networks

by

Tannous B. Frangieh

B.E., Computer Engineering, Lebanese American University, 2005

M.S., Computer Engineering, University of New Mexico, 2009

Abstract

The last decade has seen many advances in high-speed networking technologies. At the Layer 1 fiber-optic level, *dense wavelength division multiplexing* (DWDM) has seen fast growth in long-haul backbone/metro sectors. At the Layer 1.5 level, revamped next-generation *SONET/SDH* (NGS) has gained strong traction in the metro space, as a highly flexible "sub-rate" aggregation and grooming solution. Meanwhile, ubiquitous Ethernet (Layer 2) and IP (Layer 3) technologies have also seen the introduction of new *quality of service* (QoS) paradigms via *the differentiated services* (Diff-Serv) and *integrated services* (Intserv) frameworks. In recent years, various control provisioning standards have also been developed to provision these new networks, e.g., via efforts within the IETF, ITU-T, and OIF organizations.

As these networks technologies gain traction, there is an increasing need to internetwork multiple domains operating at different technology layers, e.g., IP, Ethernet,

SONET, DWDM. However, most existing studies have only looked at single domain networks or multiple domains operating at the same technology layer. As a result, there is now a growing level of interest in developing expanded control solutions for *multi-domain/multi-layer* networks, i.e., IP-SONET-DWDM.

Now given the increase in the number of inter-connected domains, it is difficult for a single entity to maintain complete "global" information across all domains. Hence, related solutions must pursue a *distributed* approach to handling multi-domain/multi-layer problem. Namely, key provisions are needed in the area of inter-domain routing, path computation, and signaling. The work in this thesis addresses these very challenges. Namely, a hierarchical routing framework is first developed to incorporate the multiple link types/granularities encountered in different network domains. Commensurate topology abstraction algorithms and update strategies are then introduced to help condense domain level state and propagate global views. Finally, distributed path computation and signaling setup schemes are developed to leverage the condensed global state information and make intelligent connection routing decisions. The work leverages heavily from graph theory concepts and also addresses the inherent distributed grooming dimension of multi-layer networks.

The performance of the proposed framework and algorithms is studied using discrete event simulation techniques. Specifically, a range of multi-domain/ multi-layer network topologies are designed and tested. Findings show that the propagation of inter-domain tunneled link state has a huge impact on connection blocking performance, lowering inter-domain connection blocking rates by a notable amount. More importantly, these gains are achieved without any notable increase in inter-domain routing loads. Furthermore, the results also show that topology abstraction is most beneficial at lower network load settings, and when used in conjunction with load-balancing routing.

Contents

List of Figures

Glossary

AS	**A**utonomous **S**ystems
ASON	**A**utomatically **S**witched **O**ptical **N**etwork
ASTN	**A**utomatically **S**witched **T**ransport **N**etwork
ATM	**A**synchronous **T**ransfer **M**ode
BBR	**B**andwidth-**B**locking **R**ate
BGP	**B**order **G**ateway **P**rotocol
CBR	**C**onstraint-**B**ased **R**outing
CSPA	**C**losest-**S**ignle-**P**ath Approach
CSR	**C**oncatenated **S**hortest **P**ath
CWS	**C**omputation **W**hile **S**witching
DES	**D**iscrete **E**vent **S**imulation

$Diff-Serv$	**Diff**erentiated **Serv**ices
$DWDM$	**D**ense **W**avelength **D**ivision **M**ultiplexing
$E2E$	**E**nd-to-**E**nd
EGP	**E**xterior **G**ateway **P**rotocol
$E-NNI$	**E**xterior **N**etwork-**N**etwork **I**nterface
ER	**E**xplicit **R**oute
ESP	**E**thernet **S**witched **P**ath
$GigE$	**Gig**abit **E**thernet
$GMPLS$	**G**eneralized **M**ulti**p**rotocol **L**abel **S**witching
GUI	**G**raphical **U**ser **I**nterface
HIR	**Hi**erarchical **R**outing
HT	**H**old-down **T**imer
$IETF$	**I**nternet **E**ngineering **T**ask **F**orce
IGP	**I**nterior **G**ateway **P**rotocols
IHT	**I**nter-domain **H**old-off **T**imer
$I-NNI$	**I**nterior **N**etwork-**N**etwork **I**nterface

$IntServ$	**Int**egrated **Serv**ices
IP	**I**nternet **P**rotocol
$IS-IS$	**I**ntermediate-**S**ystem to **I**ntermediate-**S**ystem
$ITU-T$	**I**nternational **T**elecommunication **U**nion **T**elecommunication Standardization Sector
LAS	**L**ightpath **A**ggregation **S**cheme
LMP	**L**ink **M**anagement **P**rotocol
LR	**L**oose **R**oute
LSA	**L**ink-**S**tate **A**dvertisement
LSA	**L**ink-**S**tate **A**ttribute
$LSDB$	**L**ink **S**tate **D**ata**B**ase
LSP	**L**abel **S**witched **P**ath
$MPPBCA$	**M**ultiple-**P**ath-**P**arameters-**B**est-**C**ase **A**pproach
$MPLS$	**M**ulti**p**rotocol **L**abel **S**witching
NAS	**N**ode **A**ggregation **S**cheme
NGS	**N**ext-**G**eneration **S**ONET

NNI	**N**etwork-**N**etwork **I**nterface
OIF	**O**ptical **I**nternetworking **F**orum
$OSPF-TE$	**O**pen **S**hortest **P**ath **F**irst - **T**raffic **E**ngineering
OXC	**O**ptical **C**ross **C**onnect
PCC	**P**ath **C**omputation **C**lient
PCE	**P**ath **C**omputation **E**lement
$PCEP$	**P**ath **C**omputation **E**lement **P**rotocol
$PNNI$	**P**rivate **N**etwork-to-**N**etwork **I**nterface
QoS	**Q**uality **o**f **S**ervice
RA	**R**outing **A**rea
RC	**R**outing **C**ontroller
$RSVP-TE$	**R**esource Re**s**er**v**ation **P**rotocol - **T**raffic **E**ngineering
RWA	**R**outing and **W**avelength **A**ssignment
SCF	**S**ignificance **C**hange **F**actor
SDH	**S**ynchronous **D**igital **H**ierarchy
$SNPP$	**S**ub-**N**etwork **P**oint **P**ool

SONET **S**ynchronous **O**ptical **NET**working

SPPA **S**ingle-**P**ath-**P**arameters **A**pproach

SWP **S**hortest-**W**idest **P**ath

TA **T**opology **A**ggregation

TDM **T**ime-**D**ivision **M**ultiplexing

TE **T**raffic **E**ngineering

TEDB **T**raffic **E**ngineering **D**ata**B**ase

UNI **U**ser-**N**etwork **I**nterface

WSP **W**idest-**S**hortest **P**ath

Chapter 1

Introduction

1.1 Introduction

The last decade has seen many advances in high-speed networking technologies. At the Layer 1 fiber-optic level, *dense wavelength division multiplexing* (DWDM) technology has seen fast growth in the long-haul backbone/metro sectors [1]. DWDM technology multiplexes multiple optical carrier signals on a single optical fiber by using different wavelengths, i.e. colors, of light. Typically, signals are multiplexed within the unused 1550-nm band so as to leverage the low attenuation windows which fall between approximately 1525 nm - 1565 nm (C band), or 1570 nm - 1610 nm (L band) [2]. Current DWDM standards can support over 100-channel spacings, yielding unprecedented terabits per second fiber rates. Moreover, the advancement in

DWDM sub-system component technologies has also led to the evolution of "light-path" circuit-switching paradigms, namely a wavelength can be routed in optical domain from source to destination without the need for any intermediate electronic processing.

Concurrently, many higher layer protocol technologies have also seen notable advances. For example, revamped next-generation *SONET/SDH* (NGS) (Layer 1.5) [1] has gained strong traction in the metro space as a highly flexible "sub-rate" aggregation solution. Specifically, this technology provides new features for virtual concatenation, allowing for highly-efficient transport of "non-TDM" data services. Meanwhile, ubiquitous Ethernet (Layer 2) and IP (Layer 3) layers have seen the introduction of new *quality of service* (QoS) provisions via *the differentiated services* (Diff-Serv) and *integrated services* (Intserv) frameworks. In line with these data-plane advances, related control plane standards have also emerged, most notably *multi-protocol label switching* (MPLS) for Layer 2 and Layer 3 support. Generalizations have even adapted this solution for "non-packet-switching" layers, termed as *generalized MPLS* (GMPLS) [3].

1.2 Background

As the above technologies undergo widespread deployment, a complex layering of operational networks has emerged, e.g., horizontal domains comprising multiple vertical technology layers. These segmentations are based upon various factors, such as technology, scalability, geographic, economic, administrative, etc. Now from a research perspective, a multitude of studies have looked at connection provisioning issues. For example, QoS routing in IP/MPLS networks is a widely studied topic as is the lightpath *routing and wavelength assignment* (RWA) problem in DWDM networks. Various studies have also looked at path computation across multiple network domains. Furthermore, many researchers have also studied multi-layer grooming in mixed IP-DWDM and SONET-DWDM networks.

However, given the increasingly diverse nature of networking technologies, the *distributed* multi-domain/multi-layer provisioning problem has not been addressed. In particular, as the scale and reach of services expands, there is a pressing need for "end-to-end" provisioning across heterogeneous domains, i.e., horizontal-vertical control. This need is reflected by real-world settings where different carriers and organization run separate domains at different circuit/flow granularities, i.e., gigabit wavelengths, "sub-rate" SONET/SDH tributaries, MPLS label switched paths (LSP), etc. By and larger, such inter-layer provisioning today is still done via manual provisioning of domain-specific control planes, giving high inefficiency and long lead times, i.e., hours to days [5].

1.3 Motivation

In order to exploit the full capacity of DWDM networks, it is crucial to provide distributed protocol frameworks to interface with higher-layer "grooming" networks, i.e., IP/MPLS, Ethernet, SONET [4]. Now multi-domain internetworking has long been supported in legacy IP and ATM networks. The most notable example is the distance-vector *border gateway protocol* (BGP). Moreover new standard efforts are also introducing multi-domain capabilities in optical transport networks, e.g., *Optical Internetworking Forum* (OIF) *network-to-network interface* (NNI) definition. Nevertheless the broader topic of designing provisioning algorithms across *heterogeneous* network layers has not been addressed in detail. Indeed there is a growing need to provision services over such distributed multi-domain/multi-layer networks.

Overall the multi-domain/multi-layer connection provisioning issue is a very challenging problem owing to the inherent grooming dimension and the lack of global accurate network state information. Currently, related provisioning algorithms are mostly lacking, particularly those operating in distributed (intra, inter-carrier) settings. This thesis addresses this crucial area and proposes novel distributed grooming solutions which will be applicable to IP-DWDM, Ethernet-DWDM, and SONET-DWDM networks. The proposed framework is further analyzed and tested using the $OPNETModeler^{TM}$ simulation tool.

1.4 Problem Statement

This thesis will focus on connection provisioning in distributed multi-domain, multi-layer networks, i.e., distributed grooming problem. Specifically, it will look at key open issues in inter-layer routing, topology abstraction, and distributed path computation/signaling.

1.5 Research Approach

This thesis will study traffic grooming in distributed multi-domain/multi-layer IP/SONET/DWDM networks. More specifically, it will comprise of the following key steps:

- Conduct comprehensive research survey on related work on grooming and multi-domain provisioning.
- Develop new topology abstractions for multiple domain types. Specifically

simple node and full mesh abstraction schemes will be considered and modified k-shortest paths techniques will be leveraged.

- Develop inter-layer routing update (i.e., triggering) strategies to disseminate abstracted state.

- Develop reliable path computation schemes for multi-domain/multi-layer networks. The schemes will use shortest path and/or load balancing techniques and loose routing mechanisms to achieve distributed grooming capability.

- Study the performance of the various abstractions and path computation schemes with respect to path blocking rate and routing load for various multi-domain/multi-layer network topologies.

1.6 Outline

This thesis is organized as follows. Chapter 2 presents a literature survey on multi-domain/multi-layer networks. Chapter 3 details the proposed topology abstraction, inter-layer routing, and distributed path computation/grooming algorithms. Chapter 4 then discusses the simulation tool used to evaluate the developed framework performance and the performance metrics used in the simulation study. The simulated network topologies and scenarios are then detailed in Chapter 5. Finally, Chapter 6 presents conclusions and direction for future work.

Chapter 2

Background

2.1 Multi-Domain Optical Networking Standards

Various optical standards have been developed within the IETF, ITU-T, and OIF. These efforts are now surveyed, with a particular focus on their multi-domain capabilities:

2.1.1 ITU-T: International Telecommunication Union

The ITU-T has been maturing its multi-domain capable *automatically switched transport network* (ASTN) framework for several years (G.8080, formerly G.ason) [3], which defines a hierarchical setup consisting of *routing areas* (RA). At the lowest level, an RA represents a domain comprising of physical nodes and links. At the higher level, an RA represents multiple "abstract" nodes and links. ASTN further defines component groups to setup, maintain, and release client connections, e.g., an RA can have one/more *routing controller* (RC) entities. ASTN also outlines associated component functions for tasks such as auto-discovery, auto-provisioning, restoration, etc. Note however, that network topology is not made visible to the client layer. Thus, connections are treated as *sub-network point pool* (SNPP) links, making ASTN quite flexible as each lower-layer control plane can be tailored to the particular type of equipment.

2.1.2 OIF: Optical Internetworking Forum

The OIF has largely focused on defining optical interfacing protocols, including a *user-network interface* (UNI) and a *network-network interface* (NNI) [5]. For example, UNI defines bandwidth signaling for client devices to request/release optical connections from carrier SONET/SDH or DWDM domains. Once again, resource/topology state is not propagated to clients across the UNI as no trust relationship is assumed here. Meanwhile the NNI implements inter-domain functionality for reachability/resource exchange and setup signaling and features two variants, *interior NNI* (I-NNI) and *exterior-NNI* (E-NNI). The former interfaces nodes within the same administrative area, with nodes assumed to be homogeneous, whereas the latter interfaces adjacent (possibly multi-carrier) areas. Recently the OIF has also detailed routing and signaling functionalities for E-NNI. Specifically, a hierarchical routing setup is defined (ASTN G.8080) based upon the *open shortest path first - traffic engineering* (OSPF-TE) routing protocol [6]. However the inter-carrier case has not been fully addressed yet. Overall, UNI and NNI can automate circuit setups across multiple "optical" layers, DWDM and TDM.

2.1.3 IETF: Internet Engineering Task Force

Internet Protocol (IP) data networks feature a mature multi-domain setup comprising of a hierarchy of *autonomous systems* (AS) and areas (domains). Within areas, routers run *interior gateway protocols* (IGP) such as OSPF or *intermediate-system to intermediate-system* (IS-IS) to maintain *link state databases* (LSDB) [3]. Meanwhile, the inter-AS level uses *exterior gateway protocols* (EGP), such as the distance vector *border gateway protocol* (BGP), for reachability exchange. Since the BGP is known to highly compress the exchanged information, a more capable route summary is needed for TE circuit routing. Moreover, with growing *quality of service* (QoS) needs, OSPF has defined *traffic engineering* (TE) extensions (OSPF-TE, RFC 2676) for new opaque *link state attributes* (LSA), allowing for the propagation of "QoS-related" state to support advanced *constraint-based routing* (CBR). Note that QoS destination extensions have also been proposed for BGP.

Another extension by the IETF for optical provisioning is the *generalized multiprotocol label switching* (GMPLS) framework, which extends upon *multiprotocol label switching* (MPLS). Specifically, routing this includes new OSPF-TE opaque LSA definitions for DWDM and SONET/SDH links, allowing TE databases (TEDB) to store wavelengths/usages, timeslots/usages, *shared risk link groups* (SRLG), etc. Meanwhile for signaling, *resource reservation protocol - traffic engineering* (RSVP-TE) with its *loose route* (LR) feature has been extended to support hard-state circuit setup/takedown and recovery. In addition RSVP-TE also enables LSP setup across domain boundaries-contiguous, stitched, and nested [7]. Finally, a new *link management protocol* (LMP) is also defined for resource discovery and fault localization, particularly for optical links.

Another recent IETF multi-domain standard is the *path computation element* (PCE) framework [8], which decouples TE path computation from signaling. In this setup a domain can have one/more logical (standalone or co-located) PCE entities which communicate with *path computation clients* (PCC) to resolve connection paths. All PCC-PCE communication is done via a new *PCE protocol* (PCEP) [8]. Although a PCE has access to local domain resource/policy databases, its interdomain visibility may vary [7], from local visibility (low-trust, inter-carrier) to partial visibility (high-trust, intra-carrier). Accordingly, two distributed path computation schemes are envisioned, per-domain and PCE-based [7]. All in all, the PCE framework allows policy control at domain boundaries; a crucial requirement in multi-carrier settings, on par with TE objectives. Specifically, an ingress PCE can enforce policies to determine which requests it will support along with applicable TE constraints/algorithms.

2.2 Research Survey

Despite the above detailed progress in standards, the overall area of multi-domain/multi-granularity optical networking has not seen significant research focus. Although some results can be reported, most wireline multi-domain efforts have largely focused on homogeneous packet-switching networks. Thus, in order to get a better view of the key challenges in multi-domain/multi-granularity network provisioning, a literature survey of related areas is presented below. Specifically, the existing work is classified into three broad areas: multi-domain IP networks, multi-domain DWDM networks and multi-domain IP-DWDM networks.

2.2.1 Multi-Domain IP Networks

Topology aggregation is an important technique that can reduce QoS routing protocol overhead and enhance scalability. In [9], the impact of topology aggregation on QoS routing performance in multi-domain IP networks is evaluated. Moreover, the interactions between topology aggregation and other factors such as routing overhead frequency reduction, different path computation algorithms and various network configurations is also considered for two aggregation schemes: the hybrid aggregation and the weighted aggregation with protocol overhead similar to conventional star aggregation. The results show lower bandwidth rejection rates for the hybrid aggregation compared to the weighted aggregation, and are similar to full-mesh aggregation performance. Further simulations of different size networks also reveal the heavy impact of the previously mentioned factors on the topology aggregation (full-mesh, star, hybrid and weighted aggregations) in QoS routing performance.

[13] introduces a source-oriented *topology aggregation* (TA) approach for efficient QoS-based routing in scalable networks. The goal of this approach is to eliminate the redundancy in the advertised state information by taking into consideration the relevance of this information for path selection. Three TA schemes (unified quasi-stars, source-oriented simple-node, and source-oriented star) are developed, which provide different trade-offs between compaction and accuracy. Moreover, two new approaches for computing the weights of the logical links, taking into consideration more than one QoS parameter at a time, are implemented, (*closest-single-path approach* (CSPA) and *modified-multiple-path-parameters-best-case approach* (Modified-MPPBCA)), which make a compromise between the conventional *single-path-parameters approach* (SPPA) and *multiple-path-parameters-best-case approach* (MPPBCA) approaches.

Extensive simulations for sparse and dense topologies under static and dynamic scenarios are performed. Simulations results show that under static scenarios, Modified-MPPBCA achieves almost the same success rate as the conventional MPPBCA but with lower crankback rate. CSPA performs as good as MPPBCA without causing any crankback. Moreover, the source-oriented versions of the simple-node and star schemes perform better than their conventional counterparts. On the other hand, simulations results under the dynamic scenario show that for all schemes the success rate decreases and the crankback rate increases with the length of the update interval and that the performance trends are similar to those observed in the static scenario. Moreover, increasing the update interval has more negative impacts on accurate TA schemes (i.e., full-mesh) than on lossy schemes (i.e., simple node). This is due to the fact that an accurate state advertisement gradually loses its value as the update interval increases.

Meanwhile in [10], a novel QoS representation for topology aggregation in delay-bandwidth sensitive networks is presented. Using the line segments to represent the QoS parameters of logical links, this paper follows the *private network-to-network interface* (PNNI) star topology aggregation with bypasses model to aggregate a domain topology. Here, a node is fully aware of its domain topology but only has partial information about the other domains. Algorithms are defined to compute the line segment of a link based upon the QoS constraints and to route inter and intra-domain QoS traffic in the suggested model. Metrics such as delay deviation, success ratio and crankback ratio are examined and comparisons done versus other aggregation and routing schemes, e.g., best point algorithm, worst point algorithm, and the modified Korkmaz-Krunz algorithm. Overall simulation results show that the proposed algorithms are suitable for other additive/bottleneck metric pairs and yield high success ratios with small crankback rate. However, associated compute complexities are very high here. To resolve this bottleneck, lower-end routers are allowed to forward their inter-domain routing requests to default designated routers with high computational capabilities.

More recently, [7] also presents a review of the latest interdomain path computation and setup framework for MPLS networks. Those schemes are primarily divided into two category: the PCE-based and the per-domain based path computation schemes. A per-domain path computation scheme is proposed, conforming to practical constraints of routing domains whereby the service providers do not leak routing information outside their domains for confidentiality and scalability reasons. By adding simple extensions to crankback signaling attributes while maintaining RSVP-TE scalability, the *computation while switching scheme* (CWS) extends the standard per-domain path computation scheme by continuing the quest for a better path instead of terminating the search at the first available path. Overall, this yields a "near-optimal" interdomain path without assuming the availability of complete topology information. Moreover, the CWS scheme inherently guarantees that the computed interdomain paths will traverse a minimum number of routing domains by using a simple procedure to select a path among a set of candidate paths while ensuring that the domain information remains confidential. A comparison of the CWS scheme versus the standard per-domain path computation scheme is also presented to show the improvement in terms of the total number of LSPs successfully placed (about 30%) and the number of traversed domains. Moreover, in terms of path setup latency, the CWS scheme remains comparable to other per-domain path computation schemes. However, the CWS scheme introduces extra signaling overhead as it significantly increases the number of propagated messages when finding a path. However, using an optional procedure used within the CWS scheme allows for lowering this overhead. Note that comparisons with the PCE-based computation schemes is not performed because of the sizable architectural differences.

Meanwhile, [11] studies diverse work on protection path setup in multi-domain networks. Here, the authors develop an aggregated representation that captures both the path diversity and the link-state characteristics of a path traversing a domain. A distributed routing algorithm is also defined to leverage the aggregated representation (flooded across domains using a PCE-based architecture) and find two disjoint (primary and backup) QoS paths across multiple domains. This general case of path provisioning is extended to path provisioning under the constraint of export policies where line graphs are used to capture the customer-provider and peer relationships between routing domains. However, although mathematical proofs are presented for the suggested algorithms, no simulations are detailed.

2.2.2 Multi-Domain DWDM Networks

In [12], a multi-segment optical network framework is proposed as a tool to solve the end-to-end provisioning over interconnected optical networks. This framework consisted of three key components: (1) individual segments with segment specific properties, (2) segment interconnections through gateways, and (3) traffic locality. Based upon this framework, a 5-step heuristic method is applied to solve the end-to-end provisioning problem. This heuristic method transforms the network into a multi-granularity graph and assigns link weights using one of three heuristics, i.e., finest granularity first, minimize number of multiplexer/demultiplexer nodes, or minimize hop counts. Path selection is then applied using one of three schemes: *end-to-end* (E2E) routing, *concatenated shortest path routing* (CSR) or *hierarchical routing routing* (HIR). Simulation results for a backbone connecting regional networks show the ability of the solution to handle various network conditions under different control plane architectures.

Meanwhile [14] extends the work in [12] to optical path provisioning in multi-domain networks ,i.e., *end-to-end Ethernet switched path*. Here different domains are modeled as networking segments, each with different wavelength capacity and connected to other segments by a set of gateways. Using wavelength-routed call blocking probability as a metric, the benefits of each of these algorithms are analyzed with respect to the locality of traffic, i.e., global vs. local. Moreover, different gateway adaptation capabilities are also considered, i.e., such as wavelength merging, conversion, or waveband interchange. Simulation results for multi-segment network with four regional all-optical networks interconnected over an all-optical WDM backbone show that the E2E scheme is beneficial for low blocking probability if the global traffic is dominating. In all other cases, CSR and HIR are shown to be more beneficial due to their lower computational complexity and information storage requirements. Moreover, these schemes can even approach the blocking probability of the E2E scheme by increasing the number of gateways and by their appropriate locations and adaptation capabilities.

2.2.3 Multi-Domain IP-DWDM Networks

Building on the current round of carrier Ethernet standards, [15] demonstrates how to support and implement full traffic engineering in a global-scale two-tiered native Ethernet-over-WDM optical networking architecture. To implement this vision, three key innovative components have been presented. First, a novel converged Layer 2/1 *networking model* is presented, in which all networking functionalities and intelligence is passed on down to the optical layer. This network is controlled by using unified GMPLS-based control plane, to manage both the optical and the Ethernet switches (with grooming capabilities). However only the OXC nodes (and not Ethernet switches) keep records of the network topology and resource usage information for both layers (physical and logical), making it generic and transparent to the higher layer protocols. Next, a fully *distributed integrated routing and signaling framework* for dynamically provisioning *Ethernet switched path connections* (ESP) at any bandwidth granularity including both full wavelength and finer granularity (sub-lambda) rates. The framework uses a customized version of the RSVP-TE signaling with backward reservation to accommodate for the complexity introduced by the grooming capabilities of the network. Finally, a novel notion of an integrated *link-state advertisements* (LSA) strategy based on the lightpath availability is discussed. Using this approach, the full view of the network status is generated by extracting the physical links used by the logical links (lightpath) from the initial network resource status, assuming the physical resources in the network as not as frequently added/removed as the logical resources. Simulations results for 16-node NSFNET show that connection blocking and contention probabilities for different LSA updating schemes yield better results at the expense of increased number in exchanged control messages.

Meanwhile [16] studies hierarchical routing for GMPLS-based ASON networks. Specifically, this problem is subdivided into three sub-problems: network aggregation, update policy and route selection. Next a proposed algorithm called *lightpath aggregation scheme* (LAS), based upon link delay and the number of available bandwidth as QoS parameters, is used for network aggregation. A new update policy is also suggested along with a new hierarchical routing algorithm called $ALG3\_H/k$, which addresses the routing inaccuracy problem introduced by the network aggregation. In comparing LAS to *Node Aggregation Scheme* (NAS), simulations results show reduced increment of the blocking probability (an improvement of about 2.32%) while substantially reducing the signaling overhead, thus improving ASON scalability. Moreover, heavily loaded network simulations showed an average extra improvement of 1% when $ALG3\_H/k$ and LAS are applied jointly over $NAS - FF$ and $NAS - ALG3\_H/k$.

An early study on multi-domain DWDM networks is presented in [17]. Namely, the authors address the problem of inter-domain routing in the next-generation optical transport networks from an algorithmic perspective. It suggests a multi-domain/ multi-layer network model and a routing and path selection algorithm based on the next-hop routing technique. This makes the proposed inter-domain dynamic routing scheme flexible and scalable like the scheme used in the Internet. Comparing experimental results of different cost metrics and network models show that the scheme can effectively set up end-to-end connections across multiple domains.

Finally, [18] presents a comprehensive framework for inter-domain lightpath provisioning in all-optical and opto-electronic DWDM networks. A hierarchical routing setup is presented along with two topology abstraction schemes, simple node and full-mesh. Moreover, detailed inter-domain routing/triggering policies and RWA and signaling procedures are also defined. Although the suggested framework incurs high inter-domain routing load when full-mesh abstraction is used, it clearly reduces the inter-domain lightpath blocking probability.

2.3 Open Challenges

Although the above studies represent a compelling body of work, they are not readily applicable to multi-domain *multi-layer* networks. Specifically, they do not address the inherent vertical grooming dimension. This thesis will address this open topic area.

Chapter 3

Multi-Domain/Multi-Layer Routing and Provisioning

Distributed multi-domain/multi-layer network provisioning is a very challenging problem owing to the grooming aspect across multiple layers. Herein, a comprehensive framework is developed to address this problem in two-layer IP/Ethernet and optical networks. Note that full opto-electronic wavelength conversion is assumed at the optical DWDM layer, thereby precluding the need for wavelength-selected RWA provisioning. Overall, this is a fairly realistic assumption as most operational DWDM networks employ significant amounts of regenerators and many studies have shown very close blocking performance between partial and full wavelength conversion at both the single [19] and multiple domain levels [20]. However, DWDM lightpath

RWA considerations can still be treated in future studies.

The proposed scheme addresses several key steps in multi-domain/multi-layer provisioning. Foremost, a multi-domain/multi-layer *topology abstraction* model is defined to condense domain-level state at multiple layers/granularities. Note that this favors link-state routing implementations which are anyways most suitable for handling the added dimensionality of connection routing. Subsequently, *inter-domain routing and triggering policies* are derived to disseminate link state information for various link types, i.e., physical, abstract and tunneled links. Finally, distributed *multi-domain/multi-layer path computation (grooming) and signaling* schemes are developed to setup connections. These schemes are fully compatible with the existing GMPLS protocol frameworks, including RSVP-TE and PCE. Details on these steps are now presented.

3.1 Notation Overview

Before presenting details on each of the above steps, it is first necessary to develop the required notation. Here all set and vector entities are denoted in bold and it is assumed (without loss of generality) that link connectivity is bi-directional, i.e., there are two opposite direction links between a pair of connected nodes. Now consider a multi-layer network comprising D domains, with the i^{th} domain having n^i nodes and b^i border/gateway nodes, $1 \leq i \leq D$. This network is modeled as a collection of domain sub-graphs, $G^i(\boldsymbol{V}^i, \boldsymbol{L}^i)$, $1 \leq i \leq D$, where $\boldsymbol{V}^i = \{v_1^i, ..., v_{n^i}^i\}$ is the set of physical domain nodes and $\boldsymbol{L}^i = \{l_{jk}^i\}$ is the set of physical *intra-domain* links in domain i $(1 \leq i \leq D,\ 1 \leq j, k \leq n^i)$, i.e., l_{jk}^i is the link between nodes v_j^i and v_k^i. A given link, l_{jk}^i, has capacity c_{jk}^i. Furthermore, $\boldsymbol{B}^i$ represents the set of border nodes within domain i and without loss of generality, it is assumed that these nodes are numbered as the *first* group of nodes in the domain, i.e., $\boldsymbol{B}^i = \{v_1^i, ..., v_{b^i}^i\}$.

Meanwhile for multi-domain/multi-layer routing, a *higher-level* topology is also defined to condense the domains and their associated inter-domain links. Namely, this topology is given by a graph $H(\boldsymbol{U},\boldsymbol{E})$, where $\boldsymbol{U} = \sum_i \{\boldsymbol{B}^i\}$ is the set of border nodes across all domains ($1 \leq$ i $\leq$ D) and $\boldsymbol{E}$ is the set of links. Now in this higher level topology, *three* different link types are identified, i.e., *physical* ($\boldsymbol{E}_p$), *abstract* ($\boldsymbol{E}_a$) and *tunneled* ($\boldsymbol{E}_t$), i.e., $\boldsymbol{E} = \boldsymbol{E}_p \cup \boldsymbol{E}_a \cup \boldsymbol{E}_t$. Physical links interconnect two border nodes in different domains and are denoted as $\boldsymbol{E}_p = \{e_{kmn}^{ij}\}$ where link e_{kmn}^{ij} is the n^{th} link interconnecting v_k^i in domain i with v_m^j in domain j. Additionally, c_{kmn}^{ij} is the spare/available capacity on this physical link. Meanwhile, the abstract links are generated by the full mesh abstraction (Section 3.2.2) only and represent the summarized traversal cost of a domain, i.e., denoted by $\boldsymbol{E}_a = \{e_{jk}^{ii}\}$ where link e_{jk}^{ii} is the abstract link between border nodes v_j^i and v_k^i in domain i. Similarly c_{jk}^{ii} is the computed available capacity on abstract link e_{jk}^{ii}. Finally, tunneled links represent underlying physical DWDM lightpath connections between higher layer IP/MPLS border nodes, i.e. $\boldsymbol{E}_t = \{e^{ij}_{kmn}\}$ where e^{ij}_{kmn} is the n^{th} tunneled connection link between border routers v^i_k in domain i with v^j_m in domain j (and c^{ij}_{kmn} denotes the spare capacity on this tunneled link). These links are setup during the signaling stage and can be shared by multiple higher layer connections (i.e., IP/Ethernet), as detailed in Section 3.4.2.

3.2 Topology Abstraction

Topology abstraction summarizes domain-level state. In particular the proposed hierarchical scheme designates a specific *border* node in each domain as a *routing area leader* (RAL) [3]. This entity computes an abstracted topology for the domain by transforming the physical topology into an abstract reduced graph. Specifically, two abstraction schemes are presented, i.e., simple node and full-mesh (Figure 3.1), as evolved from earlier proposals for data/cell switching networks [21]. Note that alternative star abstractions can also be considered but these are left for future study. The *abstract* state information is then flooded to border nodes across all domains in order to maintain a synchronized global virtual view of the whole network (Section 3.3). This abstract information is then used to compute/groom end-to-end inter-domain IP/Ethernet path connections (Section 3.4).

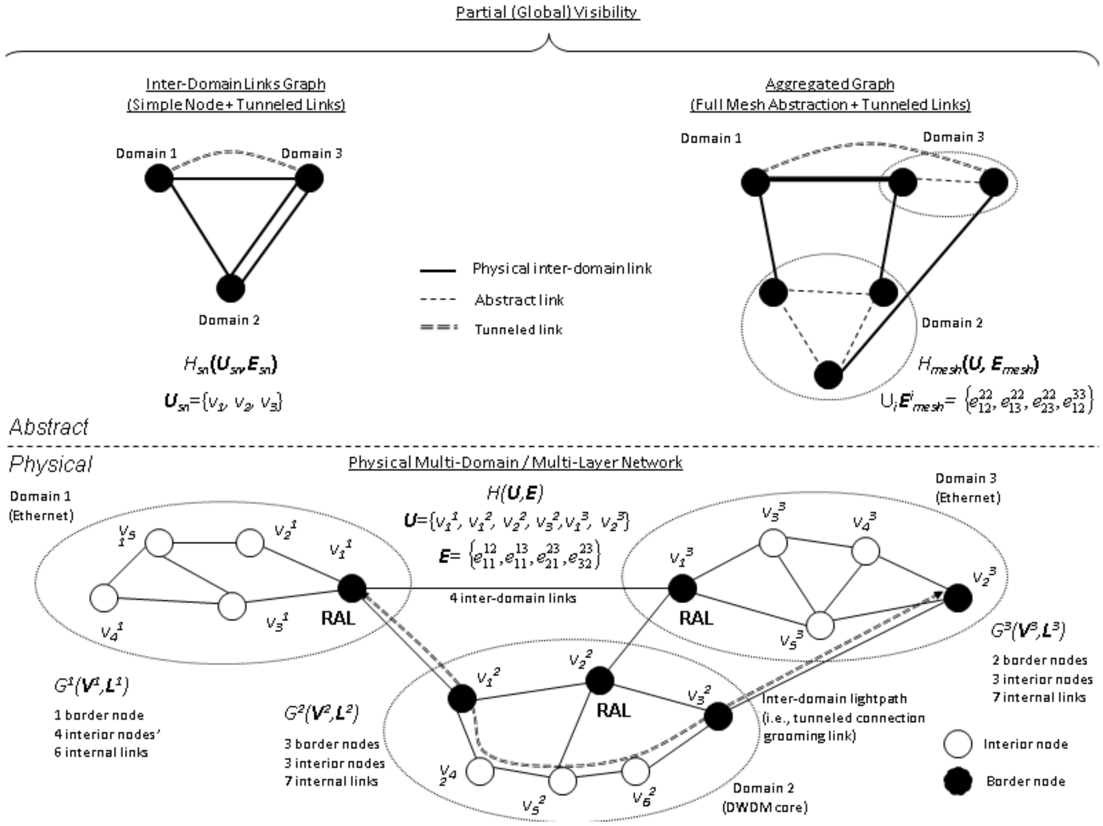

Figure 3.1: Physical/Abstract/Tunneled Links

3.2.1 Simple Node (SN) Abstraction

This is the simplest of all the abstraction schemes and condenses a domain into a single *virtual* node emanating all physical inter-domain links. For example, the three border nodes in domain 2 in Figure 3.1 are simply collapsed into a single virtual node with 3 inter-domain links. This scheme provides no visibility into domain-internal state and has low inter-domain routing overheads. Mathematically, the above transformation can be represented as $H(\boldsymbol{U},\boldsymbol{E}) \rightarrow H_{sn}(\boldsymbol{U}_{sn}, \boldsymbol{E}_{sn})$ where $\boldsymbol{U}_{sn} =$

$\{v^i\}$ is the condensed set of virtual nodes representing each domain i and $\boldsymbol{E}_{sn}$ is the set of physical inter-domain links and tunneled links only, i.e., no abstract links, $\boldsymbol{E}_{sn} = \boldsymbol{E}_p \cup \boldsymbol{E}_t$. Namely, border node set $\boldsymbol{B}^i$ is mapped to a single virtual node vertex, v^i, e.g., see Figure 3.1.

3.2.2 Full-Mesh (FM) Abstraction

The full-mesh scheme is designed to perform intra-domain state summarization. Namely, the i^{th} domain $G^i(\boldsymbol{V}^i, \boldsymbol{L}^i)$ is transformed to a sub-graph containing the border nodes interconnected via a fully meshed set of *virtual* links, i.e., $\boldsymbol{E}^i_{mesh} = \{e^{ii}_{jk}\}$. Here, available capacities are computed for all of these virtual links, c^{ii}_{jk}, to summarize the average capacity needed to traverse through the domain between the border nodes. For example, subgraph $G^2(\boldsymbol{V}^2, \boldsymbol{L}^2)$ in Figure 3.1 is transformed by connecting the border nodes with virtual links e^{22}_{12}, e^{22}_{21}, e^{22}_{13}, e^{22}_{31}, e^{22}_{23}, e^{22}_{32}. Hence this abstraction can be represented by the following transformation $H(\boldsymbol{U},\boldsymbol{E}) \rightarrow H_{mesh}(\boldsymbol{U}, \boldsymbol{E}_{mesh})$, where $\boldsymbol{E}_{mesh} = \bigcup_i \boldsymbol{E}^i_{mesh} \cup \boldsymbol{E}_p \cup \boldsymbol{E}_t$, i.e., $\boldsymbol{E}_a = \bigcup_i \boldsymbol{E}^i_{mesh}$.

Now the exact algorithm for computing the full-mesh abstraction/graph transformation is shown in Figure 3.2 and is run at the RAL node. The scheme basically loops through each border node pair (indices j, k) and computes the available capacity for the corresponding virtual link in the abstraction. Namely, the scheme first runs the

K-shortest path algorithm to generate a set of paths between each border node pair, denoted as $\{\boldsymbol{p}^i_{jkm}\}$ where $\boldsymbol{p}^i_{jkm}$ is the m^{th} path vector (node-sequence) between border nodes v^i_j and v^i_k in domain i, $1 \leq m \leq K$. These paths are then searched for the maximum bottleneck link capacity on all the route links, i.e., capacity on abstract link e^{ii}_{jk} is $c^{ii}_{jk} = max_m\{botteleneck\ capacity(\mathbf{p}^i_{jkm})\}$.

Overall full-mesh abstraction provides more accurate intra-domain usage state, albeit at the cost of significant computational complexities at the RAL nodes and higher inter-domain routing loads. Namely, inter-domain nodes must maintain additional state for $O(n^i(n^i-1)) = O((n^i)^2)$ virtual links for domain i, in addition to the physical inter-domain and tunneled links.

3.3 Multi-Layer Routing and State Dissemination

Routing protocols use triggering update policies to help disseminate link-state. In general research has shown that relative change-based strategies are the most effective, hence in terms of generating timely updates and controlling messaging overheads [6]. Hence similar strategies are adopted for the hierarchical routing framework herein. First of all, at the intra-domain routing level (first-level), LSA update messages are generated using a *significance change factor* (SCF) approach [6]. Namely a link-state update is only sent if the relative change in free capacity on the link exceeds

```
% Given domain-level sub-graph G^i(V^i, L^i), compute domain full-mesh abstraction

% Loop over all domain nodes
for j = 1 to b^i
{
  % Loop over all domain nodes
  for k = 1 to b^i
  {
  % Do not compute abstract link from node to itself
    if (j ≠ k)
    {
      Compute K shortest paths from v_j^i to v_k^i in G^i(V^i, L^i), i.e., {p_jkm^i},
          where 1 ≤ m ≤ K
      Initialize bottleneck capacity tracking variable X
      % Search paths
      for k = 1 to K
      {
        Find bottleneck link on k-shortest path
        if (bottleneck capacity > X)
        {
          Abstract link capacity c_jk^ii = X
          Update maximum bottleneck capacity and set it to X
        }
      }
    }
  }
}
Output abstract link capacities, c_jk^ii, 1 ≤ j,k ≤ b^i, j ≠ k
```

Figure 3.2: Full mesh abstraction computation algorithm

a pre-specified SCF value and the duration since the last update exceeds a *hold-down timer* (HT) [6]. The associated message indicates the free/reserved capacity on the respective link. However update triggering policies at the inter-domain OSPF-TE level (i.e., second level) require some more specific considerations to handle the three different link types involved, e.g., physical, abstract and tunneled. These policies help distribute and maintain the higher-level topologies introduced in Section 3.1.

First consider physical and tunneled links, which connect border nodes in different domains. Since these entities are associated with underlying *physical* resources, (i.e., actual links or lightpath connections), their associated updates can simply be generated using the same SCF-based approach detailed above for intra-domain physical links. Namely if the number of free capacity on a physical inter-domain link changes by more than SCF, an update is flooded by the *originating* border node to all other border nodes in its domain and directly-connected border nodes in other domains.

However abstract links do not correspond to direct physical entities, as they are generated by full mesh topology computation. Hence updates for these link types can only be sent by their generating RAL nodes as they pertain to "non-existent" computed entities. By and large, two types of abstract link triggering policies can be considered hence, periodic and relative-change based [3, 17]. The former scheme computes full-mesh abstractions at fixed timer intervals, i.e., *inter-domain hold-off timer* (IHT) and sends out abstracted link updates for all abstract links. Here a shorter IHT value will clearly yield more accurate information, albeit at the expense of excessive inter-domain messaging and computation overheads (low scalability). Conversely the latter scheme extends periodic updates by only sending abstract link updates if there are sufficient changes in domain state, i.e., combining the timer and SCF-based strategies. This "two-step" method is generally more responsive and scalable and is adopted here. Specifically, RAL nodes periodically compute domain abstractions at the expiry of the IHT but abstract link updates are only sent for the subset who have experienced significant relative change.

3.4 Distributed Multi-Domain/Multi-Layer Path Computation and Signaling

Distributed inter-domain/inter-layer path provisioning requires both path computation and signaling setup. This is a very challenging problem given the limitations of partial global state and the inherent grooming aspect of multiple domain link granularities. To resolve these concerns, a hierarchical distributed computation framework is developed using "skeleton" inter-domain route computation and intra-domain expansion. Specifically each border node uses its "global" higher-level topology (Section 3.1) to first compute a LR [3] sequence of domains to the destination. This "skeleton" route is then used to drive distributed RSVP-TE signaling and *explicit route* (ER) expansion procedures, as shown in Figure 3.3. Note that LR signaling and ER expansion are standardized techniques that are widely used in MPLS/GMPLS networks [3, 8, 18]. Now consider the details for both the LR and ER stages.

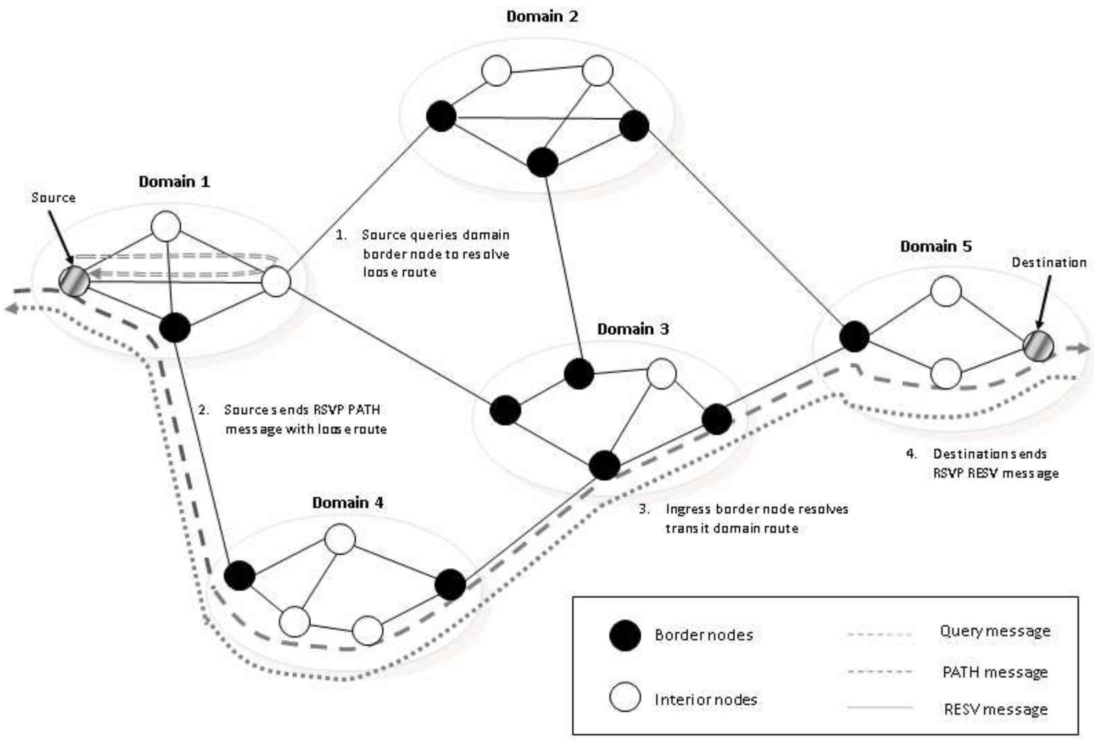

Figure 3.3: Loose route computation and path reservation

3.4.1 Skeleton Path Computation/Grooming

Generic overview of inter-domain path setup from an interior source node is shown in Figure 3.3. This node first sends a query to its nearest (or a designated) border node to compute a LR of capacity x to the destination node, i.e., as it does not have a "global" view. This node may/may not be the same as the RAL node (Section 3.2). The queried border node returns a LR sequence specifying the end-to-end border node sequence to the destination domain. In particular this sequence contains the egress border node at the source domain, all ingress/egress border nodes at intermediate domains, and the final ingress border node at the destination domain, Figure 3.3. In the special case where the source is itself is a border node, LR computation can optionally be done locally.

Now the actual LR computation is done using the global higher-level topology/ graph, i.e., $\boldsymbol{H}_{sn}$ or $\boldsymbol{H}_{mesh}$ (Sections 3.2.1, 3.2.2). The detailed pseudocode for this computation scheme is presented in Figure 3.4. First, the higher level topology, $\boldsymbol{H}_{sn}$ or $\boldsymbol{H}_{mesh}$, is modified by pruning any links without sufficient capacity (i.e., abstract links with $c^{ii}_{jk} < x$ or tunneled links with $c^{ij}_{kmn} < x$). The next step focuses on the selection of the border node pair, i.e., source domain egress node and destination domain ingress node. Now the source domain's computing node does not know the exact topology of the destination domain, here it is generally difficult to choose the "closest" border node from which to ingress. Although each source/destination border node combination can be searched, results in flat multi-domain DWDM networks show minimal blocking probability reduction versus random selection [18]. Hence the source domain egress and destination domain ingress nodes are selected in a random manner. The available path routes between these pair nodes are then searched in order to achieve a desired resource provisioning objective. Specifically, the links in **H'** are assigned appropriated weights and then the actual LR path computation is done using Dijkstra's shortest path algorithm. Note this follows a "unified" path computation/grooming approach as opposed to a two layer approach [22]. Specifically, three link weightings are considered here in order to pursue different traffic engineering objectives:

- **Minimum Hop:** In this basic approach, all links in **H'** are assigned unity cost to determine shortest inter-domain hop count path.
- **Minimum Cost:** In this modified scheme, DWDM links are assigned higher (albeit fixed) costs of α ($\alpha \geq 1$) whereas IP/Ethernet links (physical, abstract, tunneled) are kept at unity cost. The goal here is to first drive increased usage on existing IP/Ethernet "higher-layer" links before using DWDM wavelengths which are generally more costly to activate.
- **Minimum Distance:** This approach follows a load balancing objective by assigning each link a cost inversely proportional to its residual capacity (see Figure 3.4). The goal here is to prevent capacity exhaust on critical bottleneck path links and thereby better distribute connection loads across domains. However, resultant path lengths may be larger here, leading to higher resource consumption.

3.4.2 Path Signaling

Finally, the computed LR sequence is then inserted into a RSVP-TE PATH signaling message and sent downstream to resolve the explicit end-to-end path (intra,

% Compute LR of capacity x from source domain s to destination domain d.
Make a copy of higher level topology graph **H'** = $\mathbf{H}_{sn}$ or **H'** = $\mathbf{H}_{mesh}$ by pruning all links with insufficient capacity, i.e., $c_{jk}^{ii} < x$ (abstract links) or $c_{kmn}^{ij} < x (i \neq j)$ (physical, tunneled links)

% Randomly select source domain border node
i=rand(1,b^s)
% Randomly select destination domain border node
j=rand(1,b^d)
% Compute LR sequence
if (minimum hop)
{ Set cost of all links in **H'** to unity }
if (weighted link cost)
{
Set cost of all DWDM links in **H'** to α ($\alpha > 1$)
Set cost of all IP/Ethernet links in **H'** to unity
}
if (load balancing)
{
Set cost of all physical and tunneled links in **H'** to $w_{kmn}^{ij} = \frac{1}{c_{kmn}^{ij}+\varepsilon}$ where $\varepsilon \approx 0$
Set cost of all abstract links in **H'** to $w_{jk}^{ii} = \frac{1}{c_{jk}^{ii}+\varepsilon}$ where $\varepsilon \approx 0$
}
Run Dijkstra's shortest path algorithm on **H'** to compute shortest path $\mathbf{P}_{ijk}^{sd*}$
Return final LR sequence, i.e., $\mathbf{P}_{ijk}^{sd*}$

Figure 3.4: ER path computation

inter-domain nodes, see Figure 3.3). Here intermediate ingress border nodes perform ER expansion on the incoming LR sequence in order to resolve the exact sequence of nodes and links to the intermediate egress border nodes in their domain. Note that the destination domain's ingress border node must also perform ER expansion to the destination node. Consider the details of ER expansion.

First, the computed LR sequence is inserted (by the expanding border node) into the ER field, denoted by a vector $\mathbf{R}$, of a downstream RSVP-TE PATH signaling message, i.e., initially $\mathbf{R} = \mathbf{P}_{ijk}^{sd}$ (as computed in Figure 3.4). Now consider an intermediate domain i, whose ingress border node receives a PATH message with a explicit route field $\mathbf{R}$ specifying an egress border node in the domain. This ingress node must first expand the ER sequence by appropriately filling in the explicit intra-domain node sequence. This performs a local route computation, (using first level OSPF-TE topology), done using a *widest-shortest path* (WSP) approach [18] on the intra-domain virtual topology. Although alternate strategies such as *shortest-widest path* (SWP) (load balancing) can be considered, these are left for future study as the focus here is on inter-domain path computation. Specifically the K-shortest (intra-domain) paths between the given ingress/egress border nodes, i.e., $\mathbf{p}_{jkm}^{i}$ as per Section 3.2.2, are computed and then searched to find the minimum hop path having at least the requested bandwidth. If LR expansion is successful, the above-expanded intra-domain segment is appropriately inserted into the ER field ($\mathbf{R}$ vector) of the RSVP-TE PATH message. Finally the receiving destination node returns an upstream RSVP-TE RESV message towards the source. Hence, all nodes receiving this RESV message must check and reserve the required bandwidth on each link and terminate setup if the required bandwidth is unavailable, see [3] for complete RSVP-TE message processing rules.

As mentioned earlier (Section 3.1), tunneled links are also generated at the signaling phase. These links pertain to underlying DWDM lightpath segments or a computed route and play a vital role in inter-layer grooming. Basically, the goal is to allow multiple (smaller) higher layer IP/Ethernet flows to use any underlying DWDM lightpath segments that may have already been setup by earlier connection requests. Now, consider the setup of tunneled links. Namely, for a new path setup, any DWDM route segment is extracted as a direct tunneled link between the two respective IP/Ethernet ingress/egress nodes. Namely, when the RESV message is propagating back to the source, all IP/Ethernet border nodes must check to see if their *downstream* node is a DWDM border node. If so, these border nodes must extract the lightpath segment from the RESV message to the downstream egress IP/Ethernet node. This segment is used to create a "direct" inter-domain tunneled link which is then advertised (by the extracting border node) at the inter-domain level. For example, in Figure 3.1, domain 1 is the IP/Ethernet domain. Hence, when setting up the path shown, border node v_1^1 in domain 1 (IP border node) must extract/generate a tunneled link entry at the inter-domain routing level going to border node v_2^3 (domain 3). Similarly, tunneled links are also taken down during takedown signaling. Specifically when the capacity usage on a tunneled link drops to zero, the associated border node issues an RSVP PATH takedown request.

Chapter 4

Simulation and Performance Tool

4.1 Introduction

Network designers are constantly being challenged by the growing complexity of new networking technologies and the increasing scale of associated topologies deployed in the field. In order for researchers to innovate new solutions for these scenarios, robust network simulation software is needed to easily and intuitively model the intricate end-to-end behavior of protocols. These solutions must also be able to efficiently analyze the performance of these protocols and technologies in network infrastructure models of realistic scale [23].

Discrete event simulation (DES) is a methodology that can be used to emulate the behavior of real networks in response to various inputs, i.e., termed as events. This is a widely-used technique that leverages an event queue to buffer events, each of which is directed to a specific node and ordered by a timestamps. Today many different simulation tools have been developed for DES, e.g, $OPNET\ Modeler^{TM}$, $NS2$, $OMNET$, etc. However, in this study the $OPNET\ Modeler^{TM}$ tool is chosen to test the suggested framework as it provides the most complete set of features to simulate real world networks. More importantly, this tool also comes with a full C/C++ backend, allowing for detailed, specialized model development. As such, it has gained much traction and is widely used by professionals and various government departments to simulate and evaluate various networks performances. Some of the key features of $OPNET\ Modeler^{TM}$ includes:

- Robust, generic discrete event simulation engine
- Hundreds of protocol and vendor device models with source code
- Object-oriented modeling features and capabilities
- Hierarchical modeling environment
- Discrete event, hybrid, and optional analytical simulation
- 32-bit and 64-bit fully parallel simulation kernel

- Grid computing support for distributed simulation
- Optional system-in-the-loop to interface simulations with live systems
- Open interface for integrating external object files, libraries, and other simulators
- Integrated, *Graphical User Interface* (GUI) based debugging and analysis

4.2 Network Topologies

Now in order to properly investigate multi-domain/multi-layer routing and path provisioning schemes, various realistic test topologies are needed first. However, since there are really no "standard" multi-domain/multi-layer test topologies, it is necessary to design various types to cover a wide range of possible real world scenarios. Herein, two different and unique network topologies are created and tested shown below in Figure 4.1 and Figure 4.2. First, a simple seven-domain topology is devised, consisting of one DWDM domain hosting six Ethernet "client" domains (4.1). The goal here is to study the impact of connection tunneling in basic (smaller) network settings. Next, a larger nineteen-domain topology is considered, consisting of five DWDM domains hosting fourteen Ethernet "client" domains (Figure 4.2). This larger topology is taken to represent extended national (e.g, inter-carrier) and trans-national backbone settings. The key difference between the two topologies is

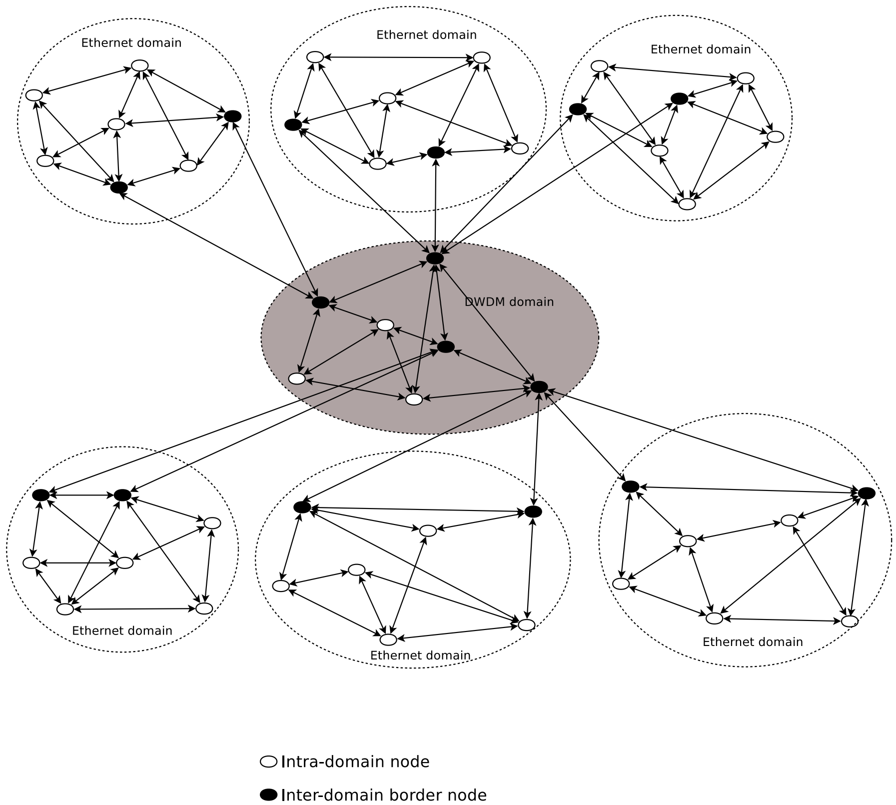

Figure 4.1: 7-domain network topology (33 intra-domain nodes and 16 inter-domain border nodes)

the inter-domain connectivity, defined as the average number of inter-domain links connecting a domain. Namely, the seven-domain topology consists of seven domains and and twelve inter-domain bi-directional physical links, resulting in domain degree of 3.43. On the other hand, the nineteen-domain topology consists of nineteen domains and fifty nine bi-directional physical links, resulting in a higher domain degree of 6.21.

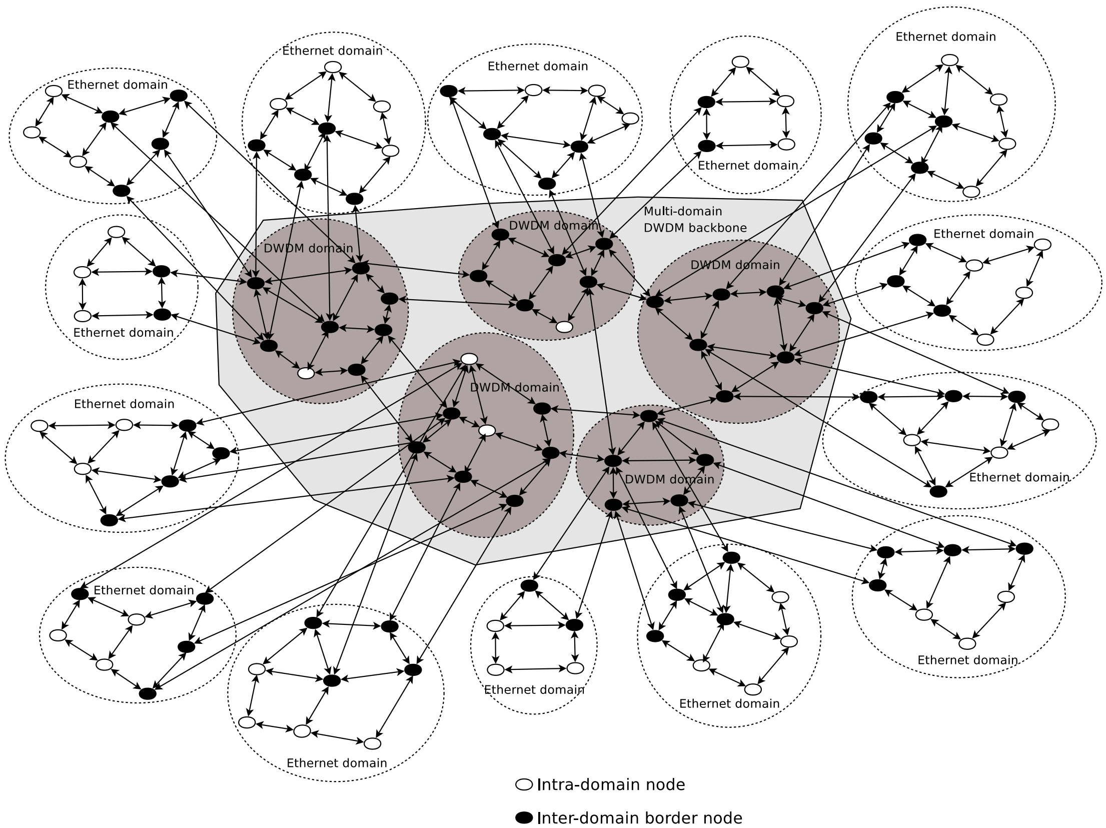

Figure 4.2: 19-domain network topology (50 intra-domain nodes and 80 inter-domain border nodes)

4.3 Performance Metrics

Various evaluation metrics are used to evaluate the performance of the proposed multi-domain/multi-layer provisioning and algorithms, including *bandwidth-blocking rate* (BBR), network load, and routing load. The BBR is defined as the ratio between the failed requested bandwidth of global inter-domain connection attempts, B_{fail}, to the total requested bandwidth of global inter-domain connection attempts, $B_{attempt}$.

The failed requested bandwidth of global inter-domain connection attempts, B_{fail}, is the summation of each failed requested path bandwidth:

$$B_{fail} = \sum_{i=1}^{m} b_i,$$

where b_i is the bandwidth of the i^{th} failed requested path, $1 \leq i \leq m$, and m is the total number of failed inter-domain connection attempts. Meanwhile, the total requested bandwidth of global inter-domain connection attempts, $B_{attempt}$, is the summation of each requested path bandwidth:

$$B_{attempt} = \sum_{j=1}^{M} b_j,$$

where b_j is the bandwidth of the j^{th} requested path, $1 \leq j \leq M$, and M is the total number of inter-domain connection attempts. Hence

$$BB_r = \frac{B_{fail}}{B_{attempt}},$$

where the above is defined as the *global* inter-domain bandwidth-blocking rate. Meanwhile, the Erlang-loading metric is used to measure the network carried load and is defined as:

$$Load\ (Erlangs) = \frac{N * T_{hold}}{IAT},$$

where N is the total number of nodes in the network generating inter-domain requests of an inter-domain connection and T_{hold} is the average connection hold time, and

IAT is the average connection inter-arrival time. Finally, a routing load metric is also defined to measure the inter-domain routing overhead comprising of LSA entities flooded throughout the network to all inter-domain nodes.

$$Routing\ Load = \frac{T_{LSA}}{Total\ Simulation\ Time},$$

where T_{LSA} is the total number of inter-domain LSA entities.

Chapter 5

Performance Evaluation Study

This chapter presents performance evaluation results for the proposed multi-domain/multi-layer provided framework and algorithms. All testing is done using the *OPNET Modeler*^{TM} discrete event simulation tool using the topologies developed in Chapter 4. Specifically, detailed node and process models are coded in C/C++ to implement all of the proposed inter-domain routing, signaling, and path computation algorithms (Chapter 3).

For all of the test cases, inter-domain connections are generated between randomly-selected Ethernet domains and within a given Ethernet domain, nodes are also chosen randomly using a uniform distribution. All connection holding times are fixed at 600 seconds mean (exponential) and associated inter-arrival times are chosen in a random manner contingent to desired loading (exponentially distributed as well). Meanwhile intra/inter domain routing timers (HT, IHT) are set to 10 second durations and the routing update load SCF value is set to 10%. Finally, all simulation runs are averaged over 25,000 connections. Results for seven and nineteen-domain topologies are now presented. Consider the details.

5.1 Seven-Domain Network Topology

First the seven-domain network is simulated and the results presented in Figures 5.1 - 5.6. Namely, Figures 5.1, 5.2, and 5.3 present inter-domain blocking probability results for the three path computation strategies; minimum hop, minimum distance and minimum cost (Section 3.4.1, Figure 3.4), for varying number of tunneled links. The goal here is to evaluate the benefits of increased inter-domain grooming levels. Along these lines, the findings show that higher tunneled link counts are extremely effective in lowering overall blocking probability performance, e.g., the average BBR decrease ranges around 16% averaged across all load-levels (except for the minimum cost scheme, Figure 5.3). Similar runs are also repeated for the full mesh abstraction scheme in Figures 5.4, 5.5, and 5.6. Again, the findings here show a notable decline in inter-domain blocking probability, e.g., 16 tunneled links giving almost 11% lower blocking than 4 tunneled links (Figures 5.4, 5.5, and 5.6).

Now it is very important to also gauge the relative performances of the different topology abstraction schemes for the seven-domain topology. To properly investigate this, Figures 5.7 and 5.8 are generated by extracting specific information from the earlier plots in Figures 5.1 - 5.5. Specifically, Figure 5.7 compares the blocking probability for simple node and full mesh abstractions for varying tunneled link counts and minimum hop path computation. Similarly, Figure 5.8 repeats the above for the minimum distance path computation scheme. These figures reveal some very interesting insights. Foremost, it is clear from Figure 5.7 that the more advanced full mesh topology abstraction scheme gives no discernible blocking probability reduction when coupled with the minimum hop path computation. This is expected as minimum hop routing ignores bandwidth resource usage information and always chooses the shortest path, i.e., full mesh abstract link information is rarely used unless it is below the requested capacity. Conversely, the results with the minimum distance scheme, Figure 5.8, show some very slight gains with full mesh abstraction at lower loads, e.g., at 70 Erlang load and 16 tunneled links, full mesh abstraction gives 3% lower blocking probability than simple node.

Further the performance of minimum hop and minimum distance path computation is also compared in Figure 5.9. Here the plots show that the minimum hop outperforms minimum distance, averaging about 10% lower blocking probability. This is likely not a general result as in this particular seven-domain topology there is only one "underlying" DWDM grooming domain. Finally, the head-to-head performance of minimum hop, minimum distance and minimum cost path computation is compared in Figure 5.10 for simple node abstraction. Results show that the BBR for minimum cost scheme is inbetween the BBR of the two other schemes.

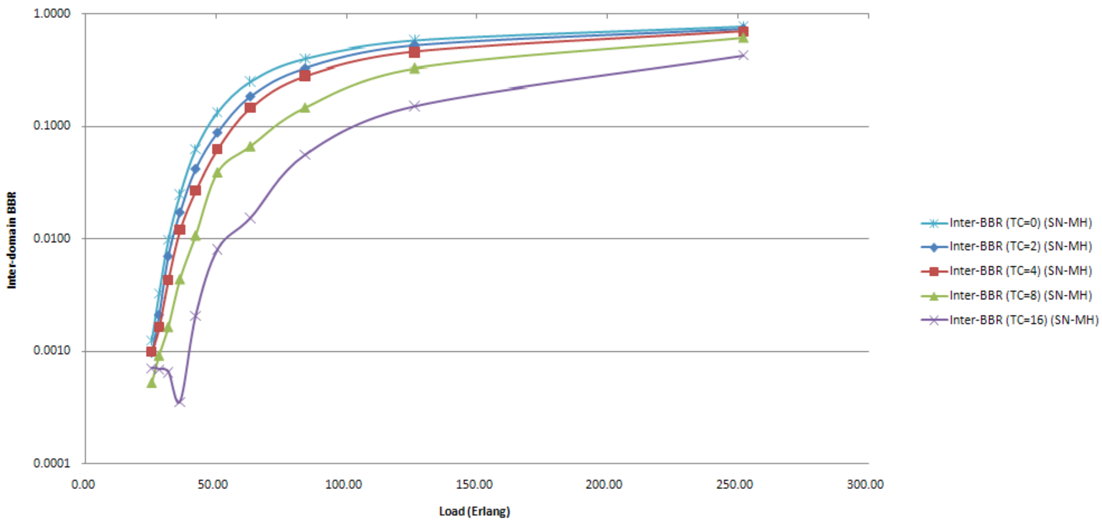

Figure 5.1: 7-domain BBR for simple node abstraction with minimum hop

In addition to blocking probability performance, control plane routing load overheads are also measured for the seven-domain topology. First, Figures 5.11 and 5.12 plot the inter-domain LSA loads for the minimum hop count scheme and varying numbers of tunneled links (simple node and full mesh abstraction, respectively). These findings indicate that the routing load is *inversely* proportional to the number of tunneled links when minimum hop computation is performed, a new and interesting result. For example, simple node topology abstraction with 16 tunneled links

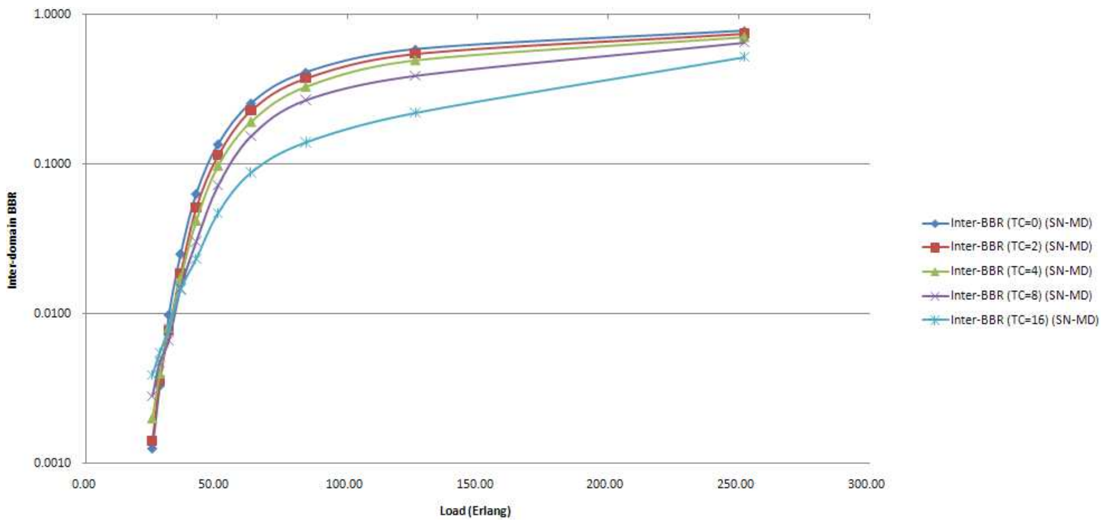

Figure 5.2: 7-domain BBR for simple node abstraction with minimum distance

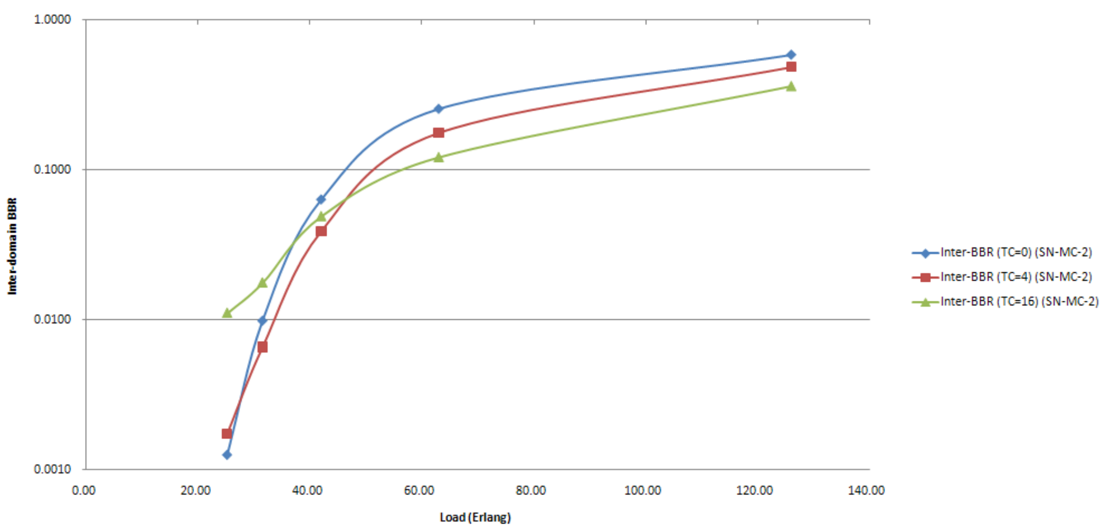

Figure 5.3: 7-domain BBR for simple node abstraction with minimum cost ($\alpha = 2$)

gives the lowest LSA load in both cases (Figures 5.11 and 5.12). This observation can be explained by the fact that increased connection grooming generally yields smaller variations in tunneled link capacities, thereby generating fewer updates. It

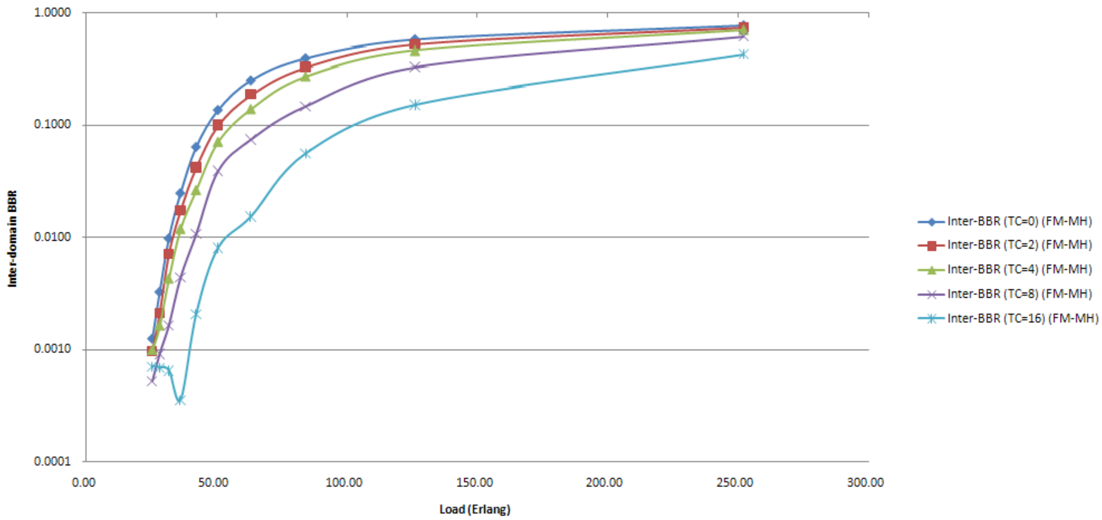

Figure 5.4: 7-domain BBR for full mesh abstraction with minimum hop

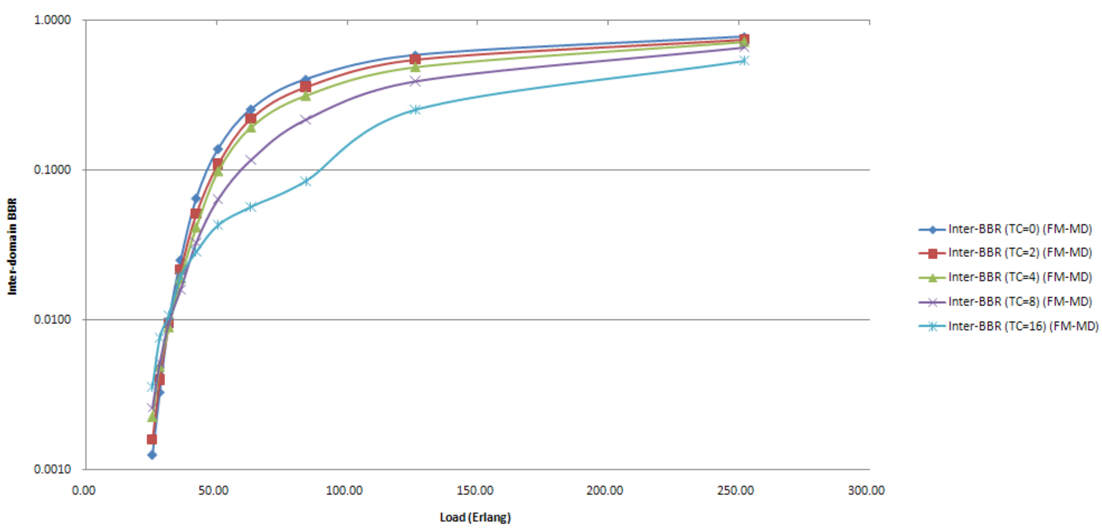

Figure 5.5: 7-domain BBR for full mesh abstraction with minimum distance

is also seen that full mesh abstraction gives significantly higher routing loads than simple node abstraction, ranging anywhere from 50%-100%(as expected).

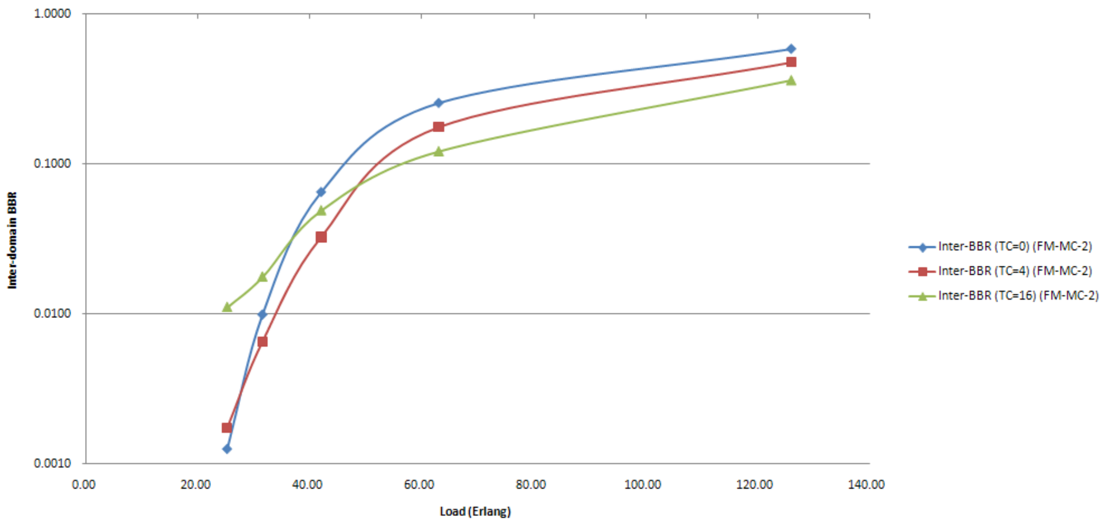

Figure 5.6: 7-domain BBR for full mesh abstraction with minimum cost ($\alpha = 2$)

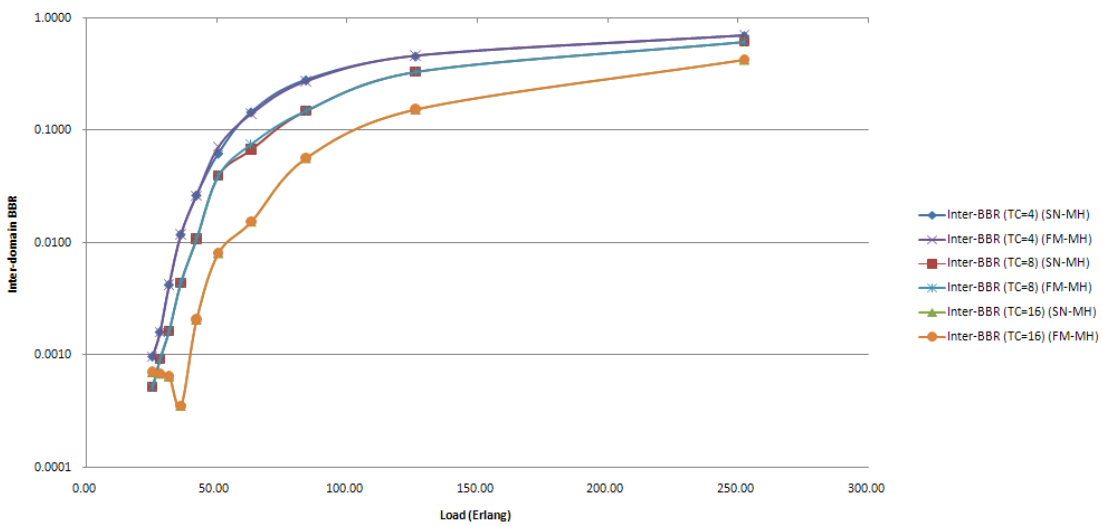

Figure 5.7: 7-domain BBR: simple node vs. full mesh abstraction with minimum hop

Meanwhile, the routing load performance with minimum distance (i.e., load balancing) path computation is notably different, as shown in Figure 5.13 and 5.14. Here, all tunneled link count values yield roughly the same level of inter-domain

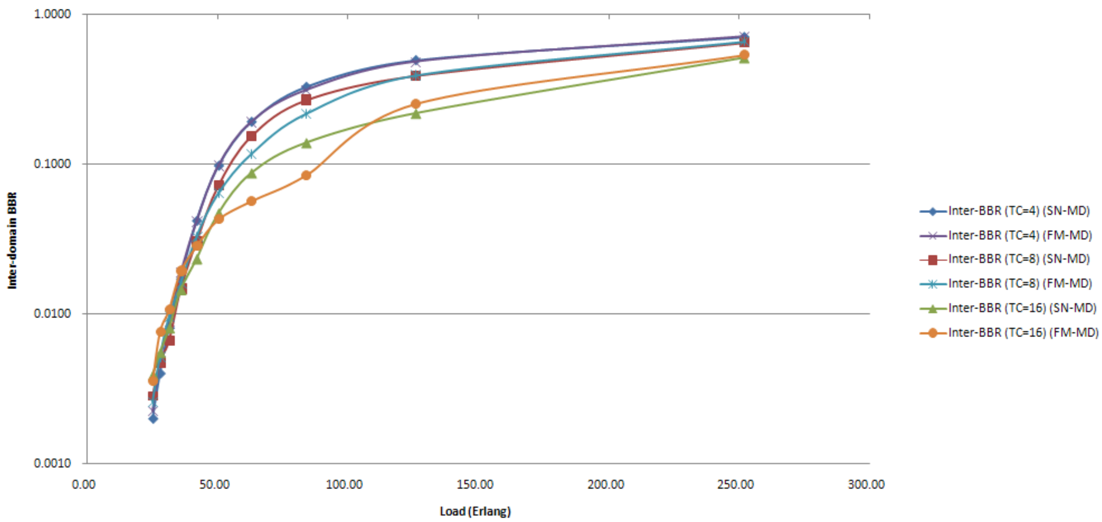

Figure 5.8: 7-domain BBR: simple node vs. full mesh abstraction with minimum distance

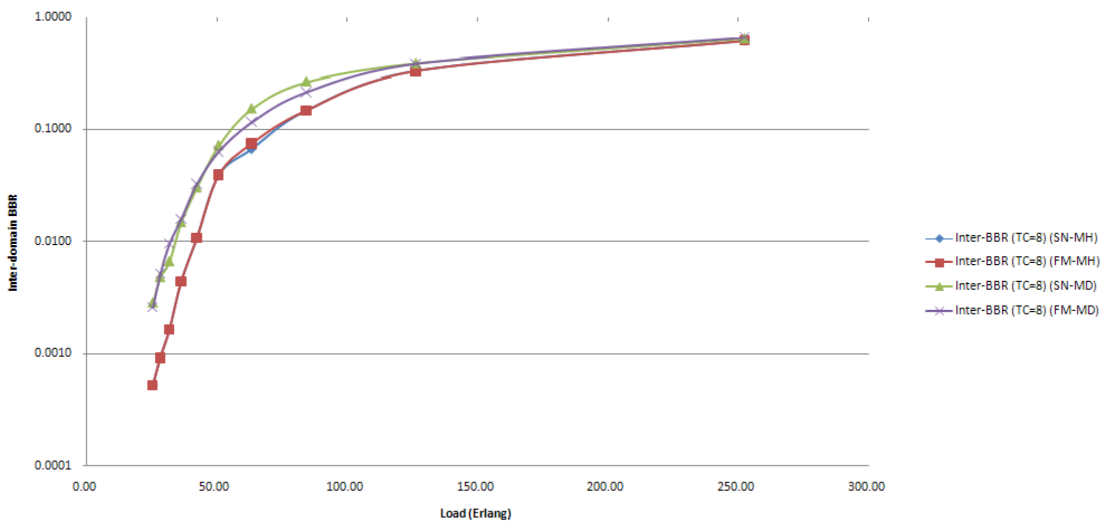

Figure 5.9: 7-domain BBR with 8 tunneled connections

routing load, i.e., clustered withing 1-2 LSA/sec across all load values. The routing load overhead for the minimum hop and minimum distance path computation schemes are also compared in Figure 5.15. Here the plots show that the minimum

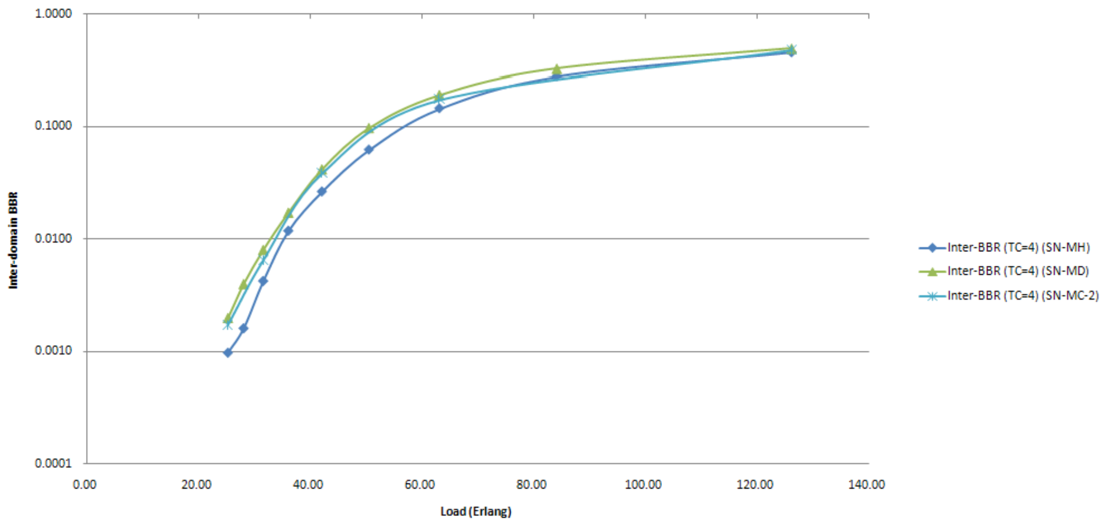

Figure 5.10: 7-domain BBR: simple node with the three suggested routing algorithms

hop scheme gives generally lower routing overheads versus the minimum distance scheme, averaging about 38% lower routing load. Overall, these findings can be explained by the fact that minimum distance, i.e., load balancing, computation tends to perturb many more links in the network, thereby yielding higher routing loads. Finally, the head-to-head performance of minimum hop, minimum distance and minimum cost path computation schemes is compared in Figure 5.16. These results show a considerable decrease in routing load for the minimum cost scheme compared to the minimum hop, averaging about 6% lower routing load.

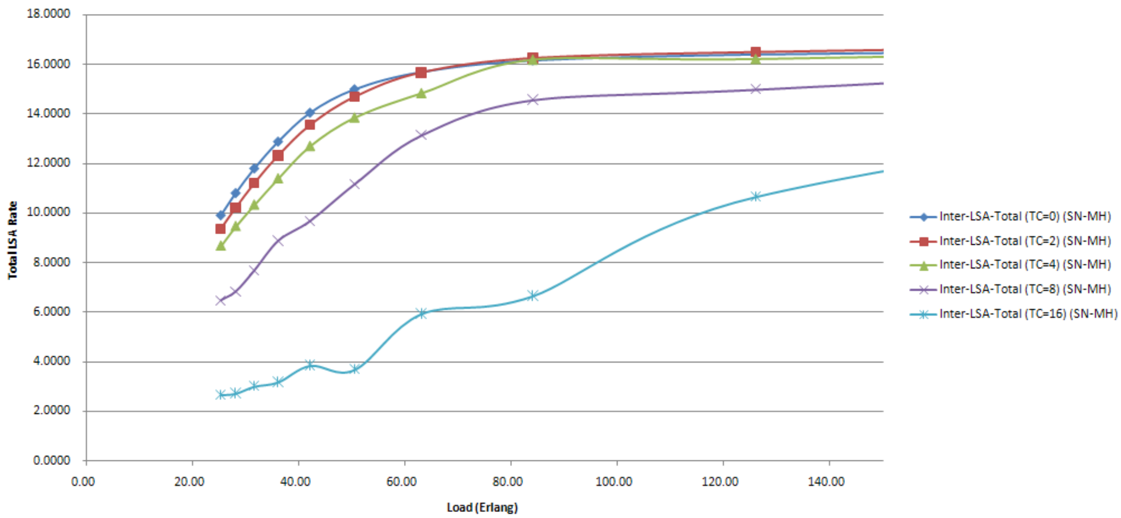

Figure 5.11: 7-domain routing load for simple node abstraction with minimum hop

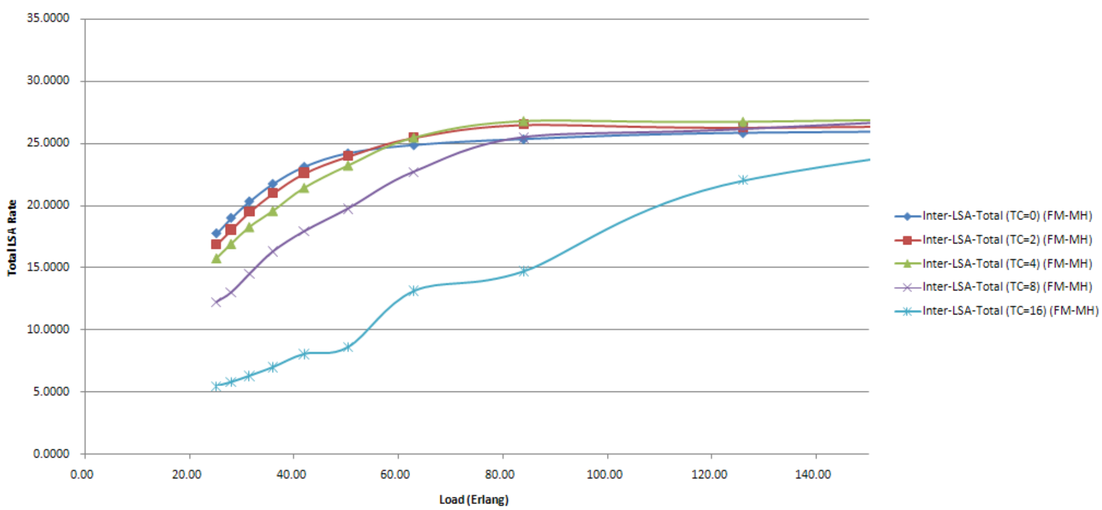

Figure 5.12: 7-domain routing load for full mesh abstraction with minimum hop

5.2 Nineteen-Domain Network Topology

Next, the more elaborate nineteen-domain network is simulated and the results presented in Figures 5.17 - 5.20. Namely, Figures 5.17 and 5.18 present inter-domain

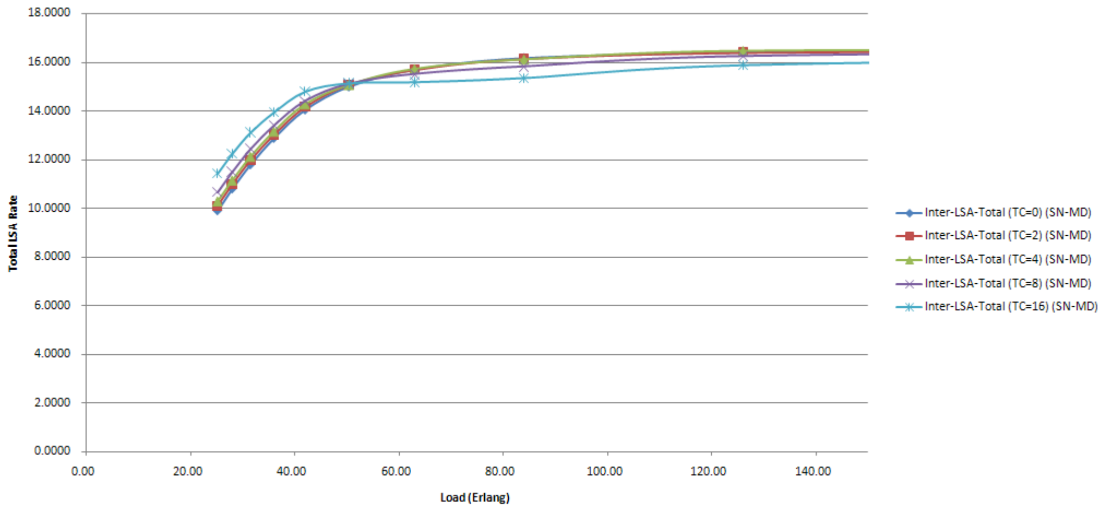

Figure 5.13: 7-domain routing load for simple node abstraction with minimum distance

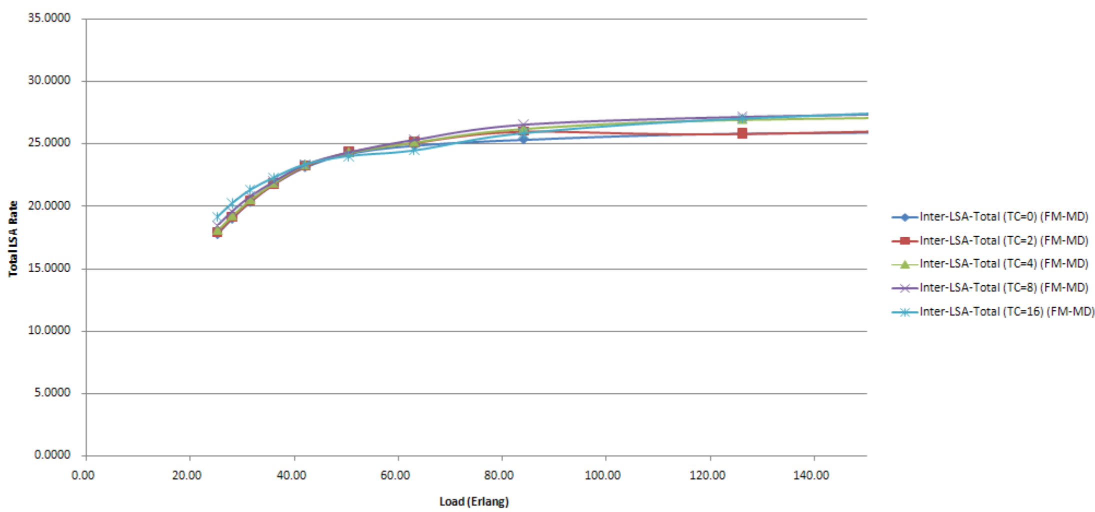

Figure 5.14: 7-domain routing load for full mesh abstraction with minimum distance

blocking probability results for the two path computation strategies, minimum hop and minimum distance (Section 3.4.1, Figure 3.4), for varying number of tunneled links. Again, the goal here is to evaluate the benefits of increased inter-domain

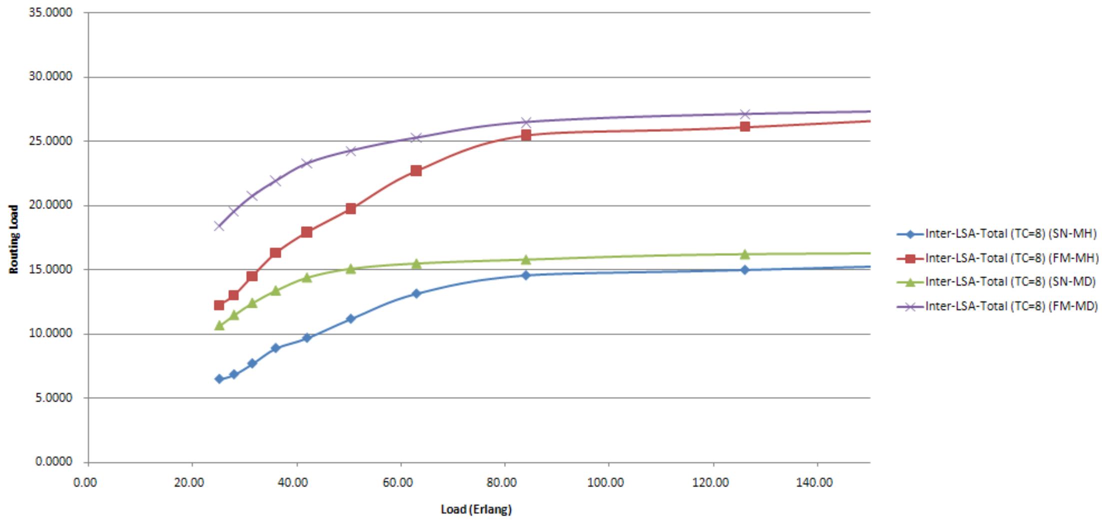

Figure 5.15: 7-domain routing load for 8 tunneled connections

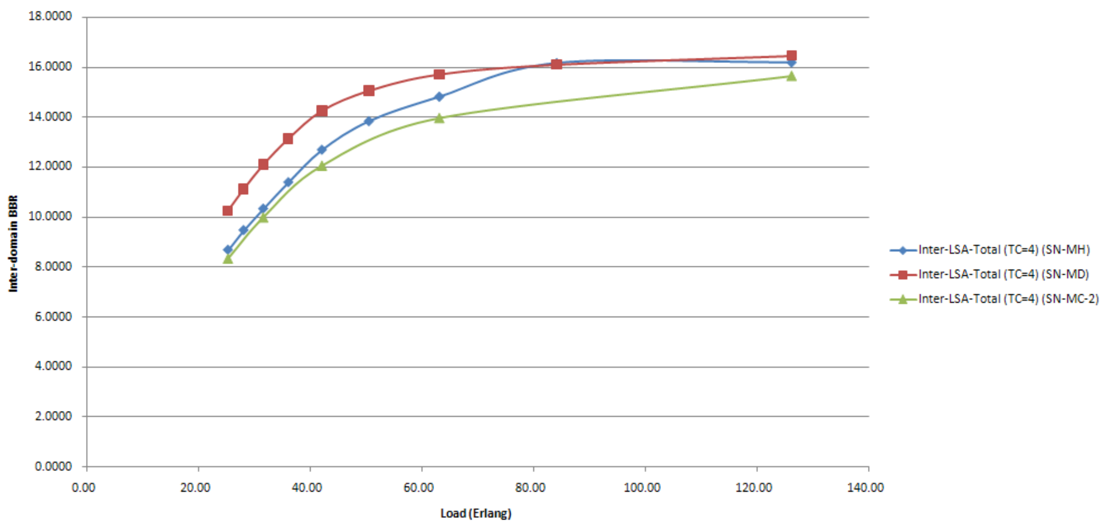

Figure 5.16: 7-domain routing load for simple node and the three suggested routing algorithms

grooming levels. Overall, the findings here are consistent with earlier results for the seven-domain network which show that increasing tunneled link counts yields notably lower blocking probabilities, e.g., the average BBR decrease ranges around

12% averaged across all load-levels. More similar runs are also repeated for the full mesh abstraction scheme in Figures 5.19 and 5.20. However, unlike the seven-domain topology, the results have shown somewhat lower blocking probability reduction with increased tunneled link counts. For example, 16 tunneled links give only a 5% lower blocking probability rate than 4 tunneled links. This indicates that the availability of condensed domain-level state has a beneficial impact on the blocking probability, lowering the need for additional tunneled link state.

Next, by extracting specific information from the earlier plots in Figures 5.17 - 5.20, Figures 5.21 and 5.22 are generated in order to study the relative performances of the different topology abstraction schemes in this nineteen-domain topology. Specifically, Figure 5.21 compares the blocking probability for simple node and full mesh abstractions for varying tunneled link counts and minimum hop path computation. Similarly, Figure 5.22 repeats the above for the minimum distance path computation scheme. These figures reveal some very interesting insights. Consistently with the previous findings, Figure 5.21 shows that the more advanced full mesh topology abstraction scheme gives no noticeable blocking probability reduction when coupled with the minimum hop path computation. Conversely, the results with the minimum distance scheme, Figure 5.22, show decent gains with full mesh abstraction at higher loads, e.g., at 96 Erlang load and 16 tunneled links, full mesh abstraction gives 6% lower blocking probability than simple node. Finally, the head-to-head performance of minimum hop and minimum distance path computation is shown in Figure 5.23. Unlike the seven-domain, the plots show that the minimum hop outperforms minimum distance for load less than 55 Erlangs, whereas the minimum distance performs better than minimum hop for larger loads.

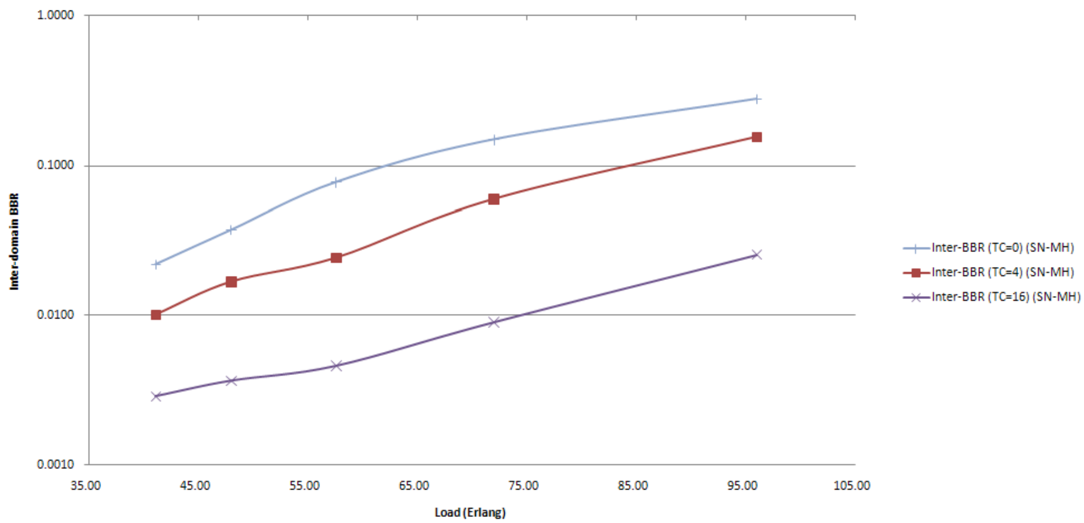

Figure 5.17: 19-domain BBR for simple node abstraction with minimum hop

In addition to blocking probability performance, control plane routing load overheads are also measured for the nineteen-domain network. First, Figures 5.24 and 5.25 plot the inter-domain LSA loads for the minimum hop count scheme and varying numbers of tunneled links (simple node and full mesh abstraction, respectively), whereas Figures 5.26 and 5.27 plot the inter-domain LSA loads for the minimum distance scheme and the varying number of tunneled links. These findings are consistent with the observations made earlier for the seven-domain topology.

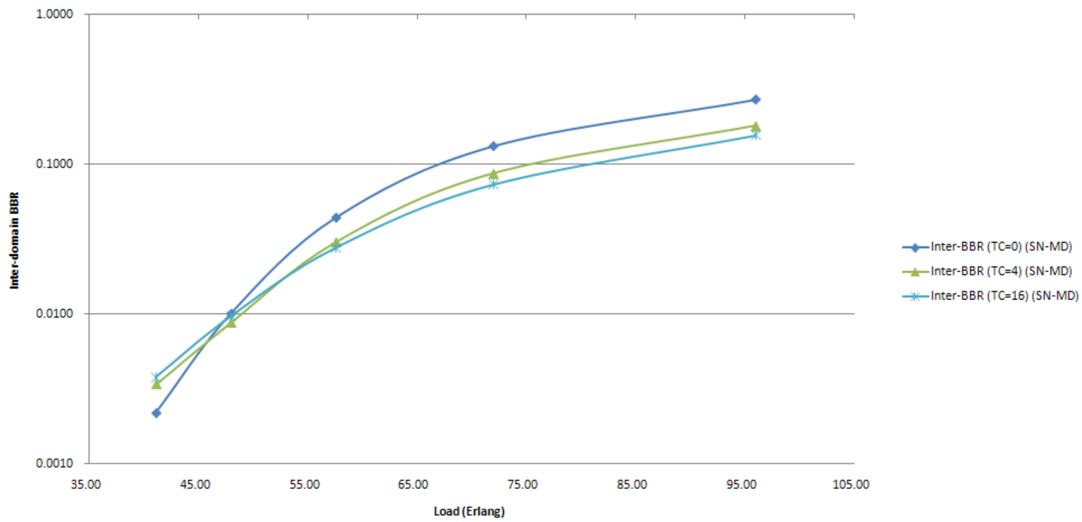

Figure 5.18: 19-domain BBR for simple node abstraction with minimum distance

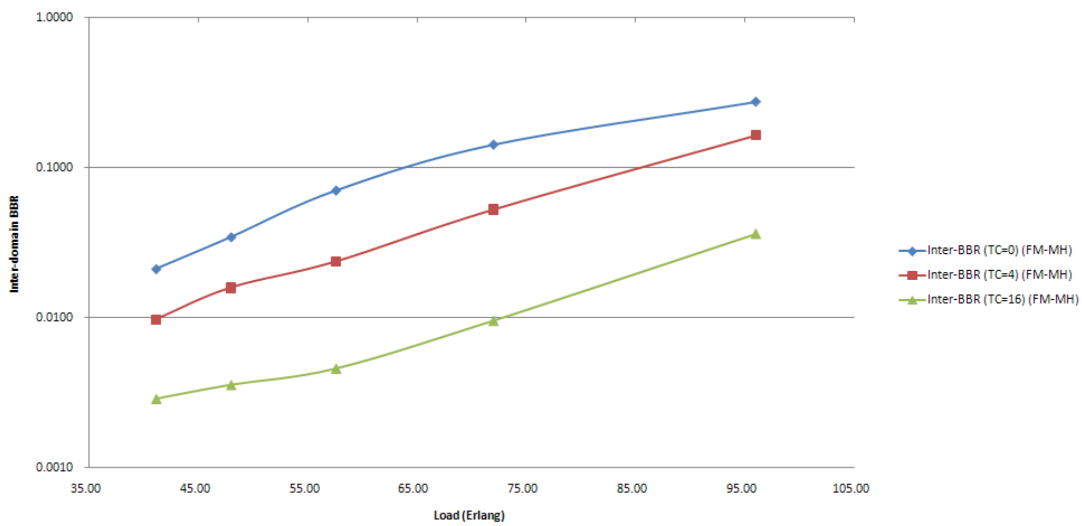

Figure 5.19: 19-domain BBR for full mesh abstraction with minimum hop

Finally, the head-to-head performance of minimum hop and minimum distance path computation schemes is also compared in Figure 5.28. Here the plots show that the minimum hop scheme gives slightly lower routing overhead versus the minimum

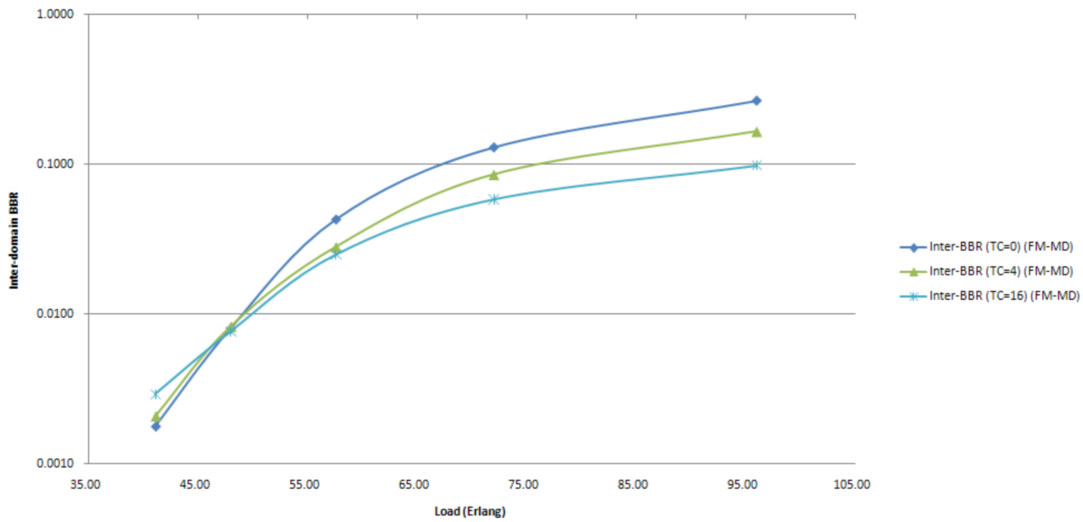

Figure 5.20: 19-domain BBR for full mesh abstraction with minimum distance

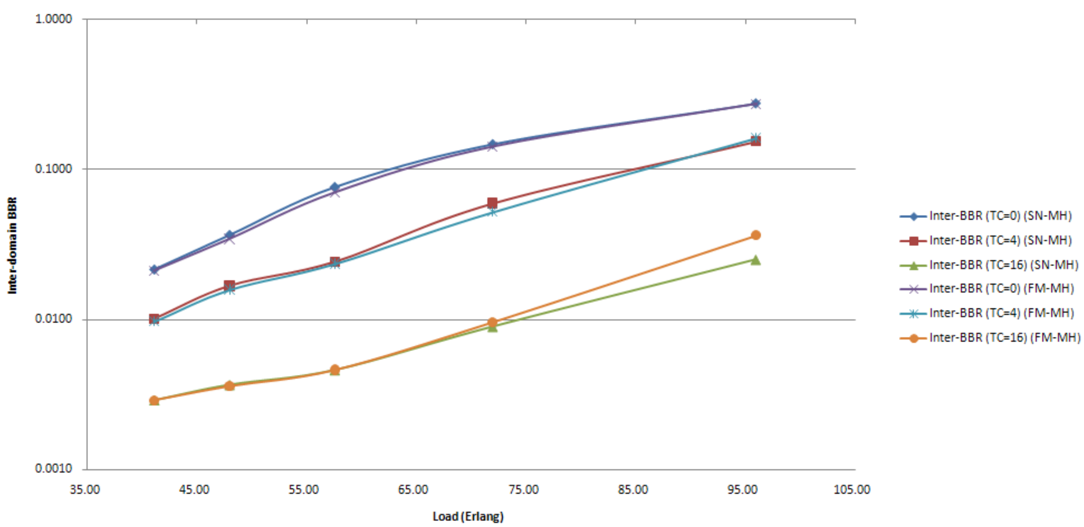

Figure 5.21: 19-domain BBR: simple node vs. full mesh abstraction with minimum hop

distance scheme, averaging about 7% lower routing load. However, full mesh abstraction has much higher overhead values, as expected, owing to the larger domain counts. Overall, these findings can be explained by the fact that minimum distance,

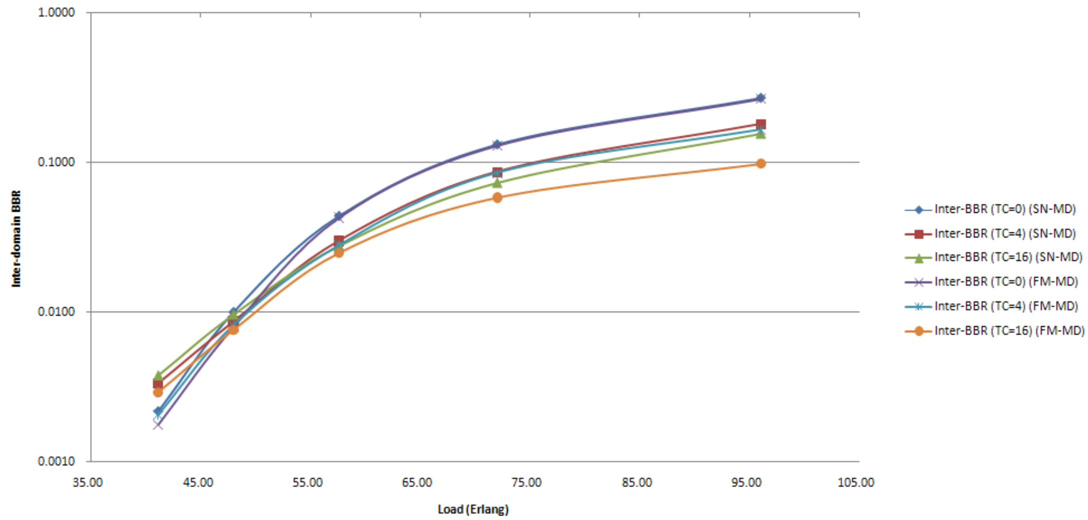

Figure 5.22: 19-domain BBR: simple node vs. full mesh abstraction with minimum distance

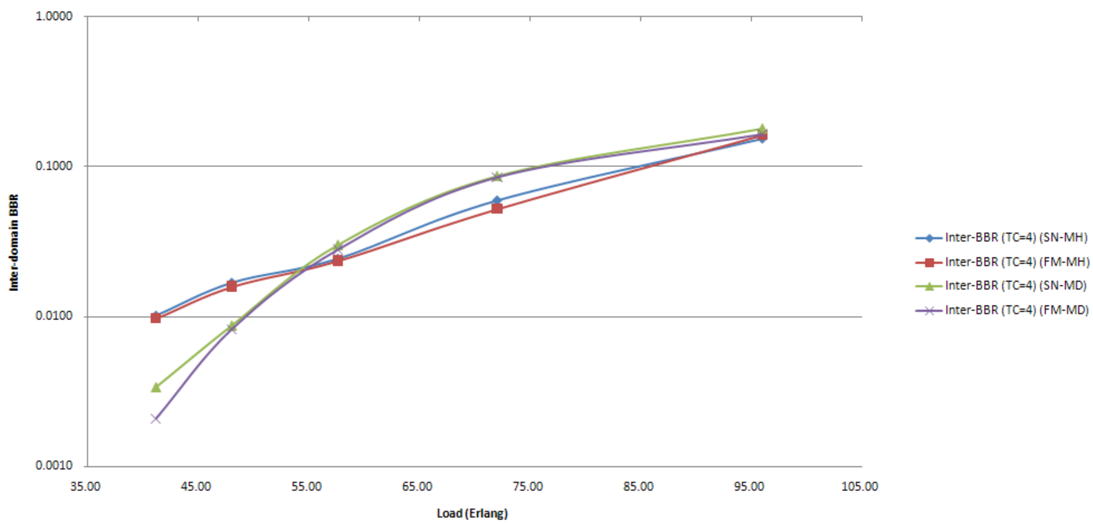

Figure 5.23: 19-domain BBR with 4 tunneled connections

i.e., load balancing, computation tends to perturb many more links in the network, thereby yielding higher routing loads.

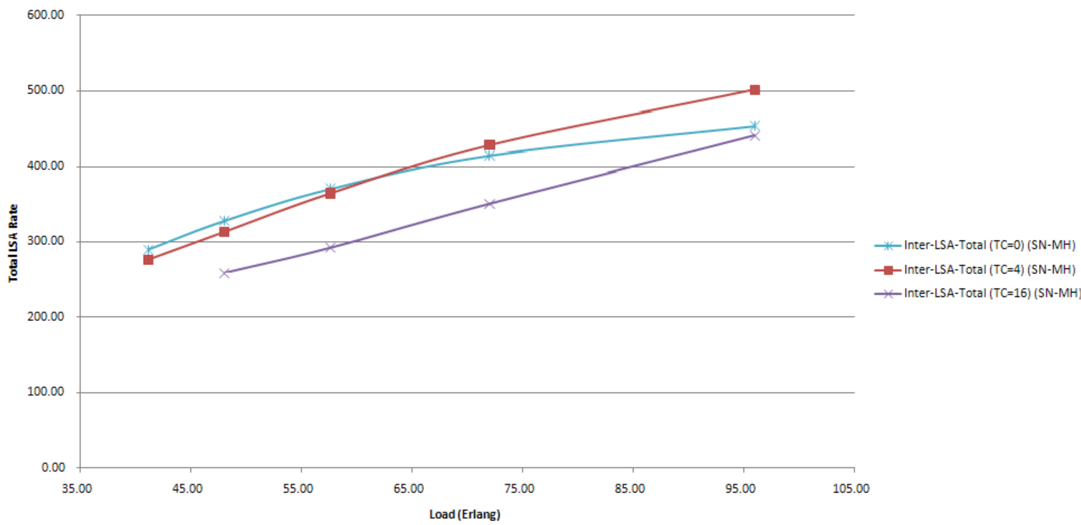

Figure 5.24: 19-domain routing load for simple node abstraction with minimum hop

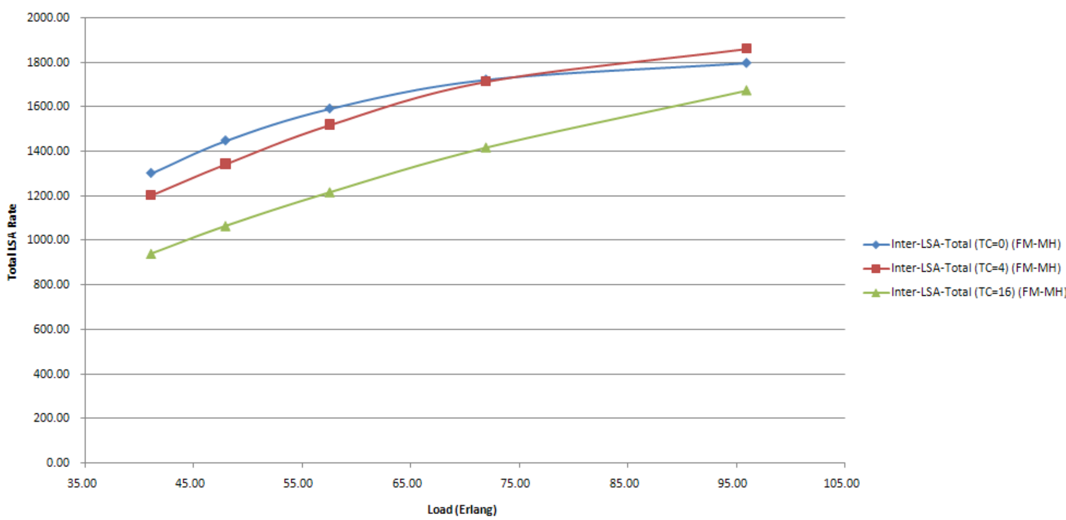

Figure 5.25: 19-domain routing load for full mesh abstraction with minimum hop

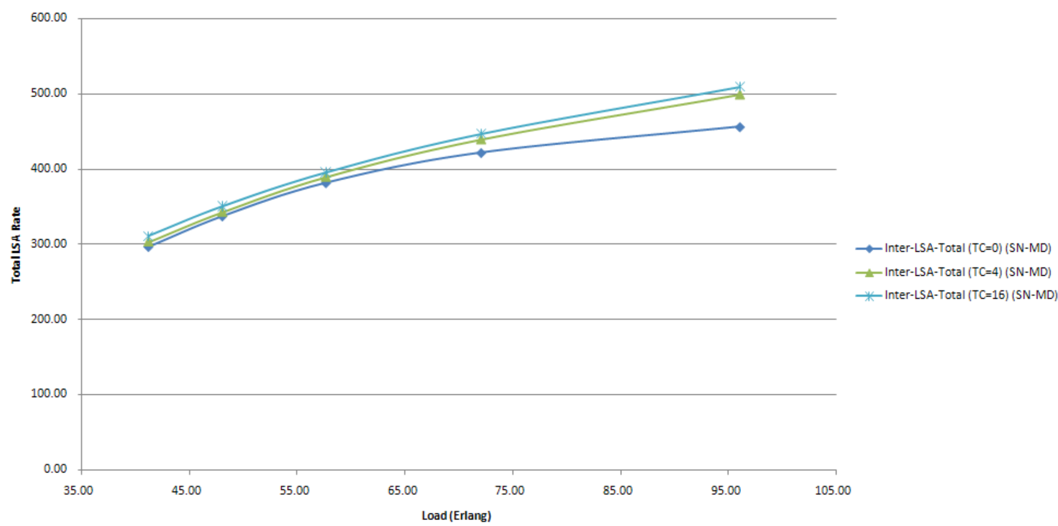

Figure 5.26: 19-domain routing load for simple node abstraction with minimum distance

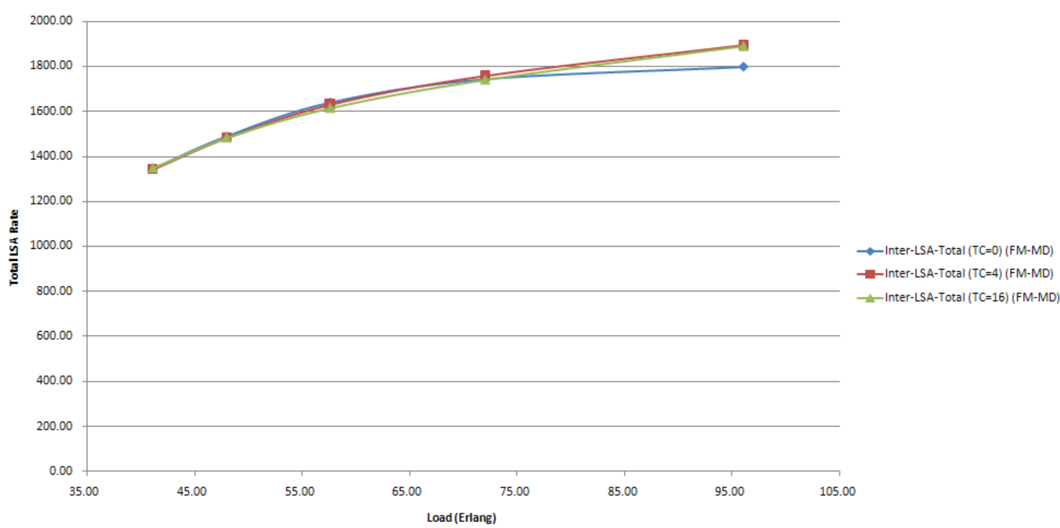

Figure 5.27: 19-domain routing load for full mesh abstraction with minimum distance

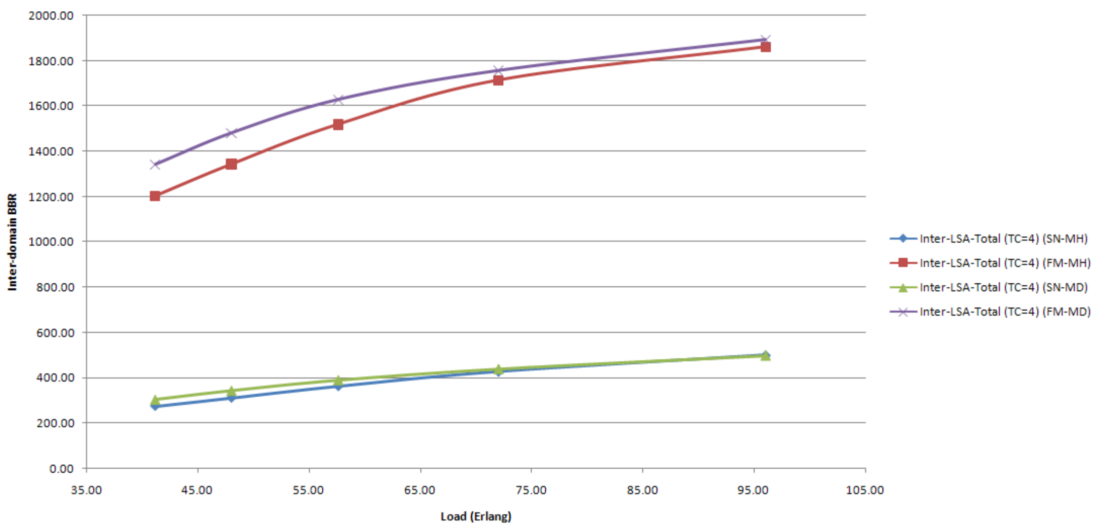

Figure 5.28: 19-domain routing load for 4 tunneled connections

Chapter 6

Conclusion and Future Work

A wide range of networking technologies are being deployed across different long-haul and metro/regional networking domains, e.g., DWDM, SONET, IP and Ethernet. As a result of these growing deployments, there is a burgeoning need for multi-domain/multi-layer control provisioning solutions. In particular, the key challenges here center around multi-domain/multi-layer routing, signaling, and path computation. Along these lines, this thesis presents a comprehensive framework for path provisioning in multi-domain/multi-layer networks. First, a hierarchical inter-domain routing setup is presented using two-differnet topology abstraction schemes, i.e., simple node and full mesh. Additionally, key provisions are added to model tunneled connection/grooming links between multiple layers. Detailed inter-domain routing/triggering policies and signaling procedures are then defined. The performance of these schemes is evaluated using discrete event simulation for different network topologies.

6.1 Conclusion

This research has successfully developed routing and provisioning algorithms for multi-domain/multi-layer networks. Overall, some key results and findings have emerged from this work:

- The propagation of inter-domain tunneled link state has a strong impact on performance, by helping lower inter-domain setup blocking probabilities. For example, increasing the number of inter-domain tunneled propagated links from 2 to 16, causes an average drop in the inter-domain path setup blocking probability of 12% for the seven-domain network, and 10% for the nineteen-domain network (simple node abstraction).
- The corresponding inter-domain routing loads are not greatly impacted by the propagation of increased amount of tunneled link state. For example, increasing the number of inter-domain tunneled propagated links from 2 to 16, causes the average increase in the tunneled link inter-domain routing load of 90%, but still yields an average decrease in total inter-domain routing loads of about 14%. This is due to the fact that with more connection groomed, a small variation

occurs at the physical/abstract available link capacity, leading to lower loading traffic flooded at the inter-domain level.

- The propagation of inter-domain tunneled link state does not yield sizable benefits for full mesh topology abstraction. Likely, the availability of detailed domain-level state results in a reduced need for tunneled link state.

- For the minimum distance scheme, all tunneled link count values yield roughly the same level of inter-domain routing load yielding a clustering for the values across all load values.

- In general, the full mesh abstraction combined with an increase in tunneled links and the minimum distance path computation scheme generates the lowest BBR at the expense of an increase in the inter-domain routing load. This general case did not apply to the seven-domain network because of the special topology with only one "underlying" DWDM grooming domain.

6.2 Future Research Directions

This thesis studies routing, signaling and path computation in multi-domain/ multi-layer networks and presents some key findings. This work provides a strong base from which to conduct further, more indepth studies on such networks. Foremost, the impact of different types of survivability on such networks can be studied.

In addition, advanced link weighting techniques, that reflects multiple QoS, and path computation strategies that leverage these techniques can be investigated. Finally, analytically models/approximations of multi-domain/multi-layer blocking probability can also be developed.

References

[1] N. Ghani, *et al*, “Metropolitan Optical Networks”, *Optical Fiber Telecommunications IV*, Academic Press, March 2002, pp. 329-403.

[2] en.wikipedia.org

[3] G. Bernstein, B. Rajagopalan, D. Saha, “Optical Network Control-Architecture, Protocols and Standards”, Addison Wesley, Boston 2003.

[4] R. Dutta, A. Kamal, G. Rouskas, “Traffic Grooming for Optical Networks: Foundations, Techniques and Frontiers”, Springer, New York 2008

[5] W. Alanqar, A. Jukan, “Extending end-to-end optical service provisioning and restoration in carrier networks: opportunities, issues, and challenges”, *IEEE Communications Magazine*, 2004, pp. 52-60.

[6] R. Alnuweri, et al, “Performance of New Link State Advertisement Mechanisms in Routing Protocols with Traffic Engineering Extensions”, *IEEE Communications Magazine*, Vol. 42, No. 5, May 2004, pp. 151-162.

[7] F. Aslam, et al, "Interdomain Path Computation: Challenges and Solutions for Label Switched Networks", *IEEE Communications Magazine*, Vol. 45, No. 10, October 2007, pp. 94-101.

[8] J. Vasseur, "Inter-Area and Inter-AS MPLS Traffic Engineering", *IETF Draft*, draft-vasseur-ccamp-inter-area-as-te-00.txt, February 2004.

[9] F. Hao, E. Zegura, "On Scalable QoS Routing: Performance Evaluation of Topology Aggregation", *IEEE INFOCOM 2000*, pp. 147-156.

[10] K. Liu, K. Nahrstedt, S. Chen, "Routing with Topology Aggregation in Delay-Bandwidth Sensitive Networks", *IEEE/ACM Trans.s on Networking*, Vol. 12, No. 1, February 2004, pp. 17-29.

[11] A. Sprintson, et al, "Reliable Routing with QoS Guarantees for Multi-Domain IP/MPLS Networks", *IEEE INFOCOM 2007*, Anchorage, AL, May 2007.

[12] Y. Zhu, *et al*, "End-to-End Service Provisioning in Multi-Granularity Multi-Domain Optical Networks", *IEEE ICC 2004*, Paris, France, June 2004.

[13] T. Korkmaz, M. Krunz, "Source-Oriented Topology Aggregation with Multiple Qos Parameters in Hierarchical Networks", *ACM Transactions on Modeling and Computer Simulations*, Vol. 10, No. 4, October 2000, pp. 295-325.

[14] Y. Zhu, et al, "Multi-Segment Wavelength Routing in Large-Scale Optical Networks", *IEEE ICC 2003*, Anchorage, AL, May 2003.

[15] A. Hadjiantonis, *et al*, "Evolution to a Converged Layer 1, 2 in a Global-Scale, Native Ethernet Over WDM-Based Optical Networking Architecture", *IEEE JSAC*, Vol. 25, No. 5, June 2007,pp. 1048-1058.

[16] S. Sanchez-Lopez, et al, "A Hierarchical Routing Approach for GMPLS-Based Control Plane for ASON", *IEEE ICC 2005*, Seoul, Korea, June 2005.

[17] X. Yang, B. Ramamurthy, "Inter-Domain Dynamic Routing in Multi-Layer Optical Transport Networks", *IEEE GLOBECOM 2003*, San Francisco, CA, December 2003.

[18] Q. Liu, et al, "Hierarchical Inter-Domain Routing and Lightpath Provisioning in Optical Networks", *OSA Journal of Optical Networking*, Vol. 5, No. 10, October 2006, pp. 764-774.

[19] C. Assi, et al, "Impact of Wavelength Converters on the Performance of Optical Networks", *SPIE Optical Network Magazine*, Vol. 3, No.2, 2000, pp. 16-23.

[20] Q. Liu, et al, "Distributed Inter-Domain Lightpath Provisioning in the Presence of Wavelength Conversion", *Computer Communication*, December 2007, Vol. 30, pp. 3362-3375.

[21] I. Iliadis, "Optimal PNNI complex node representations for restrictive costs and minimal path computation time", *IEEE/ACM Transactions on Networking*, 2000, 8 (4).

[22] R. Dutta, G. N. Rouskas, "Traffic Grooming in WDM Networks: Past and Future", *IEEE Network*, vol. 16, No. 6, November/December 2002, pp. 46-56.

[23] www.opnet.com