SOCIO-ECONOMIC IMPLICATIONS ON THE DESIGN OF TELECOMMUNICATION NETWORKS

by

Brody James Todd

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In
Engineering Management

Department of Mechanical Engineering
University of Alberta

ABSTRACT

The purpose of this work is to expand the understanding of how to efficiently and effectively plan the buildout and expansion of telecommunication networks. This was done by developing a planning framework that incorporates the full scope of inputs and outputs of the planning process as well as the stakeholders of the buildout process. In order to realize this framework a fulsome multi-period survivable network model was developed to better understand the influences of network augmentation over time. Lastly, a techno-economic planning model was developed that embodied the planning framework by incorporating input beyond just the estimated traffic demand and the corresponding network topology and capacity. The techno-economic model incorporated the user community's ability to use the network as well as the potential for network stakeholders to influence this ability.

PREFACE

This thesis is an original work by Brody Todd. Some research conducted for this thesis has been published, or submitted for publication. The papers were published with myself as the primary author, and Professor John Doucette as the second author. The work was primarily done by myself, with reviews and edits done by Dr. Doucette.

Chapter 3 was published as B. Todd and J. Doucette, "Survivable Network Capacity Allocation and Topology Design Using Multi-period Network Augmentation," *Journal of Network and Systems Management*, vol. 25, no. 3, pp. 481–507, July 2017. The work was primarily done by myself, with reviews and edits done by Dr. Doucette.

Chapter 4 was published as B. Todd and J. Doucette, "A Novel Long Term Telecommunication Network Planning Framework," *Journal of Network and Systems Management*, vol. 25, no. 1, pp. 47–82, Jan. 2017. The work was primarily done by myself, with reviews and edits done by Dr. Doucette.

Chapters 5 is slated to be submitted for publication.

TABLE OF CONTENTS

Table of Tables

TABLE OF FIGURES

List of Abbreviations

APS – Automatic protection switching
Cap-ex – Capital Expenditure
EHR – Electronic Health Records
GMPLS – Generalized Multi-Protocol Label Switching
ICT – Information and Communication Technologies
ILP – Integer Linear Programming
IP – Internet Protocol
DSP – Demand-wise Shared Protection
JCA – Joint Capacity Allocation
MPLS – Multi Protocol Label Switching
MPNA – Multi-Period Network Augmentation
NPV – Net Present Value
Op-ex – Operational Expenditure
OTN – Optical Transport Network
QoS – Quality of Service
ROI – Return on Investment
SBPP – Shared Backup Path Protection
SCA – Spare Capacity Allocation
SDH – Synchronous Digital Hierarchy
SDN – Software Defined Networking
SIPOC – Stakeholder Input Process Output Consumer
SONET – Synchronous Optical Networking
VOIP – Voice over internet protocol

Chapter 1. Introduction

Advanced communication systems and networks are a vital and nearly utility-level component of modern society's infrastructure [1]. These systems are pervasive in their impact on how the world operates. From education[2] to healthcare [3] to business [4] to basic social interactions [5], reliable, secure and capable communication systems provide a foundation from which new and better services can be provided. The impact of improving this underlying infrastructure can be seen across political, economic and cultural boundaries. Developments in the aftermath of the recent Iran election demonstrate that the Internet and mobile data services can have a significant effect on how political events unfold [6]. Telecommuting is revamping offices, and how and where work is carried out. E-Health systems are increasing in their efficiency, as well as bringing a higher level of care to remote or underserved regions [3]. There are also major changes in how we consume media and interact with one another due to advances in communication technologies. All of these are predicated to some degree by having an effective and reliable communication system [7].

Communication systems can essentially be broken down into two categories, backbone infrastructure, and access infrastructure. Access infrastructure includes technologies such as ADSL, cable, cellular, Wi-Fi, and other technologies that end users utilize to connect to the greater system. Backbone infrastructure collects and aggregates all of the sources of traffic and delivers it to the access infrastructure near the traffic destination. These backbone systems are high capacity links that are generally capital intensive to deploy, and since they carry a large amount and variety of aggregated traffic sources, they are extremely sensitive to disruptions, expensive and require significant long-term planning.

Although the majority of the systems and applications rely on a network capable of meeting relevant demand requirements, what is considered to be capable and reliable is not fixed, but is relative to the application utilizing the network. This research articulates the various factors in telecommunication network planning, the outputs of the planning process, and a mathematical model that incorporates the entire expanse of the planning ecosystem. The challenges addressed include survivable network planning over multiple time periods, the coherent articulation of the network planning ecosystem (as existing literature contains very little incorporating techno-economic network planning), and the mathematical models required to support decision making around the trade-offs in the broader network planning ecosystem.

An extensive background section including a broad literature search is presented next. Many of the topics touched on in the research presented here have been topics of study for an extensive period of time. While there was little on the fundamentals of how techno-economic considerations influence and are impacted by backbone infrastructure networks, there was significant material on either side [8]–[10]. The goal of this research was to create a framework that was capable of incorporating socio-economic considerations into the technical network design and optimization process in order to viably build out robust telecommunication infrastructure in areas that have proven challenging such as rural and remote areas, or in poorer regions of the globe [11], [12]. This framework is laid out progressively through three papers.

The first paper (Chapter 3) developed a novel optimization technique (multi-period network augmentation) that was capable of optimizing the topology and capacity of a network over time. What differentiates telecommunication planning from other resource planning techniques is the need for pre-planned routing over spare capacity allocated throughout the network in such a manor that the network is resilient to likely failure scenarios (and with thousands of kilometers of fiber in the ground, even failure rates of 1 per 1000 kilometers a year would mean incidents with significant social and economic impacts would happen multiple times per year). This challenge of the allocation of spare capacity throughout the network is compounded when looking at multiple time horizons as the structure and strategy of the most efficient topology and traffic routing can change significantly over time. This multi-period survivable network design scheme enables planning to go beyond a singular time horizon, and start to look at how to grow and expand the network.

Following the articulation of multi-period network optimization scheme, a comprehensive network planning framework is presented that uses a systems approach to articulate the key influencers and inputs, outputs and decisions, and primary groups of stakeholders surrounding the network planning process. This network planning framework is provided with significant literature supporting the structure presented. This cornerstone framework articulates the network planning ecosystem that must be accounted for in order to efficiently and effectively build out and expand network infrastructure.

Pulling the network planning framework into a survivable network design, Chapter 5 demonstrates the trade-offs between some of the key decision points. This chapter brings together long range, multi-period planning into the broader context of building and deploying a network with users at varying levels of ability to utilize the network and different abilities to pay for or fund the network. In this paper, the complex nature of the trade-offs in the network planning framework are explored and articulated through a genetic algorithm incorporating socio-economic factors such as the speed of innovation adoption [13] and the cost of marketing or enabling users (for example telemedicine) [14] with multi-period survivable network design.

Chapter 2. Background

Literature focusing on telecommunication network planning has followed a couple of paths, the business side of planning including pricing, estimation of demand and competition, and the technical design of the network to meet the estimated demand. On the technical design, there are also a variety of contexts for technical design, such as fixed capacity traffic routing [15], survivable network design [7], and others.

The first steps, and the most permanent, of network design is the topology, capacitation, and routing of the network. This level of planning is more complex than a simple multi-commodity optimization because of the availability requirements of the network and supporting restorability in light of failures [16]. There have been years of research into the design of different strategies for spare capacity allocation and the related optimization schemes [8]. Recent advances in resilient network planning include software defined networking (SDN) [17], multi-layer survivability schemes [18], and class based survivability [19].

Effective survivable network planning balances a number of factors, with the various schemes trading off among them. These factors include efficiency (amount of spare capacity required), complexity (detecting and reconfiguring the network when failures occur), impact on availability (multi-failure survivability), speed of restoration and the ability to deal with changing traffic patterns [20], [21]. A comparison of the various network survivability schemes can be found in [7]. With these trade-offs, survivable network design provides the basis for designing the topology (potentially) and link capacity of a network.

The challenge that many operators now struggle with is the next level of growth and connectivity. This involves many of the regions of the globe and many industries that prove to be challenging for a number of reasons including lack of resources, skills, training, and even the aversion to risk. The network planning paradigm developed and laid out in the chapters ahead provide a framework to combine the socio-economic factors that challenge the next phase of telecommunication network buildouts, and the core network design models. By solving this challenge, networks can be deployed more efficiently and effectively and bring high speed networks to rural, remote, and underserved regions. This requires more than good network design, but network design that can account for environment in which the network will operate and adapt to the social and economic challenges that exist. In many of these networks, whether brownfield or greenfield involve networks without established usage trends. As such the primary tools used to model the economic and user side of the problem were innovation diffusion concepts [13], [22].

This background provides an overview of telecommunication network planning, demand forecasting as it relates to telecommunications, and a summary of the study of the diffusion of innovations.

2.1 Network Planning

Network planning has been a topic of research for many years [23]. The planning process has generally been viewed as the translation of demands and technical capabilities into a proposed optimal network configuration. This planning process operates on three time scales, short, mid and long term [9]. Each of these time scales require different approaches to network planning and management.

The problems being addressed in this work focus on long term decisions, with some application to mid-term decisions. These decisions revolve around new capacity placement and physical network augmentation, as opposed to reconfiguration of existing networks. This network reconfiguration is in the domain of short and mid-term planning, as it can be done more rapidly, and usually does not require provisioning additional infrastructure.

Figure 1 – The traditional network planning and design model.

The most common network planning paradigm today involves input of the current network conditions and forecasted demand matrices. These current conditions and forecasted demands are then used to come up with an optimal or near optimal network capacity design (Figure 1). This process can utilize a number of different design methodologies including ring based schemes, and a broad range of mesh based schemes (include hybrid schemes like p-cycles). The design parameters typically include topology layout, capacity placement, and path allocation.

Topology allocation involves the design and layout of the spans in a network. This can either be done in an augmentative fashion, or greenfield. Augmentative topology design involves the possibility of adding spans to an already established network. These potential spans could include brand new installs, or they could involve leasing capacity from other networks. Greenfield planning starts with a set of network nodes, demands and potential spans. The optimization scheme instantiates spans in accordance to the protection scheme's requirements. Because instantiating spans is a lengthy and costly process (the Alberta Supernet took over 4 years to build, and cost over $400 million [24]), topology augmentation is most applicable to long range planning scenarios.

In literature, the most common network planning optimization involves capacity placement (how much capacity should be assigned to each span) and path allocation (the route and associated capacity for working and spare paths), or in some cases where capacity is fixed, optimization involves only path allocation.

Capacity allocation comes in two general styles, joint capacity allocation and spare capacity allocation. JCA involves the optimization of both the working and spare capacity (and associated paths), while SCA optimizes the placement of spare capacity given a set of working paths (and associated capacities). The working paths are usually allocated on a shortest path basis.

Network designs, whether involving topology layout, capacity placement, or path allocation are derived optimally (or within a known degree of optimality) using integer linear programming (ILP) techniques, or near optimal using various heuristics.

ILP techniques involve converting the survivability scheme to a set of mathematical constraints and an objective. This conversion creates an ILP model which can then be solved and optimized using a number of techniques (Traditional solution methods, column generation, decomposition, etc.). While not all designs based on ILP solutions are optimal, they generally give an indication of how far away from optimal the solution is (MIP gap). It should be noted that this distance from optimality is constrained by the input to the ILP problem. For example, if the set of paths an ILP has to select from is limited by either the design of the ILP, or the input to the model, then the mipgap cannot take into account the possibility that, in a global sense, the optimal solution could require paths that the ILP is not aware of. The reason these limited path sets are used is driven by the solution time and complexity of some ILP models.

Integer Linear programing solves an NP-hard problem, and hence is inherently not scalable. By limiting the problem being solved using ILP techniques, the solution time can be brought into reasonable timeframes and computing resources. This limitation, however, does impact the confidence in the optimality of the solution. There are other methods of dealing with the complexity of network design problems, generally referred to as heuristics.

Heuristic solution techniques fall into two general categories, problem specific heuristics, and general heuristics. General solution heuristics include techniques such as genetic algorithms and taboo search. These global heuristics use a generalized approach to find a solution given a certain stopping criteria (number of iterations, improvement per iteration, etc.). Problem specific heuristics use characteristics of the survivability scheme to focus the solution search. These heuristics aim to find a solution in a timeframe that ILP techniques cannot achieve and/or using fewer resources than are required for an ILP solution. These solutions do not give an indication of how far from optimal they are, but by simulating and comparing heuristics with ILP results, a degree of confidence can be given.

In general, network planning has been viewed as a process that takes a demand forecast and the current design of the network to produce an updated network design that can accommodate the forecasted demand. Most scholarly articles and books on network planning treat network planning from a highly technical perspective, not taking into account the high degree of variability in the inputs to the planning process. J. Simmons book Optical Network Design and Planning [25] goes into technical aspects of network design with significant depth, however, does not cover network design or planning in uncertain environments. The assumption is that the traffic demands are known, and the problem is how to best allocate equipment and capacity to meet these demands. Some other planning models include some feedback in the network design process [26], while others take planning system as a more or less linear process [27].

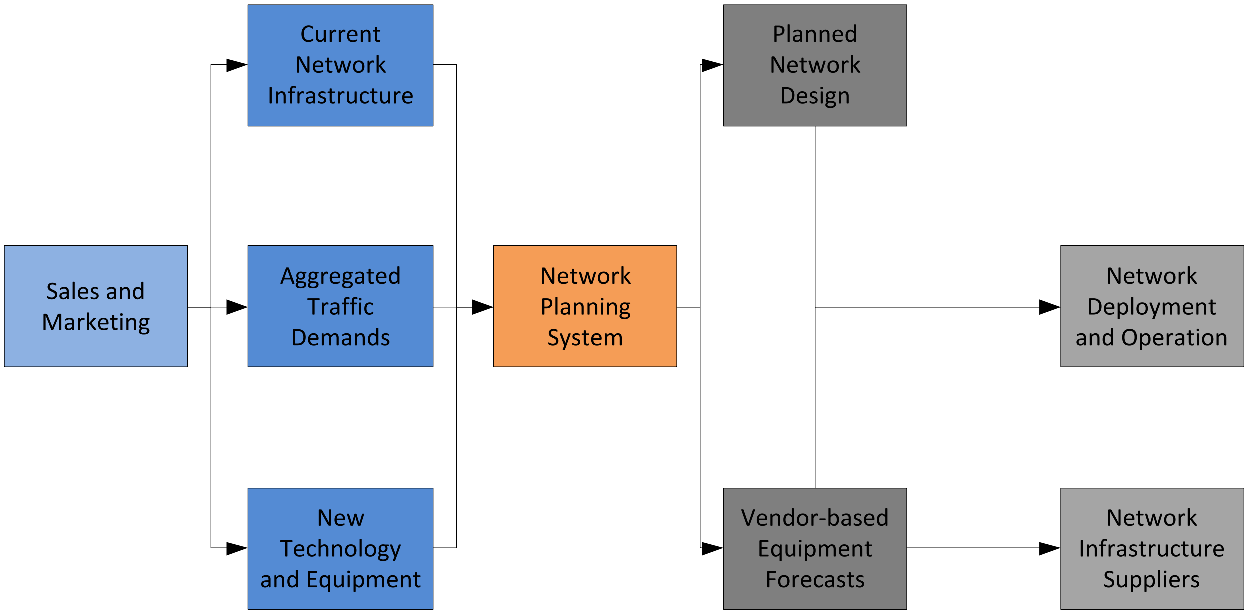


Figure 2 – A planning process presented in literature[27]

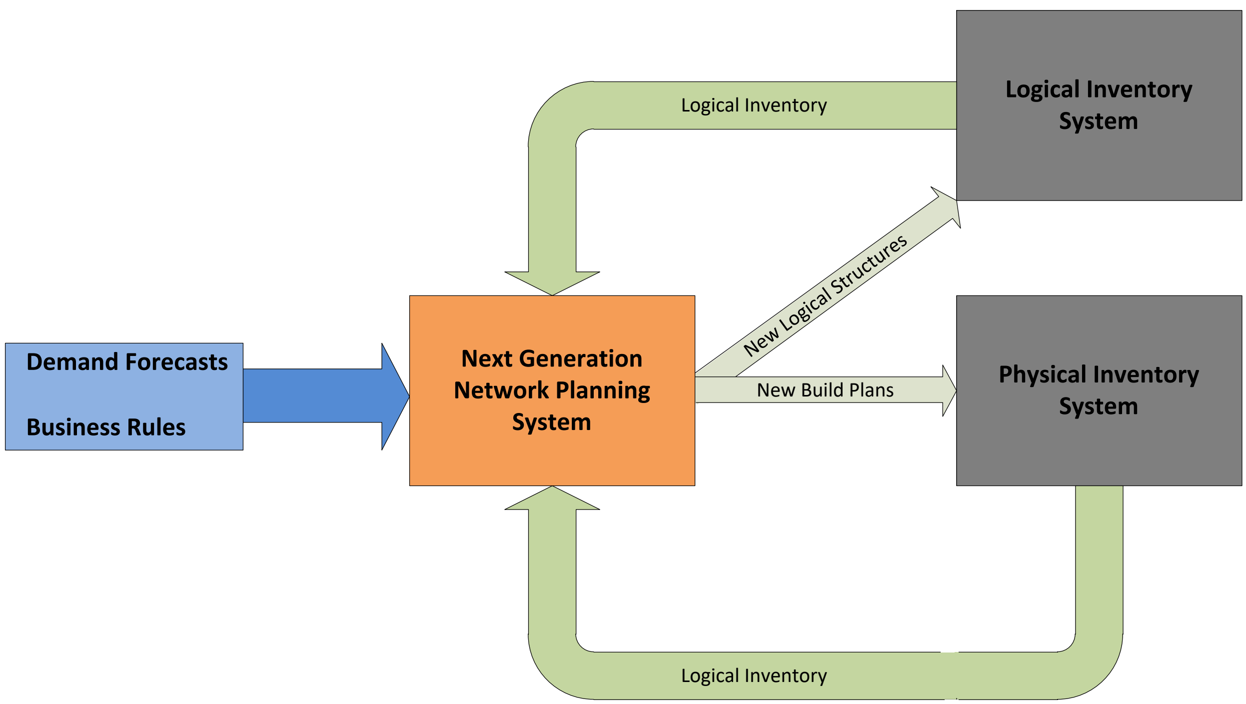


Figure 3 – A planning process in literature that includes a feedback mechanism [26]

It should be noted here that a significant amount of literature talking about network planning uses the term 'Next Generation Network' [26]–[28]. What this term signifies is a shift in underlying network technology to serve demand from higher levels in a protocol agnostic manner [29]. For example, TELUS overhauled their core network a number of years ago to run entirely using IP, rather than a conglomerate of separate networks [30], [31]. This trend toward converged networks has a number of impacts on network planning, as the importance of the integrity and availability of these networks is even more significant.

Network planning has generally not concerned itself with the types of users that will utilize the network. The assumption is that the concerns of the users can be adequately captured in QoS and bandwidth characteristics. Some work has been done to better understand and classify the types of users of a network. In the book "Optical Fiber Telecommunications" [32] users were broken down into three groups, domestic, business, and scientific (Figure 4). The usage scenarios for each of these groups were briefly discussed in the context of their usage characteristics. Although these users are identified, the need for "overarching roadmaps for future networks" was identified, as these were not currently in place.

One last point to mention with regard to network planning is with a couple of relatively new trends in literature. There has been some work done in the field of multi-period network design [20], [21], [33]–[39], and planning with uncertainty [33][40][41][42]. These attempt to expand the network planning paradigm; however, they stay within the technical domain. [33] and [40] combine multi-period planning with uncertainty for a single link, optimizing the allocation of capacity over time. This approach, while insightful, does not scale to the network planning problem. [41] provides a good overview of the application of demand uncertainty to the network planning problem, and presents a two-part stochastic programming model for span restorable networks.

To summarize, network planning has traditionally abstracted out uncertainty into the traffic demands, with optimization of the designs done almost solely to minimize the cost of meeting these demands. Some work has been done to incorporate multiple timeframes and uncertainty into the planning process, and to broaden the planning model to include legacy infrastructure and new equipment.

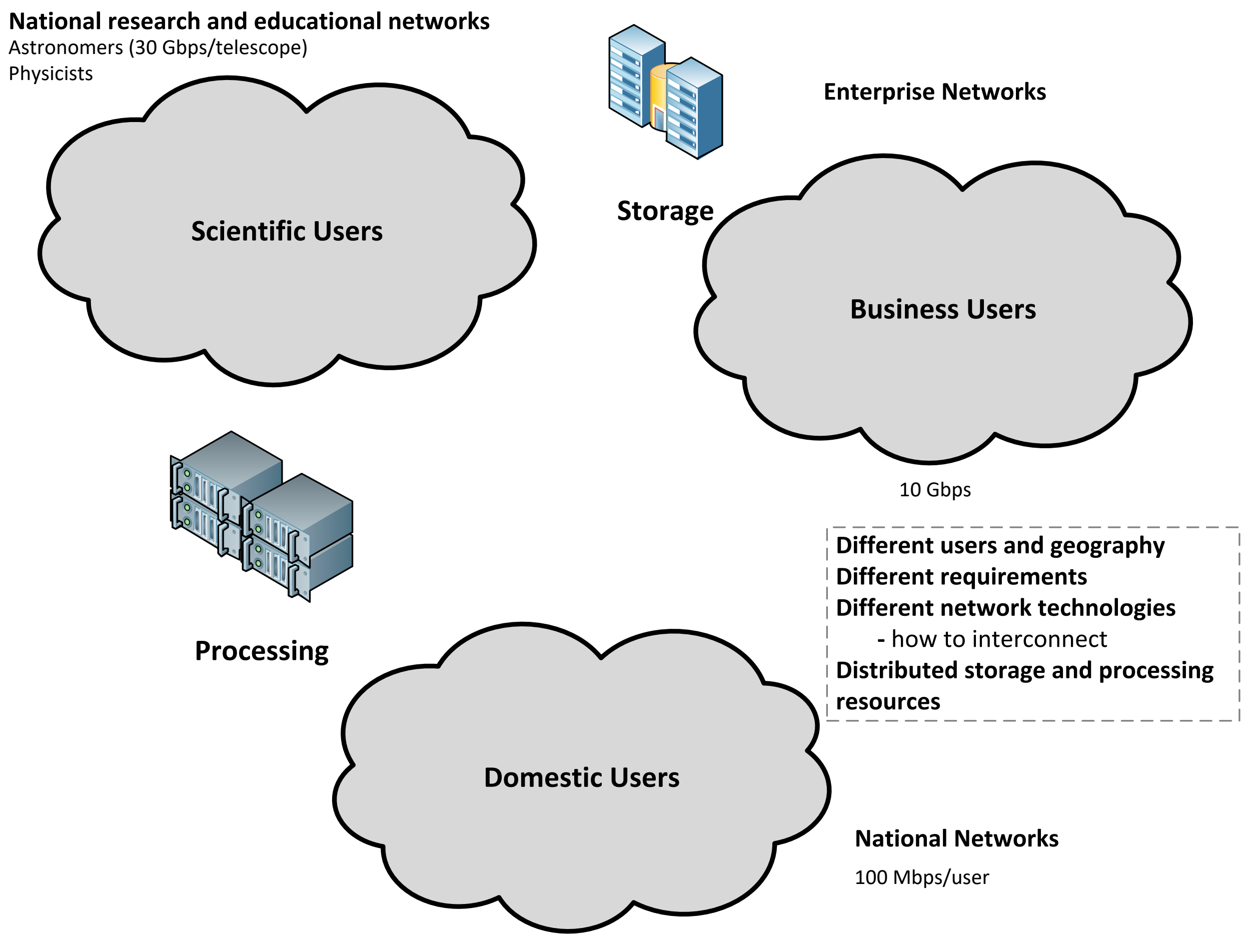


Figure 4 – An overview of types of network users as described in [32].

2.2 Traffic Demand Forecasting

While the purpose of this work is not to develop traffic demand forecasts, but rather to better align network designs with the factors that affect demands, and hence demand forecasts. Included here is a review of demand forecasting and the primary techniques used in different types of network planning.

Traffic demand forecasting is a difficult task. Even in environments with significant traffic history, shifts in application usage can happen quickly. In literature traffic demand forecasting has been discussed for a long time [43], and can be seen as a subset of the larger field of demand forecasting in general. Although forecasting telecommunication demand has been in literature for a while, it has not been extremely well documented [43]. Early papers on demand forecasting focused on predicting demand in an established market with little competition. While some of these premises were valid in the 1980's, many do no hold today. With the rise of data communications, the privatization of the industry, and many other factors, demand forecasting for telecommunications has become significantly more complex. In general, forecasting telecommunication demands has been broken into two general categories. The first is an extension early forecasting techniques, focusing on econometric techniques [44] such as linear regression analysis. The other category of forecasts emphasizes techniques that are appropriate for scenarios with significantly more variability and/or lack of historical data. The former is applicable when forecasting demand in established markets with established services. The latter is used when expanding into new regions or expanding into new service offerings. Since technology changes (both at a network, and at a service level) can significantly alter demand characteristics, it has been suggested that long term forecasting most often uses new product/service forecasting techniques, even in established markets [44].

There are a number of econometric forecasting models that have been used in network demand forecasts [43], that attempt to model factors affecting future demand. Aggregate models utilize time series data to interpolate future demand. These models incorporate price elasticity, and competition effects along with the time series data to estimate overall network traffic, and potential profits from the network. Price elasticity is a measure of how sensitive customers are to price changes for the services supplied by the network operator. Price elasticities are highly dependent on the region being served [43], and provide a way to balance pricing and demand. The effects of competition in demand forecasts using an aggregate approach use concepts such as customer churn to estimate future demand. In general aggregate models may be used in forecasting demand in established markets, however, to not adapt well to the changing nature of communication technologies in our society [43].

Another approach to demand forecasting in established markets is to model services separately using surveys, application specific historical data, and a customer choice model. Customer choice models look at three factors affecting consumer adoption decisions of a specific service. These factors are the understanding, utility, and acceptability of the product [45]. In turn network operators can influence these factors through marketing, product quality and price [45]. Forecasting models based on disaggregate service forecasts requires extensive historical and market data, and functions well in established markets (similar to aggregate approaches). However, when forecasting for the long term or in new markets, these models are generally inadequate [43].

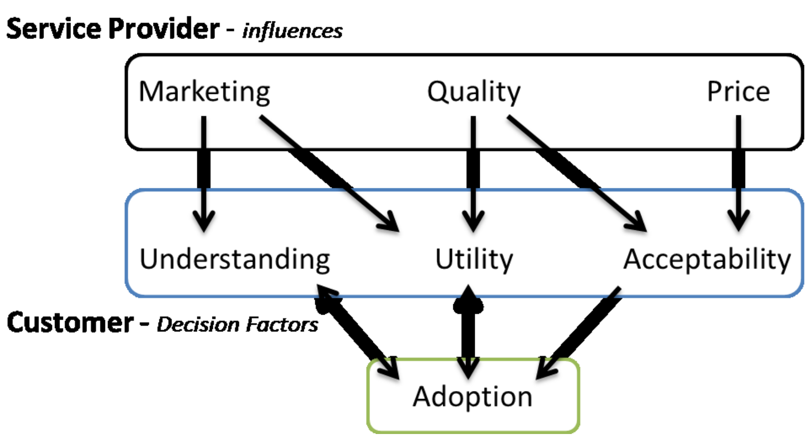


Figure 5 – Overview of the customer choice model adapted from [45]

Demand forecasting in an expansion scenario has the advantage of having historical data to base the forecasts upon. Given this trend data, forecasts can be done using time series forecasting techniques.

Time series forecasting, and other similar techniques look to extrapolate current trends into the future. These forecasts are reasonably accurate for short and mid-term forecasting in existing markets, as the effect of larger technological and social factors tend to be incorporated into recent historical data [9]. Disruptive technologies and social trends take time to propagate through society, and so the impact they have on short and mid-term forecasts is minimal.

Long term forecasting and greenfield forecasting require different techniques. These techniques must create forecasts in a highly uncertain environment, and account for underlying influencers that may be unknown, unpredictable, and/or difficult to measure. Common techniques use expert opinion, innovation diffusion models, and data from similar situations (either from jurisdictions that have already implemented similar technologies, or from previous technologies that were implemented in the same region that are deemed to have similar characteristics.

Greenfield forecasting uses a number of techniques to estimate demand, including innovation adoption models, expert opinion, and analogy from other situations that are deemed similar[43]. The challenge with this type of forecasting is the variability in potential demand. Because there is no historical data to directly base these forecasts on, demand must be inferred from indirect data, often with complex and highly variable relationships.

In summary, traffic demand forecasting for telecommunication networks has been in significant flux over the past decade, from the expected growth in demand in the late 1990's to the realization of that growth in the late 2000's. Because the underlying drivers of network demand are in such significant change, traditional forecasting methods based on historical data are no longer relevant. Some work has been done to systematically forecast traffic demand, however these methods have not yet been adequately proven [43].

2.3 Diffusion of Innovation

The last section in this background focuses on a topic that is not directly related to network planning, but provides a significant body of knowledge to enable the evaluation and integration of technical network design with the socio-economic environment in which the networks will exist. The study of the diffusion of innovations started with earnest in the 1960's, with a number of key studies being published [46] during this time period. The two studies, and their derivatives that are of concern to the current context are the works done by Rogers [22] and Bass [13]. Rogers uses the terminology diffusion of innovations, while Bass looks at the diffusion of new products, however both attempt to describe the same thing.

Rogers provides a qualitative look at how innovations move through society, while Bass provides a more quantitative model. Both provide solid basis to better integrate and evaluate new network services, and their potential impact.

The main contributions from Rogers, is an outline of factors influencing innovation adoption rates, the innovation decision process, and the perceived attributes of an innovation. Rogers also provides a breakdown of potential adopters into five categories. These categories assume adoption follows a normal distribution (which is has been called into question [46]), and are labeled innovators, early adopters, early majority, late majority and laggards. The most significant point of this breakdown is the flow of innovations from innovators, who are disconnected from the social system to early adopters, who are not risk averse, but are connected to the social context in which the innovation applies, to the early majority. The early majority, according to Rogers, are the decision makers in organizations, and are often influenced by the opinions of the early majority.

A number of traits and characteristics of both individuals and organizations that tend to adopt innovations earlier were identified. The most significant characteristic is slack resources. This can be measured in capital and in time. New innovations are risky, and sometimes fail. People and organizations that have the time and money to investigate and integrate new innovations are often those who have significant slack resources. There are other characteristics that were identified, like education levels for individuals, and organizational structure and complexity, that can affect adoption rates, however slack resources is a common influential factor.

While the other contributions Rogers has made to the understanding of the diffusion of innovations, they tend to focus on the innovation being diffused. In the context of telecommunication network planning, these innovations are not fully in the scope of the planners, and hence the perceived attributes, innovation decision process, and factors influencing adoption are not covered in this section.

As Rogers provides a general understanding for the diffusion of innovations, Bass created a simple, versatile model to quantify the diffusion of new products. Bass breaks users down into two categories, innovators and imitators, which serve as coefficients in his model (1). The innovators, and imitators represent the portion of the population show are not influenced by earlier adopters (p) and those that are (q).

$$N(t) = m\left(\frac{1 - e^{-(p+q)t}}{1 + \left(\frac{q}{p}\right) \times e^{-(p+q)t}}\right) \tag{1}$$

The Bass model has been used extensively in forecasting new product adoption, and there are a number of modifications that have been made to make it better suited to the telecommunication context [43].

The diffusion of innovations has been well studies, and although the nature of what is being studied is highly dynamic, using the concepts developed by Rogers and Bass, some predictability has been brought into this field. With the rapidly changing, and highly dynamic context of the applications that run on telecommunication networks, these concepts provide significant potential to bring this seemingly unpredictable context into the network planning framework.

2.4 Impact of communication networks

The literature survey that was done for this project was quite broad, as the scope of this research touches on a broad set of research areas including sociology, engineering, and management science. The initial topic of the PhD was broadly defined as how to better design telecom networks to align with socio-economic priorities. To understand this, a search was done to see what studies have been done to analyze the effects of telecom networks on different areas of society. To provide some structure to this search, six key areas were identified as service drivers for a network. These key service drivers were education, healthcare, social, economic, government, and military.

The purpose of the broad start was to begin to understand the relationships between telecommunication networks, and users. Dividing users or usage scenarios up by social utility allowed the literature search, and the results to focus on various societal needs. These service drivers took the components what is referred to as e-society (eHealth, eLearning, eBusiness, eGovernment)[47] and added the social component, being the usage scenarios of people outside of their work related roles focusing on entertainment and communication, and military, as national security is highly intertwined with the integrity of the communication networks that support each country. The scope of the research has been narrowed to focus significantly on healthcare, and as such, the literature review was focused accordingly. There is a body of literature relating telecommunication networks to the other key service drivers, however, healthcare will be presented here.

Part of the focus of this research is how to deploy telecommunication networks in rural and remote regions, and so the literature search emphasized articles that had specific application to these regions. Outside of literature addressing the digital divide [12], [48], [49], the literature was limited compared to more general results. Where applicable, articles relating to rural and remote areas will be highlighted.

2.4.1 Strategy and Summarized Results

The strategy that this literature review took was to focus on:

1. Articles that relate telecommunications in general to the impacts in the key services areas.
2. Articles that relate underlying infrastructure to specific outcomes in the key service areas.
3. Look at network planning paradigms, focusing on strategic network planning.
4. Advanced network design including multi-period design and design with uncertainty
5. Demand forecasting and analysis.

This strategy, and the preliminary work, highlighted another area the literature survey needed to cover. The study of innovation or new product diffusion, especially as it related to telecommunications, was found to be a foundational element and potential linkage between the socio-economics of telecommunications, and technical network design. Innovation diffusion has been mentioned in some of the demand forecasting literature, and is a key component of network design in long-term and greenfield planning.

The search started out looking at research done in the realm of studying the effects of telecommunication networks on key service drivers. The point of this was to begin to get a sense for how to measure the socio-economic impact of networks on each of the key service drivers. This first search returned very little. One of the reasons for this is most likely found in the disconnect between the user and the network in a typical value chain (Figure 6).

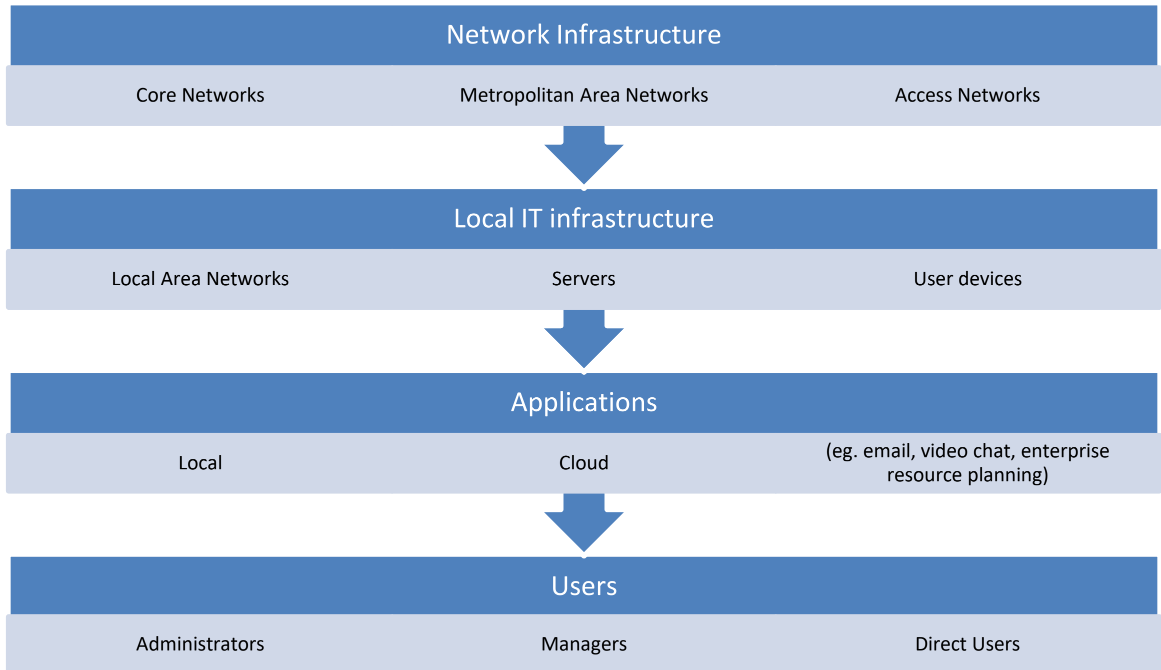


Figure 6 – A high level outline of the value chain for communication driven applications.

There were some studies done linking economic benefits and costs of specific applications [50], however all studies assumed that the supporting infrastructure was in place and adequate. One of the motivating factors in this research was from [51], where the author describe the impact of our ability to communicate ideas, and the impact that has had on human standards of living. This idea that global economics are significantly influenced by communication infrastructure has been discussed by [8]–[10]sociologists and economists [52][53][54][55][56][10], with the most significant discussion point being the causality between telecommunication infrastructure and economic growth [54][10].

One point that stood out was the understanding that it was not just the availability of ICT based tools that encouraged their use, but also human capital, in the form technical expertise, and business process re-engineering were required for their effective use [10]. This point has formed one of the primary components in the studies that I am proposing. That is, to effectively use telecommunication infrastructure, investment must be made in both the infrastructure itself, and the capacity to use it. Because increasing capacity to use the network is often driven by the perceived potential of the infrastructure, and the applications that run on it, a search was done in the available literature for works relating the impact of telecommunication networks to key deliverables. The focus was primarily on healthcare, with some emphasis on the other key areas.

The literature search used Google Scholar as a generic, cross database search tool, and specific searches were done on a number of databases and journals. Most searches were done using IEEE [57], Science direct [58], and Springer Link [59], with specific journals also targeted.

2.4.2 Healthcare

Literature linking telecommunication infrastructure to healthcare delivery was not readily available. Most articles focused on specific applications and assumed network coverage [60]–[62]. What stood out from the literature is the conflict in the economic benefits of eHealth. Some argue that there is a clear economic benefit to eHealth technologies [63], while others argue that it increases costs [64][65]. A study done in Europe attempted to quantify the impact of electronic health records in both economic and social terms [50]. The results from this study were split. In purely economic terms, there was a negative return on investment, while when the social impact is taken into account; there is a positive impact Table 1. Of significance in this study is the time horizons for which a net cumulative socio-economic impact was found. The minimum timeframe was 4 years, with an average of just under 9 years. This timeframe is significant in the study of telecommunication network planning, in that even with the telecommunication infrastructure in place, there is still nearly a decade until a net positive effect of EHR's can be seen.

	Min	Max	Avg.	Range
Annual ratio 2010	0.61	9.95	3.82	9.35
Annual ratio 2008	0.15	4.62	1.66	4.47
Cumulative ratio 2010	-0.20	1.92	0.78	2.12

Table 1 – EHR impact using a socio-economic return measurement.

Looking at the use of tele-health in rural and remote areas, [66] describes the establishment of a rural tele-health project in Wyoming. The most significant part contribution of this paper is the general lack of experience and awareness of tele-health among healthcare providers in the state. The paper alluded to the idea that this is indicative of most rural healthcare environments. Another paper looking at remote and rural tele-health in Scotland [67] found that simple data transactions, such as laboratory results, had found wide acceptance, however healthcare provides had little experience with more complicated tele-health scenarios, and approached such tools with significant skepticism. The conclusion of both these papers was that policy involving tele-health and eHealth in general needs to prioritize educational and training programs for related initiatives to be successful.

I will note here that although educational and awareness programs are part of the emphasis, this focus includes a significant innovation bias [22]. Effort needs to be put into adapting both work practices, and the technology to fit into the healthcare context [10].

There were a number of papers describing the technical design of tele-health and eHealth systems, [68][69] for example. These papers described solutions, and implementations developed on a pilot level scale. However, as noted in [70], these projects have generally not scaled into broad based adoption, as the services do not fit well into the overall needs of healthcare professionals. One recommendation from this paper [70] is that tele-health projects should incorporate innovation management theories.

2.4.3 Background Summary

Communication infrastructure is essential to social and economic growth and wellbeing around the world. The question we wish to ask is, what are the interactions between its quality and pervasiveness, and the impact those factors will have on various aspects of life, such as education delivery, healthcare, government, and economic activity? Part of the goal of this research is to define the technical, economic and political constraints affecting the deployment of backbone communication infrastructure, and the impact they have on underserved or developing economies.

Deploying the necessary backbone connectivity to provide end users with adequate levels of communication is dependent on a number of factors. What is adequate depends on the systems that can feasibly be implemented, and can either be a driver of backbone requirements, or be pushed along once the infrastructure is in place. The design, implementation, maintenance, and service provisioning can be funded by government, not-for-profit organizations, and private industry alike. Geography and political boundaries can also greatly affect how the infrastructure is set up, as well as how it is accessed.

This research will develop the understanding and feedback mechanisms between the technical and economic considerations of communication infrastructure, and the key social and economic factors they affect. The goal is to develop a model that will be able to effectively describe these interactions taking into account the plethora of external factors affecting the communication systems.

2.5 Goals, Motivation

It could be said that the current infrastructure that drives the internet in North America was built through bankruptcy law with the number of telecommunication companies that went bankrupt in the early 2000's. Although spare capacity is a secret that network operators keep hidden, it has been estimated that the capacity deployed in the dot com bubble of the late 90's and early 2000's has been sufficient to supply traffic needs until recently. The concern is that can future capacity be built at a cost that sustains much of the internet's core value (content neutrality), while not relying on another round of bankruptcies to fund it.

Another example of the troubles that are involved with building out a network can be seen in the Alberta Supernet [71][72]. The Alberta Supernet was launched in 2005 [24], however, it still is chronically underutilized, and does not live up to its purported potential [73][74][75].

The question brought forward by these examples is how do you design and deploy network capacity that is able to maximize, or at least improve, the economic and social impact. Networks alone offer very little value to end users (with large corporations leasing capacity for their own purposes being the possible exception). The value that is delivered by the end users is heavily dependent on the applications that run on the network infrastructure. These applications are vast, and are hard to predict [43]. There is a need therefore to understand and relate a given population's ability to extract value from a network through the applications that utilize it, and the network deployment activities.

The purpose of the research outlined in this chapter is to develop a network design framework that will take into account influences on network usage to better align network build outs with the ability in the intended market(s) to utilize the capacity. This alignment is intended to increase the effectiveness of the communication networks, and the capital required to deploy them. By better understanding the interactions between the current environment, network deployment and policies, and outcomes, networks, and the technologies they support, can be deployed in a more effective manner.

If an eHealth initiative (eHealth is a blanket term representing the use of ICT technologies in the delivery of healthcare) is rolled out that aims to increase in-home monitoring in a region that does not have significant network access, or does not have a population that is ready to accept such a technology, it will be difficult for the initiative to succeed. Conversely, if a government is allocating resources or incentives to build network access in a region, without a proper understanding of how the region will utilize the network, and in what kind of timeline, the network could be significantly underutilized. In general, an increased ability to align network deployment, key applications, and a population's ability to adopt these tools over time will increase the effectiveness of the key applications (or initiatives), and the capital required to deploy the networks that support these applications.

This research aims to develop a model for understanding and evaluating the interactions between the current environmental conditions with network planning activities and their outputs, in order to affect the uptake and impact of initiatives in key areas of society. This should be encapsulated in the development of a new greenfield network design paradigm that augments traditional network planning activities with considerations of the socio-economic environment and purpose in which the network would exist.

2.5.1 Research Goals

The purpose of this research is to expand the understanding of network planning by developing the techniques required to align technical network design over time with sector specific objectives and the capacity of the targeted users to utilize the network. The goals of this research are to:

- Articulate the metrics affecting the socio-economic implications (both inputs and outputs) of telecommunication networks
    - Validated through a literature search focusing on the social and economic impacts of telecommunication networks
- Develop a network planning framework that incorporates these metrics
    - Validated through literature by taking a systems approach to network planning

- Expand technical network capacity allocation schemes to incorporate this broader framework
    - Focus on the incorporation of the user's capacity to utilize the network.
    - Validated through the simulation of the technical design models

The contribution of this research is twofold, the network planning framework that articulates the broader context which technical network designs interact with, and a set of network topology and capacity allocation models that integrate key elements from this model.

Chapter 3. Survivable Network Capacity Allocation and Topology Design Using Multi-Period Network Augmentation

This first paper provides the foundation for technical telecommunication network planning capable of incorporating socio-economic factors. There have been a number of papers published on the topic of multi-period network planning, as articulated in the background section of this paper, but there was a lack of frameworks or mathematical models that integrated topology augmentation. This paper presents an integer linear program (ILP) for multi-period network augmentation that can adapt to many of the prominent network survivability schemes. This understanding of multi-period network design made it possible to include a broader set of decision variables into the planning process that included a feedback loop. Without this, there would be no mechanism to evaluate the effects of the proposed network planning framework.

[1]
B. Todd and J. Doucette, "Survivable Network Capacity Allocation and Topology Design Using Multi-period Network Augmentation," *Journal of Network and Systems Management*, vol. 25, no. 3, pp. 481–507, Jul. 2017.

3.1 Introduction

Large scale backbone communication networks have been essential for business, governance, social interactions and many other aspects of our modern society. The impact of these systems, however, is not ubiquitous. There are many areas, even in North America [76], that do not have access to the services and opportunities that these networks afford. Many rural and economically challenged areas do not have access to the networks that serve as a platform for most areas of society, despite the significant value proposition that they can offer. These areas can be difficult to plan for, and do not have the same economic assurance for network operators as larger, more populated centers have. Planning activities in these areas need to be able to adapt to changes and growth in demand as the usage patterns in these areas mature.

Planning activities can include the addition of capacity to a network, the re-allocation of existing capacity to support new traffic patterns, and network extensions and augmentation. These activities are evaluated in an attempt to optimize capacity utilization, minimize costs, and maintain an acceptable level of fault resiliency within the network.

One of the challenges is to ensure the network is capable of handling adverse events such that the end users do not experience significant disruptions in service (though admittedly, what constitutes significant is highly dependent on the end user). This can be accounted for in many ways, but for the purposes of long range planning of backbone networks, this survivable network design entails the allocation of spare capacity throughout the network to accommodate for disruptions in the network. Survivable network design aims to enable networks to continue to fulfill its function in the presence of unplanned events such as natural disasters, an errant backhoe, or a malicious attack [8][2].

The incorporation of survivability design, capacity allocation, and topology augmentation into an optimization model enables planners to evaluate strategies, options, and risks in the design process.

With planning horizons of many years [45], this evaluation will evolve, and needs to adapt to changing environments [25].

The work presented here adds the time dimension to this process. The motivation for time variant design for large scale survivable networks was to enable the evaluation of topology augmentation and capacity expansion when demand and technology may be uncertain. By delaying expansion decisions, the risk of overbuilding may be mitigated, however this has to be balanced with the realities of expanding networks, economies of scale, and other network service level obligations.

This risk is especially prevalent when looking at network expansion into rural and remote areas. The "digital divide" still exists even in some of the most advanced areas of the globe [12]. Addressing the needs of these areas are extremely difficult, with very poor economics and significant uncertainty. However, expanding high speed network services into these areas has been identified as crucial in increasing their economic and social conditions [24].

A background of survivable network design and multi-period planning is provided in section 3.2. A novel multi-period survivable network *integer linear programming* (ILP) model is outlined in section III that optimizes both topology augmentation and capacity placement over multiple time horizons. This was implemented and run using a test case network, with the experimental setup presented in section 3.4. We present and discuss the results in section 3.5.

3.2 Background

There has been a significant amount of work done in the field of survivable network design, which addresses many of the concerns around the provision of reliable communication infrastructure [7]. Much of this work focuses on long term infrastructure provisioning, requiring dedicated allocation of spare and working capacity throughout the network [41], [77], [78]. Many strategies have been developed to efficiently dimension topology and capacity, focusing primarily on single failure resiliency [7]. While many factors affect network availability, it is span failures that are the most significant contributors to network outages [7]. The most common method for comparing the various schemes in designing survivable networks is the amount of redundant (spare) capacity that is required to provide full single span failure resiliency [78]. In this work, two survivability schemes are used, *shared backup path protection* (SBPP**)** [24] being the more efficient of the two, and *demand-wise shared protection* (DSP) [79].

Both SBPP and DSP were implemented in this work using integer linear programming. When designing a survivable network using this method, there are a number of assumptions typically made in order to reduce the time it takes to reach an optimized design, as well as to maintain linearity [7]. A typical assumption is that costs to add capacity is linear (or at least piecewise linear) [80]. This approach was taken in order to ascertain the level of optimality when evaluating network designs and provide a consistent metric to compare designs.

3.2.1 Topology Design

Whether designing a network for a new region, or expanding one that already exists, understanding when and where to expand the network is a core activity. When looking at the degree of investment required to build out networks, understanding how to expand the network efficiently is essential. This is accentuated when looking at many underserved regions. These areas typically have challenging economics where capital efficiency is one of the key factors for sustainable operations. Given the time horizon of deploying networks, building the original network [81] and augmenting it over time [82] both need to be evaluated, and options need to be understood for different demand growth scenarios.

When considering topology design, or augmentation of a network, the ratio of establishing a span versus the addition of extra capacity is a critical factor. In this work, this ratio is referred to as the span establishment cost multiplier. This ratio doesn't have to be constant through time or across the network, and can be affected by ease of access to land, sharing agreements between operators, technology available, and the presence (or lack thereof) of existing rights of way. Whatever contributes to this ratio, varying it when examining topology design provides a method to evaluate the trade-off between a more diverse network, and excess spare capacity. Survivability schemes that take better advantage of the diversity of available routing paths will benefit more from a diverse network.

There are many other factors that could affect topology design, such as how demand is distributed, and its modularity, or the degree to which nodes in the network are clustered. This work considered the effect of the span establishment cost multiplier, but not those other effects on topology design.

3.2.2 Demand-wise Shared Protection

There are two main classes of survivable networks, ring and mesh. The earliest survivable networks used ring-based routing, however, these required a significant amount of redundant spare capacity. Mesh networks were enabled by advances in technology that allowed more complex and diverse routes through the network. The trade-off with mesh networks was the complexity of the network designs, and the speed of restoration. To compromise between the efficiency of mesh network survivability design, and the simplicity of ring networks, DSP limits the sharing of capacity between routes servicing the same end nodes in the network [79]. There are a number of other hybrid approaches that attempt to blend ring and mesh networks [19], [83], however, DSP was selected because of its straight forward routing, and its place in the capacity efficiency scale [7].

As mentioned, DSP allows paths servicing the same node pairs in a network to share backup capacity (with the caveat that each working path is failure independent). In its most basic form, DSP is equivalent to 1+1 *automatic protection switching* (APS) with one working and one backup path per pair of nodes in a network [7].

The efficiencies gained through DSP can be significant if path diversity is available. Figure 7 shows the routing with a single working path that uses a total of 20 λ's, including the capacity required to provide full single failure redundancy. If the working traffic is dispersed through three working paths, all sharing the same spare capacity (and maintaining full single failure redundancy), the required number of spare paths is reduced 60% as seen in Figure 8.

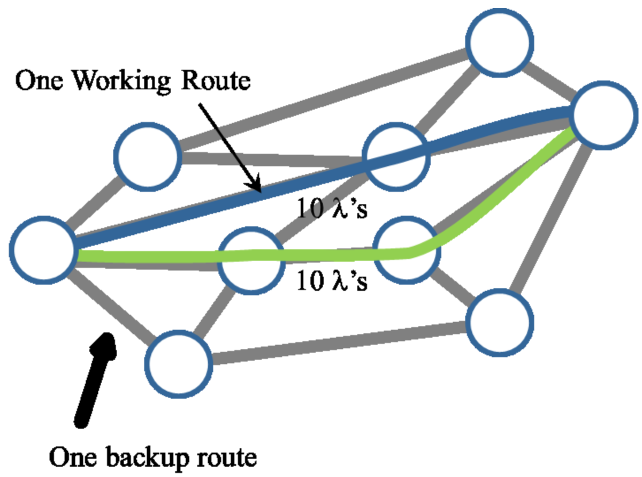


Figure 7 – DSP with traffic routed on a single working route and protected by a single backup route, using a total of 20 λ's

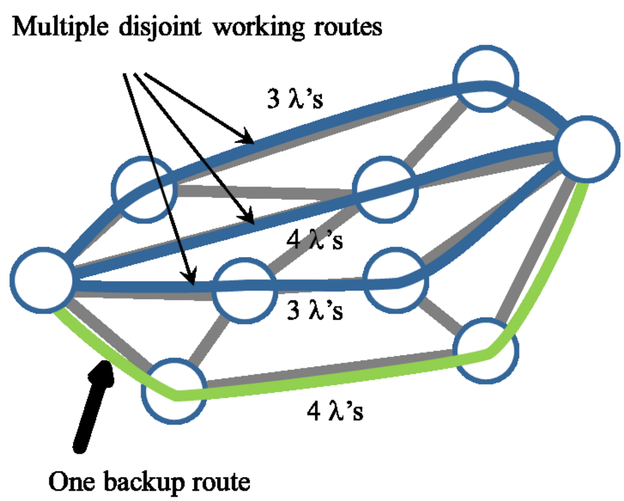


Figure 8 – DSP with traffic routed on three working routes sharing a single backup route, using a total of 14 λ's

Again, DSP provides a simple spare capacity assignment mechanism, especially when compared to many mesh network survivability mechanisms. The cost of this simplicity is that a diversity of disjoint working paths must be available, and the sharing of spare capacity is limited to a single demand. A more detailed discussion of DSP can be found in [79].

The basis for the multi-period network augmentation model presented later in this work was an ILP formulation of DSP that incorporated topology augmentation along with capacity allocation [84]. The ability to augment topology was a key piece of multi-period network design, as it provides a basis for understanding how to expand a network by balancing the affects the spare capacity efficiency with new span instantiation costs.

With network topology design the cost of topology and the diversity of paths available is evaluated, and trade-offs are evaluated. DSP requires a diverse topology for the efficiency gains it purports. It defaults to 1+1 APS at its most rudimentary design, and would theoretically maximize its efficiency if each demand pair could have d+1 disjoint paths (where d is the units of capacity required for the pair). Obviously, the best design is somewhere in between, depending on the environment where the network exists.

In this work, DSP was modeled using an *arc-flow* approach that doesn't use predefined paths [85]. This allows a dynamic typology to be evaluated within the mathematical evaluation of the design of the network. The most significant issue with this approach is the complexity of the problem. DSP was designed to be simple, but even such, the number of binary variables in this formulation can make it difficult to solve. This formulation uses capacity conservation and path disjointedness throughout the network (outside of the end nodes of each demand pair), and enforces single failure survivability.

Our formulation uses the following notation:

Sets:
N is the set of all nodes in the network.

D is the set of all node pairs with traffic demands between them.

$A_n \in N$ is a subset of *N* and represents all nodes that are connected to node *n* by a single hop.

Parameters:

$c_{i,j}^{link}$ is the cost of adding one unit of capacity to span *i*.

$c_{i,j}^{cap}$ is the cost of implementing a span connecting nodes *i* and *j*.

O_r is the origin node for demand *r*.

T_r is the destination node for demand *r*.

d_r is the number of units of traffic required by demand *r*.

M is a sufficiently large number (in our case, $M=\sum_{\forall r \in D} d_r$).

Variables:

$\omega_{i,j}^r \geq 0$ is the traffic flow from node *i* to node *j* allocated to demand *r*.

$f_{i,j}^r \in \{0,1\}$ is a binary variable indicating whether capacity allocated to demand *r* is allocated on the span between node *i* and node *j* where $j \in A_i$.

$f_{i,j} \in \{0,1\}$ is a binary variable indicating whether any capacity is allocated from node *i* to node *j* or from node *j* to *i*, where $j \in A_i$

The DSP capacity allocation and topology optimization ILP is as follows:

Minimize

$$\sum_{i \in N} \sum_{j \in A_i} \sum_{r \in D} \left(\omega_{i,j}^r \times c_{i,j}^{link}\right) + \sum_{i \in N} \sum_{j \in A_i} \frac{f_{i,j} \times c_{i,j}^{cap}}{2} \quad \textbf{(1)}$$

Subject to

$$\sum_{k \in A_i : k \neq j} \omega_{i,k}^r \geq d_r \qquad \forall r \in D, i \in O_d, j \in A_i \qquad \textbf{(2)}$$

$$M \times f_{i,j}^r \geq \omega_{i,j}^r \qquad \forall r \in D, i \in N, j \in A_i \qquad \textbf{(3)}$$

$$M \times f_{i,j} \geq \sum_{r \in D} f_{i,j}^r + f_{j,i}^r \qquad \forall i \in N, j \in A_i \qquad \textbf{(4)}$$

$$f_{i,j}^r + f_{j,i}^r \leq 1 \qquad \forall r \in D, i \in N, j \in A_i \qquad \textbf{(5)}$$

$$\sum_{(j \in A_i)} f_{i,j} \leq 1 \qquad \forall r \in D, \; i \in N \mid i \neq O_r, i \neq T_r \qquad \textbf{(6)}$$

$$\sum_{(j \in A_i)} f_{j,i} \leq 1 \qquad \forall r \in D, \; i \in N \mid i \neq O_r, i \neq T_r \qquad \textbf{(7)}$$

$$\sum_{j \in A_i} \omega_{i,j}^r - \sum_{j \in A_i} \omega_{j,i}^r = 0 \qquad \forall r \in D, \; i \in N \mid i \neq O_r, i \neq T_r \qquad \textbf{(8)}$$

In this formulation of the DSP survivability scheme, the objective function combines the capacity costs ($\omega_{i,j}^r \times c_{i,j}$) and the costs of instantiating a span ($\frac{f_{i,j} \times c_{i,j}}{2}$). The costs of instantiating a span are divided by two such that they are not double counted (assuming that the costs are accounted for whether the traffic moves from node *i* to *j* or *j* to *i*). Survivability considerations are accounted for in equation (2). This states that if any one span around the origin fails, there must be enough capacity allocated to other spans to support the traffic requirements for each node pair. This is sufficient in establishing dual failure restorability because constraints **(6)** and **(7)** ensure each path is node disjoint. Equations **(3)** and **(4)** ensure that if capacity ($\omega_{i,j}^r$) is assigned to span (*i,j*) for any demand, then $f_{i,j}$ is 1 (indicating that the span is used in the topology design). Capacity cannot flow in both directions on a given span for any traffic flows supporting the same node pair, and so **(5)** ensures that capacity is allocated in at most one direction for each span and demand pair combination. Each node that is not an origin or destination can have no more than one outgoing flow **(6)** and one incoming flow **(7)**, and the sum of capacity entering the node must equal the sum leaving **(8)**.

A fundamental part of topology design is the balance between adding capacity and adding new spans to the network. There are many characteristics that influence this balance, with a primary one being the ratio between the cost of capacity and the cost of instantiating additional spans. While this distinction is not always clear in practice, it serves as a good metric in characterizing the type of network being developed.

Other factors affect topology design, including modularity, both at the span level and the units of capacity level [86]. While modularity hasn't been directly accounted for in this work, it could be seen as the step sizes available for adding capacity to a span (measured in number of capacity units). Modularity can also be accounted for in the unit size of the demands, which is especially important when dealing with multi-path routing of demand between two end nodes. For example, a demand of just two lightpaths leaves little opportunity to take advantage of multi-path efficiencies of DSP, while a demand of ten lightpaths can be divided into many different paths to better take advantage of backup capacity sharing. If both the demands are divided to use two working paths, and share the backup path, on a relative scale the reduction in spare capacity is equivalent (half of what was required with one working path), the cost savings of the reduced spare capacity relative to the cost of establishing a new span to support the added path is a lot greater for the demand of ten lightpaths versus a demand of two lightpaths.

3.2.3 Shared Backup Path Protection

DSP provides a simple capacity sharing mechanism, but because capacity is only shared between traffic flows for a single demand, capacity redundancy is still high compared with other mesh survivability schemes [87]. On the other end of the capacity efficiency scale is shared backup path protection. SBPP shares backup capacity among any disjoint working paths, not just those between a single pair of end nodes, and is significantly more efficient with regard to spare capacity [7]. SBPP is quite efficient in its use of spare capacity, and hence the benefits of a better connected network are more pronounced (greater path diversity provides more opportunity for backup capacity sharing).

In the classic SBPP, each demand is routed on a single working route, and a single backup route is allocated with sufficient spare capacity for each working lightpath on that working route. Spare capacity can be shared on spans common to multiple backup routes, as long as the associated working routes are disjoint. This is demonstrated in Figure 9, where the working paths (in blue) between node pairs A & B and A & D are disjoint, and their respective backup paths (in green) are therefore are able to share backup capacity on span A-C. The spare capacity required for any span is the maximum spare capacity required for any failure scenario. The more balanced the spare capacity requirements are between failure scenarios, the more efficient the overall network will be with regard to spare capacity vs network traffic demands [88].

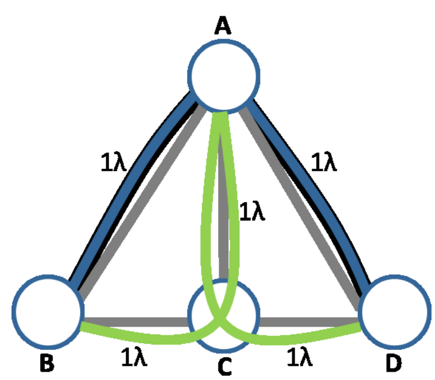


Figure 9 – Example of capacity sharing in SBPP

The SBPP strategy is computationally intensive when implemented using integer linear programming. Strategies have been developed that utilize multiple working and/or backup paths per demand, which has been shown to be quicker to solve [88]. However, to incorporate network augmentation in a strictly linear mathematical model (i.e., ILP), only one backup route can be utilized per working route (or at least we have been unable to find a means of doing it with multiple backup routes per working route), and hence the classic form of SBPP is used here.

In order to more easily incorporate network augmentation, we use an arc-flow implementation for the SBPP ILP model. This is not commonly used in literature, as it increases the computational complexity. ILP models for SBPP network design typically use an arc-path ILP formulation [77], [89].

The following is an arc-flow implementation of SBPP that enables topology augmentation. This formulation respects the SBPP capacity sharing capabilities, while determining the optimal topology and capacity allocation. Readers should note that where possible, notation used in the DSP ILP model above was also used here, and is not repeated.

New Sets:

S is the set of spans in network.

N_s is the set of end nodes from span s

New Variables:

$\boldsymbol{c_s^{link}}$ is the cost of adding one unit of capacity to span s.

$\boldsymbol{c_s^{cap}}$ is the cost of implementing span s.

ω_s^r is the working capacity allocated to span s for demand r.

β_s^r is the backup capacity allocated to span s for demand r.

$\gamma_{s1,s2}^r$ is the backup capacity allocated to span *s1* for demand r if span *s2* fails.

$\omega_{i,j}^r$ is a binary variable (as opposed to an integer in DSP) indicating whether working capacity for demand r is assigned from node i to node j.

$\beta^r_{i,j}$ is a binary variable (as opposed to an integer in DSP) indicating whether spare capacity for demand *r* is assigned from node *i* to node *j*.

Minimize

$$\sum_{(s\in S)}\left((\omega_s + \beta_s) \times c_s^{link} + f_s \times c_s^{cap}\right) \quad (9)$$

Subject to:

$$\omega_s = \sum_{r\in D}\sum_{i,j\in N_{s_1}|i\neq j}\left(\omega^r_{i,j} + \omega^r_{j,i}\right) \times d_r \qquad \forall s \in S \quad (10)$$

$$\beta_{s_2} \geq \sum_{r\in D}\gamma^r_{s_1,s_2} \times d_r \qquad \forall s_1 \in S \quad \forall s_2 \in S|s_1 \neq s_2 \quad (11)$$

$$\omega_s + \beta_s \leq M \times f_s \qquad \forall s \in S \quad (12)$$

$$\gamma^r_{s_1,s_2} \geq \sum_{i,j\in N_{s_1}|i\neq j}\left(\omega^r_{i,j} + \omega^r_{j,i}\right) + \sum_{i,j\in N_{s_2}|i\neq j}\left(\beta^r_{i,j} + \beta^r_{j,i}\right) - 1 \qquad \forall r \in D, \quad \forall s_1 \in S \quad \forall s_2 \in S|s_1 \neq s_2 \quad (13)$$

$$\sum_{(j\in A_i)}\beta^r_{i,j} = 1 \qquad \forall r \in D, i \in O_r \quad (14)$$

$$\sum_{(j\in A_i)}\omega^r_{i,j} = 1 \qquad \forall r \in D, i \in O_r \quad (15)$$

$$\sum_{(j\in A_i)}\beta^r_{j,i} = 1 \qquad \forall r \in D, i \in T_r \quad (16)$$

$$\sum_{(j\in A_i)}\omega^r_{j,i} = 1 \qquad \forall r \in D, i \in T_r \quad (17)$$

$$\omega^r_{i,j} + \beta^r_{i,j} \leq 1 \qquad \forall r \in D, \quad i \in N| \quad (18)$$

$i \neq O_r, i \neq T_r,$

$j \in A_i$

$$\sum_{j \in A_i} \beta_{i,j}^r - \sum_{j \in A_i} \beta_{j,i}^r = 0 \qquad \forall r \in D, \quad i \in N \mid i \neq O_r, i \neq T_r \qquad \textbf{(19)}$$

$$\sum_{j \in A_i} \omega_{i,j}^r - \sum_{j \in A_i} \omega_{j,i}^r = 0 \qquad \forall r \in D, \quad i \in N \mid i \neq O_r, i \neq T_r \qquad \textbf{(20)}$$

The objective of this model is to minimize the total cost of the network **(9)**, which includes the capacity costs, as well as the instantiation costs. As in the DSP ILP model, instantiation costs represent the costs incurred to secure the rights of way, and install or appropriate the physical infrastructure required. There is an assumption that these costs are incurred only once, at the time of commissioning the associate span. This ILP formulation would have to be modified to represent other cost realities. Equation **(10)** calculates the working capacity required for each span, while the constraints in **(11)** convert the backup span allocation into the required spare capacity. Equation **(11)** is an inequality as it has to ensure that there is enough spare capacity to route traffic for all failure scenarios. This inequality emphasizes the fact that it is not always the shortest working/backup path pair that provides the optimal capacity allocation. There will be some cases that a carefully chosen working path will have a backup path that can be routed on spans that will not require extra spare capacity (in some cases).

Equation **(12)** is used to detect if a given span has any capacity routed on it (working or otherwise). If there is capacity routed on span s, then f_s will be equal to 1, and 0 otherwise.

The constraints in **(13)** are used to calculate whether a backup path is used for a given working path failure. If span s_1 is used in the working path and span s_2 is used in the backup path, then γ_{s_1,s_2}^r is equal to 1. These constraints, along with those in **(11)**, implement capacity sharing in this implementation of the SBPP survivability scheme.

The constraints in **(14)** to **(17)** ensure that a backup and a working path are implemented for each demand node pair in the network. Equation **(18)** forces the working and backup paths to be node disjoint by only allowing one of them to be set for any node outside of the origin or destination. Constraints **(19)** and **(20)** ensure that all nodes except the origin and destination are transiting nodes (for each demand pair).

3.2.4 Multi-period network design

Planning a core network involves long planning horizons generally measured in years, with a user demand and regulatory environment that is not certain. A common approach to this type of uncertainty, as mentioned earlier, is to break up the planning process into multiple time horizons. Multi-period design has been discussed in the literature for many years, with the work in [90] being among the first discussion of this topic, in this case for SONET interoffice networks. This paper, like most on the topic, describes an algorithmic approach to multi-period capacity allocation. A mathematical programming approach to multi-period design was presented in [91]. This work covered 1+1 APS with network expansion and includes a thorough discussion on the impacts of the various factors affecting these network design models. Specifically, the factors discussed were the network size, the number of time periods, and the evolution of cost parameters. Due to the intensive computing requirements of the ILP presented, a heuristic approach was used to solve the multi-period network capacity and topology augmentation problem. The multi-period approach resulted in a 4.4% decrease in overall costs.

The work in [38] provides an overview of the various approaches to multi-period network design, with the approach presented in this work labeled as “all-periods”. This means that all time periods are optimized concurrently, with the impact of capacity and topology decisions fully known to earlier time periods. This is the most computationally intensive approach to multi-period design, but provides assurance to the optimality of the solutions. [38] provides an ILP for all-period optimization of equipment sizing without topology augmentation or survivability concerns. It included aspects of demand uncertainty through the use of stochastic programming. The work found cost improvements of a similar order of magnitude as [91], with results ranging from little improvement up to approximately 15%.

A discussion of the issues of multi-period optimization for network planning is given in [34], [92] compares the performance of an ILP to a cost expectation model, concluding that while the expectation model speeds calculation time considerably, actual cost estimation was done poorly, hiding the temporal costs in multi-period planning. Also discussed in the literature is the effectiveness of multi-period planning [39] against end-of-life approaches. This found that end-of-life planning increased initial costs and investment risk due to the uncertainty in potential future demand requirements.

We aren't aware of any prior work that addressed different optimization techniques in light of multi-period planning, or in comparing and contrasting the impact on the relationship between capacity costs and span establishment costs to optimize topology evolutions. These topics are addressed in the present work.

3.3 MultipPeriod Network Augmentation with Survivability

One of the most significant implications in taking an all-periods approach, when including topology expansion in the design problem, is the complexity that topology design introduces. Topology design necessitates an arc-flow ILP design, and significantly increases the computational complexity of the optimization problem. We found, however, that small-scale problems can be solved using this approach, and the result can serve as a basis to evaluate heuristic approaches to the multi-period problem. Similar to [38], an understanding of optimized topology augmentation and survivable network demand routing provides a better understanding of how changing demand patterns and uncertain growth projections can be accounted for.

3.3.1 Economic Considerations

One of the primary motivations for multi-period network design is economic (with the other being risk). The time value of money directly implies that costs in the future are better than costs today [93]. The costs considered in this work were assumed to be affected by the time value of money, dictated by an organization's weighted average cost of capital, and the general trend that per-unit capacity costs decrease over time as the supporting technology costs do not scale directly with the capacity they support.

Discounting cost over time has a significant effect on the viability of a network expansion. With a discount rate of 10%, costs pushed 5 years into the future will decrease by 61% in nominal terms. There is the potential to also delay revenue by the same amount, making the delay irrelevant. However, if the disconnect between capital expenditures and revenue generation can be decreased, then network expansions can be made more readily.

There are other mechanisms and techniques to evaluate the impact of delaying costs, such as real option valuation, however discounted cash flow is the most common, and most applicable to this work.

With large scale backbone style networks, delaying investments is not always feasible due to the modularity and economies of scale that exist either in the establishment of spans between network nodes, or in the modularity of the transmission technology selected [80]. Newer technology is more granular in the capacity steps it offers (or at the very least keeps the similar step sizes while offering the ability to scale up significantly in overall capacity) [94].

Balancing the effect of the time value of money to push costs into the future is the savings provided through economies of scale. When we refer to economies of scale, we are referring to the idea that per unit costs go down when larger overall capacities are deployed. The cost profiles for economies of scale are well documented [95], and follow a non-linear curve. This provides a challenge when implementing an ILP representation of the problem, however, that can be solved by discretizing the curve into steps (as we do below).

In this work, we use both factors to implement multi-period network design. While there are many other economic considerations that could be modeled, these provide insight into the fundamental question of how network augmentation functions in a multi-period approach.

3.3.2 ILP Formulation for Multi-Period Capacity Allocation with Topology Augmentation

In order to understand and model the optimal evolution of network topology and capacity allocation over time, a multi-period formulation was developed that enables arc-flow survivability models to be evaluated over multiple time periods.

The model fragment described below contains the constraints necessary to optimize survivable network designs by breaking apart constraints on link establishment and capacity allocation from the survivability and traffic routing constraints. While not covered in this work, the ILP structure presented here makes it ideal to utilize decomposition methods for optimization [96].

In addition to the notation used above, we use the following new notation:

New Sets:

T is the set of time periods.

T_t is the set of time periods up and including to $t \in T$.

New Parameters:

V_t^s is the cost of implementing span s at time t.

C_t^s is the capacity unit cost at time t for span s.

d_t^r is the demand volume at time t for demand $\boldsymbol{r} \in \boldsymbol{D}$.

M is a sufficiently large number (in our case, M=$\sum_{\forall r \in D} d_r^t$).

New Variables:

α_t^s is the capacity on span $s \in S$ at time t.

ω_t^s is the working capacity on span $s \in S$ at time t.

β_t^s is the spare capacity on span $s \in S$ at time t.

y_t^s is 1 if span s changes to > 0 from 0 at step t.

$\boldsymbol{v_t^s}$ is the cost of implementing span s at time t .

$\boldsymbol{c_t^s}$ is the capacity cost at time t for span s.

The multi-period DSP capacity allocation and topology augmentation ILP is as follows:

Minimize

$$\sum_{s \in S, t \in T} c_t^s + v_t^s \tag{21}$$

Subject to:

$$\sum_{t_c \in T_t} (\alpha_{t_c}^s - M \times y_{t_c}^s) \leq 0 \qquad \forall s \in S, \forall t \in T \tag{22}$$

$$v_t^s \geq y_t^s \times V_t^s \qquad \forall s \in S, \forall t \in T \tag{23}$$

$$c_t^s \geq C_t^s \times (\alpha_t^s - \alpha_{t-1}^s) \qquad \forall s \in S, \forall t \in T, \tag{24}$$

$$\alpha_t^s \geq \alpha_{(t-1)}^s \qquad \forall s \in S, \forall t \in T \tag{25}$$

$$\alpha_t^s \geq \omega_t^s + \beta_t^s \qquad \forall s \in S, \forall t \in T \tag{26}$$

The objective function (21) is the sum of capacity and span instantiation costs for all time periods. Because multiple time periods are involved, there are two options to be able to calculate discounted costs brought to a common time frame. The first option would be to multiply c_t^s and v_t^s by a discount ratio that has been pre-computed for each time step, or costs C_t^s and V_t^s could be pre-calculated to the appropriate discount rate. We chose the latter in this implementation. Equation (22) is used to set y_t^s to be 1 if there is any capacity assigned to span s in the current time period, and none before. In this equation, if any capacity has been assigned in a previous time period t_i, then the corresponding constraint would require $y_{t_i}^s$ to be 1. Now in a later time period, t_j, the indicator $y_{t_j}^s$ would not have to be set to 1, as $y_{t_i}^s$ would be sufficient to keep the left hand side of this constraint below 0. Equation (23) is used to calculate the cost of implementing a given span. Equation (24) calculates the incremental capacity costs for each span. Equation (25) enforces the idea that a span's capacity cannot be reduced in the future. Lastly, if the survivability formulation defines working and backup capacity separately based on the demand profile of the network (such as the SBPP formulation presented above), then the constraints in (26) set the span capacity to be at least the sum of the working and backup capacity assigned to the span.

If there is current infrastructure, then the initial capacities ($\alpha_{t_{-1}}^s$) and instantiation indicators ($y_{t_{-1}}^s$) can be set accordingly.

These multi-period constraints can also be augmented to include economies of scale, or modularity. We refer to the cost steps as economies of scale, as they represent reductions in per unit cost based on larger capacity volumes. These constraints were found to significantly increase computational complexity, but are included for completeness in this discussion, as modularity concerns are an important factor in network design.

These economies of scale can be modeled in a number of different ways. We have chosen to implement them as different technologies for different capacity intervals, with the implication that if the economies of scale interval changes on a given span (or at an overall network level depending on the modeling assumptions) at a given time period, costs are incurred for the total capacity (rather than the incremental capacity between time periods). In this formulation, we have not assumed a cost for a given technology or economies of scale interval, but rather, set a minimum capacity required. We also made the assumption that capacity increments between minimums are unimodular. This is a significant assumption, but could be modified to represent any level of modularity [80]. This implementation also assumes that the transmission technologies for each step of the economies of scale are compatible.

New Sets:

E is the set of possible economies of scale, ordered in increasing scale.

New Parameters:

$V_t^{s,m}$ is the cost of implementing span s at time t using economies of scale m.

$C_t^{s,m}$ is the capacity unit cost for span s at time t using economies of scale m.

E_m is the minimum capacity required for economies of scale level m.

New Variables:

$e_t^{s,m}$ is 1 if economies of scale of level m is used for span s at time t, and 0 otherwise.

$y_t^{s,m}$ is 1 if economies of scale of level m is used for the first time on span s at time t and 0 otherwise.

Additional Constraints:

$$\sum_{m \in E} E_{m+1} \times (e_t^{s,m}) \geq \alpha_t^s \qquad \forall s \in S, \forall t \in T \qquad (27)$$

$$\sum_{m \in E} E_m \times (e_t^{s,m}) \leq \alpha_t^s \qquad \forall s \in S, \forall t \in T \qquad (28)$$

$$\sum_{t_c \in T_t} (e_{t_c}^{s,m} - M \times y_{t_c}^{s,m}) \leq 0 \qquad \forall s \in S, \forall t \in T \qquad (29)$$

$$-(c_t^s - C_t^{s,m} \times \alpha_t^s) \leq M \times (1 - y_t^{s,m}) \forall s \in S, \forall t \in T, \forall m \in E \quad (30)$$

$$-\left(c_t^s - C_t^{s,m} \times (\alpha_t^s - \alpha_{t-1}^s)\right) \leq M \times (1 - e_t^{s,m}) \forall s \in S, \forall t \in T, \forall m \in E \quad (31)$$

These constraints, in combination, enable the model to select different per unit costs based on the overall capacity for each step. Equations (27) and (28) enforce the selection of economies of scale stepped capacity range to encompass the capacity on span s at time t. Note that α_t^s could be increased from the minimum required to meet total demand requirements in order to take advantage of the savings provided through the next step in the economies of scale. Equation (27) says that the capacity of span s at time t (α_t^s) cannot be more than the minimum capacity of the next step in the economies of scale set. This constraint does not necessarily need to be enforced, as materially, there are a number of different factors that could negate its motivation. Conversely, equation (28) says that economies of scale greater than the capacity selected for the span cannot be used. Span capacity could be increased to reach the next cost profile step if it proves to be beneficial overall.

Equation (29) is similar to equation (22), in that it is used to set the binary variable $y_t^{s,m}$ to 1 if it is the first time that economies of scale step m is used for the given span. This variable is then used to enforce span capacity costs for that span in the given period to be based on the total capacity in equation (30), and not just the incremental capacity from the previous time period, equation (31). If moving between steps in the economies of scale curve doesn't require incurring the total capacity cost of that span, then equations (29) and (30) can be removed from the ILP formulation.

As mentioned, economies of scale considerations in a multi-period model increase solution time dramatically. They provide a counterbalance to the effect of the time value of money, and provide a mechanism to understand the transition points between transmission technologies against growth prospects of network traffic.

3.3.3 Survivability with Multi-Period Capacity Allocation and Topology Augmentation

Presented above were the formulations for DSP and SBPP in single period form. For completeness, this section describes the explicit changes required to integrate with the multi-period constraints. Most variables, parameters, and constraint indices require an additional index representing a time period.

3.3.3.1 DSP with Multi-Period Network Augmentation

In order to include the DSP survivability mechanism into a multi-period model, we start with the DSP model presented in equations (2) through (8), modifying it by adding a time dimension to all relevant of variables and constraint indexes. We then added an additional constraint to convert node indexed flows into span indexed flows (32).

Variables $\omega_{i,j}^{r}$, $f_{i,j}^{r}$ and $f_{i,j}$ all require a time dimension, and become $\omega_{i,j}^{r,t}$, $f_{i,j}^{r,t}$ and $f_{i,j}^{t}$ respectively, in all the constraints in which they appear. Furthermore, the existing constraints all have an additional time index ($t \in T$) added.

The ILP to implement MPNA with a DSP survivability scheme uses equations (21) through (26) (or (31) if modularity is included) plus the equations below, being modified with time indices.

New Variables:

$\boldsymbol{\omega_{i,j}^{r,t}} \geq \mathbf{0}$ is the traffic flow from node *i* to node *j* for time period *t*. Note that $i \,\square\, N$, but $j \in A_i$

$\boldsymbol{f_{i,j}^{r,t}} \in \{\mathbf{0}, \mathbf{1}\}$ is a binary variable indicating whether any capacity is allocated on the span between node *i* and node *j* for time period *t*.

Additional Constraint:

$$\sum_{r\in D} \omega_{i,t}^{r,t} + \omega_{j,i}^{r,t} = \omega_t^s \qquad \forall i \in N, j \in A_i, \forall s \in S_i \cap S_j, t \in T \qquad (32)$$

3.3.3.2 SBPP with Multi-Period Network Augmentation

The formulation for SBPP defined above combined with the multi-period constraints create a formulation for multi-period shared backup path protection. This formulation is a direct combination of constraints (10) to (20) with the multi-period constraints (22) through (26) (or (31) if economies of scale are included). This formulation is computationally expensive and contains many binary variables, making it difficult to optimize. For example, $\omega_{i,j}^r$ and $\beta_{i,j}^r$ would become $\omega_{i,j}^{r,t}$ and $\beta_{i,j}^{r,t}$ and have a count of $2 \times s \times n \times (n-1) \times t$ each (where n is the node count and s is the span count). For an 8-node network, this results in 3192 variables for working path design and the same for the backup paths.

3.4 Experimental Methods and Computational Considerations

The SBPP and DSP models presented above were implemented and run on a sample 8-node network (Figure 10). We used three time periods as the planning horizon in this work. In general, it was found that the models were quite computationally expensive, on average taking hours to days to run. DSP models could solve to a 23% optimality gap in three and a half days for a 15-node network. However, the SBPP model was run for over 50 days on a 10-node network with an optimality gap of 36%. Due to the challenges of solving the SBPP model, an area for future study is to use path restoration [7] in multi-period network design. The computational cost of solving more complex networks was a limiting factor, and the reason one small network was used in this study. This is consistent with other multi-period network planning work [21], [34], [38], [92]

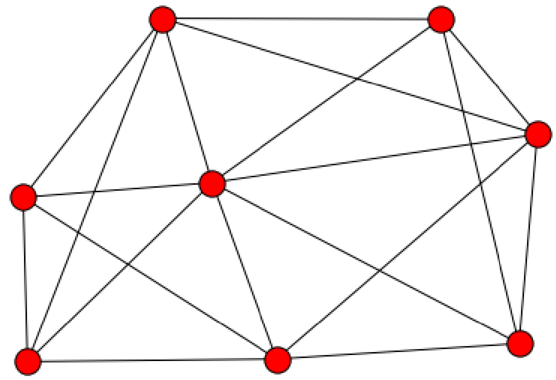


Figure 10 – 8-node network used as a base set of spans for topology design

The network used in this work was created using a pseudo-random distribution of nodes with little clustering. Each node was connected to at least the closest four nodes in the network. A few nodes, such as in the center of the network, happen to be in the closest four nodes of all the others, and connected to more than just four nodes. This network formed the set of spans that the topology design constraints in the ILP had to choose from.

The DSP MPNA model was also run on 10, 12, and 15 node networks, with the results presented for reference on the design characteristics as the networks grow. These reference networks were designed using the same nodal and demand volume methodology employed in the 8 node network.

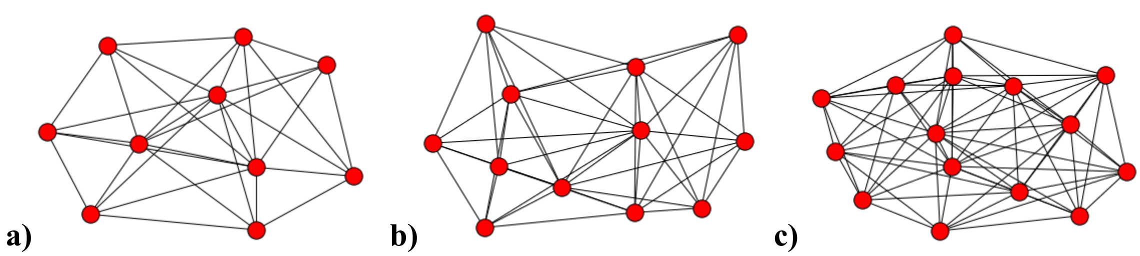


Figure 11 – Topology designs for the 10 node network (a), 12 node network (b), and the 15 node network (c)

All models were solved using Gurobi 5.6 on a 12 core 2.5 GHz computer with 96 GB of RAM. As an aside, it was found that the memory usage in these models did not exceed 10 GB, and were largely CPU limited. The solution times for these problems range from minutes to days.

In this work we made a number of assumptions around the characteristics of the multi-period context.

The demand growth rate was assumed to double on average every time period. The initial demand was normally distributed among the node pairs of the graph to values between 1 and 10. These demands were then increased over the three time periods such that demand doubled every period on average. This was accomplished by randomly distributing the increase in capacity with a mean of 2 and a standard deviation of 0.5. The average increase in capacity for the second time period was 111% and the third time period was 156%.

As is standard in many business practices, we used a discount rate of 10%, although, it may vary between industries and organizations. This discount rate represents the cost of capital, and is applied to all future costs in the model.

The assumption that per-unit costs of adding capacity to the network decreased over time has a similar effect on capacity costs as the discount rate. For capacity costs, the discount rate and the per-unit cost decrease were combined for an effective discount rate of 25%. We also investigated using a higher discount rate but did not observe an effect on topology design or capacity placement in the range of values tested. Theoretically, a large deviation between the effective discount rate for capacity and the discount rate applied to span establishment costs would distort the ratio of the span establishment costs versus the capacity costs. This makes the use of extra capacity on existing spans more favourable, rather than augmenting the network to make it more efficient at later time periods. However, this larger discount rate would have to be outside of what would reasonably exist.

The factor that was examined in this work was the span establishment cost multiplier. This ratio was incremented on a logarithmic scale from 10 to 1000.

3.5 Results and Discussion

This work examined and isolated the effects on the time value of money and the decreasing trend in unit capacity cost from span establishment cost ratios. As discussed above, several test case scenarios were run, with various discounting effects and establishment cost multipliers. More specifically, we ran the same models with two different discount rates in order to determine if discount rates affected the solution. We did not find any variation between a discount/effective discount rate set of 10 and 25 with a rate set of 15 and 40. While these discount rates could be carried further, these were chosen to be opposite ends of a reasonable range. We find that the span establishment cost ratio played a much bigger role in influencing network topology. Although discount rate could encourage capacity to be added in the future, we did not see evidence that this affect was strong enough to delay span establishment.

The numerical results from the simulations are presented in Table 2 and Table 3. Table 2 summarizes the normalized results from the optimizations of an 8-node network with a span establishment cost ratio of 10, 100 and 1000 for the DSP-MPNA survivability model. In this table, we can see how the establishment costs relate to the capacity costs, where the establishment costs in the 1000 times span establishment cost ratio network overwhelm the capacity costs, resulting in a ring configuration for the first time period, and nearly the same for the second and third time periods.

The results from the SBPP-MPNA optimized designs did not augment the topologies over time. These designs were also more sparse when compared to the topologies optimized for the equivalent DSP-MPNA networks. Even with these sparse networks, the capacity efficiency of SBPP is obvious. This indicates that what influencing factor over the augmentation of the networks in the DSP-MPNA results was not the growth of network demands. If it was, then we would expect to observe topology augmentation with SBPP-MPNA designed networks as well. If demand growth was the driver for topology augmentation in this case (and since SBPP can more efficiently utilize network diversity), the extra capacity should have encouraged spans to be added at later time periods. Since they weren't for SBPP-MPNA network designs, there was another factor affecting topology augmentation over time in the DSP-MPNA networks.

The fact that topology and topology growth changes significantly between the two survivability strategies is significant, because this implies that topology designs and strategies for network augmentation should include survivability strategy considerations.

Imp. Ratio	Period	Span Disc.	Cap Disc.	Estab. Costs	Cap. Costs	Total Costs	Span Count	Total Cap.
10	p1	10	25	0.434	1.285	1.718	19	447
10	p2	10	25	0	1.055	1.055	19	915
10	p3	10	25	0	2.109	2.109	19	2050
100	p1	10	25	2.052	1.745	3.798	11	724
100	p2	10	25	0.682	1.000	1.682	14	1168
100	p3	10	25	0	2.408	2.408	14	2614
1000	p1	10	25	14.105	3.019	17.123	8	1328
1000	p2	10	25	1.630	1.448	3.078	9	2108
1000	p3	10	25	0	3.884	3.884	9	4748

Table 2 – Normalized cost breakdown for the DSP-MPNA results. Normalization was against the minimum span cost for any time period and any of the establishment cost multipliers.

Imp. Ratio	Period	Span Disc.	Cap Disc.	Estab. Costs	Cap. Costs	Total Costs	Span Count	Total Cap.
10	p1	10	25	0.530	1.174	1.704	15	338
10	p2	10	25	0	1	1	15	700
10	p3	10	25	0	2.100	2.100	15	1627
100	p1	10	25	3.508	1.434	4.942	11	447
100	p2	10	25	0	1.248	1.248	11	929
100	p3	10	25	0	2.609	2.609	11	2184
1000	p1	10	25	24.110	2.323	26.433	8	767
1000	p2	10	25	0	1.954	1.954	8	1576
1000	p3	10	25	0	3.799	3.799	8	3532

Table 3 – SBPP-MPNA results from 8-node network optimization with an implementation factor of 10, 100 and 1000

To illustrate the benefits that the additional spans brought the network, Figure 12 presents the average capacity increases between time periods for SBPP designs (red left column), DSP networks for time periods where topology did not change (green middle column), and time periods where topology did change (blue right column).

The fact that the SBPP-MPNA designs did not augment topology over time begs the question as to what conditions would cause the SBPP-MPNA network designs to augment topology over time. Looking at overall span counts over the various span establishment cost multipliers, it is obvious that this is a driver on topology. As mentioned earlier, increasing the difference in how much capacity costs are discounted compared to the costs of establishing new spans in the network effectively increases the span establishment cost multiplier. Conversely there is a point when the total volume of demand is great enough that the reduction in spare capacity requirements from adding a span to a network would overcome the effects of the establishment costs of that span.

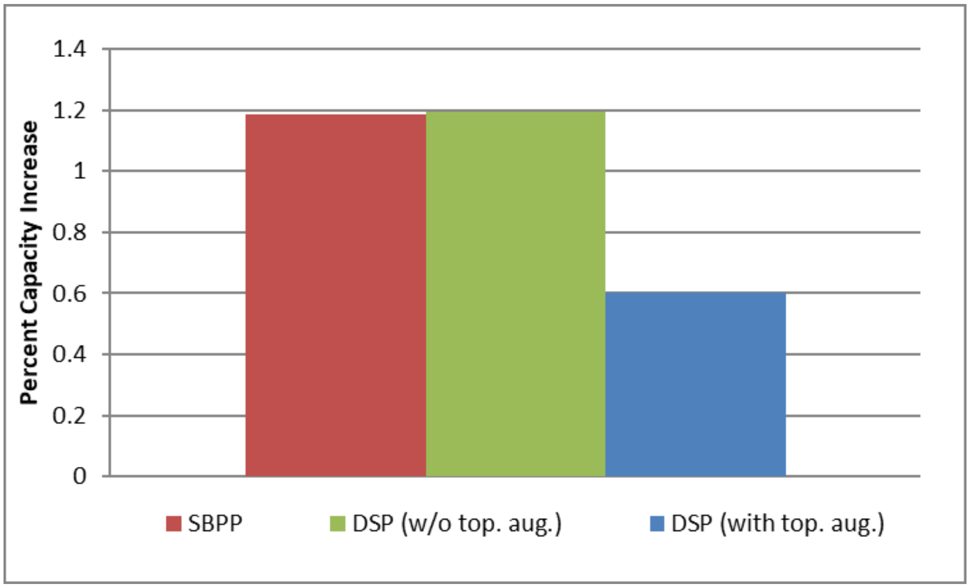


Figure 12 – Average relative increase in capacity compared with previous time period for SBPP networks, DSP networks for time periods where there were no topology augmentations, and DSP networks where there were topology augmentations

Another set of SBPP-MPNA network designs were run, this time with a mean capacity increase of quadruple the previous time period demand. The same probability of increase was used in these designs; however the mean capacity increase was adjusted from two to four times. The results are documented in Table 4.

With capacity quadrupling every time period, the SBPP-MPNA network designs did augment topology for each time period except when span establishment costs were set to 1000 times span capacity costs.

Imp. Ratio	Period	Span Disc.	Cap Disc.	Estab. Costs	Cap. Costs	Total Costs	Span Count	Total Cap.
10	p1	10	25	0.576	1.000	1.576	17	312
10	p2	10	25	0.035	1.632	1.667	18	942
10	p3	10	25	0.000	4.661	4.661	18	3184
100	p1	10	25	3.141	1.284	4.425	11	447
100	p2	10	25	0.646	1.837	2.483	13	1195
100	p3	10	25	0.361	4.901	5.262	14	3678
1000	p1	10	25	21.587	2.080	23.667	8	767
1000	p2	10	25	0.000	3.422	3.422	8	2349
1000	p3	10	25	0.000	9.090	9.090	8	7596

Table 4 – SBPP-MPNA results from 8-node network optimization with an implementation factor of 10, 100 and 1000 with an average demand quadrupling each time period

In Figure 13 through Figure 16 we present the topologies optimized with each span establishment cost multiplier at each period. With only minor differences across time and even between survivability schemes, it is obvious that the driving factor in these designs, as in actual networks, is the difference in cost between adding capacity to current spans, and augmenting the network.

From the simulated designs here, network topologies are most affected by capacity growth when the cost of establishing a span is around 100 times the cost of adding a unit of capacity to the span. When span establishment costs approach 10 times capacity unit costs, there is significant incentive to have a high span count in the topology. The savings from better capacity sharing, and even shorter paths, quickly outweighs the costs of adding spans. Both SBPP and DSP designs used all, or nearly all, of the available spans.

When the cost of establishing spans approaches 1000 times the unit cost of adding capacity, the cost reduction from capacity sharing does not reach the required level to offset the massive costs of adding spans to the network.

From the results run here, the span establishment cost multiplier of 100 was close to the balancing point between the benefits of adding more capacity or more efficiently routing traffic using additional spans. This meant that the effects of additional capacity demands more easily met the threshold of the cost of augmenting the network topology.

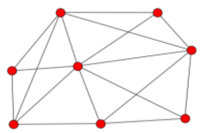
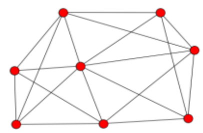
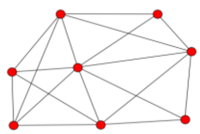
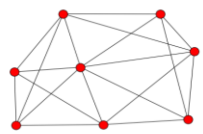
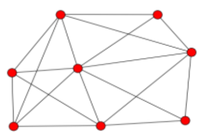
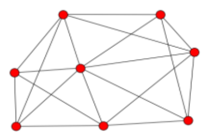
Scenario	SBPP-MPNA Design	DSP-MPNA Design
10-p1		
10-p2		
10-p3		

Figure 13 – Span designs from DSP-MPNA when capacity doubles each time period, and SBPP-MPNA when capacity quadruples for a span establishment cost multiplier of 10

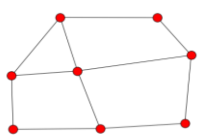
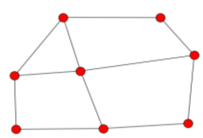
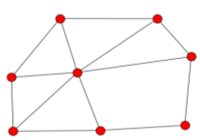
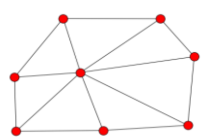
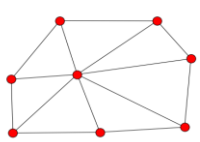
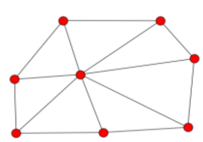
Scenario	SBPP-MPNA Design	DSP-MPNA Design
100-p1		
100-p2		
100-p3		

Figure 14 - Span designs from DSP-MPNA when capacity doubles each time period, and SBPP-MPNA when capacity quadruples for a span establishment cost multiplier of 100

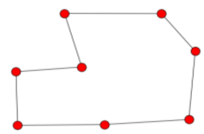
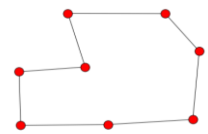
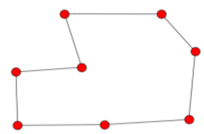
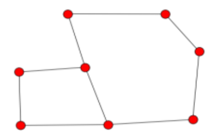
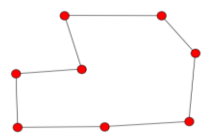
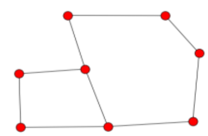
Scenario	SBPP-MPNA Design	DSP-MPNA Design
1000-p1		
1000-p2		
1000-p3		

Figure 15 - Span designs from DSP-MPNA when capacity doubles each time period, and SBPP-MPNA when capacity quadruples for a span establishment cost multiplier of 1000

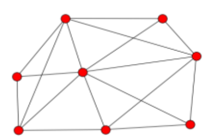
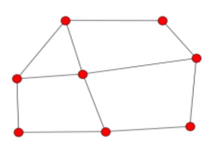
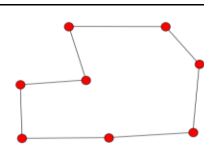
Scenario	SBPP-MPNA Design
10	
100	
1000	

Figure 16 - Span designs from SBPP-MPNA when capacity doubles each time period for span establishment cost multiplier

The costs and span counts generally behaved as would be expected. The networks with a higher implementation ratio had fewer spans and a higher total capacity. The total capacity was higher for networks with higher span implementation factors because in general, paths between demands are longer and there is less opportunity to leverage DSP's capacity sharing.

One aspect that wasn't expected was the change in combined (working and spare) capacity (excluding topology) costs over time. We expected capacity costs to remain relatively flat, or decrease slightly (due to the effect of time value of money) for the second time period. However, as we can see in Figure 17, there was a substantial reduction in capacity cost in period 2 relative to the capacity cost in period 1. To take a closer look at this, we can examine the capacity allocated in period 1 of the simulation with an implementation ratio of 1000 and the second period of the same network (Figure 15). The first period topology has a total of 1328 units of capacity to route 166 units of demand, while the second period had 2108 units of capacity for 338 units of demand. This means that routing in the second period required 172 more units of demand but required only 780 more units of capacity. By adding span S9, the capacity to demand ratio decreases from 8 to 4.5. This drastic decrease in capacity required per demand offset the cost of implementing the span. This reduction in required capacity, coupled with the time value of money, caused the drop in capacity costs between the first and second periods.

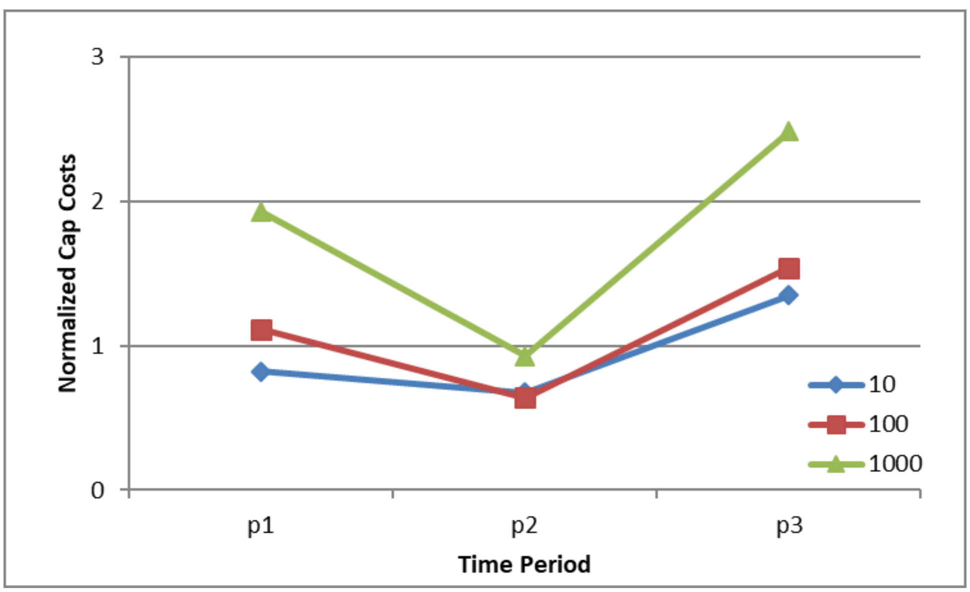


Figure 17 – Capacity costs over time for each network for DSP-MPNA network designs

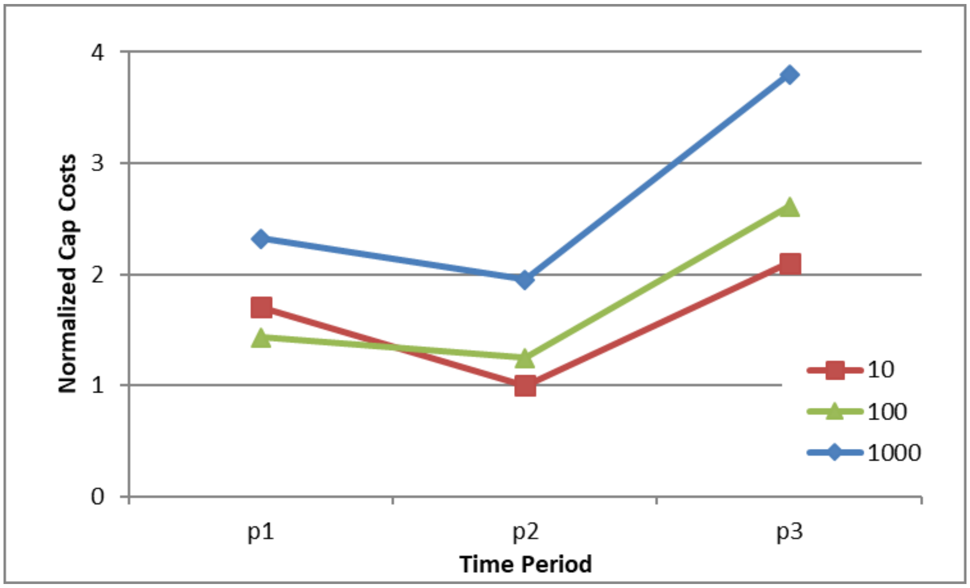


Figure 18 – Capacity costs over time for each network for SBPP-MPNA network designs with the capacity growth rate set to doubling

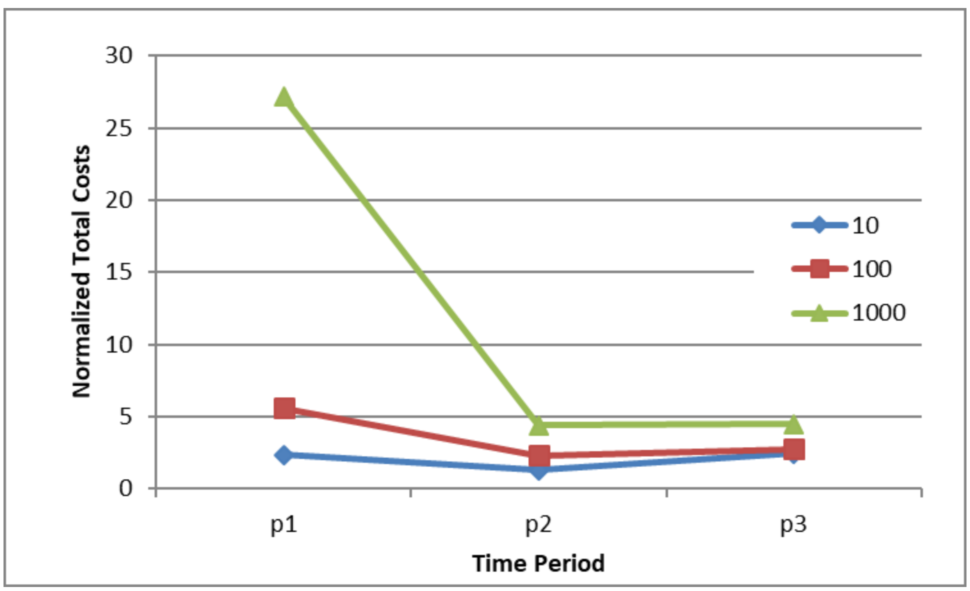


Figure 19 – Normalized total costs over time for the DSP-MPNA network designs

While these results were run using a small 8 node network, the DSP model was also run on a representative 10, 12, and 15 node networks (network designs using the SBPP model did not converge on a confidence interval of less than 36% in one month). The capacity costs for each of these network designs (Figure 20 - Figure 22) exhibited a similar pattern where the costs decreased in the second time interval before again increasing in the third.

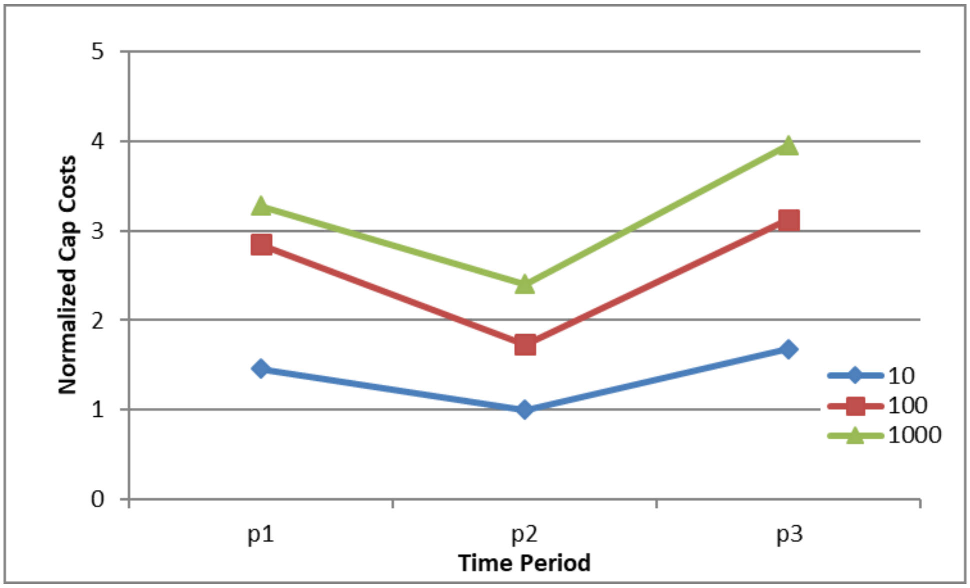


Figure 20 – Capacity costs over time for each network for 10 node DSP-MPNA network designs

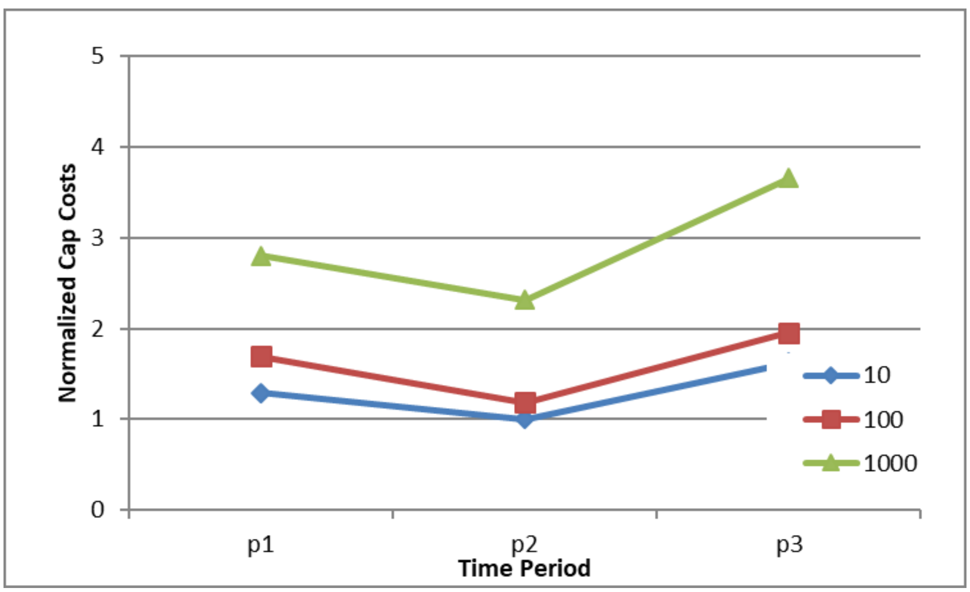


Figure 21 – Capacity costs over time for each network for 12 node DSP-MPNA network designs

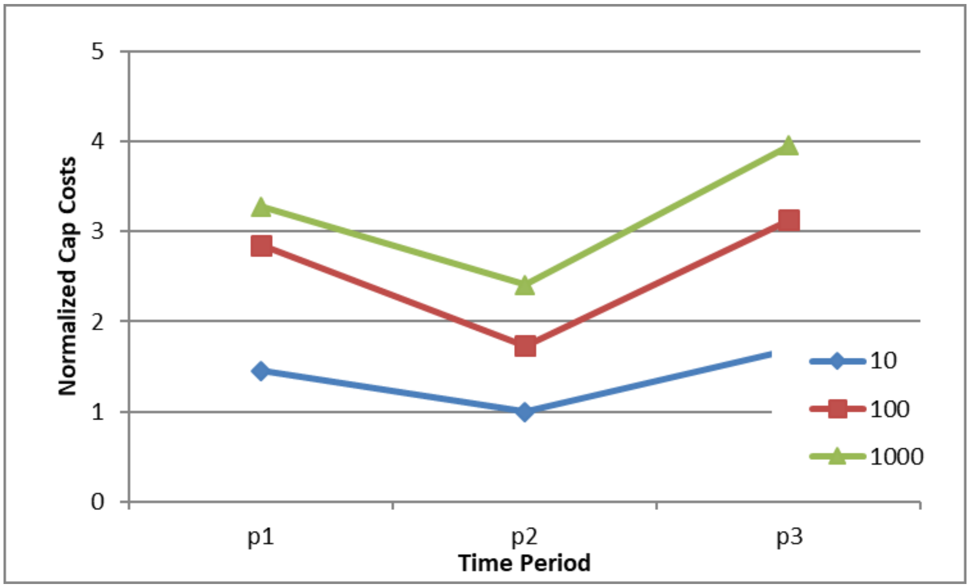


Figure 22 – Capacity costs over time for each network for 15 node DSP-MPNA network designs

We should also note that modularity and economy-of-scale factors were investigated for this work but were omitted for time considerations. In order to implement economies of scale into the model, the savings must be made piecewise linear. We found that the granularity of the piecewise linear function representing economies of scale was insufficient with the computing power available. The resulting ILP solution would take months per design using the reasonable powered computational server listed above.

3.6 Concluding Discussion

Network topology design and augmentation over time, with capacity growth and the time value of money, provides insight into how to adapt a network. While multi-period planning is difficult, and optimization is computationally expensive, having an optimal reference point (under the described assumptions) provides insight into how to build and extend network infrastructure in an efficient manner.

This work presented a novel ILP formulation that optimizes DSP and SBPP networks over multiple time frames to provide a globally optimal capacity and topology augmentation schedule. By varying the span establishment cost multiplier the obvious effect of having sparser networks was observed, however, when compared across the two survivability mechanisms, the results were not so obvious. There was an observed trade-off between the cost of adding capacity, and the cost of instantiating new spans. This trade-off is difficult to assess intuitively, but has a significant impact on the capacity efficiency of the network. Additionally, DSP networks, which do not utilize spare capacity as efficiently as SBPP networks, generally had network designs with higher span counts. It could be initially assumed that because SBPP can better utilize spare capacity in a more populated network, it would result in network topologies with higher span counts than DSP. Based on the results found here, it appears that more effective use of spare capacity reduced the benefits of having a better-connected network.

By better understanding the trade-offs in the selection of survivability schemes and the factors affecting network growth, better decisions can be made in expanding current networks, as well as bringing networks to underserved regions, where the economics are already challenging, and efficient planning is a necessity.

Chapter 4. A Novel Long-Term Telecommunication Network Planning Framework

Building on the enabling capabilities of multi-period network augmentation, this next paper articulates the core concept from the research. It is a network planning framework that reaches out beyond the traffic demand, survivability requirements and cost. It articulates the distinct sectors of society that uniquely draw value from telecommunications. These sectors are important in that they each have unique value drivers and impetus to utilize telecommunication technologies. The framework laid out in this paper formulates the large and varied factors that influence a sectors ability to pay for, use, and derive value from the applications that rely on telecommunication network. Lastly the decision vectors that come out of the network planning process, including but not limited to technical designs. These decision vectors form the basis of the feedback loop that the multi-period design optimization enables.

B. Todd and J. Doucette, "A Novel Long Term Telecommunication Network Planning Framework," *Journal of Network and Systems Management*, vol. 25, no. 1, pp. 47–82, Jan. 2017.

4.1 Introduction

Telecommunication network planning and design is a difficult task, especially when considering long term planning in the infrastructure and backbone network domains, which are relatively immutable. Time horizons are long, technology changes rapidly, and customer demand is difficult to predict. In the late 1990's the paradigm was to build as much as you can, because demand was supposed to be constantly greater than supply (due to technologies like video on demand) [97]. However, that era ended, at least in the United States, in a bang with the collapse of the "dot-com bubble" in the early 2000's, taking with it major corporations [97]. Since then there has been little need for network build-outs considering the vast amount of excess capacity that could still be utilized. Today, networks are growing again, spurred on by changes in customer usage patterns and wireless access technologies (Wi-Fi, LTE, etc.) [98]. There are also regions that are underserved by current infrastructure, such as rural and remote locations, even in developed countries. There have been a number of initiatives that have aimed at reducing the so-called digital divide between urban and rural locations, as this (lack of network infrastructure) is seen as a significant economic retardant for these areas [99]. The challenge in addressing these new opportunities is to not repeat the mistakes of the past, but rather to better align the network planning and design context with the social and economic paradigms that exist within the regions being targeted.

With wireless technologies, and remote data links through satellites, the cost of delivering network access is changing rapidly. There is significant potential to be able to deliver internet connectivity to many places that are hard to reach (or impossible through conventional methods). A challenge with providing access to these locations is to ensure that the users are capable of using the network, and have the motivation to do so. It is not about building the network and everyone will jump on board. The challenges and opportunities in many underserved regions require a broader approach to network build outs.

It can be argued that network planning and construction should not concern itself with social and economic concerns, that it is an enabling technology; with it, people will naturally find uses and application for the network's capabilities. With that point of view, demand forecasting should account for the socio-economic dynamics, but these uncertainties should not extend further into the planning process. This framework can be sustained when demand is growing rapidly, and the potential customer base is dynamic enough to incorporate new network technologies, or when the financial backers of the network expansion can withstand a significant period of underutilization. However, in general, this framework can lead to a significant misallocation of resources, as happened in the early 2000's in the US, and many other locations.

This paper presents a new planning framework that expands the common planning context to account for the various influencing factors and decision vectors that will better align network planning with the socio-economic conditions in which it will exist. This framework does not address specific design or planning techniques, but rather, serves as a tool to systematically incorporate and integrate multiple inputs and decision vectors.

Current practices and literature are review in section 2 including other planning frameworks and demand forecasting in the telecommunications environment. Section 3 presents the new planning framework, and goes into some depth in the various components of this framework. Section 5 provides a discussion of some current research areas, and how they fit into this framework, and an analysis of what the authors see as the portions of the framework that currently are poorly understood. Lastly, section 5 includes a decision model based on the network planning framework presented.

4.2 Background

In order to build the context in which a new network planning paradigm would exist, an understanding of current network planning activities must be established, as well as some background on how to assess and quantify the environment the network would exist in, and its predicted uses. Network planning activities include demand forecasting and capacity allocation. This view of network planning has been well established [25], [100]. Assessing the environment in which the network would exist and its predicted uses is more difficult. In a greenfield situation, or a situation where the new network is a step change from what is already in place, many forecasting models are not adequate, as historical trends are not established. The main tools that will be used to assess the impact of the context and purpose of the network will be innovation diffusion concepts [13], [22]. Networks enable and enhance most aspects of society and therefore require users to be able to adopt and utilize many new innovations in order to leverage new opportunities that these networks enable. The purpose of the research is to better manage network deployment and expansion from the point of view of delivering value to the end network users. This background provides an overview of telecommunication network planning, demand forecasting as it relates to telecommunications, and a summary of the study of the diffusion of innovations as it relates to telecommunication network planning.

4.2.1 Network Planning

Network planning has been a topic of research for many years [23]. The planning process has generally been viewed as the translation of demands and technical capabilities into a proposed optimal network configuration. This planning process operates on three time scales, short, medium and long term, [9], and each time scale requires a different approach to network planning and management.

The problems being addressed in this work focus on long-term decisions, with some application to mid-term decisions. These decisions revolve around new capacity placement and physical network augmentation, as opposed to reconfiguration of existing networks. This network reconfiguration is in the domain of short- and mid-term planning, as it can be done more rapidly, and usually does not require provisioning additional infrastructure.

Figure 23 – The traditional network planning and design model.

The most common network planning paradigm today involves input of the current network conditions and demand forecasts. These are then used to arrive at an optimal or near optimal network capacity design (illustrated in Figure 23). This process can utilize a number of different design methodologies including ring-based schemes, and a broad range of mesh-based schemes (including hybrid schemes like *p*-cycles). The design parameters typically include *topology allocation* [82], *capacity placement* [41], [77], [101], and *path allocation* [102], [103].

Topology allocation involves the design and layout of the spans in a network. This can either be done in an augmentative fashion, or greenfield. Augmentative topology design is the addition of new spans to an already established network; these new spans could include new installs (i.e., trenching new fibre), or they could involve leasing capacity from other networks or carriers. Greenfield planning assumes only a set of network nodes and demands, with an optimization scheme instantiating spans in accordance to the protection scheme's requirements. Because installing entirely spans is a lengthy and costly process (the Alberta SuperNet took over 4 years to build and cost over $400 million, [24]), topology augmentation is most applicable to long-range planning scenarios if some underlying infrastructure can be utilized as a basis for the new network.

Network designs, whether involving topology layout, capacity placement, and/or path allocation can be derived optimally (or within a known degree of optimality) using *integer linear programming* (ILP) techniques [77], [85], [104], or using various near-optimal heuristics [7], [105]. The network design problems we consider here are NP-hard problems [7], and hence they are inherently not scalable. In practice, however, we can typically solve many such problems by limiting the problem in some fashion (i.e., considering only a subset of eligible paths, etc.), thereby obtaining in reasonable timeframes with computing resources generally available to even small research groups [106]. This limitation, however, does impact the confidence in the optimality of the solution, which is why we often resort to heuristic approaches. Meta heuristics such as genetic algorithms and tabu search are common [105], [107]–[109], but problem specific algorithms also find a niche [110].

In general, network planning can be thought of in the most general case as a process that takes a demand forecast (and perhaps some pre-existing network infrastructure) to produce an updated network design that can accommodate traffic forecasts in a manner that makes efficient use of capacity and provides some desired level or survivability, reliability, and/or availability. Most scholarly articles and books on the topic view network planning from a highly technical perspective, not taking into account the high degree of variability in the inputs to the planning process. Some works go into detailed technical aspects of network design with significant depth, but do not cover network design or planning in uncertain environments [25]. The assumption is that the traffic demands are known, and the problem is how to best allocate equipment and capacity to meet these demands. Some other planning models include some feedback in the network design process, as illustrated in Figure 24 [26], while others take planning system as a more or less linear process, as illustrated in Figure 25 [27].

It should be noted here that a significant amount of literature talking about network planning uses the term "Next Generation Network" [26]–[28]. What this term signifies is a shift in underlying network technology to serve demand from higher levels in a protocol agnostic manner [29]. For example, TELUS overhauled their core network a number of years ago to run entirely using IP, rather than a conglomerate of separate networks [30], [31]. This trend toward converged networks has a number of impacts on network planning, as the importance of the integrity and availability of these networks is even more significant [111].

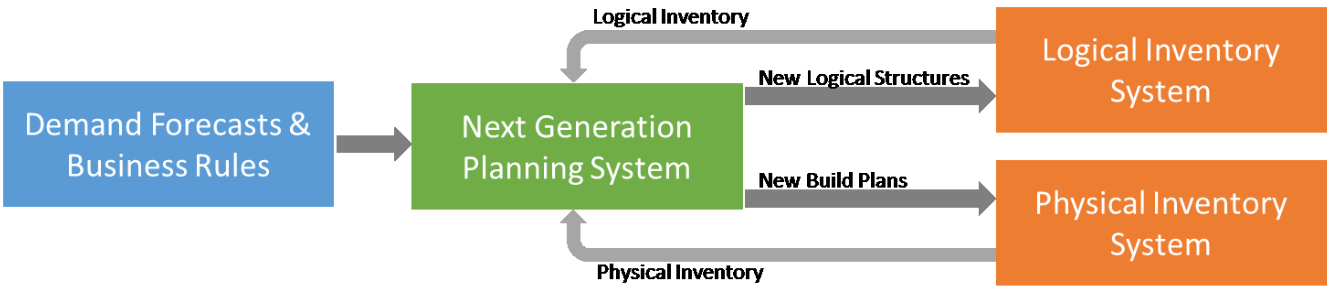


Figure 24 – M. H. M. Moh's network planning process with a feedback mechanism, [26]. This figure is a duplicate of figure 3.

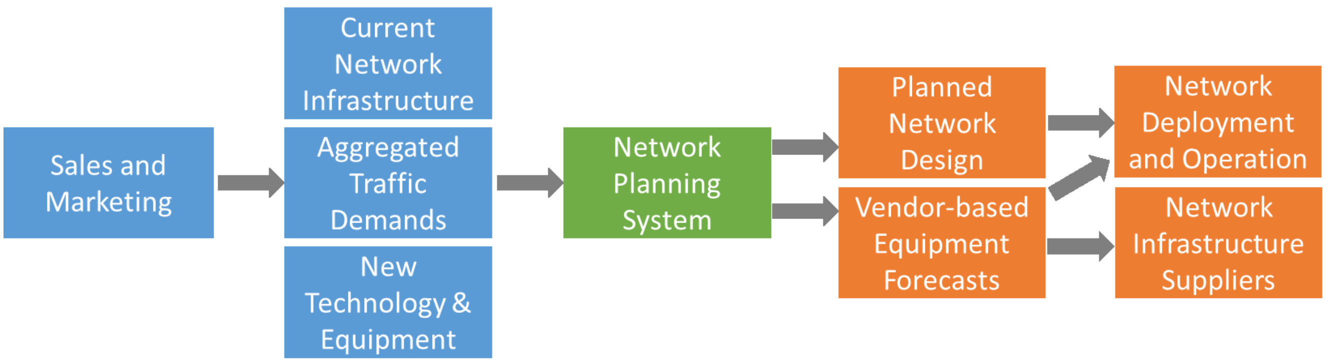


Figure 25 – S. K. Mohapatra's linear network planning process, [27]. This figure is a duplicate of figure 2.

Network planning has generally not concerned itself with the types of users that will utilize the network. The assumption is that the concerns of the users can be adequately captured in QoS and bandwidth characteristics. Some work has been done to better understand and classify the types of users of a network. In [32], users were broken down into three groups: domestic, business, and scientific, as shown in Figure 26. The usage scenarios for each of these groups were briefly discussed in the context of their usage characteristics. Although these users are identified, the need for "overarching roadmaps for future networks" [32] was identified, as these were not currently in place.

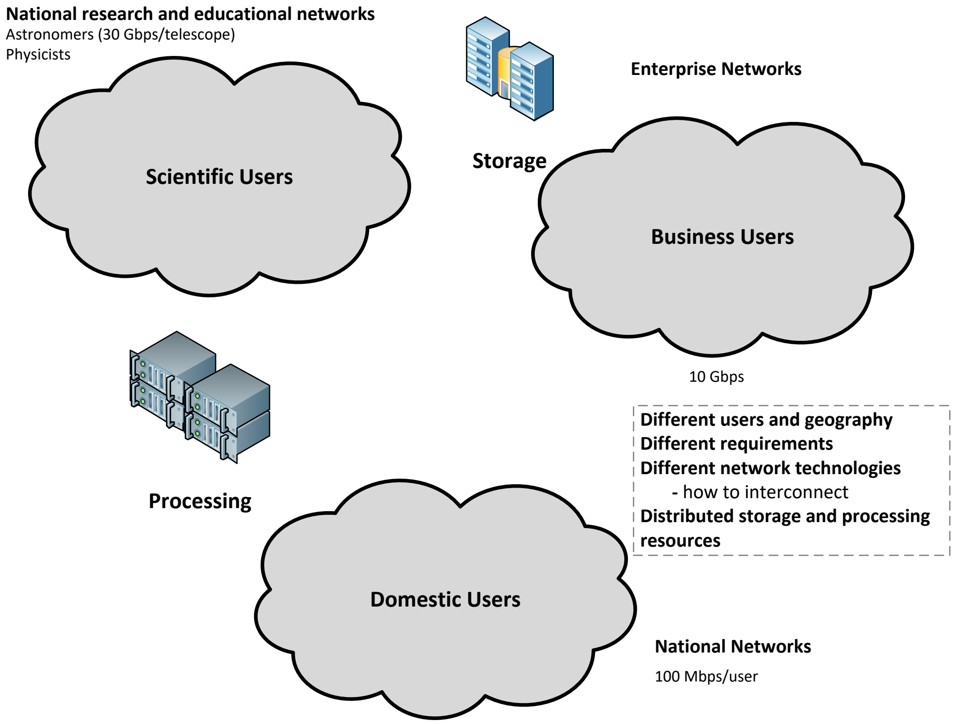


Figure 26 – An overview of types of network users as described in [32]. This figure is repeated from figure 4.

In addition to what was covered in [32], the category of machine to machine communications in the context of the internet of things has become a growing concern [112]. With the estimated pervasiveness and the corresponding quantity of network devices, the concepts behind the internet of things has the potential to alter how telecommunication networks are used, and what is required of them. The network devices typically involved with the internet of things are low power, low bandwidth, but typically require a significant set of access points. The required structure and characteristics of a network supporting the internet of things on a large scale is not well defined, as the applications could vary from customized consumer advertising to self driving cars, to telehealth. The concepts behind the internet of things are highlighted here because of the potential impact they have with the machine to machine structure of these supporting telecommunication networks

Whether it is the internet of things, social networking, mobile computing, or some other trend, a solid reliable network underpins it all. The challenge is to effectively and efficiently build out both the applications and the networks that support them, given the uncertain nature and high cost of delivery.

One last point to mention here is with regards to relatively new trends within literature. There has been some work done in the field of multi-period network design [20], [21], [33]–[39] and planning with uncertainty [33], [40], [42]. These approaches effectively serve to expand the network planning paradigm, but they stay within the technical domain. The work in [41] provides a good overview of the application of demand uncertainty, and presents a two-part stochastic programming model for span restorable networks. The work in [33] and [40] even combines multi-period planning with uncertainty for a single link, optimizing the allocation of capacity over time. However, these approaches, while insightful, do not scale well within the general network planning problem.

4.2.2 Traffic Demand Forecasting

The purpose of this work is not to develop methods for network demand forecasts, but rather to better align network designs with the factors that affect demands, and hence drive demand forecasts. Included here is a review of demand forecasting and the primary techniques used in different types of network planning. Network traffic forecasting is a difficult task. Even in environments with significant traffic history, shifts in application usage can happen quickly [113]–[115]. Network traffic forecasting has been discussed in the literature [43], and can be seen as a subset of the larger field of customer demand forecasting in general. Early papers on demand forecasting focused on predicting demand in an established market with little competition [45], [116], [117]. While some of these premises were valid in the 1980's, many do no hold today [43], [118], [119]. With the rise of data communications to a ubiquitous utility-type service and the privatization of the industry, traffic forecasting in telecommunication networks has become significantly more complex. In general, forecasting telecommunication traffic demand has been broken into two general categories. The first category is an extension of early forecasting techniques, focusing on econometric techniques, [44] such as linear regression analysis. The other category of forecasting emphasizes techniques that are appropriate for scenarios with significantly more variability and/or lack of historical data. The former is applicable when forecasting demand in established markets with established services. The latter is used when expanding into new regions and/or developing new service offerings. Since technology changes (at a network level and at a service level) can significantly alter traffic characteristics, it has been suggested that long term forecasting most often uses new product/service forecasting techniques, even in established markets [44].

There are a number of econometric forecasting models that have been used in network demand forecasts [43], that attempt to model factors affecting future demand. Aggregate models utilize time series data to interpolate future demand. These models incorporate price elasticity and competition effects along with time series data to estimate overall network traffic, and future profits from the network. Price elasticity is a measure of how sensitive customers are to price changes for the services supplied by the network operator. Price elasticities are highly dependent on the region being served [43], and provide a way to balance pricing and demand. The effects of competition in demand forecasts using an aggregate approach utilize concepts such as customer churn to estimate future demand. In general, aggregate models may be used in forecasting demands in established markets, but they do not adapt well to the changing nature of communication technologies in our society [43].

Another approach to demand forecasting in established markets is to model services separately using surveys, application specific historical data, and a customer choice model. Customer choice models, illustrated in Figure 27, look at three factors affecting consumer adoption decisions of a specific service. These factors are the understanding, utility, and acceptability of the product [45]. In turn, network operators can influence these factors through marketing, product quality, and price [45]. Forecasting models based on disaggregate service forecasts require extensive historical and market data, and functions well in established markets (similar to aggregate approaches). However, when forecasting for the long term or in new markets, these models are generally inadequate [43].

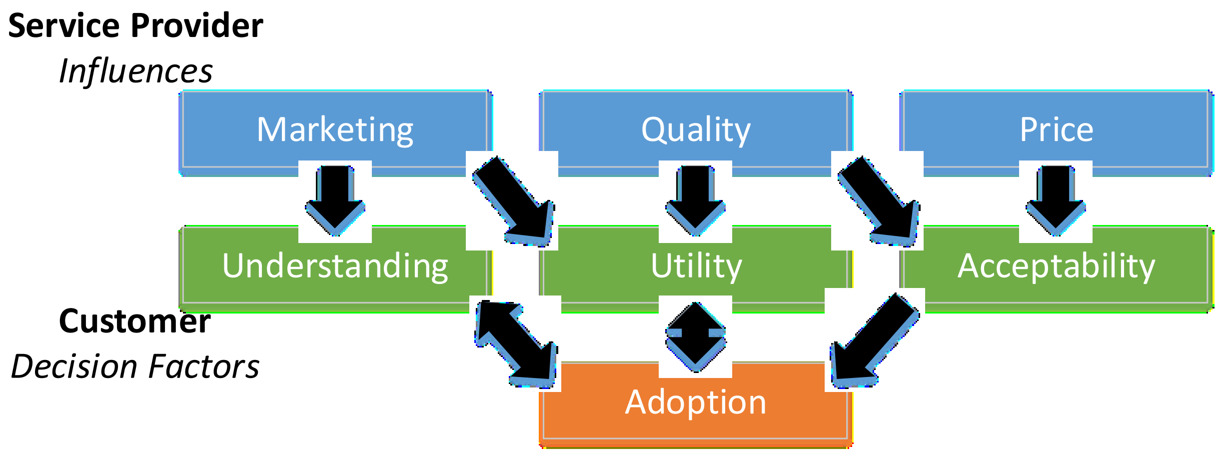


Figure 27 – Overview of the customer choice model, [45]. This is a duplicate of figure 5

Demand forecasting in an expansion scenario has the advantage of having historical data to base the forecasts upon. Given this trend data, forecasts can be developed using time series forecasting and other similar techniques. Time series forecasting techniques look to extrapolate current trends into the future. These forecasts are reasonably accurate for short- and mid-term forecasting in existing markets, as the effect of larger technological and social factors tend to be incorporated into recent historical data [9]. Disruptive technologies and social trends take time to propagate through society, and so the impact they have on short- and mid-term forecasts is minimal.

Long-term forecasting and forecasting for an entirely new network (i.e., greenfield design) require different techniques. These techniques must create forecasts in a highly uncertain environment, and account for underlying influencers that may be unknown, unpredictable, and/or difficult to measure. Forecasting in these scenarios use a number of techniques to estimate demand, including innovation adoption models, expert opinion, and analogy from other situations that are deemed similar [43]. The challenge with this type of forecasting is the variability in potential demand. Because there is no historical data to directly base these forecasts upon, demand must be inferred from indirect data, often with complex and highly variable relationships.

In summary, traffic forecasting for telecommunication networks has been in significant flux over the past decade, from the extremely high growth expectations in the late 1990's to its delay and then eventual the realization of that growth in the late 2000's. Because the underlying drivers of network traffic demand are in such significant change, traditional forecasting methods based on historical data are no longer relevant. Some work has been done to systematically forecast traffic demand, however, these methods have not yet been adequately proven [43].

4.3 Proposed Network Planning Framework

In order to build a better understanding of how to carry out network planning, and align these activities with broader socio-economic objectives, we developed an expanded view of the network planning system shown in Figure 28. This framework used a modified systems approach [120] to analyze stakeholders, inputs, processes, and outputs. The stakeholders are the primary sectors of society that utilize the network. These stakeholders were broken down by general purpose of use and value to society as a whole. Stakeholders are also referred to as service drivers, as it is to solve their needs that network services are ultimately deployed.

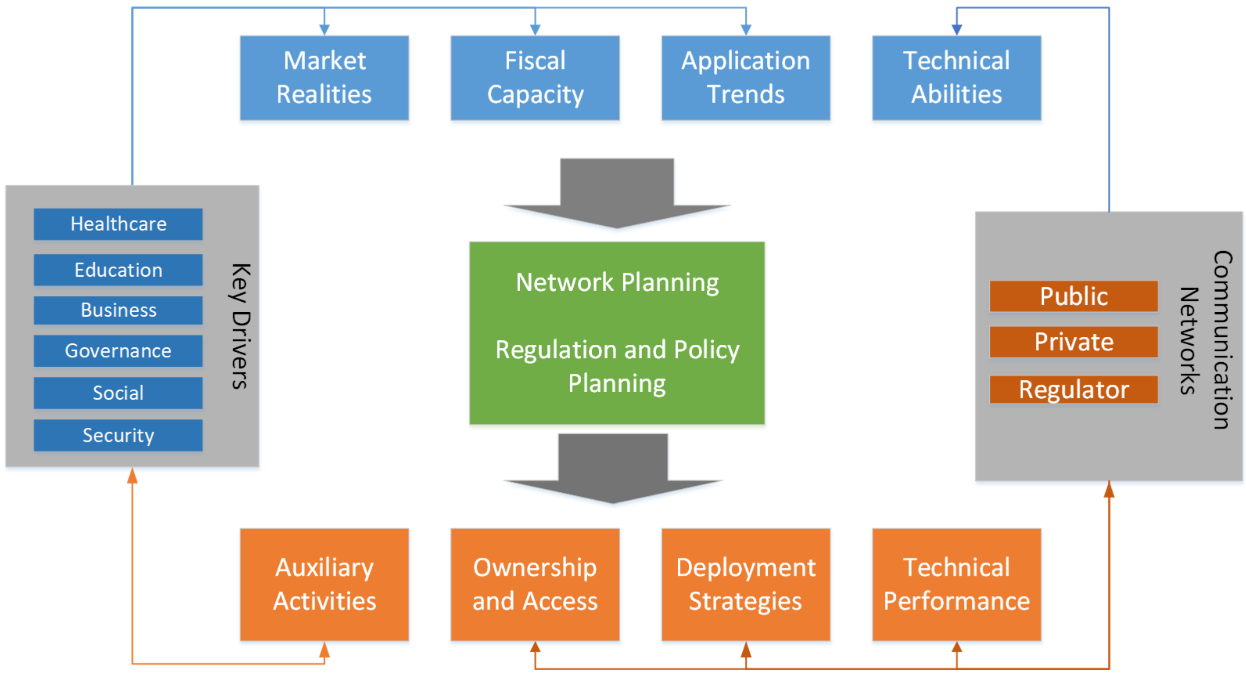


Figure 28 – An outline of our proposed network planning framework.

We developed this framework to serve as an outline to related work, and to better focus future research. Some areas of this research have been well developed in the network planning field (technical network design with fixed demand matrices, for example, [7], [21], [79], [104]); others have been explored from other disciplines [13], [121], [122]. A broad framework like this has not been found in literature (except in part in a high-level overview in [123]), however, so we felt it is timely to present to the research community for discussion and in hopes that it will move the network planning processes forward in a manner that will benefit carrier, network operators, and other stakeholders. By developing this framework, and focusing our work on a few key components that have not been well developed, the intent is to advance the effectiveness of network planning, especially in the dynamic environment we find ourselves today.

This proposed framework uses a broad view of inputs and outputs to the network planning process. This broad view includes secondary and tertiary elements, not directly related to the strict network planning process, [26], but deemed important in understanding how resources are deployed and used in the network, and ultimately how the service drivers are able to extract value. The inputs look at social trends, application trends, financial capabilities of the key players in a network's deployment, and the technical reality of the current network. The outputs include the technical design of the network, how the network will be deployed, its ownership and access structures, and auxiliary activities to enhance the ability to derive value from the network. The network planning system focuses on various network planning steps, however, regulatory and policy makers have also been included, as this also has a significant impact on the ability of end users to derive value from the system.

As discussed in section 4.2, there have been a number of network planning models presented in literature. These are perfectly acceptable approaches and are widely used in cases where networks have been well established and usage patterns are predictable, [43]. However, with the rapid changes occurring in our society (e.g., the social networking, widespread use of broadband application, etc.) and in network technologies, many of these models are not well suited for the telecom environment as it is today. As such, the network planning framework we are proposing herein attempts to provide a broad overview of the fundamental components influencing telecom networks, and hence the planning process, so that the dynamics of this environment may be accounted for to some degree.

This framework does not quantitatively connect the various components, but identifies them and their broad relationships to each other. With this broader view in mind, it provides a framework to expand current network planning models to better incorporate the dynamics and realities of the present networking climate, some of which have already started [28], [107], [111]

. The intended scope of the process is the high-level, long-term network planning. The level of planning targeted by this framework concerns itself with problems such as node location, span implementation, capacity allocation, and to some level of granularity, path design. Other network planning activities, such as wavelength assignment, equipment procurement, and other technical considerations rely on the network topology, capacity, and path allocation decisions included in this network (see the discussion on output, in section 4.4, for more details).

The next steps from this model are to identify which inputs and outputs can be quantified, and incorporate them into new network design schemes.

4.4 Detailed Discussion and Justification

4.4.1 Proposed Network Planning Framework Service Drivers (Stakeholders)

The areas of society that are affected by or encourage communication networks have been broken down into six key areas (as seen in Figure 29), which we call *key service drivers*. These areas are healthcare, education, governance, economic development, social development, and national security. In general, each of these areas has somewhat unique services that they deliver, and therefore provide the possibility of unique interactions with the communication network.

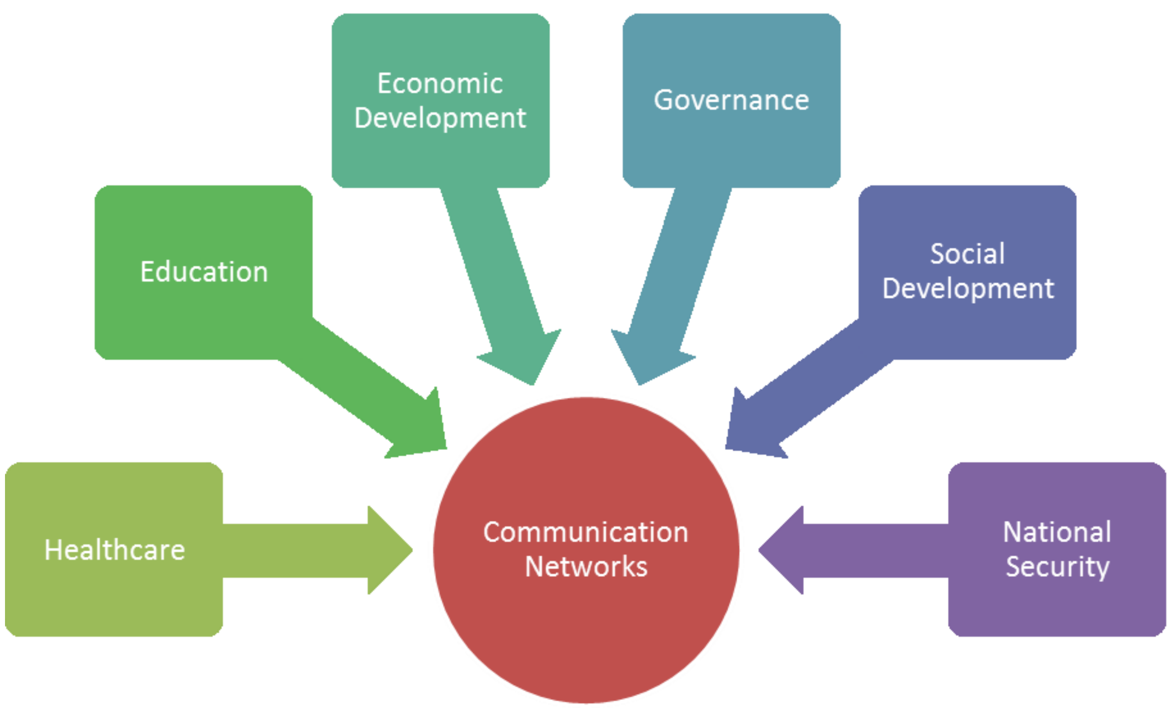


Figure 29 – Key areas serviced by communication networks.

With simple tools such as video conferencing as well as more advanced network requirements arising from remote diagnostic equipment, the healthcare industry is undergoing a transformation. These changes are especially relevant to regions that have limited access to healthcare, as eHealth tools can bridge the physical distance between doctors and specialists, and those that require care, but reside in regions too poor or too remote to have sufficient local healthcare. ICT has the potential to drastically increase the level of care accessible in a cost effective manner.

Education is another pillar of society that has the ability to be drastically altered, especially in developing regions. In the developed world, access to high quality primary and secondary education is taken for granted. This is not the case in developing countries, or even in remote regions of developed countries. ICT has the potential to deliver low cost quality education anywhere that has adequate connectivity. The impact of ICT on education does not have to be constrained to traditional elementary and secondary schooling. ICT can help disseminate advanced concepts such as soil management, construction techniques, and other practical knowledge. Communication systems can provide a platform for better and more consistent education anywhere.

The ability of a modern telecommunications network to deliver economic development is multi-faceted and broad in scope. Mechanisms range from advanced management tools, to home business capabilities, to the ability for a farmer to order and receive a new plough. There are specific tools that utilize ICT to enhance economic development (such as supply chain management), and there are general tools that can have equal or even greater impact on the efficiency of doing business (such as a cell phone). All these tools rely on pervasive connectivity to be effective.

Another key service that benefits from network connectivity is governance. This can be twofold, with the government using ICT to better interact with people, as well as for people to better monitor their government. ICT can enhance the judicial system, and in democratic societies it can provide better tools and more accurate information during elections, for example. On the other hand, as has been made evident with recent events in Egypt [124], Tunisia [125], and Iran [6], ICT can be used to make government more responsible to the people. Although government may try to control the flow of information through the communication channels opened up through ICT, it is significantly more difficult than with less advanced communication tools.

Many people use the Internet, and the tools that utilize it, to communicate in a primarily social capacity. With increased communication, and access to information, biases and bigotry that damage the social fabric are harder to hold on to. Powerful concepts such as gender and racial equality, the value of human life, and other basic human rights are more easily integrated into a local culture when there is more opportunity to interact with the world around them. Increasing human rights is one of the impacts that ICT can have on social development, and there are many more, some tangible and others less so. ICT has a profound influence on how we interact in our daily lives.

National security and defence also are large users of communication networks, however their scope and priority is significantly different from the key service areas previously mentioned. Advanced ICT services relating to national security and defence applications are the subject of a great deal of study by others, and are not included in the scope of this work.

By breaking down applications and tools in a network into the key service areas mentioned above, it is easier to align service objectives with network requirements. To be effective, each area has different requirements on the network. For example, networks designed to encourage economic growth may be extended to key centers of economic activity with high bandwidths. However, if the objective is to ensure a minimum level of healthcare throughout the population, a lower capacity network that is more pervasive could be more effective. A framework for analyzing the impact and demands of communication networks is provided by breaking down network users into these key areas.

4.4.2 Proposed Network Planning Framework Inputs

Inputs into the network planning process were taken to be quite broad, and are broken into four main categories consisting of market realities, technical abilities, fiscal capacity, and trends in applications (as shown in Figure 30). The market, technology and finances are core elements to any business planning process [126], and therefore are included as fundamental inputs into the network planning process. Application trends are also included as these trends directly influence the value that can be derived from the network.

Telecommunication networks are a platform technology, and do not have much inherent value to the end user; it is the tools and applications that utilize the networks that provide value to the end user. In the case of telephony, or cable TV networks, the network operator controls the vertical market, meaning that the value the network generates can easily be attributed to a single, or a limited set of applications (i.e., the telephone or the television). For more than a decade, data traffic has outstripped voice traffic routed in core networks [127], and much of the potential value that these networks have has been realized through data communications. Basic telephone service is a mature market, and outside of a few disruptive technologies (VOIP and messaging), is well modeled and predictable. Television and video delivery is also relatively predictable in mature markets. These applications, however, now present a limited part of the value in the network. Through IP and other low level protocols (e.g., MPLS), the core network can support a vast array of applications, delivered through many devices and mediums. This is what has made demand forecasting significantly more unpredictable in recent years [43], [128].

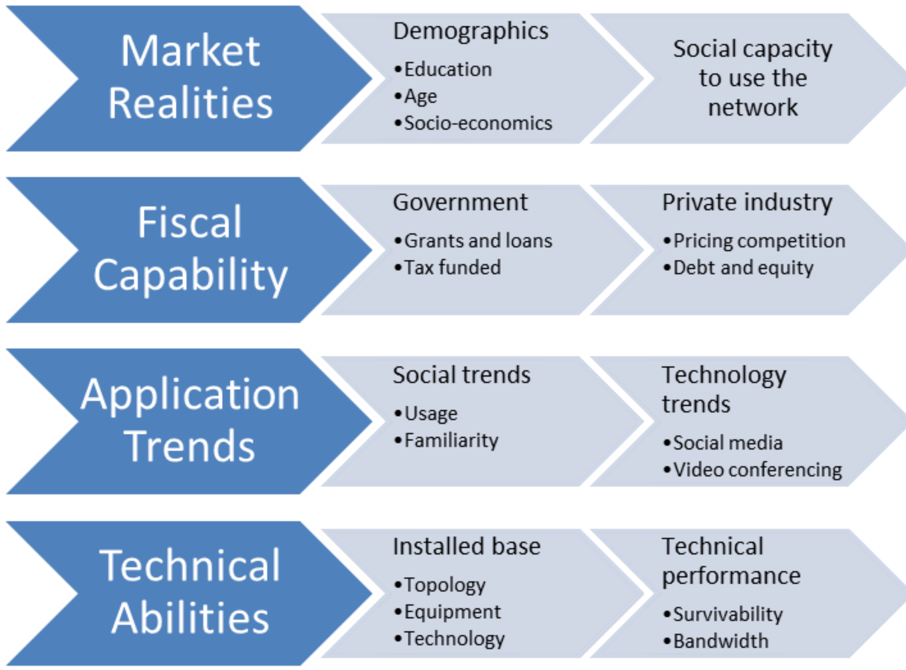


Figure 30 – Overview of the inputs to the network planning process

In expanding telecommunication networks, it is other applications that can unlock the value in network capabilities. As such, when considering network planning, trends in user applications have been included in order to account for broader shifts in usage patterns. This could be a subsection of market data, but because of its importance in the use of telecom networks, it has been highlighted as a primary input.

4.5 Input 1: Market Realities – Details and Justification

One of the most difficult but fundamental questions in greenfield network design, and/or significant network augmentation, is the capacity of the potential user base to derive value from the increased ability to communicate [22], [129]. How this value is derived can vary significantly between service drivers. For example, business analytics and intelligence can provide significant value for businesses and operations, however, does not provide significant value to education users. In education, access to vast amounts of research and educational material is of great significance, but may not provide a compelling value proposition to other categories. The measure, then, in assessing market realities, or potential, is how readily the target market will adopt new tools and applications that are of value to them. While it is not a straightforward process to incorporate this information into the network planning model, it nonetheless has a significant impact on the impact of the proposed network.

In mature markets, and when the applications are well known and understood (such as telephony, or television), the market can be evaluated via econometric analysis, [43], attempting to draw relationships from the various data sources to reliably forecast demand, and return on investment in network expansion. The techniques involved in econometric analysis rely on good historical data and relevant market data that can be used to interpolate future demand. These models have room for the effects of competition (customer churn), and substitution products, as long as there is the historical data to define the model parameters. These models and tools, however, have difficulty analyzing and predicting demand with new products and in long term forecasting scenarios [43]. New products have rapidly changing adoption characteristics, and long-term forecasts must necessarily deal with fundamental shifts in society's relationship to the networks (such as the recent social networking phenomenon).

Competition is not highlighted in this network planning model; because telecommunication networks are extremely expensive to deploy, they often operate in a monopolistic environment, similar to other utilities [130]. Competition is severely limited in these environments, with significant legislative powers preventing network operators from exploiting their monopolies. Network planning may incorporate competitive analysis when forecasting the impact of certain applications (like TV through phone lines or telephone through cable), however, competition at the infrastructure level is limited, especially when planning new networks. Because of this, competition effects are not emphasized.

Since econometric market analysis tools used in network planning in the past are not reliable in the current context, with rapid change and flux, a different view is required. As suggested in [128], a closer measure of the drivers of user demand could be applied to the evaluation of the target market for network expansion. With no reliable historical data, panel data, such as demographics, socio-economic conditions, and other marketing data can be used to measure the potential usage of the network. The challenge here is to make minimal assumptions regarding the specific services that drive network usage in order to not tie the forecast to a specific application, but rather tie network usage to the target market's ability to adapt and derive value from the network. This ability to derive value from the network could be seen in proxy as the ability of the target market to adopt new innovations.

This measure, of how readily new applications will be adopted, has been attempted to be answered through the study of the diffusion of innovations (or new products). The foundations of the study of the diffusion of innovations was created largely in the 1960's, with the works from Rogers, [22], and Bass, [13], having the most impact as related to telecommunication planning [43]. The two take rather different approaches to the topic, with Rogers providing a good overview, and Bass providing a way to quantify the uptake potential. It is extensions from these works that provide the most promise in enabling a measure of the market's capability to derive value from the network as an input to the planning process.

Both Rogers and Bass provide methods to forecast demand, with Rogers assuming a normal distribution, and Bass using a special case of the shifted Gompertz distribution [13]. In practice, the Bass model has been more commonly used. In tying market realities to the network planning process, these diffusion models provide a solid analytical base. The application of these models will take specific insight into the market the network is being deployed in, as their parameters must be tuned to each product and market. It should be noted that there has been work done to develop Bass models without historical usage data [131].

In order to optimize a network's design, the outputs of the network planning process must have some feedback mechanism into the inputs. For example, if the design calls for the network expansion plan to take a focused approach on a few locations, and expand to new locations slowly, as opposed to creating a lower quality network that reaches more locations, there will be implications in user uptake. Also, if there is a decision to allocate resources to application development or user training, rather than to network facilities, it will require feedback into the demand matrix to be able to analyze and optimize these decisions.

It is therefore the goal of the market realities input into the network planning process to not just statically forecast demand, but to create a mechanism for feedback of design and other output decisions into the system. Because econometric forecast models have been deemed inadequate in forecasting demand for new networks and for long-term planning, models that are based on new product diffusion are suggested in order to quantify the market dynamics.

4.6 Input 2: Fiscal Capacity – Details and Justification

One of the primary considerations or constraints in network planning is financial. The financial considerations in network planning include the ability to fund the expansion, and the expected return on investment. While it seems that these two are invariably intertwined, as they often are, we delineate them in our planning framework to highlight the distinction in use cases, and network owners. Looking at the service drivers presented above, one can imagine the diverse ownership/access/ROI requirements that each could represent. The difference can be manifest as to whether the network is seen as a cost center, or a profit center for an organization. Network operators will view voluntary network expansion as a potential profit center, and filter the expansion from a direct ROI perspective. From a government perspective, telecommunication networks have a much broader social impact. These impacts can be as diverse as rural economic diversification to judicial processes (such as having a defendant be present at a trial through video conference technology) to educational advances. Because governments and possibly non-governmental organizations (e.g., the World Bank, the International Monetary Fund, etc.) evaluate the justification for network expansion and augmentation from a broader perspective than direct revenue and expenses, the fiscal aspects of network planning can be more complex.

Network planning can be done on a strict capital cost minimization basis, or, similarly, it can be done by combining capital and operational costs [132]. This perspective is common in literature, and most network planning models approach the actual network design from this perspective. This cost minimization perspective is due to the linear planning approach that has been common in literature. Since the general network planning and design process has considered inputs as fixed, and the output of the planning process to be primarily a technical network design, the objective has been to minimize cost while ensuring that the input requirements are met. This view is limiting, and because of the dynamic nature that has been discussed in the market realities, may not be adequate to increase, maximize, or reasonably account for the derived value of the network.

The impact and evaluation of the financial component of the network planning process is amplified as it is applied to regions more sensitive to the economic costs of the network (such as much of the developing world). The Alberta Supernet has been praised in its scope and objective of connecting the whole province to high-speed network infrastructure. While this goal may be lauded, the actual impact of this network has come into question. The $400+ million cost of the project was significant, but not overly so if it did not meet expectations (which has been argued [72]). If the project was undertaken in a region without the financial resources available in Alberta, the consequences could have been significantly more dire. As such, the motivation for this network planning framework is emphasized when looking at the financial inputs to the network planning process.

As discussed previously, the two categories of financial inputs to the network planning process are the capabilities to fund the project and the expected return on that investment. Network infrastructure is capital intensive and the ability to fund the network expansion is not trivial. The World Bank and the ITU both have departments dedicated to assisting the deployment of telecommunication networks in developing countries [133], [134]. In the developed world, the capital expenses involved with network expansions can be significant. In the widespread expansion of communication networks in North America in the late 1990s, many companies involved in building the networks went bankrupt [135]. It has even been alluded to that the current networks in the United States were built using bankruptcy law [136]–[139]. The capital cost, and the ability of the network owners or funders (if grants are used) is of concern to the network planning process, and if included, could be instrumental in reducing the capital risk of network expansions.

Part of the purpose in including the capital capabilities of the network owner would be to better manage the gap between projected demand growth and projected capacity. As shown in Figure 31, early phases of a network's lifetime can have large gaps in the capacity and usage. (It should be noted that Figure 31 was adapted from [45], but is not an exact representation of the data.) These gaps represent significant risk, as the income or value derived from the network has not represented a positive return on investment. If the network owner cannot sustain this gap, the network will become financially unviable, and will either become another statistic in bankruptcy law, or fall into disrepair. If the network owner can sustain this debt level, but this gap could be reduced, this becomes in inefficient outlay of capital.

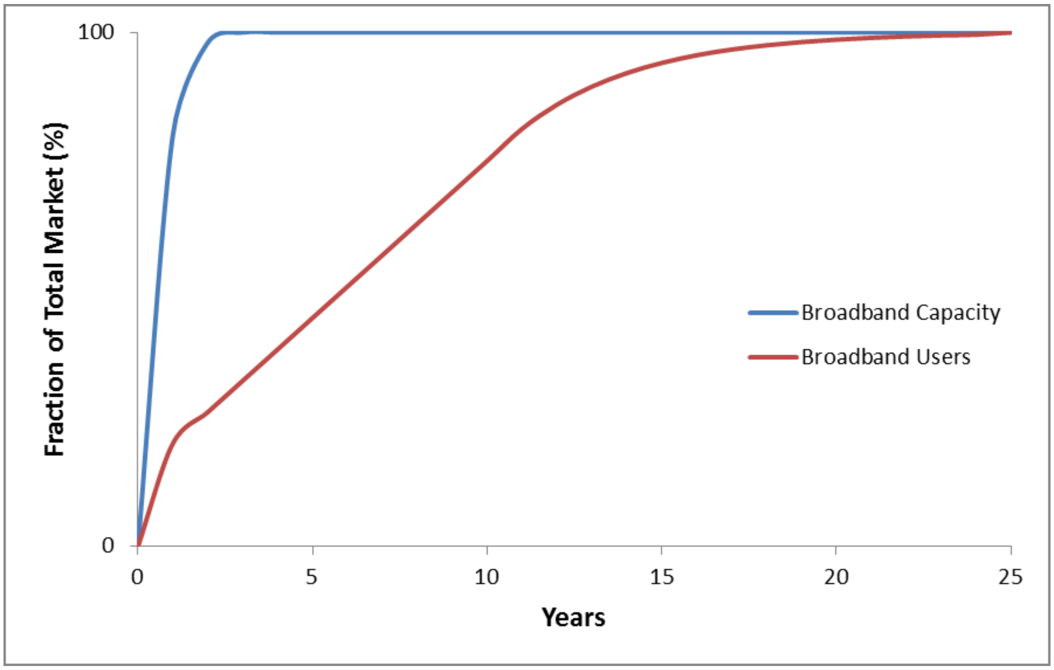


Figure 31 – Comparison of broadband usage (customers), and deployed capacity adapted from [49].

The point of the proposed network planning framework is to be able to better align network designs and the overall planning process to better manage the gap between usage (assuming this is a proxy for derived value from the network), and deployed resources. For this to be done, the capital constraints should be clear, both in terms of the capital costs, and the expected return.

The expected return on the network investment is generally measured in monetary terms, as that is a straightforward quantitative measure that is easy to relate to the cost of the network. For profit network operators, this is arguably the only measure that matters. This changes, though, when dealing with society as a whole, as it would with governments and non-governmental organizations. From this standpoint, the value needs to be evaluated in a broader sense. What is the economic value of increased access of information in schools, or better support of e-health services? These may be difficult to measure, but there is a proxy measurement that could represent the derived value that each group sees from the network, and that is usage.

The expected or desired usage rates would depend heavily on the types of applications and use cases for the network. Incorporating video conferencing in healthcare delivery would increase network usage from a bandwidth perspective significantly more than access to text based business analytics for a business, but the value for the end user may not be proportional to the bandwidth. Nevertheless, it could be assumed that people will use the network if they find actual value for the respective application in the scope of their responsibilities. The return on investment is a real and critical aspect of the network planning process, but in situations where the value derived from the network is not directly from subscribers, but rather through the use of applications that impact broader goals and purposes, its evaluation can be more difficult.

4.7 Input 3: Application Trends – Details and Justification

Application trends are a difficult input to accurately describe and quantify. These trends, however, are very important in the network planning process; their input into demand forecasts can be significant. If consumers are using the Internet to stream video, and content providers are moving to provide their content online to meet demand, it would be expected that a jump in bandwidth usage would occur. Outside of the obvious, however, there are subtler effects of application trends. These can be brought to light by looking at key aspects on innovation adoption as articulated by Rogers [22]. Rogers articulates five attributes of an innovation that affect its adoption rate; complexity, observability, trialability, compatibility, and relative advantage. Application trends can address the first three attributes (complexity, observability, and trialability), and their impact on the network can be better understood.

Consider a nurse nearing retirement. Like many in her demographic, she has limited computer skills. Video conferencing and similar interactive tools have been available in her workplace for a while, but because she has no exposure to them, she does not see them as useful and will not use them. In her personal life, however, her children life in other cities and have been encouraging her to use Skype® to keep in touch. Although not related to her work, she is becoming more familiar with the technology, and usage patterns required to take advantage of similar technologies. This exposure to a video conferencing technology has indirectly increased the observability and trialability of similar technologies, while also reducing complexity, making it easier to adopt video conferencing in the workplace.

Any new technology has a diffusion curve, and is dependent on social systems to propagate through society. By looking at broad application trends, insights can be gained into the potential network usage demands, as well as the rate of change in these demands. Quantifying these impact may be difficult, and there are other factors affecting adoption rates, but by including application trends, both demand requirements and derived value (ROI) can be better evaluated.

4.8 Input 4: Technical Abilities – Details and Justification

One of the fundamental inputs into the network planning process is the pre-existing network, and the capabilities (and cost) of the equipment available to create or augment the network. These technical capabilities include the actual network layout, where the cables are placed (spans), and where they terminate (nodes), the capacity of the equipment in the networks, and the technology they utilize. The technologies used are relatively static, and significant effort is made to ensure there is compatibility between generations of transmission technologies. This section provides an overview of the technical considerations that affect the network planning process.

It should be noted that high-level logical networks can use quite diverse and sometimes incompatible network protocols. For example, IP networks use a packet switching routing mechanism, while ATM networks use a circuit switched routing methodology. These two protocols are not inherently compatible, as they approach the traffic routing problem from significantly different paradigms. There has been significant work in recent years to provide a lower level platform that can accommodate many different protocols, the most popular being MPLS [140] (or GMPLS) with SDN growing in prominence [17], [111]. When planning networks at an infrastructure level, it is assumed that these technologies are in place, and are generally capable of grooming to fit the transportation protocol with reasonable efficiency [7], [141].

Network design, by its nature, happens on many different layers [142]. This means that path configuration, capacity allocation, topology expansion and survivability design can have many different levels of scope and planning horizons. For example, different layers provide different methods to be resilient to failure, and while there is a strong argument for survivability design at the lightpath level, there has been work in integrating survivability design between various network levels [37]. For the purpose of our network planning framework, it is the bottom two layers (representing the physical network and the communication paths, typically light paths, routed directly on it). The assumption is that the fundamental requirement is to have physical connectivity between locations, and the design considerations of this framework (node placement, span implementation, and capacity allocation) are within this scope of planning. It is shorter range planning that generally concerns itself with the design of these higher layers [9], assuming the lower layers are already fixed.

The technical capabilities input into the network planning process include the already installed equipment as well as the performance and compatibility of new equipment. This is a fundamental part of the network planning process, and in most cases is relatively straightforward to integrate into the design process.

Incorporating the technical abilities into the design process has two primary considerations. The first impact in the design process is in the representation of the technology (or technologies) at the design optimization stage. The second is when translating the optimized network designs into actual plans [26]. It is necessary to have the essential constraints and costs from the technology included in the optimization model, however, much of the details should be abstracted out to ensure the planning process does not get overwhelmed with details (either due to human limitations or computing capacity limitations). In many network planning models, the costs involved are taken as a function (usually linear) of capacity and distance of the spans (cables) in the network [83]. With some work being done in greenfield network design, there is a separate instantiation cost for a given span in the network. There has also been some work done in attempting to model and optimize costs with more granularity, and incorporating operational costs [132].

There are many ways in which current infrastructure can be represented in the model used for network planning. The current network equipment can be used in the network design process by setting the costs associated with the network to zero (or some relevant low cost). In some cases, there may be other infrastructure that can be leveraged for less cost than installing new cables and equipment. For example, when Tanzania was looking at deploying an optical core network in their nation, the economic constraints of building a network from nothing did not warrant the investment. However, there was a certain amount of government and military infrastructure that could be leveraged, even though the node equipment was out of date. By leveraging this current infrastructure, the network became economically feasible [143]. Another example of current infrastructure affecting the network model is when there is a competitor's infrastructure in the geographic region. It may be more feasible to lease capacity on this infrastructure where fully implementing your own infrastructure does not make sense. This would eliminate the capital cost of that portion of the network, but would lead to higher operational expenditures. Current infrastructure can impact the network design process in many ways and is a fundamental part of the network planning process.

As discussed above, there have been recent advances in enabling the core network to serve a wide variety of traffic sources through technologies such as MPLS. Technologies like this can go a long way in increasing the lifespan of infrastructure while enhancing and enabling the network to meet new demands. An example of this is the new OTN protocol [141] that accommodates both SONET/SDH and gigabit Ethernet. One of the biggest challenges in managing changes in network technologies is matching the line rates (or speeds of transmission). One can imagine a scenario where the core technology had a nominal line rate of 1, and the traffic that was interfacing with it had a line rate of 0.51. This would mean that 49% of the capacity would be lost due to line rate incompatibility. This was the case in relating gigabit Ethernet, which has been growing in popularity for metropolitan area networks, to the SONET/SDH protocols of the core long haul networks, which has line rates originally based on voice traffic. One of the aims of the OTN standard was to reduce this incompatibility and increase the efficiency of serving both types of traffic.

In general, the technical considerations in input to the network planning process consist of the current available infrastructure, and the choice of transmission technology standard. These considerations are influenced by some outside concerns, such as the cost of equipment, expertise of the organization, etc. Within the scope of the design process, these considerations are relatively static with simple trade-offs in the choices.

4.8.1 Proposed Network Planning Framework Decision System

The network planning and decision system has been described in the background section of this chapter, at least from a technical point of view. Network planning and design, of relevant scope to the framework being discussed here, typically follows a process of translating predicted demands to network capacity allocations, routing plans, and equipment specifications. This view addresses some of the components of this framework, but to be more effective with network planning and design as presented in this framework, a broader view is needed.

In general, network planning activities involve the allocation of resources, and the design of policies and regulations as shown in Figure 32 using the relevant inputs to address the outputs of the planning process (as described in section 4.4) in order to enhance or maximize the value that the key service drivers are able to derive from the network. Most of the work described in the introduction attempts to solve this problem in a linear fashion. Some of them attempt to deal with the uncertainty of the process, and others attempts to align the design of higher network layers to the core infrastructure design being discussed here [37].

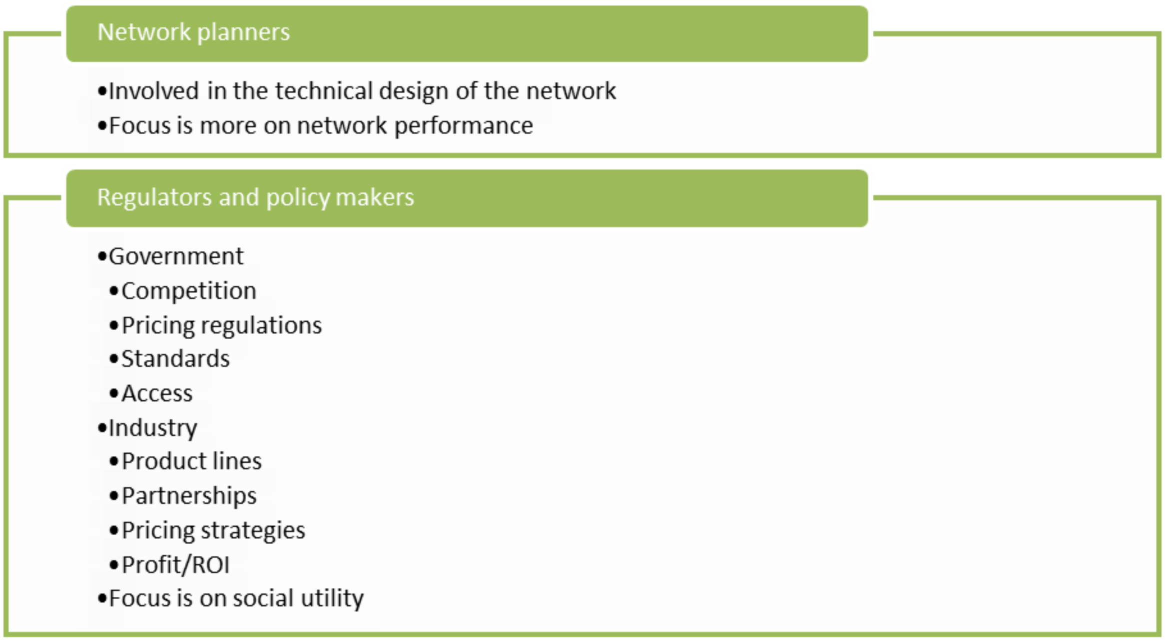


Figure 32 – Outline of the core network planning decision maker and their responsibilities

The purpose of this framework, as it is currently presented, is to define the context in which these planning processes occur. This awareness will identify the primary weaknesses, and potential opportunities to enhance the network planning ecosystem. The current shortcoming seen in most network design models in the literature is the narrow scope of inputs and outputs (limited to pre-set demands as inputs, and static resource allocations as outputs). Future network design and network planning paradigms should be able to incorporate the dynamics and uncertainty of the planning environment using models that address the broad spectrum of resources and policies required to effectively deploy network resources. These paradigms should not just focus on the technical design of networks, but include the social utility for the key service areas for which the networks exist.

4.8.2 Proposed Network Planning Framework Outputs

Defining the output of the network planning process on a technical level can be straightforward; at the highest level it is simply an arrangement of nodes, spans, capacities, and lightpath, though more detailed technical planning would also define wavelength assignments, equipment specifications, and other technical details. As was seen in the input section of this proposed network planning framework, this technical view misses some of the other aspects which affect the end goal of the planning framework, to enable and enhance the value that the service drivers are able to extract from the network. As such, output has been broken down into four categories (Figure 33); two dealing with the technical aspects of the network, and two that consider the environment the networks will exist in. The technical categories are the performance requirements of the network, as in quality of service, compatibility of the equipment, and capacity placement, and the deployment strategies associated with this equipment. The non-technical outcomes of the network planning process attempt to capture the variables that can affect user adoption from an operator/owner perspective. The first of these is ownership and access looking at construction and operations of the network. The second non-technical output category has been labeled "Auxiliary Functions" and is intended to capture activities and decisions that are not directly related to the deployment and operation of the network, but rather, help drive usage of it.

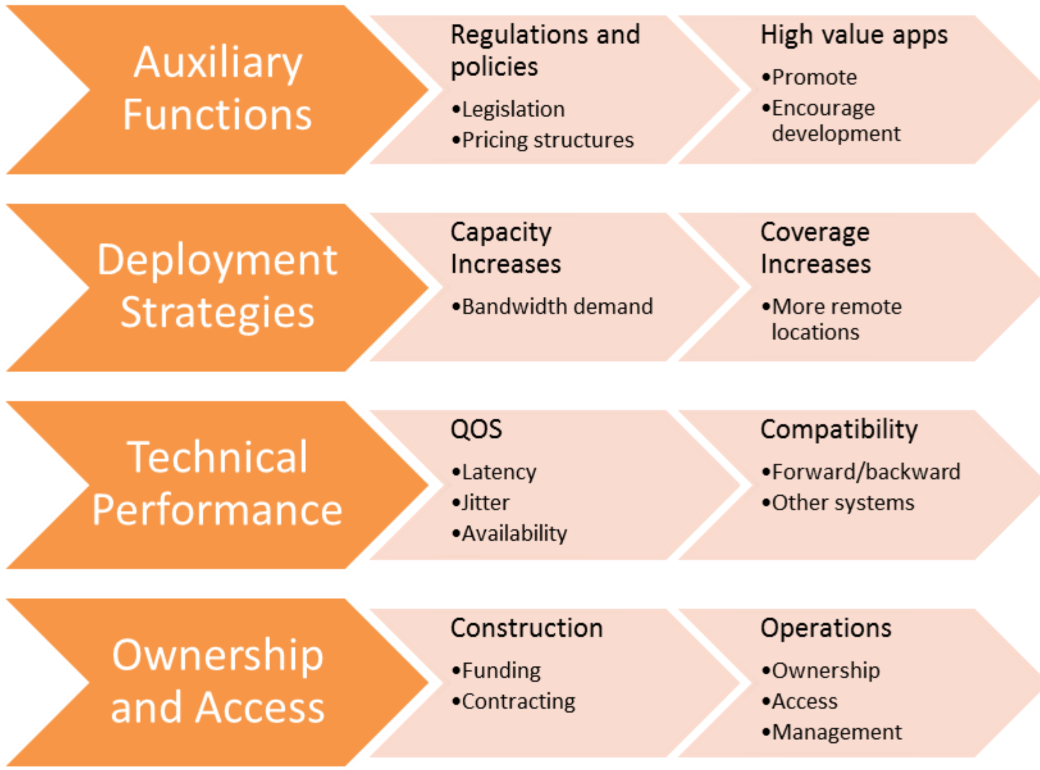


Figure 33 – Overview of the outputs of the network planning framework

4.9 Output 1: Auxiliary Functions – Details and Justification

The last output defined in this framework is a bit of a catch-all to represent demand side stimuli. Network planning and deployment often treats the end user's ability to utilize the infrastructure as something that cannot be influenced (outside of maybe marketing effort) [26], [27]. In a broad perspective, however, there are things that can be done to enhance users' ability to derive value from the network. From a policy perspective, promoting or encouraging network development and expansion could be seen as a supply side intervention. This intervention must be balanced with the demand side. A "build it and they will come" philosophy may not be effective, especially if the network is targeted at key service drivers who are not readily able to use the network.

To delve into another concept from the diffusion of innovation concept, the primary indicator of an organization's innovativeness is their slack resources. Slack resources provide the ability to explore, learn and adopt new technologies for that organization. New technologies can be seen as a high risk, high reward endeavour, even from an adopter's point of view. If education and healthcare are looked at, the slack resources available in these sectors of society is often very limited, especially if they are operated as public services. Their uptake on new technologies has been historically slow, and the un-served value potential in these areas often takes some kind of stimulus to unlock in a timely manner.

The output decisions in auxiliary functions are the broad demand side stimulations, taking the form of regulations and policy, or the promotion and training of high value applications. Regulation and policy decisions can be as broad as tax breaks for companies incorporating certain technologies, to stimulus for the development of new services and applications. Promotion and training consists of activities that enhance the user base in a key service area ability to use key services. These interventions often already occur, but by aligning efforts to use key services, and the deployment of network services.

An example could be taken from the healthcare industry. In a family practice, doctors do not have the resources to disrupt their current processes and procedures in order to adopt new advances, even though in the long run it may pay off. Ludwick depicts a drop in patient volume when implementing electronic healthcare records in a doctor's office [144] (Figure 34). If support is offered to mitigate this drop in productivity, there will be faster uptake of network related technologies, and a shorter payback period for the investments into the network infrastructure.

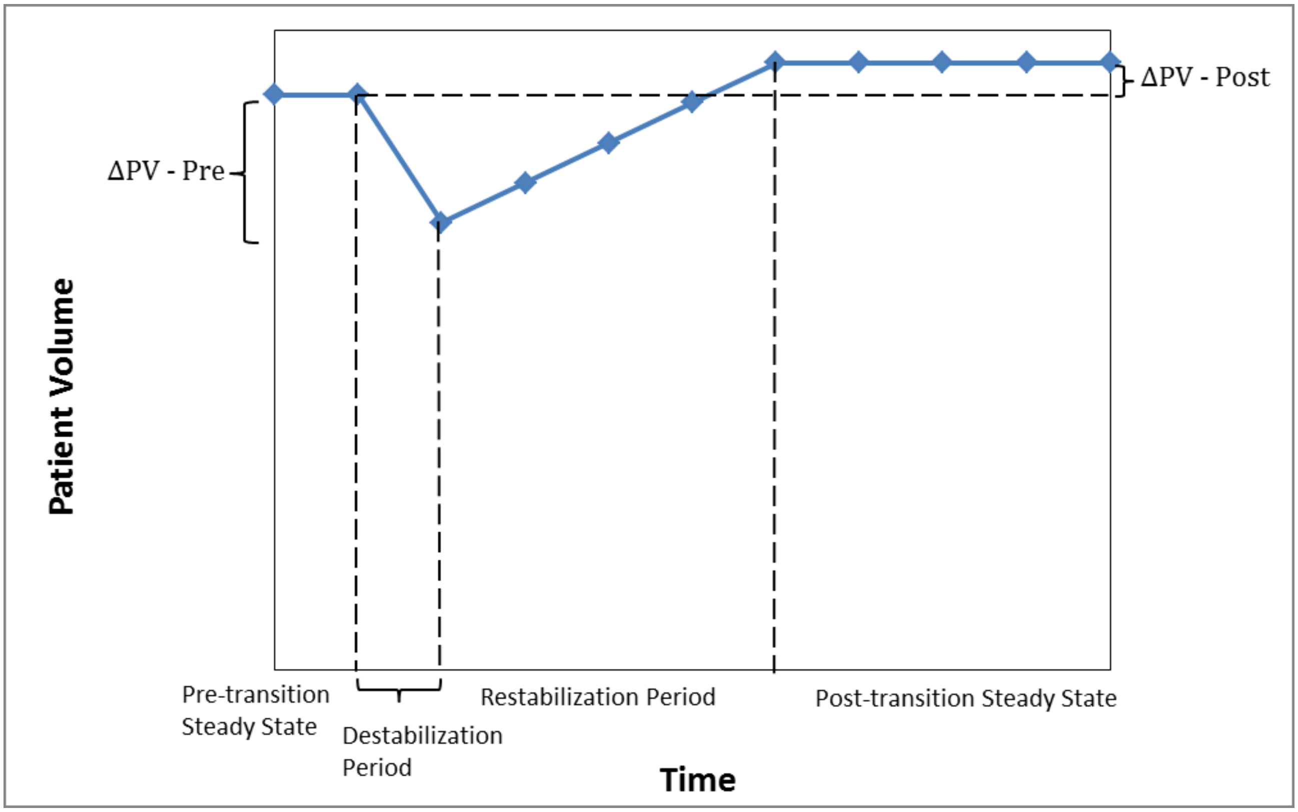


Figure 34 – The impact of adopting electronic health records on patient volume in a doctor's office. Adapted from [144].

There is an obvious trade-off in allocating resources to auxiliary functions, as these will not be put into developing the network. Actually allocating resources to these types of functions requires insight into the user's ability to derive value from the network, the bottlenecks that are present in this ability, and a methodology to balance the allocation of resources.

4.10 Output 2: Deployment Strategies – Details and Justification

By including deployment strategies as an output to the network planning framework, phasing decisions can be incorporated into the design process. As described in the input section of this chapter, there can be a significant gap between deployed bandwidth and usage early on in the network's lifecycle. By phasing deployment of resources, capital can be better allocated, and there can be better management of risk.

Deployment phasing can be broken down along two dimensions, geography, and capacity. Geographic phasing can emphasize some regions over others. Technology adoption does not usually occur evenly across different regions. By evaluating jurisdictions on the ability to adopt innovation, and focusing on those that tend to be earlier adopters [22], resources can be focused on where they will be used first. As these regions successfully leverage the network, others will be better equipped to follow suit.

Phasing the deployment of capacity on a per link basis can allow better management of equipment costs. The point of allowing phased deployment of resources is to prevent excessive over-provisioning in a network. Some degree of over-provisioning can be desirable in order to accommodate to a certain degree peak usage scenarios (sometimes referred to as the "Mother's Day effect"), but too much capacity is an inefficient allocation of resources.

By incorporating different deployment strategies into the network planning output, concepts from the diffusion of innovations can be incorporated into the network planning process. This will allow the possibility of network resources to be better deployed. The actual phasing may be incorporated into the design optimization, or afterward, depending on the complexity of the network design model, the computing resources available, and the experts involved in the network planning process.

4.11 Output 3: Technical Performance – Details and Justification

The technical specifications of the network planning process can vary in detail, according to the scope of the process. It can include a high level definition of the node and span locations, and the associated span capacities. Most survivable network design models also define the path routes between each node in the network, and the division of capacity between working and spare for each span. From there, the technical design is expanded into detailed specifications. For the purpose of this framework, it is assumed that the primary decision variables are the node and span allocations, along with the span capacities and routing specifications. The further technical design is a refinement of these specifications.

There may be a case to include more details in the planning process when attempting to optimize both capital and operational expenses [132]. Models that optimize both components use a more detailed technical design to better reflect the costs of building the network and of operating it. These details though, may not be relevant in the scope of this network planning framework where the allocation of resources attempts to address value delivered to the key service drivers.

4.12 Output 4: Ownership and Access – Details and Justification

Network infrastructure at a regional and national level is difficult to manage from a policy and operational perspective. At this level (regional and national scale networks), the costs of deployment and operations move the network into a utility like industry, with redundant infrastructure required for competition not often feasible. Especially when deploying networks in remote and underserved regions, the economics of creating competing infrastructure can be prohibitive and unreasonable [139]. As such, there are a number of choices that can be made that affect the ability of the end user to derive value from the network; these choices involve ownership, management, and access to the network [145]–[147].

There has been a trend in the last 20 years in developed countries to move from state owned telecom providers to a privatized industry with legal provisions to prevent abuse of the potential monopolies these former government owned companies would receive. This model works where competition can be simulated, and government can put into place adequate measures that ensure the monopolies (or near monopolies) are not abused. These policies however, are not favourable to the objectives of privatized network operators (see the net neutrality debate in the US, and the controversies in Canada around usage based billing [148]). Ownership and access is a balance of effective cost allocation between the investors in network infrastructure and the users. Push the balance too far toward the network owners, and the cost of using the network could become cost prohibitive to many potential users. If the balance is too much in favour of users, then there is little incentive to invest in new networks, and enhanced capacity. This chapter will touch on this briefly here, but this is a significant, complicated, politically charged decision set.

One of the primary characteristics of the Internet is that it separated layers of the value chain in network operations (Figure 35). This stands in quite stark contrast from other telecom related industries (namely telephone companies, and cable television providers). When the NSFNET was converting to a privatized system, care was taken to separate the long-haul network providers, the access network providers, and the services run on the network. This separated model was effective in creating competition at each of these layers, providing users a broad range of competitive services. The power of this separated model can be seen more recently in the mobile market. One of the most disruptive elements of the iPhone was that it made wireless operators a "dumb pipe", no longer in control of the data transferred through their networks, and no longer able to charge based on the individual value of that data. This differentiation has led to an explosion of user facing services available in a mobile context.

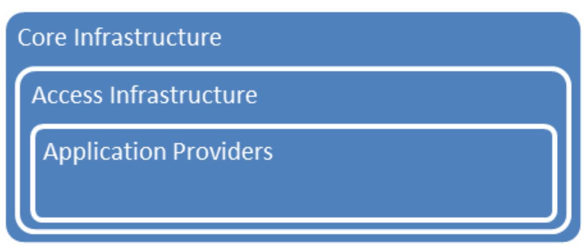


Figure 35 – Levels in the telecom network value chain.

This separation of the network value chain can produce significant results, but requires an environment that is conducive to competition. There are many different models in place that allow different levels of vertical integration and access to the network, and the services that run on them.

The decisions around ownership and access are a balance between who builds (and pays for) the infrastructure, who manages it, and who can access it. The impact on the user is not always direct, but as described in the mobile industry, can have a significant impact on user's ability to derive value from the network.

4.13 Analytical Model of the Proposed Network Planning Framework

To illustrate the network planning framework in action, we have included a model that connects the inputs of technical abilities, fiscal capacity, and market realities to the planning outputs of network reliability, capacity, and user engagement (auxiliary functions). The decision variables in the model were kept to a capacity target (measured in this case by the amount of spare capacity added relative to estimated demand), and a cost for user engagement (advertising, training, change management, etc.). In this case, these were assumed to be constant for the entire planning horizon. The model looks at the effects of the two decision variables over a set period of time to estimate an optimal allocation of resources between the two based on their impact to the time discounted value of the network.

The model presented here is appropriate at the scoping stage of planning to get directional correct guidance of where to focus time and effort. This was done by capturing (as described above) two significant assumptions around network reliability (correlated to spare capacity) and expected usage growth. Growth estimates and reliability requirements vary across all of the key stakeholder groups described in section 4.1 [121]. By using a model such as what is presented here operators can quickly evaluate the impact of where strategic effort it focused.

The model uses a multi period demand growth profile that was derived from a common new user model for technology growth. The model used is the Generalized Bass Model [149], which uses a modified s-curve style user growth model. The Bass model has been used to forecast demand growth in a large range of industries and product classes including telecommunications [43]. The Generalized Bass Model was introduced to incorporate decision variables into the model while still having a closed form and the ability to reduce to the original Bass model [149]. There have been many different models designed to incorporate the effect of actions like advertising and pricing, however, the Generalized Bass Model offers a consistent method to incorporate generic decision variables that affect the rate of adding new users to the network [46][91].

The Bass Model takes two parameters, p and q, to represent the innovators and imitators in a target market. We used a general estimate of these for telecommunication networks of 0.056 and 0.5660 for p and q respectively [150]. In practice these would be tuned to the market in which the network would exist, as there can be a large number of factors influencing the uptake of new products [22]. Studies have found that the sum of these two parameters typically lie between 0.3 and 0.7, with an average p of 0.03 and an average q of 0.38.

In addition, network design and capacity allocation is not a continuous activity and therefore, the Generalized Bass Model was discretized, giving the number of new users as below:

$n(t)$ is the new growth at time t
$N(t)$ is the cumulative users at time t
$x(t)$ is the market intervention at time t
$X(t)$ is the cumulative market intervention effort up to time t
m is the market potential
q is the imitator coefficient
p is the innovator coefficient
α_i is the market intervention factor(s)
β_i is the market intervention coefficient

$$n(t) = \big(m \times p + m \times q \times N(t-1)\big) \times \big(1 - N(t-1)\big) \times x(t) \tag{1}$$

The Generalized Bass Model is most simply represented as a hazard rate $h(t)$ (i.e., the probability that those not using the network will start using it at time t):

$$\mathrm{h}(t) = \big(p + q \times F(t)\big)x(t) \tag{2}$$

As discussed in the paper describing the Generalized Bass Model [90], the cumulative market intervention (X(t)) was assumed to follow a linear growth rate, $X(t) = cT$, where $x(t) = c$, and $c = 1 + \beta_1 \ln(\alpha_1) + \beta_2 \ln(\alpha_2)$.

For the example presented here, the reliability metric is the ratio of total capacity to demand, where demand can be under served (retarding demand growth) or over served (representing a degree of survivability in the network). The ratio of capacity to demand is captured in α_1, which is the ratio of capacity to demand (i.e., α_1=capacity/demand). The coefficient β_1 provides a mechanism to weight the factors, and was assumed to be 1 for the purposes of this work.

The other decision was assigned to the effort given to supporting activities. This could include advertising, but could include promotion of key uses (such as equipment and facilities required for remote diagnostics in telehealth applications). This effort (or cost) for supporting activities was set to have α_2 equal to 1 when there was no effort assigned to supporting activities, , and 1 plus the cost of supporting efforts in general.

The actual cost of supporting activities is not relevant to this study, it is the relative to the other parameters in the model. The goal of this study was to look at the tradeoff between investing in the network (deployment and technical capabilities) captured in α_1 versus investing in the capacity of the user base to derive value (supporting activities) ennumerated as α_2. In order to evaluate this tradeoff, an economic model was built around the revenue, capacity costs, and supporting costs. The point of this economic model is not to capture the full set of costs, but to put quantitative metrics around the framework presented in this work in order to examine the tradeoffs in key decision points. In the decision model presented here, the trade-offs affecting technical investment and supporting activities were the revenue per demand, R_{demand}, the costs of the supporting activities, $C_{support}$, and the cost of adding capacity, $C_{capacity}$.

i is the discount rate used for the economic model (10%)
$C_{capacity}$ is the cost for each unit of capacity
$C_{support}$ is the cost for each unit of supporting activity
R_{demand} is the average revenue for each unit of demand
$C(t)$ is the total units of capacity in the network at time t
$D(t)$ is the sum of the network demands between all nodes in the network at time t
$S(t)$ is the supporting activity effort at time t
$m_{y,z}$ is the total potential demand for each node pair in the network (y,z)

$d_{y,z}(t)$ is the new traffic between nodes *y,z* during time period *t*
$D_{y,z}(t)$ is the cumulative traffic at the start of time *t* between network nodes *y,z*
$C_{y,z}$ is the distance required to route one unit of capacity between nodes *y,z*

The objective metric of the model is the net present value of the revenue minus the costs (3) where we used a standard 10% discount rate. The users added each period were added using the Generalized Bass Model (4), assuming each individual node pair in the network behaved like an isolated market. It was assumed that new users came on at the end of every period, and therefore the demand for any period t was calculated by adding the new users from the previous period to the cumulative users from that period (5). Demand symmetry was also assumed, and the total demand used for calculating revenue and costs was the sum of $D_{y,z}$ and $D_{z,y}$ divided by two (6). The market intervention factor (7) was calculated as a constant across time as discussed earlier. The capacity for each time period was calculated as a ratio of the demand (8), and the supporting activity effort was defined in (9).

Maximize:

$$NPV = \sum_{t=1\ to\ T} \frac{R_{demand} \times D(t) - C_{capacity} \times C(t) - C_{support} \times S(t)}{(1+i)^t} \tag{3}$$

From the Generalized Bass Model, the new demand and cumulative demand for each time period was calculated as such:

$$d_{y,z}(t) = \left(p \times m_{i,j} + (q-p) \times m_{y,z} \times D_{y,z}(t) - \frac{q}{m_{y,z}} \times D_{y,z}(t)^2\right) x(t) \tag{4}$$

$$D_{y,z}(t) = D_{y,z}(t-1) + d_{y,z}(t-1) \tag{5}$$

$$D(t) = \sum_{(\forall i,j|i \neq j)} \frac{D_{y,z}(t)}{2} \tag{6}$$

$$x(t) = 1 + \beta_1 \ln(\alpha_1) + \beta_2 \ln(\alpha_2) \tag{7}$$

$$C(t) = \alpha_1 \times \sum_{(\forall i,j|i \neq j)} \frac{C_{y,z} \times D_{y,z}(t)}{2} \tag{8}$$

$$S(t) = \alpha_2 - 1 \tag{9}$$

The primary constraint prevents the market intervention factor ($x(t)$) from being negative:

$$x(t) \geq 0 \tag{10}$$

Lastly, while capacity could be below demand ($\alpha_1 < 1$), it doesn't make sense to have a negative effort or cost for supporting activities. Therefore, the following constraints hold for the decision variables:

$$\alpha_1 > 0 \tag{11}$$

$$\alpha_2 \geq 1 \tag{12}$$

The decision model is non-linear and as such a genetic algorithm was used for optimization. The genetic algorithm used a population of 1000, a mutation rate of 0.5, and stopping criteria of one minute between improved solutions.

The factors affecting the decision points of adding capacity versus investing in supporting activities were varied across a number of simulations. R_{demand} was kept constant, while $C_{capacity}$ and $C_{support}$ were varied across a range of values:

$$100 \leq C_{support} \leq 1000 \tag{13}$$

$$0.1 \leq C_{capacity} \leq 3 \tag{14}$$

These can be seen as relative to R_{demand} (and the total potential user base of the network, $\sum m_{y,z}$), and the actual numbers have no physical meaning in themselves. The results have been broken into two charts (Figure 36 and Figure 37) where we see a distinct relationship between adding capacity to the network, supporting activities, and the potential revenue (or whatever other metric is of interest to the user). Figure 36 shows the effort allocated to supporting activities versus the cost of this effort, broken down by the capacity costs. Conversely, Figure 37 shows the capacity ratio versus the capacity costs broken down by the cost increments of the supporting activities. Based on the ratios between revenue, capacity cost, and the cost of supporting activities (relative to their effect on increasing demand) the allocation of resources ranges from building out the network with capacity to demand ratios with no supporting activity, to limiting technical performance of the network to a bare minimum and investing significantly in supporting activities.

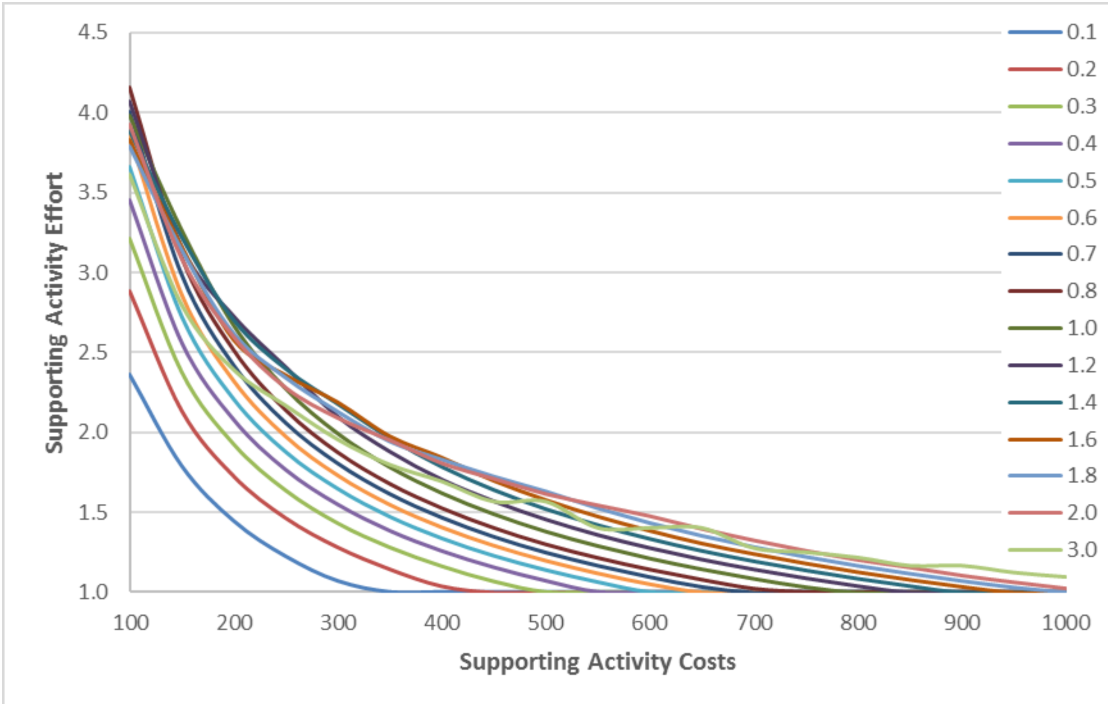


Figure 36 – Supporting activity efforts broken down by capacity costs ranging from 0.1 to 3 (relative to per unit revenue).

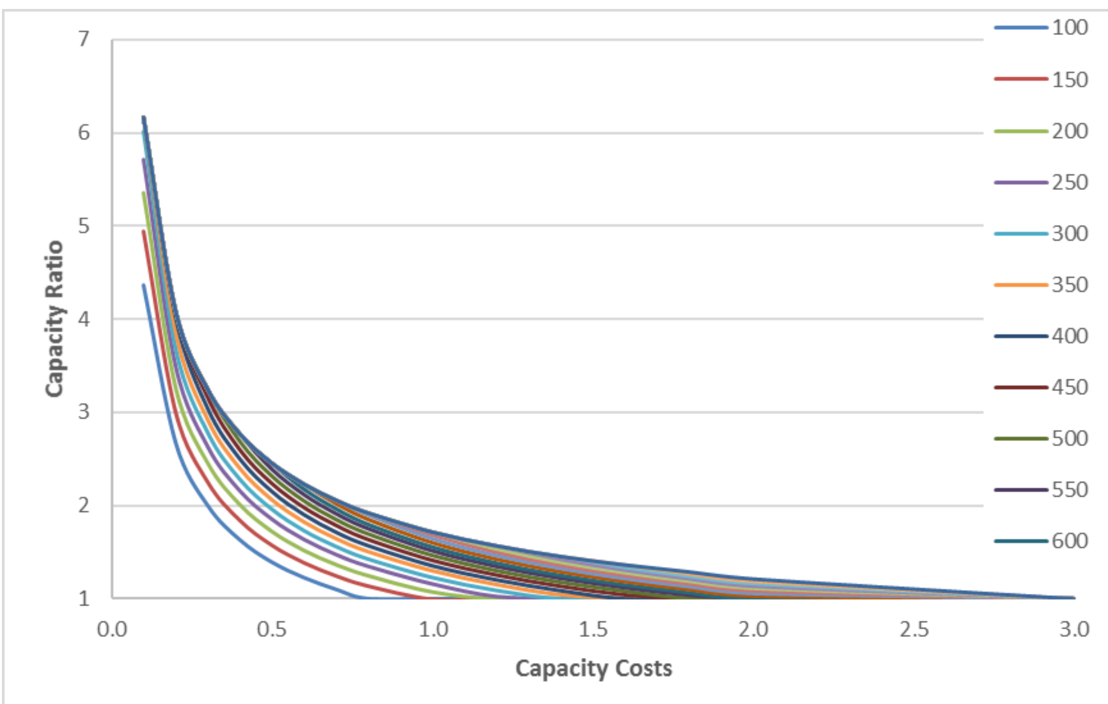


Figure 37 – Excess capacity ratio broken down by the cost levels for supporting activities.

Because the results are across two dimensions (supporting activity costs, and capacity costs), the results in Figure 36 with higher capacity costs appear to have anomalous results. When thought of as a three dimensional surface, it can be seen that the advertising factor eventually decreases as costs grow. Under expensive scenarios, network growth is not as accelerated as quickly. As capacity costs grow, user demand growth shifts to be based on supporting activities and less on excess capacity. There hits a point where growth is too expensive for the diminishing returns, and the supporting activities should be reduced as well. What this demonstrates is that the trade-offs in these factors are lost if a purely linear and disconnected planning process is used.

To re-iterate, this is a model to illustrate how the proposed network planning framework can be used to enhance planning of telecommunication networks, but is limited in scope. One enhancement that is of potential for future work is to incorporate more sophisticated survivability requirements, where reliability needs to be enhanced based on the overall penetration of the market (assuming that a certain percentage of users are fault tolerant, while the majority of users are increasingly not fault tolerant). Another area for further work is to model the buildout in different regions before others to incorporate more complex deployment strategies. By incorporating the network planning framework proposed herein, decision parameters can be quantified and evaluated in a systematic manner that enable a more complete understanding of the impact and opportunities of various network rollout decisions.

4.14 Concluding Discussion

The framework presented, as mentioned in the introduction, is an attempt to provide a systematic way of incorporating the various influencers and decision vectors in the broader network planning and design process. The challenges left unaddressed by this framework are how to assimilate the inputs into the planning process, and how to optimally or even analytically direct and control the identified outputs of the planning process.

A few network planning and design research thrusts address these concerns. Network planning with uncertainty [41] provides a way to incorporate the variability caused by uncertainty in the various inputs into the planning process. Capital and operational expenditure (cap-ex/op-ex) design models [132] can be modified to account for fiscal considerations, although the scope of the financial inputs into the system is broader than just cap-ex and op-ex. Technical capabilities have traditionally been incorporated into network planning and design, as this is directly part of the planning and design process. There are components of technical capabilities that have not been readily addressed, such as technology changes and advances, and interoperability. Application trends, and market realities have been assumed to be dealt with in the demand forecasting process, however as noted in [117], these are not well developed or understood. There appears to be significant room to better develop the understanding of how to integrate aspects of these inputs into the planning and design process.

There are a number of network design paradigms that attempt to provide better mechanisms in network design and planning outputs that can account for variations in the inputs. Recent work in multi-class availability design is one example. This increases granularity in availability design to account for different usage scenarios, and customer needs. Another active area of research is multi-period planning. By using a multi-period planning horizon, networks could be made to be more adaptive to changing demands and technologies. There has been, however, limited discussion into the effects of deployment and ownership strategies. Some of the ownership considerations mimic considerations in railways, power grids, and other systems that tend to be monopolistic in nature, but have not been studied in depth with regard to telecommunication infrastructure (although this starts to get into net neutrality type issues, which there are a plethora of opinions, and research on). There is debate as to the role and impact of various ownership structures, however this is still limited [147]. Impacts or implications of different deployment strategies has received little study, but when looking at reaching underserved regions, is significant.

The impacts and implications of decisions described above as auxiliary functions has had little study. These auxiliary activities can be quite diverse and undertaken by parties quite removed from network planning and operations. They, nonetheless, have an impact on network users' ability to derive value from the network, and hence an impact on the network planning process.

In general, the network planning framework presented broadly covers the influencers and outcomes of long term infrastructure level network planning. Some aspects of this framework have a solid understanding and research base. There are some research initiatives that are attempting to fill gaps in this framework, however there are some key components of this framework that remain poorly understood. By clearly and systematically identifying these gaps, further research work can be better targeted, and diverse research topics can be better brought together to align network planning with the socio-economic realities and goals of the user base for which they are deployed.

Chapter 5. Survivable Network Design with Techno-Economic Growth Effects

This chapter brings together the network planning framework presented in chapter 4 with the design optimization model presented in chapter 3 to explore and articulate the trade-offs of a few of the decision variables articulated in the framework. In particular the decisions around supporting users to utilize the network (through advertising, promoting network based applications, telehealth training, or any other end user based activity) and the fundamental capabilities of the network (the ability to service the estimated current and future traffic demands throughout the network and their required resiliency).

[1] B. Todd and J. Doucette, "Survivable Network Design with Techno-Economic Growth Effects," *To be submitted*.

5.1 Introduction

This chapter outlines an approach to long term network planning that incorporates both network topology design and augmentation with investments in influencing traffic demand growth. When building out networks in areas that have marginal economics such as rural and remote areas, or supporting sectors that may not have strong incentives to utilize the network, such as healthcare, education, or government support, understanding the tradeoff between expanding and augmenting the core network and providing incentive and capability to utilize the network helps maximize the value of the network. In order to do this technical network design has been combined with socio-economic factors that influence the growth and uptake of the network.

One of the motivating factors behind this work was the poor utilization of significant investments made in building a large backbone network that connected every town, village, school, government office and medical center in the province of Alberta, Canada [24]. One of the fundamental aims of this network was to address the digital divide providing high speed access to citizens throughout the province. It was found, however, that the uptake and utilization of the network underperformed expectations[71]. The expansion of the network was not matched with the incentivization and the last mile investments required to cause the intended audience to start using the network. This predicament is not unique to this location. Network demand growth requires the ability and motivation to utilize the network.

With traditional network survivability planning, demand volumes are an input to the process. Capacity and in some cases topology are designed to efficiently provide transport capacity to the estimated demand. This has been taken a step farther within literature to include multiple time periods. In these works, efficient planning is given another dimension. Whether the planning horizon is over a single period or over a longer planning horizon, survivable network design takes in assumptions of demand behavior. This work presents a network planning and optimization model that takes into account the behavior of the users in terms of demand growth. This model allocates resources between network capacity and increasing the capacity of the target users to derive value from the network.

5.2 Background

To take a step away from technical network planning, there are a number of other factors that influence network capacity planning. A framework outlining the influences, stakeholders, and key decisions has been presented here [151]. The inputs affecting the network planning include market realities, technical abilities, fiscal capacity and application trends. Of these, the goal of this work was to include factors in the market realities of the target audience and the technical capabilities and requirements of the network. Market realities describe the target audience of the network, the number of users, and their potential to use the network. Section II.C provides the background on methods used to estimate a user base's potential. Demand forecasting is not generally of concern for technical network planning. However, when the planning problem looks at the allocation of resources between network capabilities and Encouraging or enabling users to use the network, Demand forecasting becomes an important part of the planning design model.

5.2.1 Survivable Network Design

Survivable network design the ability to survive failures and continue to supply all or most of the required traffic demands. Most survivable network designs focus on single failure survivability as this is adequate for most users. Survivable Network design can involve capacity routing with fixed capacity assumptions, capacity allocation with fixed topology assumptions, or full topology design with route planning and capacity allocation.

There are a number of different strategies for allocating spare capacity within the network broken down into two broad categories ring and mesh based strategies. The work here uses path restoration, a mesh based strategy. Path restoration is one of the most efficient mesh based survivability strategies allowing the rerouting of all traffic under each failure scenario to provide the most efficient allocation of spare capacity. There are other strategies such as shared backup path protection, p-cycle, and demand-wise shared protection [7], but path protection was used for it represents the limits of efficiency and it uses a straightforward spare capacity allocation strategy.

The defining characteristic of path restoration is the ability to completely reroute traffic under each failure scenario. What this means is that under every failure scenario the most efficient routing of traffic is used. The implication of this strategy is that it is possible that every demand could experience short outages while traffic is rerouted for an unrelated failure.

Mathematically, path restoration can be described as such:

N – Set of nodes

S – Set of spans (edges) in the network, denoted by the end nodes of the span *(i,j)* where $i, j \in N$

A_n – Set of nodes adjacent to node $n \in N$

D – Set of demands (node pairs)

F – Set of failure scenarios

F_f – Set of spans that are unavailable under failure scenario f

O_r – origin node for demand $r \in D$

D_r – Destination node for demand $r \in D$

Parameters:

C_s – Cost of adding one unit of capacity to span s

d_r– demand volume for demand $r \in D$

Variables:

ω_s – Capacity allocated to span s

$\omega^f_{i,j,r}$ – Capacity allocated to the span connecting nodes i and j for demand r under failure scenario j

$n^f_{i,j,r}$ – Binary variable indicating whether span connecting nodes (i, j) was used (in the direction of node i to node j) for demand r in failure scenario f

Minimize:

$$\sum_{\forall s \in S} C_s \times \omega_s \tag{2}$$

Such that:

$$\omega_s = \sum_{r \in D} \sum_{i,j \in N_{s_1} | i \neq j} \left(\omega^r_{i,j} + \omega^r_{j,i}\right) \times d^t_r \qquad \forall s \in S \tag{3}$$

$$\sum_{j \in A_i} \omega^f_{i,j,r} \geq d_r \qquad \forall r \in D, \forall f \in F, \forall i \in O_r \tag{4}$$

$$\omega^f_{i,j,r} \leq n^f_{i,j,r} \times M \qquad \forall i \in N, \forall j \in A_n, \forall r \in D, \forall f \in F \tag{5}$$

$$\sum_{j \in A_i} \omega^f_{i,j,r} - \sum_{j \in A_i} \omega^f_{j,i,r} = 0 \qquad \forall i \in N | i \neq O_r, D_r, \; \forall r \in D, \forall f \in F \tag{6}$$

$$\sum_{j \in A_i} n^f_{i,j,r} - \sum_{j \in A_i} n^f_{j,i,r} = 0 \qquad \forall i \in N | i \neq O_r, D_r, \; \forall r \in D, \forall f \in F \tag{7}$$

$$\omega_s = 0 \qquad \forall f \in F, \forall s \in F_f \tag{8}$$

This path restoration formulation uses an arc flow approach instead of pre-calculating potential path candidates for each demand. This approach is required for topology augmentation to scale the problem space with even small networks [84]. The objective of this model is to minimize the cost of capacity allocated in the network (26). The capacity for each span is the maximum required on the span across all failure scenarios (3). This model uses a multi-flow approach [88] where multiple paths can be used to transport capacity so long together all of the paths for a given demand are sufficient to route the demand's traffic volume (4). When routing each path, they must remain node disjoint ((5)-(7)). Finally, to ensure the path(s) chosen do not include failed spans for each failure scenario, the capacity allocated to that span must be 0 (8)

The work presented here is not dependent on path restoration for its survivability algorithm, and the behavior of key decision variables under different approaches to survivability is an area for further study. As mentioned, path restoration was selected because it provides a simple spare capacity allocation algorithm along with a high level of capacity sharing.

5.2.2 Topology Design

A challenge of network planning is the augmentation of a network. Adding a new physical link can be quite costly [152], but can greatly reduce the amount of spare capacity in a network. Finding the balance of investing in physical connections of a network against adding capacity elsewhere in the network, and in the case of the work presented here, increasing user demand, is not simple. There are a number of factors that affect topology design, such as the ratio of capacity cost to the cost of establishing a new span, the modularity of capacity, the distribution pattern of nodes in the network, and the relative distribution of demand volumes [153].

Incorporating topology augmentation into survivable network optimization has a significant impact on the scalability of the algorithm. It was effectively impractical to pre-calculate paths within the network, as these would have to take into account all of the different topology configuration options.

Topology optimization requires an arc-flow approach where paths are dynamically computed as part of the optimization process. This necessitates a large number of binary decision variables (9) covering each potential span (10) and each demand (11) that could route traffic over it that that grows in the order of $O(n^4)$ in order to ensure simple path routing and no demand splitting at transit points.

B(n) – number of decision variables for traffic routing in a network with *n* nodes

d(n) – number of node pair demands in a network with *n* nodes

s(n) – Number of potential spans that can be added to a network with *n* nodes

$$B(n) = d(n) \times s(n) \tag{9}$$

Where the number of potential spans are also in the same order of magnitude

$$s(n) = (n \times (n-1))/2 \tag{10}$$

And the number of demands in a network can be estimated at:

$$d(n) = (n \times (n-1))/2 \tag{11}$$

Practical considerations can limit the number of candidate spans that go into a topology optimization. The number of candidate spans may not be exactly as calculated in (10), but grow at a rate of $O(n^2)$. [82] provides a discussion about the complexity of topology design algorithms.

Topology optimization is a difficult problem and typically requires heuristic optimization [82] or limited network augmentation scope. Most network buildouts are based on some infrastructure already installed so having a limited number of potential new spans within a network is a reasonable assumption. This work, however, focuses on a greenfield network design problem. This topology augmentation requires a few additional decision variables and a constraint added to equations (3)-(8), and replaces the cost function (2).

Sets:

$S_{i,j}$ – Spans connecting nodes *i* and *j*

Parameters:

V_s – Cost of establishing a span *s*

Variable:

n_s – Binary variable indicating whether span connecting nodes (i, j) was included in the topology

Minimize:

$$\sum_{\forall(i,j)\in S} (C_s \times \omega_s + V_s \times n_s) \tag{12}$$

Where:

$$\sum_{\forall f\in F} \sum_{\forall r\in D} n_{i,j,r}^{f} \leq M \times n_s \qquad \forall i \in N, \forall j \in A_i, \quad \forall s \in S_{i,j} \tag{13}$$

The updated cost function (12) includes the cost of the capacity $(c_s \times \omega_s)$ and the cost of adding spans into the network $(x_s \times n_s)$. A constraint was added to the model in order to ensure that if a span is used in any failure scenario for any demand, that span must be included in the network (3), and included in the cost calculation.

5.2.3 Demand Growth Models

Modeling and estimating demand growth is a well-established discipline, with numerous techniques incorporating a multitude of factors [46]. Like most areas of business, the demand for a given telecommunication services has a degree of uncertainty with it. With the large capital investment and long payback time for telecommunication networks [136] reasonable estimates of demand growth predicates good network design and helps ensure the long-term viability of the network.

[154] provides an overview of telecommunication demand forecasting, including the key factors that go into producing a forecast. These include the typical marketing mix of price, product, promotion, and place, as well as an additional parameter, permission. But beyond the marketing mix, the challenge is creating a demand forecast model on which these variables would exert their influence. [43] provides a summary of mathematical models for both existing growth and new product expansion.

p – Coefficient of innovation

q – Coefficient of imitation

$h(t)$ – Probability of adoption at time t

$N(t)$ – Cumulative number of adopters at time t

$n(t)$ – Number of new adopters at time t ($\frac{dN(t)}{dt}$)

m – Total number of potential adopters

$$h(t) = p + \frac{q \times N(t)}{m} \quad (14)$$

$$n(t) = p \times m + (q - p)N(t) - \frac{q \times N(t)^2}{m} \qquad (15)$$

The most commonly used mathematical models used for forecasting demand are based off of the Bass model, first published in 1969 [13], [43]. This model (and most others) label new users or consumers of a product or service as adopters. In the bass model, adopters are broken down into two categories innovators (denoted with the variable p) and imitators (denoted with the variable q). The probability of a new adopter at any point in time is a function of the innovator coefficient the and plus the imitation coefficient multiplied by the cumulative number of adopters at that point in time divided by the total potential adopters (14). This function leads to a first order differential equation representing the number of new adopters over a period of time (15).

The shortcoming of this model is that there is no room to predict the impact of interventions (typically advertising, but comprehensively any action) on the rate of adoption. In 1994 Bass, Krishnan, and Jain published a general bass model [149] that incorporates a concept they call "marketing efforts" into the model (16). In the context of the paper, the adoption rate is modified by this marketing effort, however, they demonstrate that this is generalizable into any effort that impacts the adoption rate.

$x(t)$ –Marketing effort at time t

$$h(t) = \left(p + \frac{q \times N(t)}{m}\right) x(t) \qquad (16)$$

In the case where each decision variable, α_1, in $x(t)$ maintains a constant proportion to the value in the preceding time, $x(t)$ can be modeled as a summation of natural log of each decision variable multiplied by a coefficient representing the effectiveness, β_i, of each decision variable (26) [149]. This form uses a diminishing returns assumption for the magnitude of the decision variables.

DV – Set of decision variables

α_i – decision variable *i*

β_i – co-efficient of effectiveness for each decision variable *i*

$$x(t) = 1 + \sum_{i \in DV} \beta_i \times \ln(\alpha_1) \tag{17}$$

Another demand forecasting model of note to telecommunications is the rate of growth in the presence of supply restrictions [155]. Because networks can take a significant amount of time to build out, user growth can be restricted by the actual availability (in the sense of the infrastructure being in place) of the network. This model allows for the buildup of user demand if there are timeline restrictions in the deployment of the network. The work presented below makes the assumption that the network is built out in step with growing user demand. Incorporating supply restrictions is an area open for further study.

There are several other demand growth models that could be utilized [43]. For this work the Generalized Bass Model was chosen for its quality of forecasting and its simplicity in calculation.

5.2.4 Network planning

The work done in this paper falls within the context of multi-period network planning. We do not provide a novel multi-period planning model, there are a number out there ([20], [21], [35], [36], [38], [84], [156]), but the purpose of this work is to model and design network capacity and topology in the larger framework of network planning as discussed in [151].

This framework (Figure 38) articulates the key inputs, outputs and stakeholders in planning and deploying telecommunication networks. Inputs into network planning include the ability and capacity of the potential users to use and pay for the network (market realities), the ability to pay for the construction of the network (fiscal capacity), the capabilities of current telecom technology (technical abilities), as well as broader trends in application usage and maturity (application trends). Broadly speaking the decisions being made in the network planning process include the network design (technical performance), the staging of when the network gets built (deployment strategies), who owns and controls access to the network (ownership and access), and lastly the set of activities that will encourage or facilitate use of the network (auxiliary activities).

The planning framework acknowledges the broad range of users and their related purposes for using the network by categorizing six key drivers. Each of these groups are important, as users in each will exhibit unique innovation adoption profiles (p and q values in (14)).

The techno-economic network growth model presented in this paper incorporates several these inputs and outputs articulated in this framework.

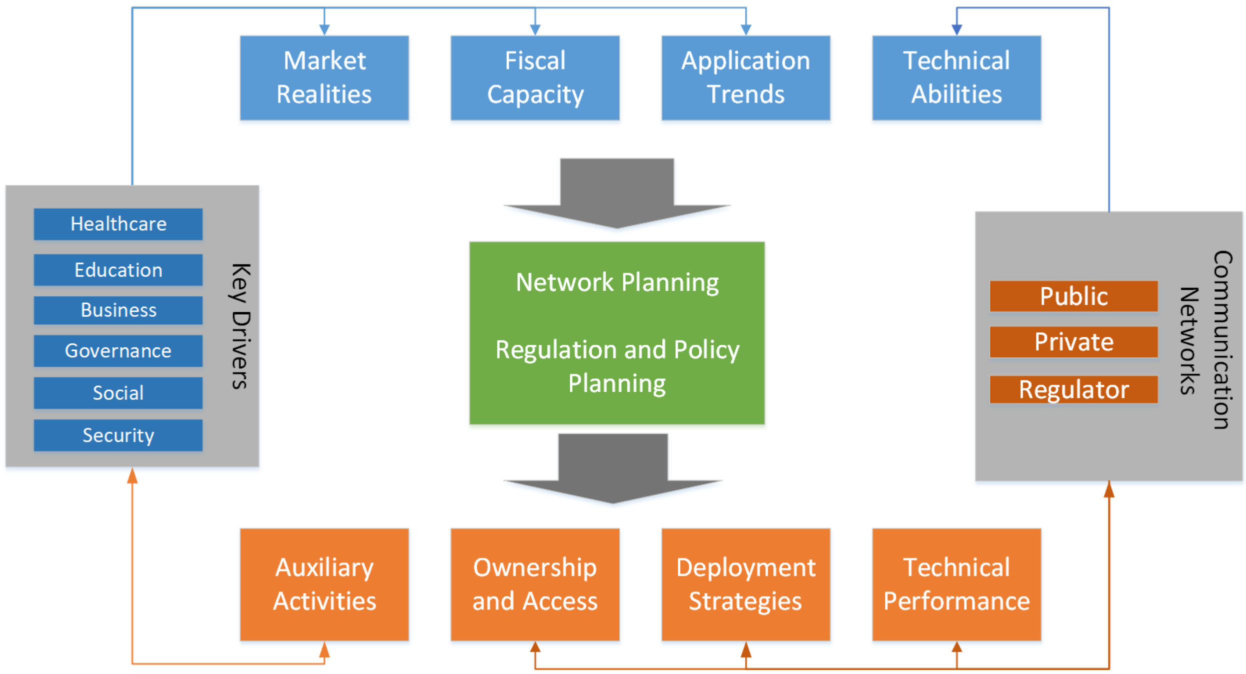


Figure 38 – Outline of the influences, inputs and outcomes of telecommunication network planning [151]. This is a duplicate of figure 28.

5.3 Incorporating Techno-Economic Factors into Reliable Network Planning

The purpose of this paper is to provide an integrated optimization model that combines reliable telecommunication network planning over time with design parameters that account for and influence network usage. This integrated model is referred to the techno-economic network growth model throughout this paper. We have based the design parameters in this model after the network planning framework presented in [151], with a focus on the market realities that the network is being deployed into (the number of users and the rate of growth) on the inputs side, and a combination of technical performance of the network and auxiliary activities on the output side.

The Generalized Bass Model provides decision variables on which the investment in the capacity of users to use the network over time can be balanced with the appropriate investment in network infrastructure.

5.3.1 Articulate Techno-Economic Factors

An underlying assumption in the techno-economic network growth model is that users will adopt network enabled technologies in a manner that causes network traffic growth to follow the behavior modeled in the Bass model [13]. The innovator and imitator coefficients model the market realities and account for the rate that users will adapt these network-enabled technologies.

It is acknowledged that not all technologies will drive network traffic in the same way. The assumption being made is that the overall traffic demand will follow the s-curve in the Bass model. There may be multiple epoch's or sets of technology that can be modeled,

The other two factors that are incorporated into this model represent an aggregate of what [151] refers to as auxiliary activities and investment in the technical performance of the network. The auxiliary activities encompass the activities that a network stakeholder could undertake that would encourage end users to use network enabled applications more and/or sooner. These can be thought of as advertising or awareness campaigns, but could also include training, support, or the investment in building out specific applications. This is generic, as the set of auxiliary activities depends on the key drivers behind the network deployment. The impact of increasing these auxiliary activities (referred to as the *user capability factor,* U_t) was included in the calculation of user demand growth as a part of the Generalized Bass Model .

The technical performance of the network in this case encompassed the degree to which capacity in the network matched user demand. The option was given to underserve or overserve user demand with a corresponding impact on future user growth (referred to as the *network capacity factor* N_t). The assumption was made that an under provisioned network would slow user demand growth, and conversely overprovisioning would increase demand. For example, if medical professionals attempting to implement a tele-health system consistently had connection issues, the adoption of telehealth would slow down significantly. Conversely, if users know that connectivity would not be an issue, they would be more likely to not just adopt the telehealth system, but other offshoot technologies, accelerating traffic demand in the future.

U_t – User capability factor at time t

N_t– Network Capacity factor at time t

x_t – Combined effort to influence user growth at time t

$$x_t = 1 + \alpha_1 \times \ln(U_t) + \alpha_2 \times \ln(N_t) \qquad \forall t \in T \qquad (18)$$

The Generalized Bass Model for product adoption accounts for the user capability and network capacity factors in what was referred to in the paper as the current marketing effort, along with calibration factors to adjust the weighting between these factors. The amplitude of these factors are limited by the nature of the factor and structure of the Bass model. Combined, the factors of x(t), p and q cannot be more than one (19) or else the Generalized Bass Model would be invalid (16). The user capability factor cannot be below 1 (20) as this represents no investment into auxiliary activities. It does not make sense that investments would be made to dissuade users from using the network. On the other hand, the network capacity factor could be below 1 (21), representing a scenario where the network was under provisioned. Because of the logarithmic nature of the impact of the factors (18), under provisioning network capacity has a significant retardation of future capacity growth.

$$x_t \times (p + q) < 1 \qquad \forall t \in T \qquad (19)$$

$$U_t \geq 1 \qquad \forall t \in T \qquad (20)$$

$$N_t > 0 \qquad \forall t \in T \qquad (21)$$

Other outputs such as ownership and access, and deployment strategies were not included in the techno-economic network growth model. Ownership and access decisions do not have a direct impact on network design as these decisions involve business or policy factors. Deployment strategy decisions could be modeled within the framework where reliability levels and even whether and when a node is connected into the network become part of the decision variables where the traffic demand growth profile would augment the Generalized Bass Model [149] with the presence of supply restrictions [155].

5.3.2 Growth Model

In order to include dynamic demand growth based on the decision variables of user capability (U_t) and network capacity (N_t), the model used the concepts from the Generalized Bass Model for demand growth. This dynamic growth makes the model non-linear, and ineligible for integer linear programming techniques. The calculation of traffic demand, capacity requirements, the cost of increasing user capability, and the revenue are presented below.

Sets:

T – Set of time periods

Parameters:

R – Unit revenue per unit of traffic demand

Variables:

d_r^t – Demand volume at time $t \in T$ for demand $r \in D$

a_r^t – Capacity requirement at time $t \in T$ for demand $r \in D$

Uc_r^t – Cost of user capability at time $t \in T$ for demand $r \in D$

R_r^t – Revenue at time $t \in T$ for demand $r \in D$

i – Discount rate

$$d_r^t = d_r^{t-1} + \left(m \times p + (q-p) \times d_r^{t-1} + \frac{q}{m} \times (d_r^{t-1})^2\right) \times x_t \qquad (22)$$

$$a_r^t = d_r^t \times N_t \qquad \forall t \in T \qquad (23)$$

$$Uc_r^t = (U_t - 1) \times \frac{Uc}{(1+i)^t} \qquad \forall t \in T \qquad (24)$$

$$R_r^t = d_r^t \times \frac{R}{(1+\ i)^t} \qquad \forall t \in T \qquad (25)$$

Demand was calculated by transforming the conditional probability of using the network (16) into an estimated user demand for each time frame of the model (22), where x is calculated with (18). [149] provides more details behind this formulation. The cost of the network capacity decision variable was included by modifying the capacity requirements for the network to over (or under) build based on estimated demand. The capacity requirements for each demand was calculated based on the capacity multiplier N_t and the estimated traffic demand (23). The cost of increasing user capabilities was calculated based on a constant cost based on the assumption that increasing user capability is dependent on the entire potential set of users, rather than current users, discounted appropriately for each time period (24). This user capability cost could be modified to be more representative of the type of activities this would entail (i.e. training sessions, or subsidized equipment for new users, general advertising, or constant support for all current users). The last part of the growth model was a simple revenue calculation with the discounted per unit revenue multiplied by the number of users (25). There are several assumptions put into this calculation where each unit of traffic demand generates the same revenue over time (a metered based approach) where this is obviously could be tweaked for other scenarios. This formula was used as a generic representation of the value users receive out of the network (as there are many scenarios where the core network does not directly generate revenue) and assumes that each unit of traffic represents a similar value to the users of the network.

5.3.3 Incorporating the growth model into survivable network design with a genetic algorithm

The model presented in section 5.2.1 was augmented to include time-period dependencies. This was done using the strategy laid out in [153]. This modification to the model ensured capacity for each time period was sufficient to route demand under each failure scenario, as well as to ensure that capacity was maintained or grown for each span over time. Costs were calculated using the discount rate i.

This model is non-linear due to the ability to change capacity and demand requirements through a decision variable

Sets:

T_t – Set of time periods up and including to $t \in T$.

P – Set of spans denoted by a span identifier (rather than the pair of end nodes)

Modified time dependent parameters:

V_t^s - Cost of implementing span s at time t.

C_t^s – Capacity unit cost at time t for span s.

Modified time dependent variables:

ω_s^t – Capacity allocated to span s at time t

$\omega_{i,j,r}^{f,t}$ – Capacity allocated to the span connecting nodes i and j for demand $r \in D$ under failure scenario j at time t

$n_{i,j,r}^{f,t}$ – Binary variable indicating whether span connecting nodes (i,j) was used (in the direction of node i to node j) for demand $r \in D$ in failure scenario f at time t

n_s^t – Binary variable indicating whether span connecting nodes (i,j) was included in the topology at time t

Additional variables that govern multi-period optimization:

y_t^s – 1 if span s changes to > 0 from 0 at step t.

$\boldsymbol{v_t^s}$ – Cost of implementing span s at time t .

$\boldsymbol{c_t^s}$ – Capacity cost at time t for span s.

Maximize:

$$\sum_{r \in D} \sum_{t \in T} R_r^t - Uc_r^t - \sum_{\forall s \in S} (c_s^t + v_s^t) \tag{26}$$

Multi-period constraints:

$$\sum_{t_c \in T_t} (\omega_s^{t_c} - M \times y_s^{t_c}) \leq 0 \qquad \forall s \in S, \forall t \in T \qquad (27)$$

$$v_s^t \geq y_s^t \times V_s^t \qquad \forall s \in S, \forall t \in T \qquad (28)$$

$$c_t^s \geq C_t^s \times (\omega_t^s - \omega_{t-1}^s) \qquad \forall s \in S, \forall t \in T, \qquad (29)$$

$$\omega_s^t \geq \omega_s^{(t-1)} \qquad \forall s \in S, \forall t \in T \qquad (30)$$

Path survivability constraints:

$$\omega_s^t = \sum_{r \in D} \sum_{i,j \in N_{s_1} | i \neq j} \left(\omega_{i,j}^{r,t} + \omega_{j,i}^{r,t}\right) \qquad \forall t \in T, \forall f \in F, \forall i \in N, \forall j \in A_n \qquad (31)$$

$$\sum_{r \in D} \omega_{i,j,r}^{f,t} \leq \omega_{(i,j)}^t \qquad \forall t \in T, \forall f \in F, \forall i \in N, \forall j \in A_n \qquad (32)$$

$$\sum_{j \in A_i} \omega_{i,j,r}^{f,t} \geq \alpha_r^t \qquad \forall t \in T, \forall r \in D, \forall f \in F, \forall i \in O_r \qquad (33)$$

$$\omega_{i,j,r}^{f,t} \leq n_{i,j,r}^{f,t} \times M \qquad \forall t \in T, \forall i \in N, \forall j \in A_n, \forall r \in D, \forall f \in F \qquad (34)$$

$$\sum_{j \in A_i} \omega_{i,j,r}^{f,t} = \sum_{j \in A_i} \omega_{j,i,r}^{f,t} \qquad \forall t \in T, \forall i \in N | i \neq O_r, D_r, \forall r \in D, \forall f \in F \qquad (35)$$

$$\sum_{j \in A_i} n_{i,j,r}^{f,t} - \sum_{j \in A_i} n_{j,i,r}^{f,t} = 0 \qquad \forall t \in T, \forall i \in N | i \neq O_r, D_r, \forall r \in D, \forall f \in F \qquad (36)$$

$$\omega_{(i,j)} = 0 \qquad \forall t \in T, \forall f \in F, \forall (i,j) \in F_f \qquad (37)$$

$$\sum_{\forall f \in F} \sum_{\forall r \in D} n_{i,j,r}^{f,t} \leq M \times n_{i,j}^t \qquad \forall t \in T, \forall i \in N, \forall j \in A_i \qquad (38)$$

$$n_{i,j}^t \leq n_{j,i}^t \qquad \forall t \in T, \forall i \in N, \forall j \in A_i \qquad (39)$$

The constraints (18) through (25) are also included in the model to govern the growth of traffic demand based on the decision variables U_t and N_t.

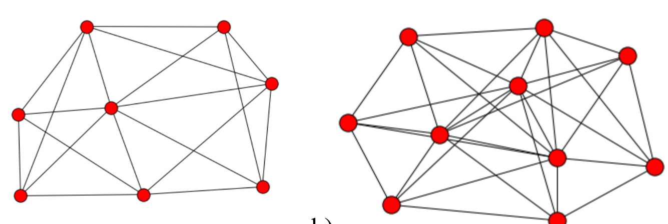
The multi-period constraints are described in detail in [153]. They ensure that the cost of the span is accounted for only in the period that the span is first used with equations (22) and (23). Span costs are only the incremental amount between time periods (24), and a span cannot reduce its capacity (25).

The path restoration formulation is included again with two changes. The first is that every equation is now indexed by the period and the second is that the demand volume is now the calculated volume based on user growth and the network capacity factor (33).

This mathematical formulation of the techno-economic network growth model is significantly complex with non-linear constraints. Heuristic implementations of this model are therefore required. The implementation of the model using a genetic algorithm [7] is presented below.

5.4 Simulation Setup

The genetic algorithm presented above was implemented and run using the DEAP (Distributed Evolutionary Algorithm for Python) evolutionary computing framework [157] on four networks with node counts of 8, 10, 12 and 15 (Figure 39). The computational complexity was found to be a limiting factor, with the 8 node network taking near 4 hours to reach the end conditions for the algorithm and the 15 node network running for over a full day. The calculation time for a 25-node network was in an order of a couple weeks. We found that the 15 node network provided a balance between computational capabilities and the ability to meaningfully provide insight and validation into the techno-economic network growth model. All results presented in this paper were from the 15 node network and were in line with the results from other networks.



a) b)

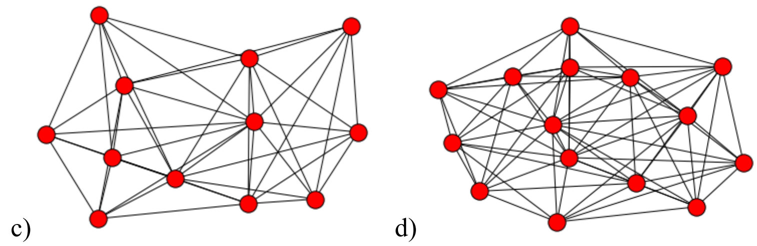


Figure 39 - Topology designs for the 8 node network (a), 10 node network (b), 12 node network (c), and the 15 node network (d)

The algorithm was run using a single core on a 3.5GHz Intel i7 computer with 32 GB of memory. There are a number of options that could be taken to improve performance (like enabling the algorithm to run in parallel, and for longer periods of time), however the results from these simulations were deemed to be sufficient to evaluate the impact of the cost of the network against the cost of encouraging usage growth in the target audience of the network. There were three variables that were evaluated relative to one another, the cost of adding capacity to the network, the estimated revenue per user, and the cost of influencing the user base. The values for these factors were not deemed as important, as they could change drastically between deployments. Rather the significance of these values was in reference to one another. As such, we kept the cost of influencing the user base constant and varied the cost of the network and the per unit revenue (Table 5).

Table 5 – Capacity, revenue and unfluence factor run for each network

Per Unit Capacity Cost	Per Unit Revenue	Cost of Influencing Users
10,000	10000	1000000
1,000	10000	1000000
100	10000	1000000
10	10000	1000000
10,000	1000	1000000
1,000	1000	1000000
100	1000	1000000
10	1000	1000000
10,000	100	1000000
1,000	100	1000000
100	100	1000000
10	100	1000000

The growth of demand for the network used a coefficient of innovation (p) was 0.06 and the coefficient of imitation (q) was 0.3, as discussed in [150]. With these parameters, demand growth becomes minimal after approximately 20 time periods across a range of combined marketing efforts, $x(t)$ (Figure 40).

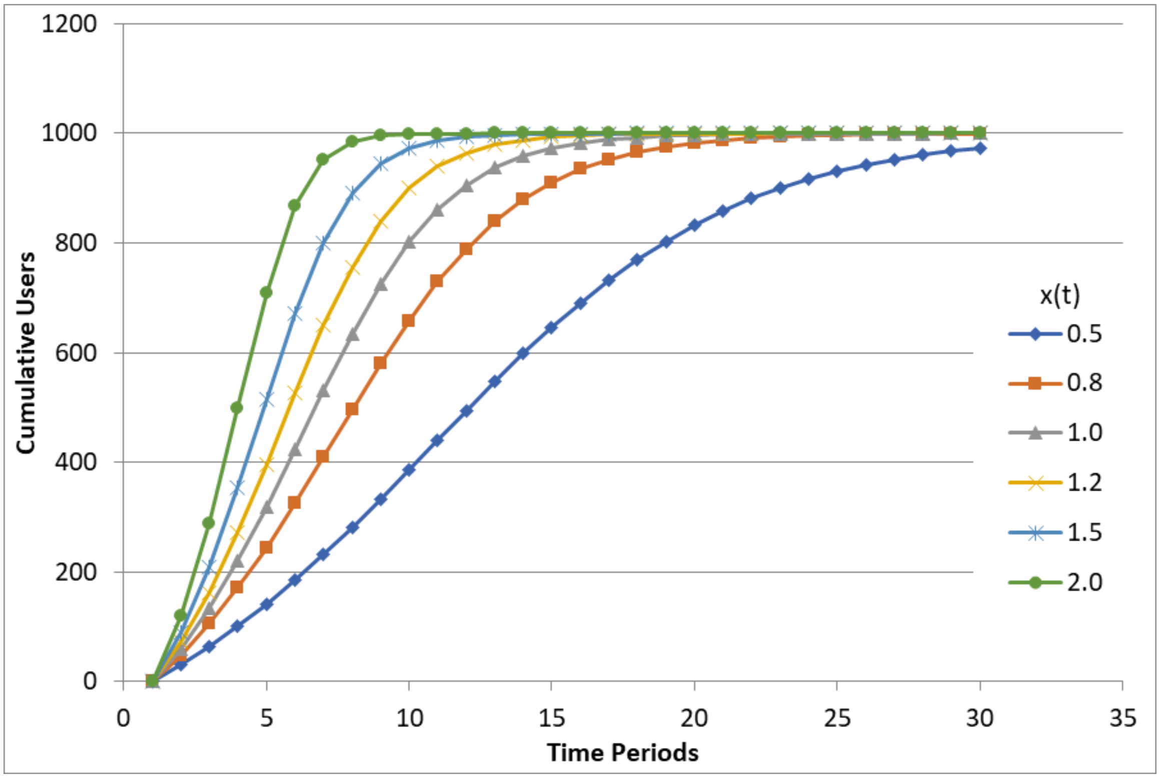


Figure 40 – Cumulative users over time with for the modified Bass model (16) with p=0.06, q=0.03 $x(t)$ ranging between 0.5 and 2

This range of factors and time periods provide sufficient scope to evaluate the effects of mixing the investment in growing the network and building the capacity of the user base to use it.

The cost of adding capacity was set to be proportional to the distance between the end nodes of the span. This represents the case where the number of repeaters or power of transmission increased with distance and had an incremental impact on cost. There may be cases where this assumption is not the case. The study of the designs under alternative cost assumptions and scenarios is left for future work. If cost of capacity was not proportional to the distance of the network, then it is recognized the topology of the networks could change, possibly in a significant manner as modern data traffic does not exhibit significant locality [7].

5.4.1 Genetic Algorithm for the Techno-Economic Network Growth Model

The techno-economic network growth model presented in section 5.3 was implemented using a genetic algorithm. The value function of the algorithm is the discounted revenue minus the discounted user capability, capacity and span establishment costs (26).

This implementation of the GA used a binary genetic encoding where the variables indicated whether a candidate span was in a network topology ($n_{i,j}$). The penalty function was a measure of the number of paths that could not be routed with a given topology plus the number of invalid factors. The fitness function of the GA was the cost of establishing a span between two nodes plus the cost of adding capacity to the network (26). The capacity requirements for each span were calculated such that the constraints (27) through (39) were met. The demand routing was done using Dykstra's shortest path algorithm for each demand in each failure scenario (for practical purposes, these were cached between iterations of the GA and only recalculated when needed). This simplification inevitably reduced the efficiency of the network design but was done to make the algorithm computationally.

Variables:

y – Genetically encoded individual

s_i – Integer representing the period span *i* was added to the network

U – User capability factor

C – Network capacity factor

$$y = \{s_i | i \in S \ + (U, C)\ \} \tag{40}$$

$$0 \le s_i \le |T| \qquad \forall i \in S \tag{41}$$

The genetic algorithm defined the individual in two parts. The first was a set of integers that represent the period that the span was implemented (26). The integer ranged between 0 (implemented in the start of the network) to the cardinality of set of time, where this maximum value represented the span not being implemented (41). The second part captured the user capability and the network capacity factors as two floating number values (26).

It was found that by having purely random mutations for if and when spans were implemented, it was difficult to find feasible topologies. The mutation process was modified to randomly select a span to modify, that span was then moved to different, randomly selected period. If the network had a node with a degree less than two at any point in time, a random span adjacent to the node was then added to the network at that point in time. The factor mutation process involved randomly selecting the first factor to mutate to set to a randomly selected but potentially valid value. Next the combined factor was randomly selected between 0 and 1. The other factor was then calculated based on those two random values.

The crossover function took a random crossover point in the parent individuals for the set of integers and separately the user capability and the network capacity factors. Because the factors had completely separate meaning, and a simple crossover would automatically mean that the child was infeasible (because the floating number factors would have been representing spans), the results and the variability of using this two-part crossover function was found to be most useful.

The feasibility of an individual was determined by whether the network capacity (C) and user capability (U)factors were valid (resulting in a value of x between 0 and 1), and whether every demand was routable under every failure scenario. A number of population sizes, mutation probabilities, crossover probabilities and size of surviving individuals were tested to balance computational complexity, memory usage and optimality progress.

A mu + lambda approach was taken where the best fit individuals for the next generation come from both the offspring and the previous generation [157]. The population was set at 50 with 20 offspring generated for 100 generations. The crossover and mutation probabilities were initially set to 0.3 and 0.2 respectively with a uniform distribution. After initial trials, it was found that taking a phased approach to optimization with three different mutation probabilities was most effective. The mutation probability of the network topology was separated from the probability of mutation of the network capacity and user capability factors. The first phase had a topology mutation probability of 0.3 and a factor mutation probability of 0.2. The second phase only mutated topology and the third phase only mutated the factors. Each phase consisted of 100 generations. The initial population was seeded with feasible individuals as it was difficult for reasible solutions to be found with a random initial population. This was due to the biconnected requirements for single failure restorability.

To ensure the final results were at least difficult to find better solutions, the number of generations were varied, with 3 simulations run for 1000 generations. It was found that the results did not improve significantly after 75 generations. The choice of 100 generations provided a buffer to ensure stability of the final solution. Randomly selected individual optimizations were repeated to ensure the repeatability of the results. While there some differences in the repeated simulations, they did not materially change the final solution.

5.5 Results

The purpose of this work was to examine the relationship between investment in the expansion of the network and the capacity of the user base to use the network and to build an understanding of some of the key metrics that would influence balance of this allocation. As mentioned, the factors that were varied were the costs of adding capacity to the network and the revenue per unit of capacity. Through the simulations we were able to observe the limits where these factors would saturate the resulting optimal allocation of resources. The results for the simulations with constant capacity costs (and variable span establishment costs) are presented first, followed by the results where the entire cost profile (capacity and span establishment costs) was manipulated.

For each of these scenarios there are 4 things we want to look at:

1. User capabilities factor
2. Network capacity factor
3. Ratio of user capabilities to network overbuild
4. x(t)

The factors, ratios and results presented here are based on the assumptions in the Generalized Bass Model with no calibration in the degree of influence between these factors (16). Specific network designs would require calibration across all the factors in the model. The insights and observations drawn from the simulations are intended to provide validation that the techno-economic network growth model makes intuitive sense, give the reader an overview of the tradeoff analysis that can be done with the model, as well as to highlight conditions where major changes in the structure of the network or the allocation of resources takes place under the assumptions used.

5.5.1 Simulation results with static per unit capacity costs

As mentioned, we ran simulations over several scenarios, varying the cost of the network and the cost of influencing the user capabilities. The first set of simulations varied the span establishment costs while keeping per unit capacity costs static. The second set of simulations kept a span establishment cost ratio to unit capacity costs to a constant 100 (seen to be where a balance between these costs are typically found [153]. The total value of each simulation can be seen in Figure 41 and Figure 42, where Figure 42 is truncated to display the differences in value for simulations with a unit revenue of 100 and 1000. In general, the model followed as expected, with the unit revenue driving value (as seen by the relatively flat values across span establishment cost multipliers for each level of unit revenue). The span establishment cost did play a part with the general downward trend as costs increased, but only a minor one as the revenue or capacity costs dwarfed the span establishment costs in some scenarios. The authors note this is generally unrealistic, but are included as this work is to validate the accuracy of the model.

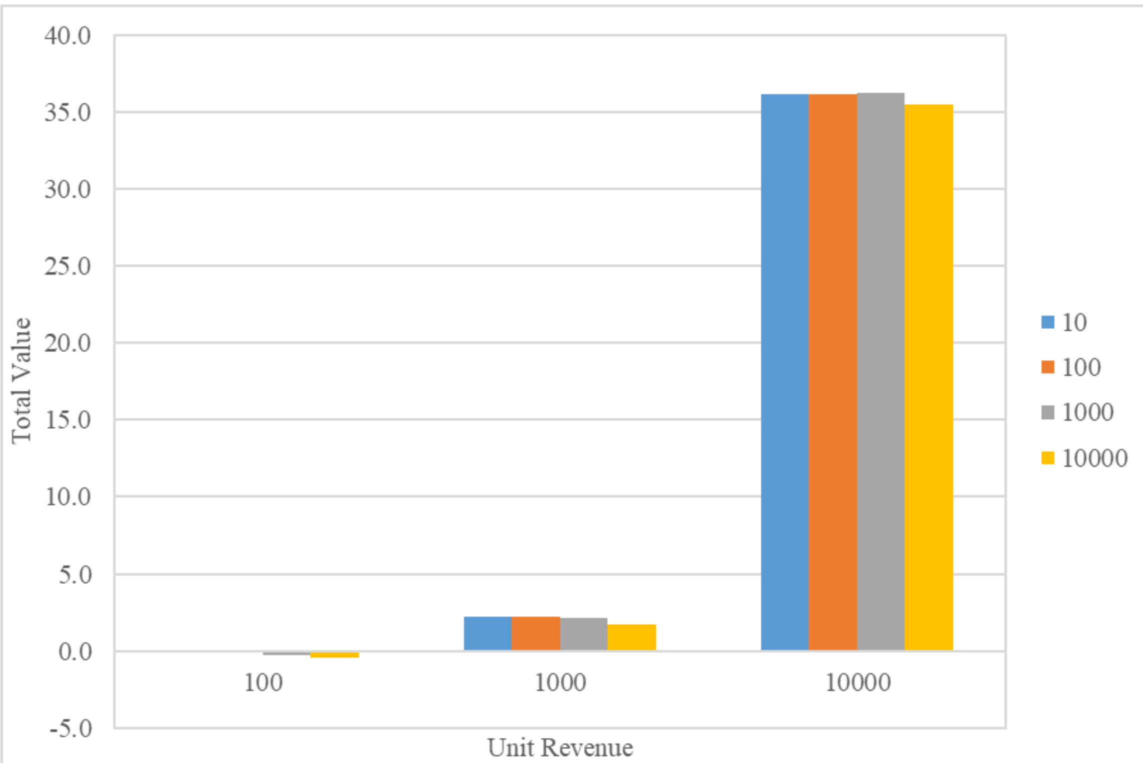


Figure 41 - Total value for simulations with a static per unit capacity cost for span establishment costs of 10, 100, 1000, and 10 000 times the length of the span

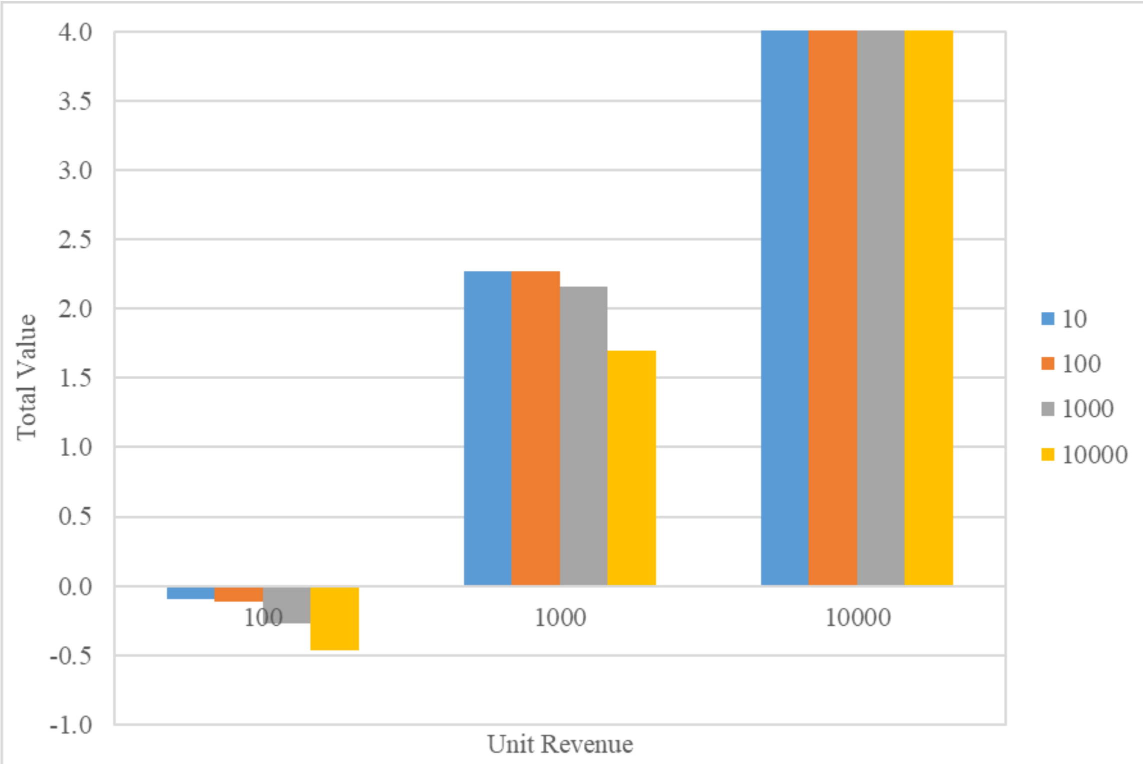


Figure 42 – Total value for simulations with a static per unit capacity cost with the scale truncated to 4.0×10^9

The details of concern are the allocation of resources to user capability vs. network expansion. This was analyzed by looking at the ratio of the two respective factors in the model, the network capacity and user capability. While there is some variation in this ratio due to the stochastic nature of the genetic algorithm, some key trends are obvious. In the simulations with a unit revenue of 100, the emphasis was on user capability, whereas with a unit revenue of 1000, the emphasis was on network capacity. What was observed when networks are highly valued (represented by unit revenues of 10 000), was that both the network capacity and the user capability were emphasized equally. This was because the weighting factor in the model provided equal ability of increasing user capability and network capacity in increasing network demand. This was a presupposition of the model used, but it provides a mechanism to alter this emphasis [149] by adjusting the weighing of each factor.

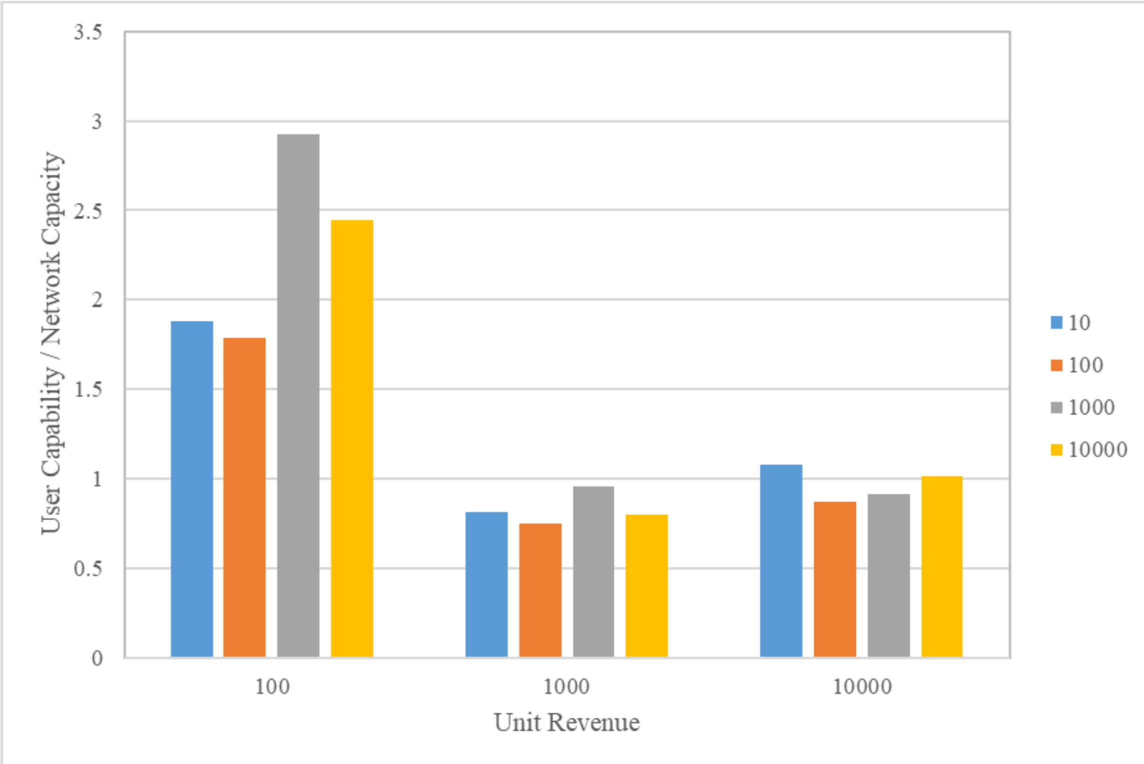


Figure 43 – The ratio of user capability vs network capacity for simulations with a static per unit capacity cost.

The next chart used to understand the simulation results is the total combined modifier *x* (see equation (16)) across the variations of unit revenue and span establishment costs (Figure 44). With low per unit revenue, the modifier was below zero, meaning that capacity in the network underserved demand. Combined with the results in Figure 43, we can see distinct trends as to when there was an emphasis on user capability and network capacity, or to maintain a balanced emphasis on both (as seen with the simulations using a unit revenue of 10 000).

While all four simulations with a unit revenue of 100 had a total value below zero (Figure 42), the combined modifier did go down as the costs increased. The reason it wasn't as low as it could go for all four scenarios was that with lower span establishment costs, the per period value (revenue – costs calculated at each time-period) became positive in later time periods. This is significant because the results show that the model provides a mechanism to understand the minimal subsidy required to get a network into a self-sustaining state as well as the ratio of investment into user capabilities and network capacity.

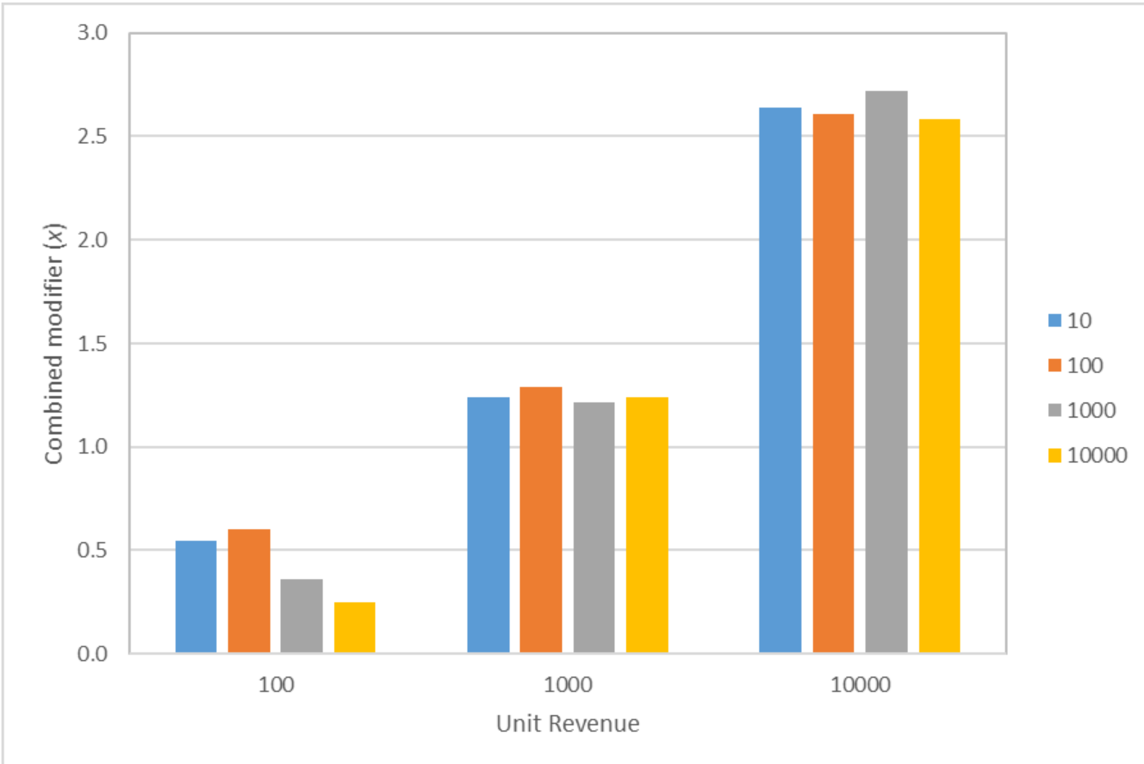


Figure 44 – The combined modifier, x(t), of user capability and network capacity on the Bass model

Before a discussion of the results with a varying capacity cost, the results from the low revenue and high costs scenario are examined here, as this represents many rural or remote areas. All the low revenue simulations kept a low (but not none) investment in user capability (Figure 45) for unit revenue of 100. This is combined with a significant reduction in capacity in the network (Figure 46) that retard the user growth of the network. In general, the results show that for networks with challenging economic value, some investment should be made into the potential network users to be capable of utilizing services on the network while dramatically under provisioning the network. The techno-economic network growth model does not account for changing population adoption rates but does account for the ratio of individuals who will use the network because of its perceived benefits and those that will use the network based on others using the network. This study did not look at the differences in investment ratios across a variety of adoption values (p and q in (14)).

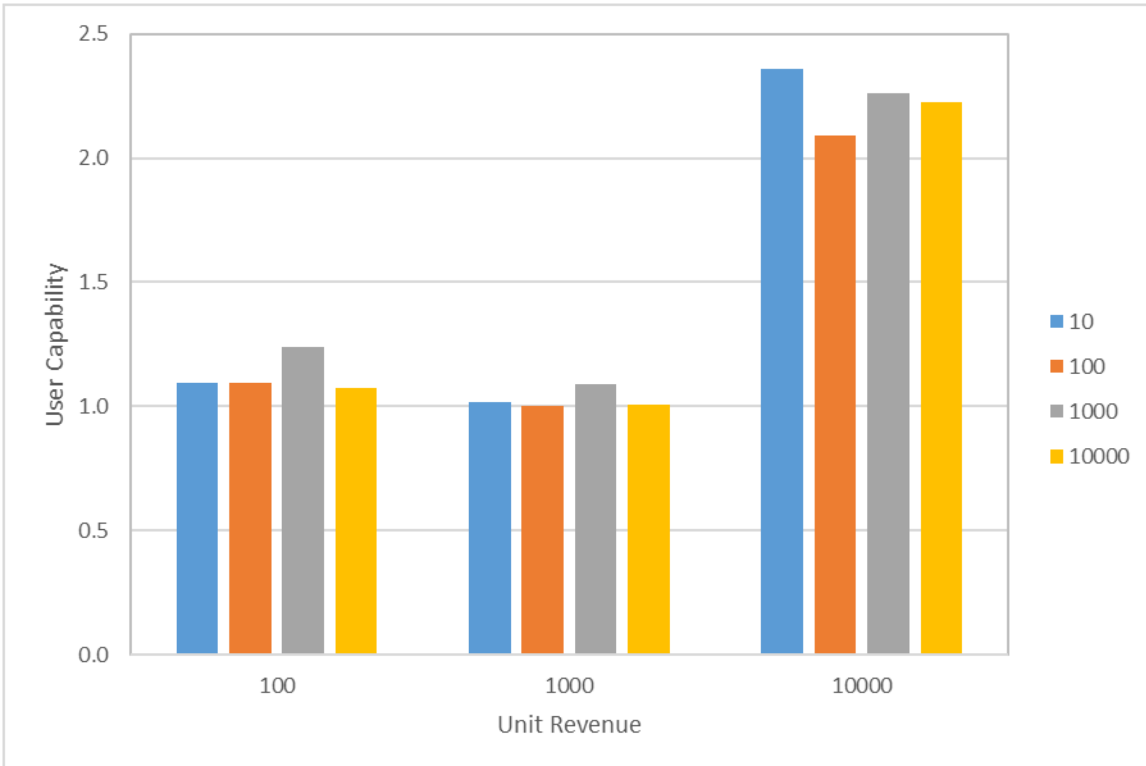


Figure 45 – User Capability factor representing the investment in a user's capability to utilize the network across span establishment cost multipliers.

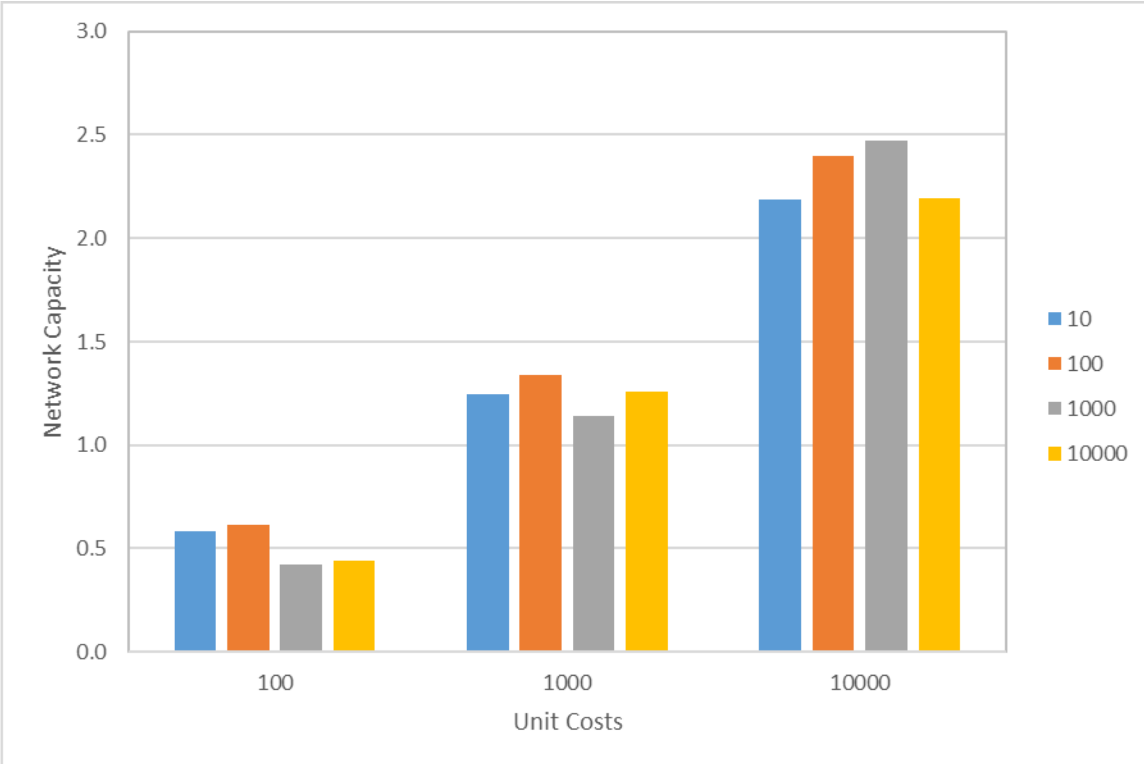


Figure 46 – Network capacity overbuild factor representing how much spare capacity against estimated demand should be placed in the network across span establishment cost multipliers of 10, 100, 1000, and 10 000.

By increasing user capability to use the network while reducing capacity, the network would be overloaded. The impact of this has been accounted for in the techno-economic network growth model. Under these conditions (with a combined factor, *x*, below 1), demand growth is reduced and muted compared to scenarios where the cost of the network was better aligned with the value of the network and had a positive combined adoption impact factor, *x*. Common sense and practice is to delay significant investment in the network in such cases where there is a cost-value mismatch. What the techno-economic network growth model presented in this paper demonstrates an analytic mechanism to evaluate the delay in investment based on the cost of the network, cost of enabling user demand, and the overall value the network provides.

Figure 47 demonstrates the expansion of the network was delayed for the high cost simulations. It wasn't until there was sufficient user demand that it made sense to expand the network. The incremental value of the network simulation with a span establishment cost multiplier of 100 and a unit revenue of 100 is shown in Figure 48. The incremental value doesn't become positive until the 10th time period, but does become positive. The cumulative value at that point represents the required subsidy for the network to be self-sustaining.

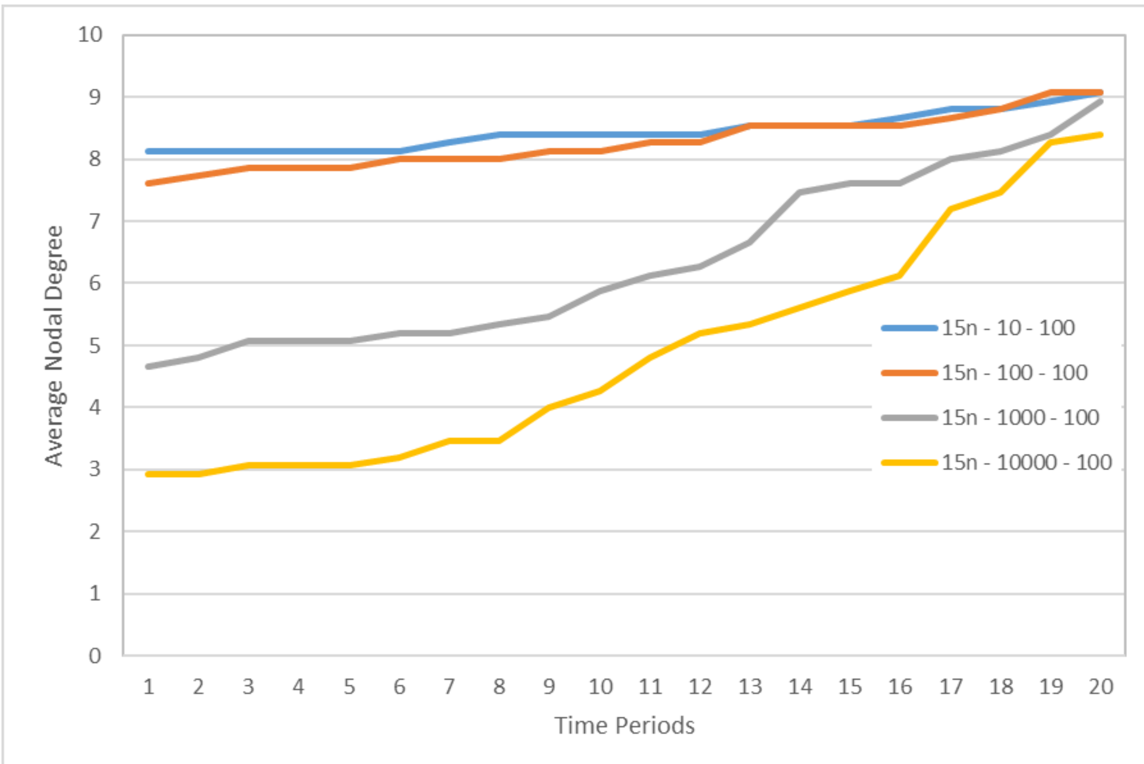


Figure 47 – Average nodal degree over the 20 time periods included in the simulations of span establishment costs of 10, 100, 1000, and 10 000 and a revenue of 100.

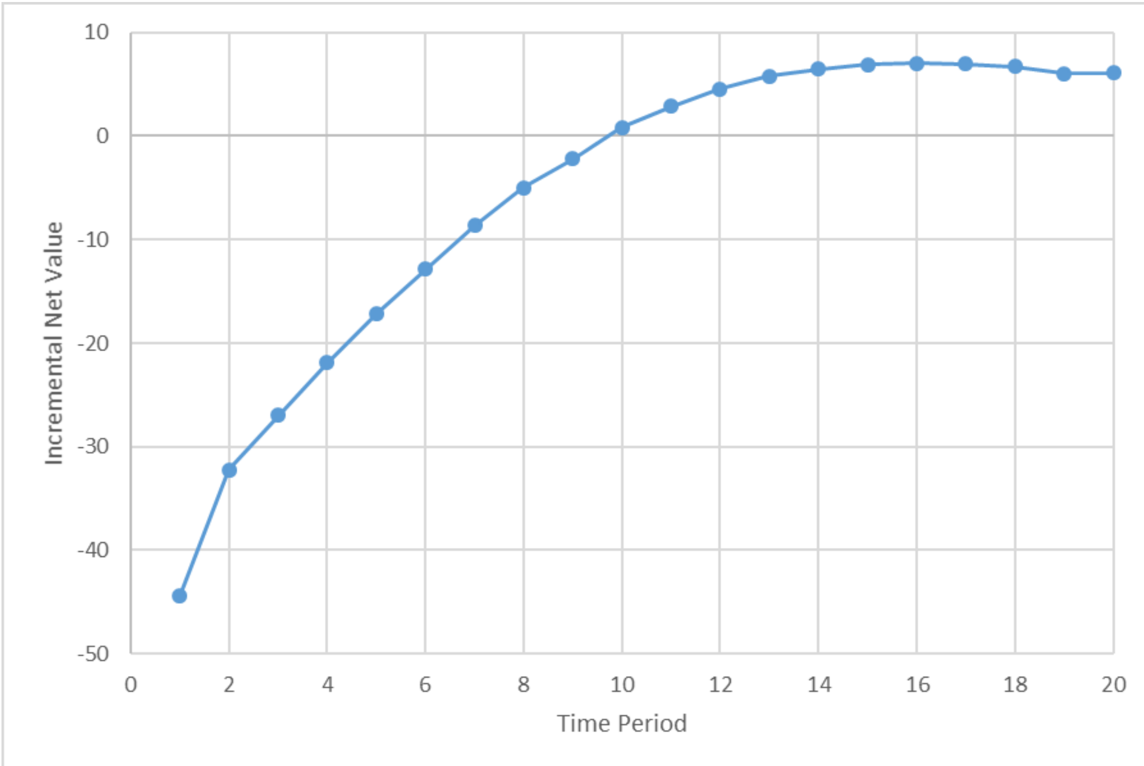


Figure 48 – Total value at each time period for a span establishment cost of 100 and a per unit revenue of 100.

5.5.2 Simulation Results with a Constant Span Establishment Cost / Capacity Cost Ratio

The simulation results presented above kept the cost of capacity constant while varying the cost of stablishing new spans. The simulations in this section kept a fixed ratio between capacity and establishment costs of 100 (this value was chosen as previous studies indicated that this was a balanced ratio between minimizing the number of spans as they were too expensive and maximizing the number of spans outright because they were too cheap [153]). The results are presented in context of the ration between the key inputs, as these illuminate the trade-offs, impact and sensitivity of the network planning framework.

The total value for simulations with a constant span establishment cost ratio is presented in Figure 49 and Figure 50. As expected the value is impacted by the unit cost multiplier more significantly when the cost multiplier applies to both span establishment and capacity costs. Figure 50 highlights the details of the simulations with a per unit revenue of 100 and 1000, where the results transition from positive to negative value. While these do not demonstrate anything that is not intuitive (value goes down when costs go up), they provide a reference as the selection of network design drives these values and are influenced by whether the value is positive or negative.

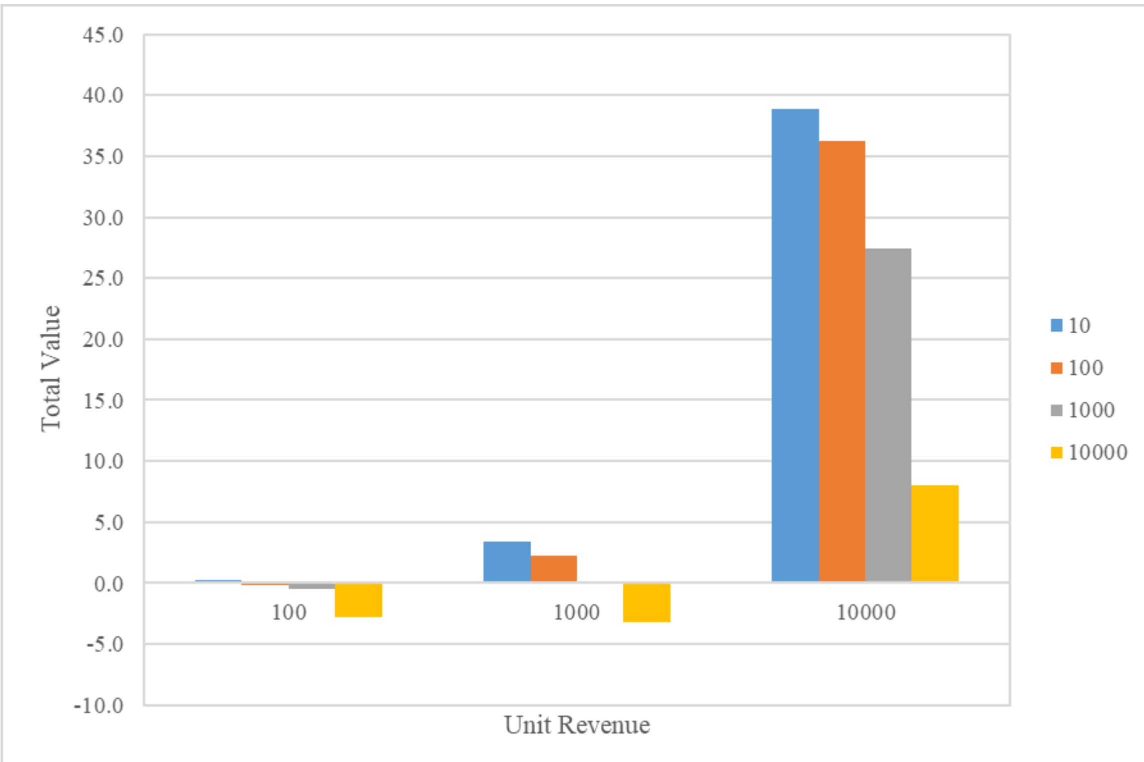


Figure 49 – Total value for simulations with a constant span establishment cost ratio of 100

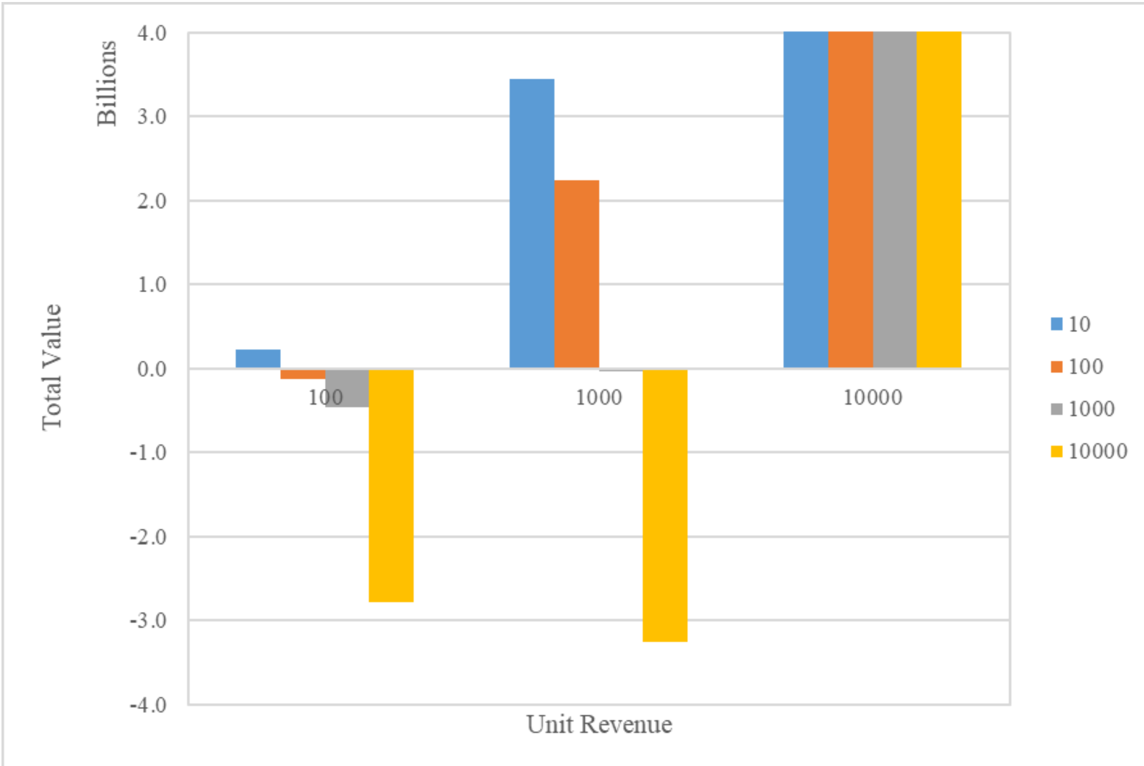


Figure 50 – Total value for simulations with a constant span establishment cost ratio of 100 truncated to $\mathbf{4.0 \times 10^9}$

The network capacity factor was influenced by the range of span establishment costs (10 – 10 000) as seen in Figure 51 significantly more than when the simulations with a fixed capacity cost. If the alternate view of the data is examined (looking at costs across multiple levels of per unit revenue), at high cost levels the network capacity factor remained in a similar range across varying revenue. This demonstrates a set of circumstances where the balance between cost and revenue isn't as important as cost. In general, if the ratio between capacity costs and span establishment costs were constant, the unit revenue and cost to span length factors together that influenced the capacity of the network.

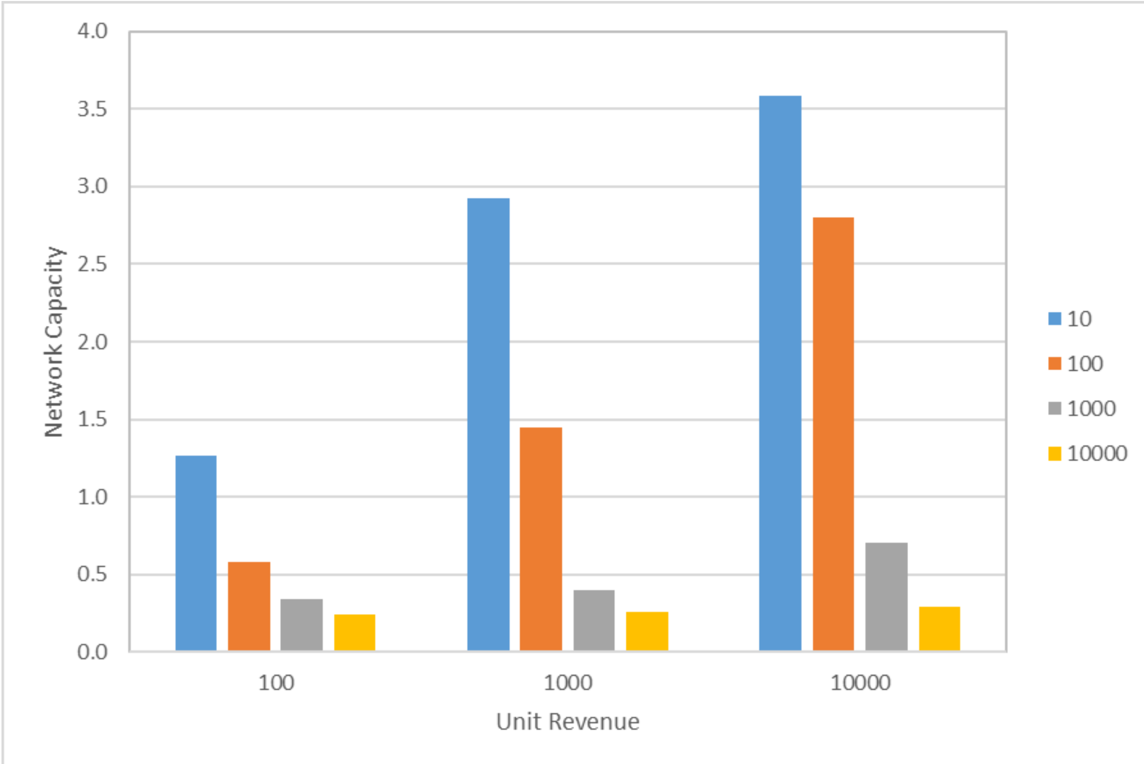


Figure 51 – A comparison of the network capacity overbuild factor with varying unit revenue for span implementation costs multipliers of 10 (blue), 100 (orange), 1000 (grey), and 10 000 (yellow)

The influences of the cost of the network had a similar impact on investing in user capability as network capacity, although not as dramatic (Figure 52). As costs increased, there was more of an emphasis on user capability. Because the cost of increasing capability was kept constant in these simulations, increasing span establishment and capacity costs shifted the value impact of influencing user capability. This focus on user capability increased even when the value of the networks decreased.

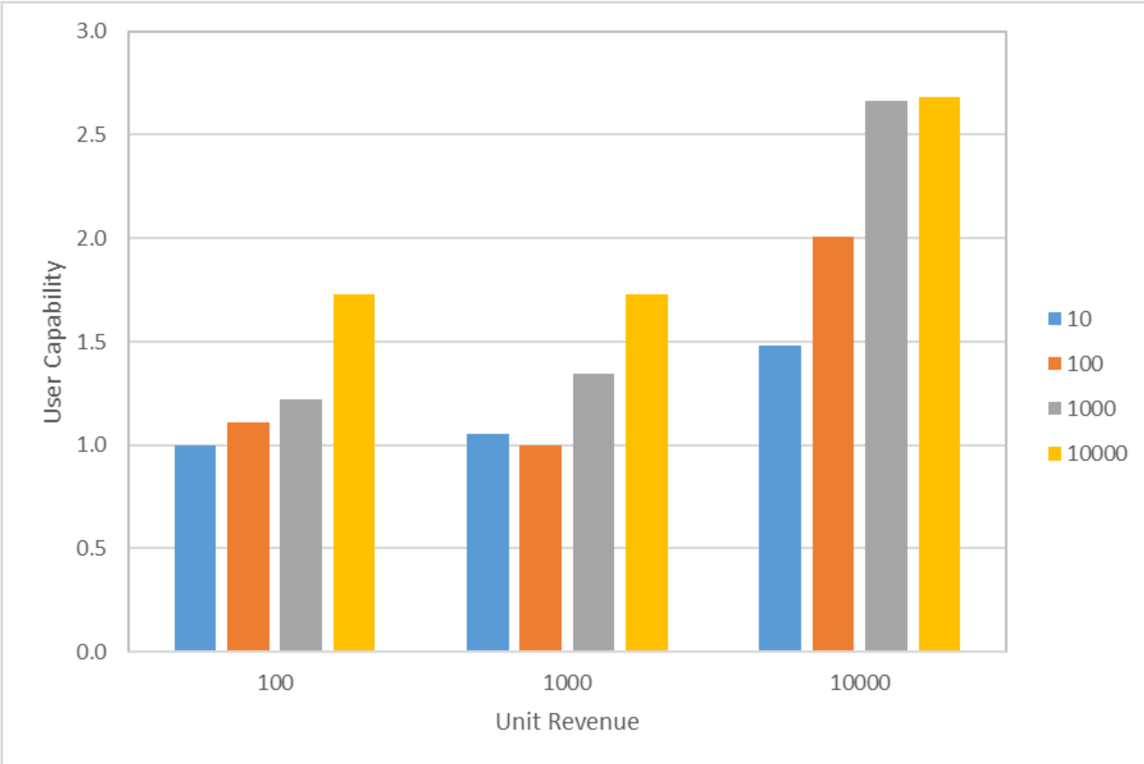


Figure 52 - A comparison of the user capability investment with varying unit revenue for span establishment costs multipliers of 10 (blue), 100 (orange), 1000 (grey), and 10 000 (yellow) and a constant capacity to span establishment cost ratio of 100

While looking at the behavior of the two factors individually provides some insight into the optimization of investing in the expansion of a network, the comparison of the two factors provides insight into the optimal allocation between expanding the network and expanding the user's ability to derive value from the network. What was observed in the results of the simulations was that the ratio of investing in user capability vs network capacity was highly correlated with costs with less correlation to unit revenue (Figure 53). It is intuitive that as costs increase, investing in user capacity would be emphasized over excess capacity. Figure 53 highlights that there is significant variability in these ratios, and the cross-over point between emphasizing network capacity vs user capability is not an obvious thing.

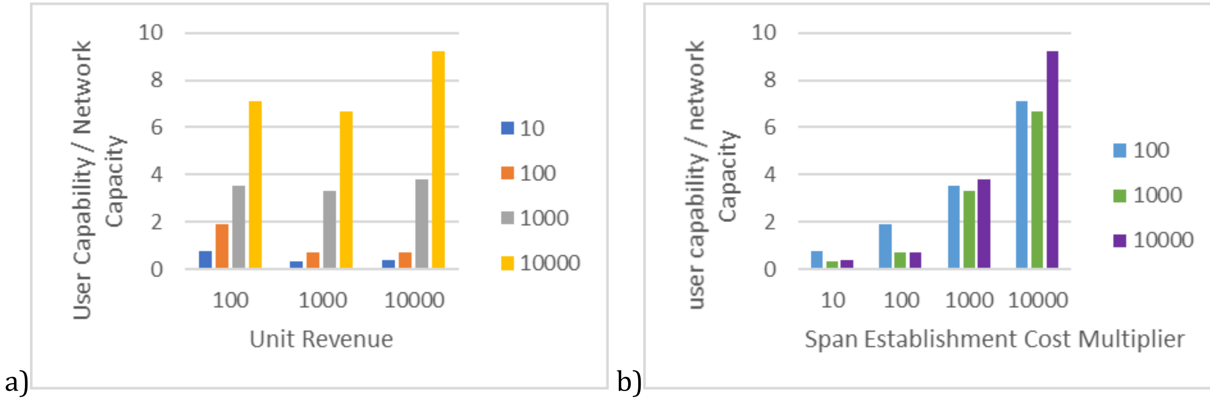


Figure 53 – The ratio of the user capability and network expansion factors for span establishment multipliers of 10, 100, 1000 and 10 000 where the capacity costs were kept constant.

What is interesting looking at the ratio of user capability and network capacity is that the ratio is quite different even when the ratio of unit revenue and cost are the same Figure 54). What did change across the three simulations presented was the ratio between unit revenue (and cost) and the cost of increasing user capabilities. The decision variable representing the overbuild of network capacity had these costs captured in the capacity and span establishment costs. The user capability factor cost was not varied across the simulations as the variability in revenue and network costs in relationship to this cost covered this variability, as mentions in section 133. What is demonstrated in Figure 54 is that the techno-economic network growth model accounts for the variability in the costs of building up or enhancing user capability.

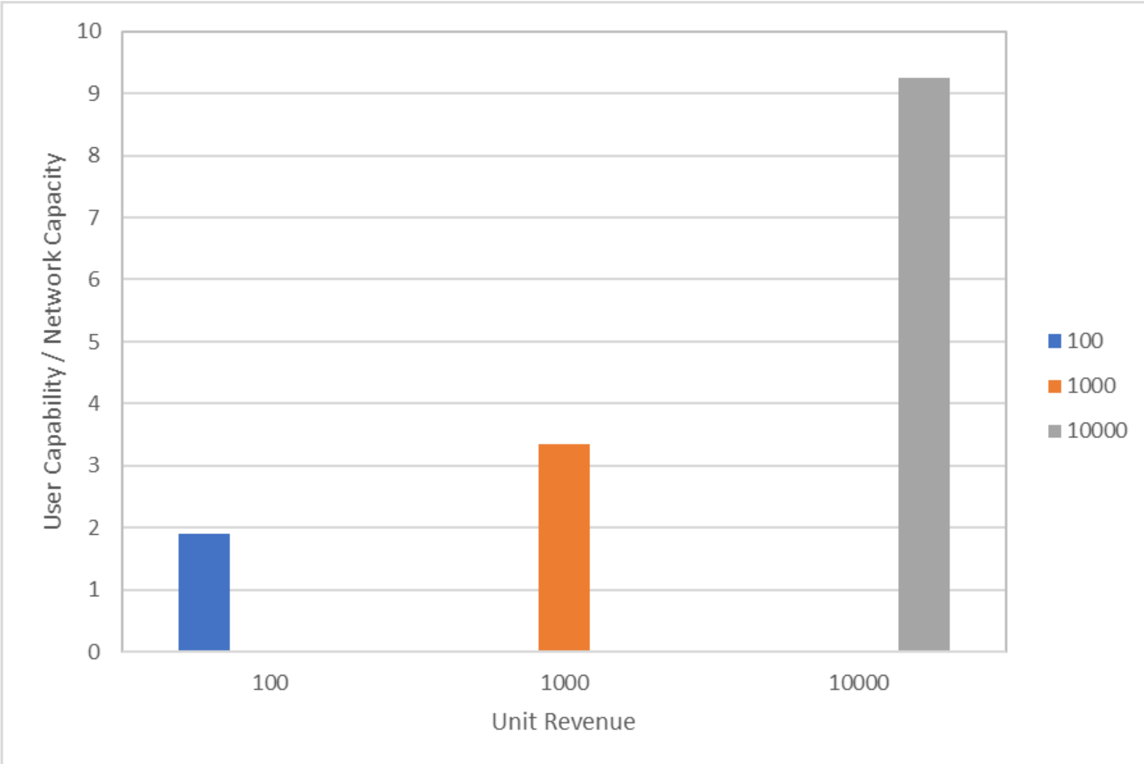
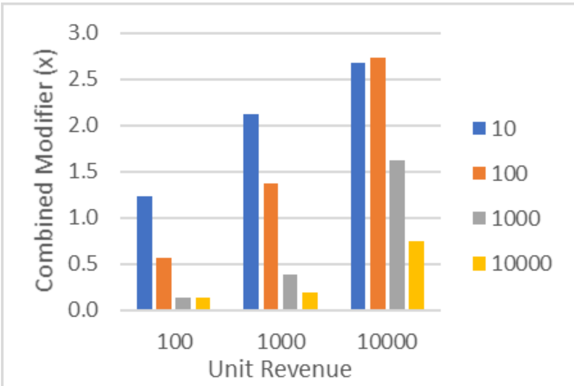


Figure 54 – User capability divided by network capacity for simulations with the same ratio of unit revenue and costs (with the revenue and cost being 100, 1000, and 10 000)

While the ratio of the user capability and the network capacity factor was more dependent on cost than revenue, the combined modifier representing the intervention in user growth had significant variation across both revenue and cost factors (Figure 55). Most of the simulations followed expected patterns where negative value meant that the combination of user capability and network capacity factors were below 0. The exception is the simulation with high cost and high revenue (a value of 10 000 for each) where the overall value was positive (Figure 49) but the combined modifier was negative (Figure 55).



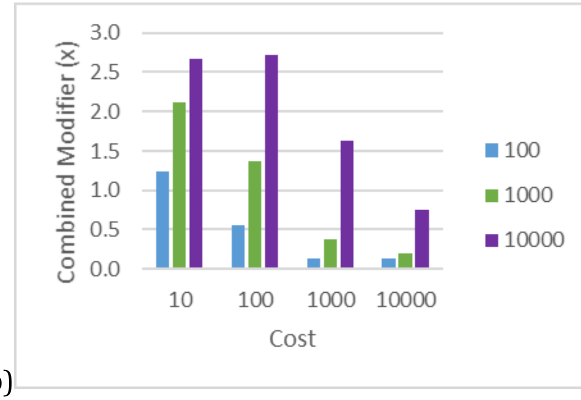


Figure 55 – The combined growth multiplier x used in the formula (16) for each span establishment cost multipliers of 10, 100, 1000 and 10 000 where the capacity costs were kept cons

When looking at the simulation results there are a couple that stand out. The networks with a unit revenue of 10 000 and a cost factor of 1000 and 10 000 had positive value (Figure 49) with both having a network capacity factor below 1 (Figure 51) and the simulation with a cost of 10 000 had a combined factor below (Figure 55). Both of these cases represent instances where it would be beneficial to invest heavily in the users of the network while under provisioning the network. The case with unit revenue and costs of 10 000 was the only solution with a positive value but a combined factor that is negative. The period by period value of the simulation is shown in Figure 56. The solution from the simulation with a cost of 1000 and a unit revenue of 10 000 was also calculated using the cost of 10 000 and a unit revenue of 10 000 (Figure 57). This recalculation had a significantly negative total value, compared with the genetic algorithm results using those same parameters (which had a positive total value), implying that in challenging economic conditions, the techno-economic network growth model suggests slowing down growth, minimizing network buildout costs, and investing heavily in users to utilize the network can turn a project that would otherwise destroy value into a project that is economically viable.

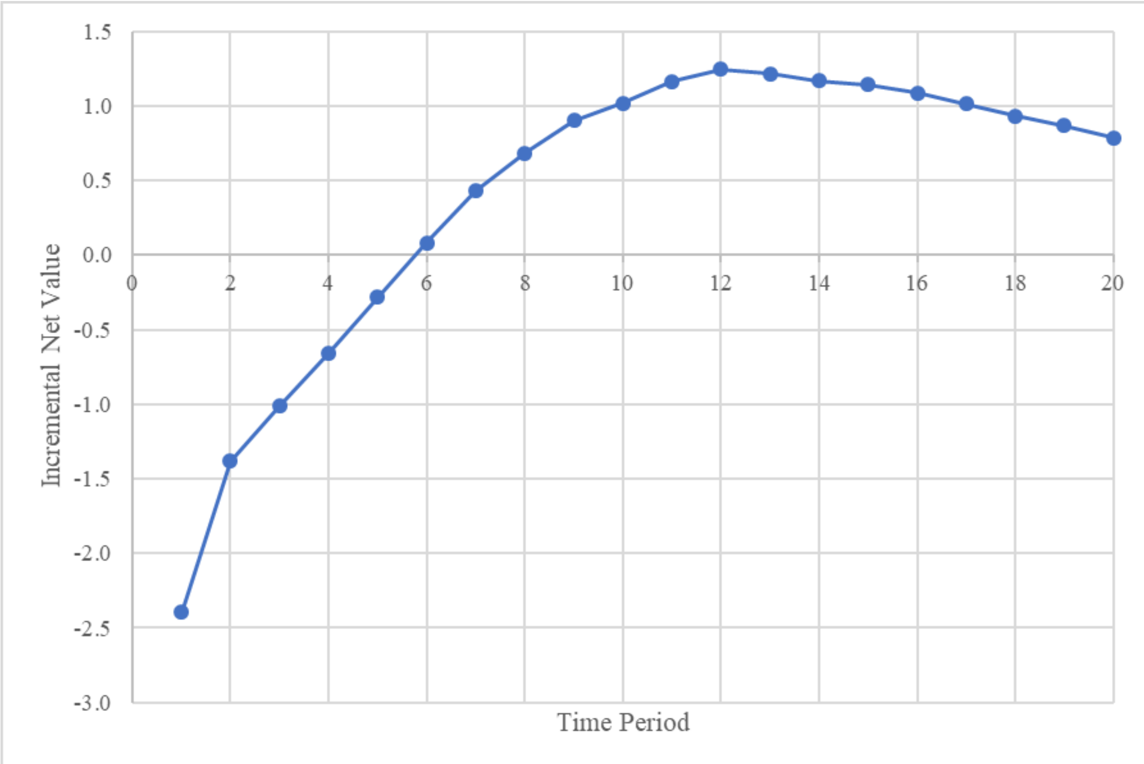


Figure 56 – Value for each period for the fixed span capacity to establishment cost ratio simulation with a span establishment cost multiplier of 10 000 and unit revenue of 10 000

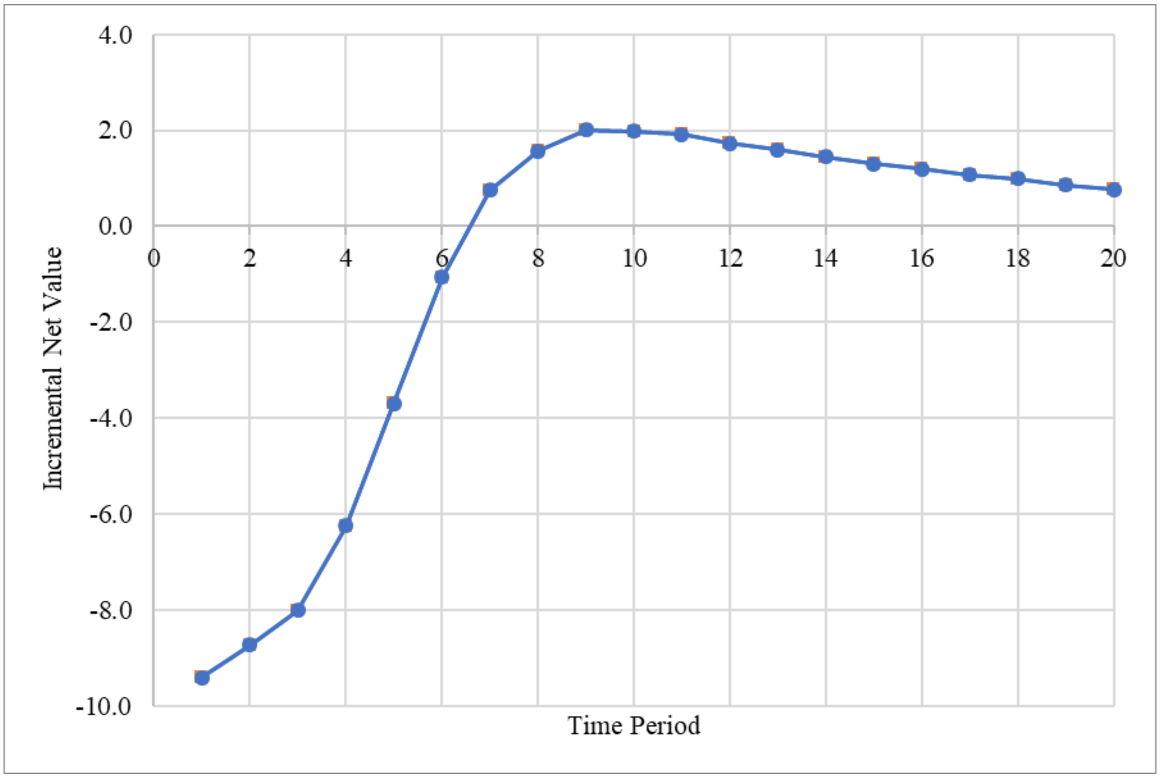


Figure 57 - Value for each time period using the network topology and augmentation profile, the user capability factor, and the network capacity factor from the simulation using a unit cost of 1000 and a unit revenue of 10 000 but calculated value with a span establishment cost multiplier of 10 000 and unit revenue of 10 000

The topology results of the techno-economic network growth model provided were similar in structure to those from other topology designs [84], [153] with the caveat that due to modeling network designs over 20 time periods meant that the final capacity was significant enough to justify a highly connected network, where most networks in place at the regional level utilize an average nodal degree significantly less than the 9 that the designs from the simulations in this work ended with [7].

It should also be noted that the design optimization utilized a shortest path routing strategy rather than a traditional network optimization. As a topic of further study, the difference between the results of this work should be compared to optimized results using the demand profiles generated from the techno-economic network growth model. The focus of this work was to analyze the allocation of resources between the network infrastructure and the users of the network. As such traditional survivable network design schemes would provide an interesting analysis of the impact of these factors, but the results of these networks schemes in a stand alone basis are not directly comparable to this work.

This work focused on the analysis of the effects of cost and revenue. It is an area of future study to investigate the sensitivity of the techno-economic network growth model to parameters in the Generalized Bass Model, namely the innovator coefficient (p), the imitator coefficient (q) and the two influencing factors (β_i).

The techno-economic network growth model presented here provides a quantitative mechanism to evaluate the allocation of resources between increasing a user's capacity to utilize the network, and the acceleration of the buildout of the network itself combining new product adoption models with network design and optimization. The work done here was based on a relative scale but by varying cost and revenue, but provides clear insight into how to effectively build out networks in environments that do not fall under the favorable conditions of high revenue and low cost. The results from the techno-economic network growth model align with intuition and provide a quantitative assessment of the emphasis on the capabilities of the users to derive value from the network and the physical capacity built into the network.

5.6 Conclusion

The techno-economic network growth model provides a mechanism to evaluate the allocation of resources between the network infrastructure and the user's capability to utilize it. If uncertainties in user uptake, cost, and revenue are put run through the model to build a pareto-optimal curve, network operators and other stakeholders can make better decisions in building out and expanding networks in areas with challenging economics or where there are social or structural barriers to utilizing the network (such as tele-health, distance education, or a general lack of skills in relevant network based applications).

While the model used here focused on overall capacity and user capability. Other effects could be evaluated within the model, such as degree of restorability in the network, the timing of connecting nodes, the effect of subsidizing networks (by decreasing costs or increasing revenue at points in time).

Challenges in the solution time of the model can be partly overcome through parallelized calculations and more efficient computing structures within the implementation of the model. The genetic algorithm used [157] had the capability of utilizing parallel computing but the implementation was not able to take advantage of this ability (and alternatively ran multiple simulations at the same time).

Chapter 6. – Conclusions and Summary

Communication networks are transforming how we interact, deliver services, provide education, and the many other areas of society, and will continue to do so. As we move to the next era of a connected world, with the potential of the internet of things, automated transportation, telemedicine, augmented reality, and the many other technologies on our horizon, A secure reliable network infrastructure that connects even rural and remote communities is critical. By incorporating the human socio-economic dynamics into what has been historically a very technical process, the value that new or expanded networks present to key service drivers can be better managed from a cost perspective, and also from a usage, and capital efficiency perspective. It was the goal of this research to contribute to this aim, and to develop a framework for planning telecommunication networks that aligns technical design considerations with the context and needs of the key service drivers that will use them.

In the context of long term planning of backbone network infrastructure, the research presented here provides solves some of the major challenges that have not been addressed in previous research. It articulates a planning framework that brings together the variety of encompassing factors required to make network construction viable. Through an extensive literature search across multiple disciplines, the framework articulated four fundamental inputs into the network planning process, the capabilities of the target market for the network, the fiscal capacity to fund the network, trends in applications that rely on telecommunication infrastructure, and the technical fundamentals of the infrastructure being deployed. The outcomes of the network design, policy, and regulatory planning processes can be broken down into four key areas. The most commonly articulated output is the technical design of the network, with the appropriate level of detail for the stage of planning. In comprehensive planning, the ownership and access to the network, the phasing and deployment staging are two other key outputs. The last output the framework articulated was the auxiliary activities that were not directly related to telecommunication networks, but are essential for successful network infrastructure deployments (such as healthcare policy and regulation, development and deployment of key services and applications that rely on the network infrastructure, and the many others presented in Chapter 4).

In order to implement the planning framework, survivable network optimization should be able to incorporate the evolution of the network design over time. In Chapter 3 a mathematical formulation that optimizes survivable network schemes was presented, and run across a number of networks in order to compare and contrast the impact of the selected network scheme on the topology of the network over time. One of the primary outcomes of this research was how the two schemes (DSP and SBPP) impacted topology augmentation. The multi-period network augmentation model developed in this research augmented existing multi-period research by incorporating topology augmentation enabling network planning frameworks to evaluate the impact of other inputs and output decisions against the fundamental topology of the network.

Lastly, Chapter 5 connected the network planning framework with two key decision points, investing in the technical capabilities of the network and investing in the end user's capacity to use the network. The outcome of the simulations run through the mathematical models developed through this research highlighted some of the complexities and trade-offs that are a part of the planning process. When dealing with networks that have an overall negative value (which could be a common characteristic when deploying networks in rural, remote, or less economically affluent regions) the allocation of effort and investment between growing the network and building the user's capacity to use it was not consistent. Depending on the costs of each of the network and the cost of supporting the users, different designs and deployment strategies come into play.

6.1 Future Work

There are several areas where this work can be expanded. These include utilizing different types of networks, planning under uncertainty and trial simulations with networks already constructed.

This work targeted backbone networks with characteristics typically found at the metropolitan, regional, or national level. There are other network classes where the network planning framework would apply. Two classes that are currently experiencing a good deal of growth are local access fiber networks and cellular networks. Both of these classes of networks are characterized by different topologies, traffic patterns, and capacity allocation strategies but have the same fundamental questions of topology structure, capacity allocation. Local access networks typically utilize a star topology using GPON [ref] technology. The deployment of these networks are expensive and are sensitive to user adoption rates. Cellular networks use a coverage topology with highly variable traffic demands due to the transient nature of the connections [ref]. The planning framework can apply to these networks with modifications from the work presented in chapters 3 and 5.

Another area of investigation is the utilization of this framework with probabilistic models. As the framework deals with many inter-related decisions, utilizing probabilistic models would provide an understanding to the most robust options in light of the uncertainties that are present in the framework. A simple example of this would be to design the network under high growth and low growth scenarios and note the common elements of the topology. One of the major issues with probabilistic models is the computational complexity they present. The models presented in this work take in some cases in the order of months to solve. To analyze the probabilistic results would take significantly longer, and as such more efficient methods of optimizing the decision models are required. There have been a history of work looking at uncertainty in network planning [41], and the impact of uncertainty on network designs was described in [20].

Lastly the planning framework, the multi-period network augmentation model and the techno-economic network growth model should be validated with real world results. Network designs and costs are closely held information and are not readily available. The cost ratios used in this work are meant to represent different scenarios, such as dense urban settings where right of ways mostly exist vs remote locations or new builds where right of ways must be established. By using ratios between the key inputs, insight and guidance can be gained into the general direction network planning should take.

6.2 Summary

Together the research presented here provide a framework, and the mathematical formulations required to expand the scope of network planning across knowledge domains, such as product growth forecasting, survivable network design, and economics. The work presented here has highlighted key decision points and some of the trade-offs in these decisions. It was the goal of this research to better deploy networks across regions to enable the benefits that the internet, and all advanced communication based technologies provide to all regions and better cross the digital divide [158].

References

[1] M. Castells, *The Rise of the Network Society*. Blackwell Publishers, 2000.

[2] Routledge, “Technology, e-learning and Distance Education (Paperback) - Routledge.”

[3] D. Gans, J. Kralewski, T. Hammons, and B. Dowd, “Medical groups’ adoption of electronic health records and information systems,” *Health Aff (Millwood)*, vol. 24, no. 5, pp. 1323–1333, Oct. 2005.

[4] B. K. Atrostic, S. V. Nguyen, and C. for E. S. (U.S.), *Computer Networks and U.S. Manufacturing Plant Productivity: New Evidence from the CNUS Data*. Center for Economic Studies, U.S. Department of Commerce, Bureau of the Census, 2002.

[5] A. L. Kavanaugh and S. J. Patterson, “The Impact of Community Computer Networks on Social Capital and Community Involvement,” *American Behavioral Scientist*, vol. 45, no. 3, pp. 496–509, Nov. 2001.

[6] E. Morozov, “Iran Elections: A Twitter Revolution?,” *The Washington Post*, 17-Jun-2009. [Online]. Available: http://www.washingtonpost.com/wp-dyn/content/discussion/2009/06/17/DI2009061702232.html. [Accessed: 24-Apr-2012].

[7] W. D. Grover, *Mesh-Based Survivable Networks: Options and Strategies for Optical, Mpls, Sonet, and Atm Networking*. Prentice Hall Professional, 2004.

[8] J. Rak, *Resilient Routing in Communication Networks*. Cham: Springer International Publishing, 2015.

[9] S. Verbrugge, K. Casier, J. Van Ooteghem, and B. Lannoo, “Practical steps in techno-economic evaluation of network deployment planning part 1: methodology overview,” in *Telecommunications Network Strategy and Planning Symposium, 2008. Networks 2008. The 13th International*, 2008, pp. 1–101.

[10] R. Katz, “The economic and social impact of telecommunications output,” *Intereconomics*, vol. 44, no. 1, pp. 41–48, 2009.

[11] H. Lee, N. Park, and Y. Hwang, “A new dimension of the digital divide: Exploring the relationship between broadband connection, smartphone use and communication competence,” *Telematics and Informatics*, vol. 32, no. 1, pp. 45–56, Feb. 2015.

[12] M. R. Gilbert, M. Masucci, C. Homko, and A. A. Bove, “Theorizing the digital divide: Information and communication technology use frameworks among poor women using a telemedicine system,” *Geoforum*, vol. 39, no. 2, pp. 912–925, Mar. 2008.

[13] F. M. Bass, “A New Product Growth for Model Consumer Durables,” *Management Science*, vol. 15, no. 5, pp. 215–227, Jan. 1969.

[14] A. Martinez, V. Villarroel, J. Puig-Junoy, J. Seoane, and F. del Pozo, “An economic analysis of the EHAS telemedicine system in Alto Amazonas,” *J Telemed Telecare*, vol. 13, no. 1, pp. 7–14, Jan. 2007.

[15] S. Yang and F. A. Kuipers, “Traffic uncertainty models in network planning,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 172–177, Feb. 2014.

[16] B. Sansò and P. Soriano, Eds., *Telecommunications Network Planning*. Boston, MA: Springer US, 1999.

[17] K. Kirkpatrick, “Software-defined Networking,” *Commun. ACM*, vol. 56, no. 9, pp. 16–19, Sep. 2013.

[18] A. M. C. A. Koster, S. Orlowski, C. Raack, G. Baier, and T. Engel, “Single-Layer Cuts for Multi-Layer Network Design Problems,” in *Telecommunications Modeling, Policy, and Technology*, S. Raghavan, B. Golden, and E. Wasil, Eds. Springer US, 2008, pp. 1–23.

[19] A. Kodian and W. D. Grover, “Multiple-quality of protection classes including dual-failure survivable services in p-cycle networks,” in *2nd International Conference on Broadband Networks, 2005. BroadNets 2005*, 2005, p. 231–240 Vol. 1.

[20] C. Kronberger, T. Schondienst, and D. A. Schupke, “Impact and Handling of Demand Uncertainty in Multiperiod Planned Networks,” in *Communications (ICC), 2011 IEEE International Conference on*, 2011, pp. 1–6.

[21] R. N. de Alencar Vidal, C. Kronberger, and D. A. Schupke, “Multiperiod planning of optical protection schemes,” in *2011 IEEE GLOBECOM Workshops (GC Wkshps)*, 2011, pp. 641–645.
[22] E. M. Rogers, *Diffusion of innovations*. New York: Free Press, 2003.
[23] V. Joseph and B. Chapman, “The Evolution of Communication Systems,” in *Deploying QoS for Cisco IP and Next Generation Networks*, Boston: Morgan Kaufmann, 2009, pp. 1–22.
[24] “Alberta SuperNet up and running,” *The Globe and Mail*, 08-Apr-2009.
[25] J. M. Simmons, *Optical network design and planning*. Springer, 2008.
[26] S. K. Mohapatra, “Integrated planning for Next Generation Networks,” in *IFIP/IEEE International Symposium on Integrated Network Management-Workshops, 2009. IM ’09*, 2009, pp. 205–210.
[27] M. H. M. Moh, “A solution Framework for Next Generation Network Planning,” *IEC Annual Review of Telecommunications*, 2007.
[28] S. Chen and K. Nahrstedt, “An overview of quality of service routing for next-generation high-speed networks: problems and solutions,” *Network, IEEE*, vol. 12, no. 6, pp. 64–79, Dec. 1998.
[29] ITU-T, “ITU-T’s Definition of NGN.” [Online]. Available: http://www.itu.int/en/ITU-T/gsi/ngn/Pages/definition.aspx. [Accessed: 25-Apr-2012].
[30] “About TELUS - Next generation network,” *Our next generation network*, 24-Apr-2012. [Online]. Available: http://about.telus.com/community/english/about_us/technology_%26_innovation/telus_networks/next_generation_network. [Accessed: 25-Apr-2012].
[31] “TELUS Builds IP Next-Generation Network Cloud Services Using Cisco Unified Service Delivery Technology,” *Marketwire*, 14-Jun-2010. [Online]. Available: http://www.marketwire.com/press-release/telus-builds-ip-next-generation-network-cloud-services-using-cisco-unified-service-delivery-nasdaq-csco-1275315.htm. [Accessed: 25-Apr-2012].
[32] I. P. Kaminow, T. Li, and A. E. Willner, *Optical Fiber Telecommunications V B: Systems and Networks*. Academic Press, 2008.
[33] K. Guan and Chi Guan, “Long-term capacity planning of WDM transport networks under demand uncertainties and economy of scale,” in *2006 11th International Workshop on Computer-Aided Modeling, Analysis and Design of Communication Links and Networks*, 0-0 0, pp. 100–106.
[34] N. Geary, A. Antonopoulos, E. Drakopoulos, and J. O’Reilly, “Analysis of optimisation issues in multi-period DWDM network planning,” in *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, 2001, vol. 1, pp. 152–158 vol.1.
[35] K. Guan and C. Guan, “Multi-Period Planning of WDM Transport Networks Incorporating Demand Uncertainties,” in *23rd Biennial Symposium on Communications, 2006*, 2006, pp. 135–139.
[36] C. Kronberger, D. A. Schupke, and J. Eberspacher, “Multiperiod Planning with Controlled Shortest Path Routing,” in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, 2010, pp. 1–6.
[37] E. Palkopoulou, C. Meusburger, D. Schupke, L. Wosinska, and T. Bauschert, “Combining multi-period and multi-layer network planning: Ignored potential?,” 2010, pp. 1–3.
[38] D. A. Schupke and C. Kronberger, “Multiperiod planning for optical networks,” in *Telecommunications Network Strategy and Planning Symposium (NETWORKS), 2010 14th International*, 2010, pp. 1–6.
[39] S. Strauss, A. Kirstadter, and D. A. Schupke, “Multi-period Planning of WDM-Networks: Comparison of Incremental and EoL Approaches,” in *Internet, 2006 2nd IEEE/IFIP International Conference in Central Asia on*, 2006, pp. 1–7.
[40] Chi Guan and V. W. . Chan, “Topology design and resource dimensioning for optical mesh networks under demand uncertainties,” in *IEEE Global Telecommunications Conference, 2005. GLOBECOM ’05*, 2005, vol. 4, p. 7 pp.-2090.

[41] D. Leung and W. D. Grover, “Capacity Planning of Survivable Mesh-based Transport Networks under Demand Uncertainty,” *Photonic Network Communications*, vol. 10, no. 2, pp. 123–140, Sep. 2005.
[42] K. S. Ho and K. W. Cheung, “Network planning under uncertainties,” *Proceedings of SPIE*, vol. 7137, no. 1, pp. 713722-713722–12, Nov. 2008.
[43] R. Fildes and V. Kumar, “Telecommunications demand forecasting—a review,” *International Journal of Forecasting*, vol. 18, no. 4, pp. 489–522, Oct. 2002.
[44] P. G. Allen and R. Fildes, *Econometric Forecasting*, vol. Principles of Forecasting: A Handbook for Researchers and Practitioners. Kluwer Academic Publishers, 2001.
[45] M. H. Lyons, F. Burton, B. Egan, T. Lynch, and S. Skelton, “Dynamic modeling of present and future service demand,” *Proceedings of the IEEE*, vol. 85, no. 10, pp. 1544–1555, Oct. 1997.
[46] N. Meade and T. Islam, “Modelling and forecasting the diffusion of innovation – A 25-year review,” *International Journal of Forecasting*, vol. 22, no. 3, pp. 519–545, 2006.
[47] G. Magoulas, G. Lepouras, and C. Vassilakis, “Virtual reality in the e-Society,” *Virtual Reality*, vol. 11, no. 2, pp. 71–73, 2007.
[48] C. Fuchs and E. Horak, “Africa and the digital divide,” *Telematics and Informatics*, vol. 25, no. 2, pp. 99–116, May 2008.
[49] “Canada’s digital divide - The Globe and Mail.” [Online]. Available: http://www.theglobeandmail.com/report-on-business/canadas-digital-divide/article1521975/. [Accessed: 03-Apr-2010].
[50] A. Dobre, T. Jones, V. N. Stroetmann, K. A. Stroetmann, Y. Vatter, and K. Peng, “Interoperable eHealth is Worth it - Securing Benefits from Electronic Health Records and ePrescribing,” European Commision Information Society and Media, Feb. 2010.
[51] J. Sachs, *The end of poverty : economic possibilities for our time*. New York: Penguin Press, 2005.
[52] B. Stiller, “Telecommunication economics – overview of the field, recommendations, and perspectives,” *Computer Science - Research and Development*, vol. 23, no. 1, pp. 35–43, 2009.
[53] L.-H. Röller and L. Waverman, “Telecommunications Infrastructure and Economic Development: A Simultaneous Approach,” *The American Economic Review*, vol. 91, no. 4, pp. 909–923, 2001.
[54] K. S. Sridhar and V. Sridhar, “Telecommunications and Growth: Causal Model, Quantitative and Qualitative Evidence,” *Economic and Political Weekly*, vol. 41, no. 25, pp. 2611–2619, Jun. 2006.
[55] S. K. Majumdar and S. Venkataraman, “Network Effects and the Adoption of New Technology: Evidence from the U.S. Telecommunications Industry,” *Strategic Management Journal*, vol. 19, no. 11, pp. 1045–1062, Nov. 1998.
[56] P. Howitt, *Implications of Knowledge-based Growth for Micro-economic Policies*. University of Calgary Press, 1996.
[57] “IEEE Xplore - Home.” [Online]. Available: http://ieeexplore.ieee.org.login.ezproxy.library.ualberta.ca/Xplore/home.jsp. [Accessed: 24-Apr-2012].
[58] “ScienceDirect.com | Search through over 10 million science, health, medical journal full text articles and books.” [Online]. Available: http://www.sciencedirect.com.login.ezproxy.library.ualberta.ca/. [Accessed: 24-Apr-2012].
[59] “SpringerLink - electronic journals, protocols and books.” [Online]. Available: http://www.springerlink.com.login.ezproxy.library.ualberta.ca/. [Accessed: 24-Apr-2012].
[60] D. K. Kim and S. K. Yoo, “Design and evaluation of teleultrasound consultation system,” in *Control, Automation and Systems, 2007. ICCAS ’07. International Conference on*, 2007, pp. 2425–2428.
[61] S. Soegijoko and S. Tjondronegoro, “Development of e-Health Kiosks and telemedicine system for disaster area: experience in Aceh,” in *Enterprise networking and Computing in Healthcare Industry, 2005. HEALTHCOM 2005. Proceedings of 7th International Workshop on*, 2005, p. 24.

[62] D. S. Venkateswarlu, K. S. Verma, and K. S. R. Murthy, “e Health networking to cater to Rural Health Care and Health Care for the Aged,” in *e-Health Networking, Application and Services, 2007 9th International Conference on*, 2007, pp. 273–276.
[63] K. A. Stroetmann, T. Jones, A. Dobrev, and V. N. Stroetmann, “An evaluation of the economic impact of ten European e-health applications,” *Journal of Telemedicine and Telecare*, vol. 13, no. 5, pp. 62–64, Jul. 2007.
[64] H. J. Schau, M. F. Smith, and P. I. Schau, “The healthcare network economy: The role of Internet information transfer and implications for pricing,” *Industrial Marketing Management*, vol. 34, no. 2, pp. 147–156, Feb. 2005.
[65] A. D. Black *et al.*, “The Impact of eHealth on the Quality and Safety of Health Care: A Systematic Overview,” *PLoS Med*, vol. 8, no. 1, p. e1000387, Jan. 2011.
[66] R. E. Gantenbein, F. Cadez, and B. H. Stamm, “Establishing a rural telehealth project: the Wyoming network for telehealth,” in *Engineering in Medicine and Biology Society, 2004. IEMBS ’04. 26th Annual International Conference of the IEEE*, 2004, vol. 2, pp. 3089–3092.
[67] G. King, H. Richards, and D. Godden, “Adoption of telemedicine in Scottish remote and rural general practices: a qualitative study,” *J Telemed Telecare*, vol. 13, no. 8, pp. 382–386, Dec. 2007.
[68] H. L. Koekemoer and C. Scheffer, “Heart sound and electrocardiogram recording devices for telemedicine environments,” in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, 2008, pp. 4867–4870.
[69] S. E. Hussein, “Picture Archiving and Communication System Analysis and Deployment,” in *Computer Modelling and Simulation, 2009. UKSIM ’09. 11th International Conference on*, 2009, pp. 520–525.
[70] A. B. Cunha, M. A. M. Capretz, and L. Raptopoulos, “Support systems for Telehealth services: Critical operational and ICT complementary assets for large-scale provisioning,” in *Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference*, 2009, pp. 340–345.
[71] F. Fraser, “Only So Super,” *Alberta Venture*.
[72] D. Climenhaga, “Half a billion? Albertans need to know what the SuperNet has really cost,” *Rabble.ca*. [Online]. Available: http://rabble.ca/blogs/bloggers/djclimenhaga/2011/11/half-billion-albertans-need-know-what-supernet-has-really-cost. [Accessed: 11-Apr-2012].
[73] B. Satchwill, “Better health care needs better technologies,” *Cybera*. [Online]. Available: http://www.cybera.ca/blog/better-health-care-needs-better-technologies. [Accessed: 11-Apr-2012].
[74] K. Poscente, L. Rourke, and T. Anderson, “Perceptions Of An Impending High Speed, Broadband Network: Anticipation And Anxiety Among K-12 Teachers, Technical Support Personnel, And Administrators,” *The Journal of Applied Educational Technology*, vol. 4, no. 1, 2007.
[75] Extension Learning Solutions, Alberta. Alberta Education, and Alberta. Alberta Education. Stakeholder Technology Branch, *SuperNet snapshot : final summary report*. Edmonton, AB: Alberta Education, 2007.
[76] B. Whitacre, S. Strover, and R. Gallardo, “How much does broadband infrastructure matter? Decomposing the metro–non-metro adoption gap with the help of the National Broadband Map,” *Government Information Quarterly*, vol. 32, no. 3, pp. 261–269, Jul. 2015.
[77] M. Herzberg, S. J. Bye, and A. Utano, “The hop-limit approach for spare-capacity assignment in survivable networks,” *IEEE/ACM Transactions on Networking*, vol. 3, no. 6, pp. 775–784, 1995.
[78] M. Clouqueur and W. D. Grover, “Availability analysis of span-restorable mesh networks,” *Selected Areas in Communications, IEEE Journal on*, vol. 20, no. 4, pp. 810–821, May 2002.
[79] A. M. C. A. Koster and A. Zymolka, “Demand-wise shared protection for meshed optical networks,” in *Fourth International Workshop on Design of Reliable Communication Networks, 2003. (DRCN 2003). Proceedings*, 2003, pp. 85–92.

[80] J. Doucette and W. D. Grover, “Influence of Modularity and Economy-of-scale Effects on Design of Mesh-Restorable DWDM Networks,” *IEEE JSAC Special Issue on Protocols and Architectures for Next Generation Optical WDM Networks*, vol. 18, no. 10, pp. 1912–1923, Oct. 2000.
[81] A. Narula-Tam, E. Modiano, and A. Brzezinski, “Physical topology design for survivable routing of logical rings in WDM-based networks,” *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 8, pp. 1525–1538, 2004.
[82] W. D. Grover and J. Doucette, “Topological Design of Survivable Mesh-Based Transport Networks,” *Annals of Operations Research*, vol. 106, no. 1–4, pp. 79–125, Sep. 2001.
[83] J. Doucette, W. D. Grover, and P. A. Giese, “Physical-Layer p-Cycles Adapted for Router-Level Node Protection: A Multi-Layer Design and Operation Strategy,” *Selected Areas in Communications, IEEE Journal on*, vol. 25, no. 5, pp. 963–973, Jun. 2007.
[84] B. Todd and J. Doucette, “DSP survivable network capacity allocation and topology design using multi-period network augmentation,” in *Design of Reliable Communication Networks (DRCN), 2015 11th International Conference on the*, 2015, pp. 41–48.
[85] B. Todd and J. Doucette, “Demand-wise shared protection network design with dual-failure restorability,” in *International Conference on Ultra Modern Telecommunications Workshops, 2009. ICUMT ’09*, 2009, pp. 1–6.
[86] B. Todd and J. Doucette, “Use of Network Families in Survivable Network Design and Optimization,” in *IEEE International Conference on Communications, 2008. ICC ’08*, 2008, pp. 151–157.
[87] B. Forst and W. D. Grover, “Factors affecting the efficiency of Demand-wise Shared Protection,” in *6th International Workshop on Design and Reliable Communication Networks, 2007. DRCN 2007*, 2007, pp. 1–8.
[88] B. Todd and J. Doucette, “Multi-Flow Optimization Model for Design of a Shared Backup Path Protected Network,” in *IEEE International Conference on Communications, 2008. ICC ’08*, 2008, pp. 131–138.
[89] K. Walkowiak and J. Rak, “Shared Backup Path Protection for Anycast and Unicast Flows Using the Node-Link Notation,” in *2011 IEEE International Conference on Communications (ICC)*, 2011, pp. 1–6.
[90] T.-H. Wu, R. H. Cardwell, and M. Boyden, “A multi-period design model for survivable network architecture selection for SONET interoffice networks,” *IEEE Transactions on Reliability*, vol. 40, no. 4, pp. 417–427, 1991.
[91] M. Pickavet and P. Demeester, “Long-Term Planning of WDM Networks: A Comparison between Single-Period and Multi-Period Techniques,” *Photonic Network Communications*, vol. 1, no. 4, pp. 331–346, 1999.
[92] C. Meusburger, C. Meusburger, and D. A. Schupke, “A Multiperiod Expectation Model for Capital Expenditures in Optical Networks: Applicability and Limitations,” *Photonic Networks, 2010 ITG Symposium on*, pp. 1–5, May 2010.
[93] A. Damodaran, *Investment Valuation: Tools and Techniques for Determining the Value of Any Asset*. John Wiley & Sons, 2012.
[94] M. Jinno, H. Takara, B. Kozicki, Y. Tsukishima, Y. Sone, and S. Matsuoka, “Spectrum-efficient and scalable elastic optical path network: architecture, benefits, and enabling technologies,” *IEEE Communications Magazine*, vol. 47, no. 11, pp. 66–73, 2009.
[95] F. T. Moore, “Economies of Scale: Some Statistical Evidence,” *The Quarterly Journal of Economics*, vol. 73, no. 2, pp. 232–245, May 1959.
[96] G. Desaulniers, J. Desrosiers, and M. M. Solomon, *Column Generation*. Springer Science & Business Media, 2006.
[97] D. Sandham, “Into the light,” *Engineering Technology*, vol. 4, no. 6, pp. 70–73, Apr. 2009.

[98] L. Kubler, “Dark Fiber is Lighting Up,” *Dark Fiber is Lighting Up*, 03-Oct-2013. [Online]. Available: http://cacm.acm.org/news/168424-dark-fiber-is-lighting-up/fulltext. [Accessed: 16-Feb-2015].

[99] J. Manyika and C. Roxbugh, “The great transformer: The impact of the Internet on economic growth and prosperity,” McKinsey Global Institute, Oct. 2011.

[100] T. G. Robertazzi, *Planning Telecommunication Networks*, 1st ed. Wiley-IEEE Press, 1998.

[101] V. Y. Liu and D. Tipper, “Spare capacity allocation using partially disjoint paths for dual link failure protection,” in *Design of Reliable Communication Networks (DRCN), 2013 9th International Conference on the*, 2013, pp. 171–178.

[102] Y. Liu and D. Tipper, “Successive survivable routing for node failures,” in *Global Telecommunications Conference, 2001. GLOBECOM ’01. IEEE*, 2001, vol. 4, pp. 2093–2097 vol.4.

[103] E. Modiano and A. Narula-Tam, “Survivable lightpath routing: a new approach to the design of WDM-based networks,” *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 4, pp. 800–809, May 2002.

[104] T. Stidsen, B. Petersen, S. Spoorendonk, M. Zachariasen, and K. B. Rasmussen, “Optimal routing with failure-independent path protection,” *Networks*, vol. 55, no. 2, pp. 125–137, 2010.

[105] Y. S. Kavian, R. Rejeb, and O. Strobel, “Dual-link failure covering in DWDM optical networks using genetic algorithms,” in *2010 12th International Conference on Transparent Optical Networks (ICTON)*, 2010, pp. 1–4.

[106] B. Jaumard, “Operation research tools and methodology for the design and provisioning of survivable optical networks,” in *2013 17th International Conference on Optical Network Design and Modeling (ONDM)*, 2013, pp. 53–58.

[107] A. Konak and A. E. Smith, “A hybrid genetic algorithm approach for backbone design of communication networks,” in *Proceedings of the 1999 Congress on Evolutionary Computation, 1999. CEC 99*, 1999, vol. 3, p. 1823 Vol. 3.

[108] S. J. Koh and C. Y. Lee, “A tabu search for the survivable fiber optic communication network design,” *Computers & Industrial Engineering*, vol. 28, no. 4, pp. 689–700, Oct. 1995.

[109] X. Yang, L. Shen, and B. Ramamurthy, “Survivable lightpath provisioning in WDM mesh networks under shared path protection and signal quality constraints,” *Journal of Lightwave Technology*, vol. 23, no. 4, pp. 1556–1567, Apr. 2005.

[110] J. Doucette, D. He, W. D. Grover, and O. Yang, “Algorithmic approaches for efficient enumeration of candidate p-cycles and capacitated p-cycle network design,” in *Fourth International Workshop on Design of Reliable Communication Networks, 2003. (DRCN 2003). Proceedings*, 2003, pp. 212–220.

[111] H. Kim and N. Feamster, “Improving network management with software defined networking,” *IEEE Communications Magazine*, vol. 51, no. 2, pp. 114–119, Feb. 2013.

[112] L. Atzori, A. Iera, and G. Morabito, “The Internet of Things: A survey,” *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, Oct. 2010.

[113] A. Ghezzi, M. N. Cortimiglia, and A. G. Frank, “Strategy and business model design in dynamic telecommunications industries: A study on Italian mobile network operators,” *Technological Forecasting and Social Change*, vol. 90, Part A, pp. 346–354, Jan. 2015.

[114] J. L. Garcia-Dorado, A. Finamore, M. Mellia, M. Meo, and M. Munafo, “Characterization of ISP Traffic: Trends, User Habits, and Access Technology Impact,” *IEEE Transactions on Network and Service Management*, vol. 9, no. 2, pp. 142–155, Jun. 2012.

[115] J. Li, A. Aurelius, V. Nordell, M. Du, A. Arvidsson, and M. Kihl, “A five year perspective of traffic pattern evolution in a residential broadband access network,” in *Future Network Mobile Summit (FutureNetw), 2012*, 2012, pp. 1–9.

[116] A. M. Lago, “Demand Forecasting Models of International Telecommunications and Their Policy Implications,” *The Journal of Industrial Economics*, vol. 19, no. 1, pp. 6–21, Nov. 1970.

[117] K. Stordahl and E. Murphy, “Forecasting long-term demand for services in the residential market,” *IEEE Communications Magazine*, vol. 33, pp. 44–49, Feb. 1995.
[118] S. K. Korotky, “Semi-Empirical Description and Projections of Internet Traffic Trends Using a Hyperbolic Compound Annual Growth Rate,” *Bell Labs Tech. J.*, vol. 18, no. 3, pp. 5–21, Dec. 2013.
[119] K. Papagiannaki, N. Taft, Z.-L. Zhang, and C. Diot, “Long-term forecasting of Internet backbone traffic,” *IEEE Transactions on Neural Networks*, vol. 16, no. 5, pp. 1110–1124, Sep. 2005.
[120] I.-I. O. for Standardization, “ISO - International Organization for Standardization.”
[121] V. Mahajan, E. Muller, and F. M. Bass, “Diffusion of New Products: Empirical Generalizations and Managerial Uses,” *Marketing Science*, vol. 14, no. 3 supplement, pp. G79–G88, Jun. 1995.
[122] “ASQ SIPOC Diagram.” [Online]. Available: http://asq.org/service/body-of-knowledge/tools-sipoc. [Accessed: 04-Mar-2015].
[123] B. Todd and J. Doucette, “A novel long term telecommunication network planning framework using a SIPOC approach,” in *2013 5th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, 2013, pp. 236–242.
[124] A. Alexander, “Internet role in Egypt’s protests,” *BBC*, 09-Feb-2011. [Online]. Available: http://www.bbc.co.uk/news/world-middle-east-12400319. [Accessed: 25-Apr-2012].
[125] C. Delany, “How Social Media Accelerated Tunisia’s Revolution: An Inside View,” *Huffington Post*, 25-May-2011. [Online]. Available: http://www.huffingtonpost.com/colin-delany/how-social-media-accelera_b_821497.html. [Accessed: 25-Apr-2012].
[126] Canada Business Network, “Business planning.” [Online]. Available: http://www.canadabusiness.ca/eng/page/2865/. [Accessed: 25-Apr-2012].
[127] S. D. Maesschalck *et al.*, “Pan-European Optical Transport Networks: An Availability-based Comparison,” *Photonic Network Communications*, vol. 5, no. 3, pp. 203–225, May 2003.
[128] M. Hopkins, G. Louth, H. Bailey, R. Yellon, A. Ajibulu, and M. Niva, “A multi-faceted approach to forecasting broadband demand and traffic,” *Communications Magazine, IEEE*, vol. 33, no. 2, pp. 36–42, Feb. 1995.
[129] B. Briscoe, A. Odlyzko, and B. Tilly, “Metcalfe’s law is wrong - communications networks increase in value as they add members-but by how much?,” *IEEE Spectrum*, vol. 43, no. 7, pp. 34–39, Jul. 2006.
[130] M. O’Mahony, “Future optical networks,” in *Optical Fiber Telecommunications V B (Fifth Edition)*, Burlington: Academic Press, 2008, pp. 611–640.
[131] T. Islam, D. G. Fiebig, and N. Meade, “Modelling multinational telecommunications demand with limited data,” *International Journal of Forecasting*, vol. 18, no. 4, pp. 605–624, Oct. 2002.
[132] A. Jarray, B. Jaumard, and A. C. Houle, “Minimum CAPEX/OPEX design of optical backbone networks,” in *Ultra Modern Telecommunications Workshops, 2009. ICUMT ’09. International Conference on*, 2009, pp. 1–8.
[133] “infoDev.org.” [Online]. Available: http://www.infodev.org/en/Index.html. [Accessed: 27-Apr-2010].
[134] “Information & Communication Technologies Home - The World Bank.” [Online]. Available: http://www.worldbank.org/en/topic/ict. [Accessed: 08-Mar-2015].
[135] “The great telecoms crash,” *The Economist*, 18-Jul-2002.
[136] A. Troianovski, “Optical Delusion? Fiber Booms Again, Despite Bust,” *Wall Street Journal*, 03-Apr-2012.
[137] P. Starr, “The Great Telecom Implosion,” 08-Sep-2002. [Online]. Available: http://www.princeton.edu/~starr/articles/articles02/Starr-TelecomImplosion-9-02.htm. [Accessed: 03-Apr-2015].
[138] O. P. Malik, *Broadbandits: Inside the $750 Billion Telecom Heist*, 1 edition. Hoboken, N.J.: Wiley, 2004.

[139] E. M. Noam, “Fundamental instability: Why telecom is becoming a cyclical and oligopolistic industry,” *Information Economics and Policy*, vol. 18, no. 3, pp. 272–284, Sep. 2006.

[140] Internet Engineering Task Force (IETF), “Multiprotocol Label Switching (mpls) - Documents.” [Online]. Available: http://datatracker.ietf.org/wg/mpls/. [Accessed: 25-Apr-2012].

[141] ITU-T, “Definitions and Descriptions (OTNT, OTN, MON).” [Online]. Available: http://www.itu.int/ITU-T/2001-2004/com15/otn/definitions.html. [Accessed: 25-Apr-2012].

[142] V. Srivastava and M. Motani, “Cross-layer design: a survey and the road ahead,” *IEEE Communications Magazine*, vol. 43, no. 12, pp. 112–119, Dec. 2005.

[143] S. Pazi, C. Chatwin, R. Young, P. Birch, and Wei Wang, “Cost-Effectively ICT Backbone for Tanzania.,” *European Journal of Scientific Research*, vol. 39, no. 2, pp. 217–233, Jan. 2010.

[144] D. Ludwick, “What is the effect of information and computing technology on healthcare?,” Thesis, University of Alberta, Edmonton, AB, 2009.

[145] L. A. Andrés, J. Schwartz, and J. L. Guasch, *Uncovering the Drivers of Utility Performance: Lessons from Latin America and the Caribbean on the Role of the Private Sector, Regulation, and Governance in the Power, Water, and Telecommunication Sectors*. World Bank Publications, 2013.

[146] J. M. Alonso, J. Clifton, D. Díaz-Fuentes, M. Fernández-Gutiérrez, and J. Revuelta, “The race for international markets: Were privatized telecommunications incumbents more successful than their public counterparts?,” *International Review of Applied Economics*, vol. 27, no. 2, pp. 215–236, Mar. 2013.

[147] J. P. Doh, H. Teegen, and R. Mudambi, “Balancing Private and State Ownership in Emerging Markets’ Telecommunications Infrastructure: Country, Industry, and Firm Influences,” *Journal of International Business Studies*, vol. 35, no. 3, pp. 233–250, May 2004.

[148] “Usage-based internet billing - Technology & Science - CBC News.” [Online]. Available: http://www.cbc.ca/news/technology/story/2011/01/31/f-faq-usage-based-billing.html. [Accessed: 25-Apr-2012].

[149] F. M. Bass, T. V. Krishnan, and D. C. Jain, “Why the Bass Model Fits without Decision Variables,” *Marketing Science*, vol. 13, no. 3, pp. 203–223, Aug. 1994.

[150] T. Turk and P. Trkman, “Bass model estimates for broadband diffusion in European countries,” *Technological Forecasting and Social Change*, vol. 79, no. 1, pp. 85–96, Jan. 2012.

[151] B. Todd and J. Doucette, “A Novel Long Term Telecommunication Network Planning Framework,” *J Netw Syst Manage*, vol. 25, no. 1, pp. 47–82, Jan. 2017.

[152] B. K. Chen and F. A. Tobagi, “Network Topology Design to Optimize Link and Switching Costs,” in *Communications, 2007. ICC ’07. IEEE International Conference on*, 2007, pp. 2450–2456.

[153] B. Todd and J. Doucette, “Survivable Network Capacity Allocation and Topology Design Using Multi-period Network Augmentation,” *J Netw Syst Manage*, vol. 25, no. 3, pp. 481–507, Jul. 2017.

[154] P. McBurney, S. Parsons, and J. Green, “Forecasting market demand for new telecommunications services: an introduction,” *Telematics and Informatics*, vol. 19, no. 3, pp. 225–249, Aug. 2002.

[155] D. Jain, V. Mahajan, and E. Muller, “Innovation Diffusion in the Presence of Supply Restrictions,” *Marketing Science*, vol. 10, no. 1, pp. 83–90, Feb. 1991.

[156] C. Meusburger, D. A. Schupke, and J. Eberspacher, “Multiperiod Planning for Optical Networks - Approaches Based on Cost Optimization and Limited Budget,” in *Communications, 2008. ICC ’08. IEEE International Conference on*, 2008, pp. 5390–5395.

[157] F.-A. Fortin, F.-M. D. Rainville, M.-A. Gardner, M. Parizeau, and C. Gagné, “DEAP: Evolutionary Algorithms Made Easy,” *Journal of Machine Learning Research*, vol. 13, no. Jul, pp. 2171–2175, 2012.

[158] I. K. Rohman and E. Bohlin, “Does Broadband Speed Really Matter for Driving Economic Growth? Investigating OECD Countries,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2034284, Apr. 2012.

Appendix A– Multi-period Network Augmentation Models

a. Demand-Wise Shared Protection with Multi-Period Network Augmentation Integer Linear Program

```
set Time ordered;
# The Time periods for which to model

set SPANS;
# Set of all spans and potential spans in the network

set NODES;
# set of all nodes in the network

set Demands;
#set of all Demands that exist


# *******************************************
# Parameters
# *******************************************

param ImpTechCost{t in Time} integer;
# the cost of implementing a certain technology on a given span (dependent on span length)

param CapTechCost{t in Time} integer;
# the cost of adding capacity to a given span (dependent on span length and capacity)

param SpanLength{s in SPANS};
# This is a multiplier to get the cost of adding capacity to a span.

param DemandVolume {r in Demands, t in Time};
# the amount of demand between a node pair r for Time t

#param Restorability{r in Demands, t in Time};
# the fraction of demand volume for demand r at Time t that requires single failure restorability

param initSpanCapacity{s in SPANS};
# the initial capacity in each span

param MaxFlow := sum {r in Demands, t in Time} DemandVolume[r,t];
# Used for upper bounds on flow and capacity variables

param MinCycle{d in Demands} integer;

# *******************************************
# Variables
# *******************************************
```

```
var spanCapacity{s in SPANS, t in Time} >= 0, <=MaxFlow;
# The total capacity allocated to span s at Time t

var spanChange{s in SPANS, t in Time} >=0, <=1, integer;
# 1 if technology has changed to technology i in Time t on span s, 0 otherwise

var techCost{s in SPANS, t in Time} >=0 integer;
# represents the cost of changing technologies for span s in Time t

var capCost{s in SPANS, t in Time} >=0;
# the capacity cost on span s at Time t


# ****************************************************************
# DSP
# ****************************************************************

# ***********************
# TOPOLOGY DEFINITION
# ***********************

set NODE_SPANS{n in NODES} within SPANS;
# Set of all spans attached to a given node

set ADJ_NODES{n in NODES} within NODES;

# ***********************
# DESCRIPTION OF Demands AND THEIR ROUTING
# ***********************

set ORIGIN{Demands} within NODES;
# The origin for demand d

set DESTINATION{Demands} within NODES;
# the destination for demand d



# ***********************
# VARIABLES
# ***********************

var span_flow{r in Demands, i in SPANS, t in Time} >=0, <=MaxFlow integer;
# The amount of lightpaths assigned to span that starts at i and finishes at j.

var node_flow{r in Demands, i in NODES, j in ADJ_NODES[i], t in Time} >=0, <=MaxFlow integer;
# Whether or not span i,j has any traffic routed on it

var totalwork >=0;
```

```
var node_direction {r in Demands, i in NODES, j in ADJ_NODES[i], t in Time} >=0, <=1 integer;
# ensures traffic is not assigned in both directions on a span

# ****************************************************************
# objective function
# ****************************************************************

minimize TotalCost:
	sum{s in SPANS, t in Time}(capCost[s,t] + techCost[s,t]);

# ****************************************************************
# Multi period constraints
# ****************************************************************


subject to EoSClassChangeDetection{s in SPANS, t in Time}:
	sum{tc in Time: ord(tc,Time) <= ord(t,Time)}(spanCapacity[s,tc]- MaxFlow *
(spanChange[s,tc]))<= 0;

subject to implementationCosts{s in SPANS, t in Time}:
	techCost[s,t] >= spanChange[s,t] * ImpTechCost[t] * SpanLength[s];

#if techSelected > 0 then capcost - differential cost >= 0
subject to capacityCosts{s in SPANS, t in Time}:
	capCost[s,t] = (if t = first(Time) then (CapTechCost[t]*SpanLength[s]*(spanCapacity[s,t] -
initSpanCapacity[s])) else (CapTechCost[t] * SpanLength[s] * (spanCapacity[s,t] -
spanCapacity[s,prev(t,Time)])) );

subject to no_capacity_reduction{s in SPANS, t in Time: ord(t, Time) > 1}:
	spanCapacity[s,t] - spanCapacity[s,prev(t,Time)] >= 0;

subject to minimum_span_count{t in Time}:
	sum{s in SPANS, tc in Time: ord(tc,Time) <= ord(t,Time)}(spanChange[s,tc]) >=
card(NODES);
	#if spans form a loop, this is the minimum requirement for the graph to be bi-connected

subject to bi_connected_nodes{n in NODES, t in Time}:
	sum{s in NODE_SPANS[n], tc in Time: ord(tc,Time) <= ord(t,Time)}(spanChange[s,tc]) >= 2;

# ****************************************************************
# Survivability model constraints
# ****************************************************************

# ***********************
# CONSTRAINTS
# ***********************

subject to calculate_span_capacity {s in SPANS, t in Time}:
	spanCapacity[s,t] = sum{r in Demands} span_flow[r,s,t];
```

```
subject to single_failure_restorability {r in Demands, n in (ORIGIN[r]), j in NODE_SPANS[n], t in Time}:
		sum{k in NODE_SPANS[n]: k<>j}span_flow[r,k,t] >= DemandVolume[r,t];
		# If a single path fails (starting with span i,j), the rest of the paths must have enough capacity to route traffic on it.

subject to translate_node_direction {r in Demands, n1 in NODES, n2 in ADJ_NODES[n1], t in Time}:
   DemandVolume[r,t] * node_direction[r, n1, n2,t] >= node_flow[r,n1,n2,t];

subject to node_flow_not_both_ways {r in Demands, n1 in NODES, n2 in ADJ_NODES[n1], t in Time}:
   node_direction[r, n1, n2,t] + node_direction[r, n2, n1,t] <= 1;

subject to disjoint_paths_1 {r in Demands, n1 in ((NODES diff ORIGIN[r]) diff DESTINATION[r]), t in Time}:
   sum{n2 in ADJ_NODES[n1]} node_direction[r,n1,n2,t] <= 1;

subject to disjoint_paths_2 {r in Demands, n1 in ((NODES diff ORIGIN[r]) diff DESTINATION[r]), t in Time}:
   sum{n2 in ADJ_NODES[n1]} node_direction[r,n2,n1,t] <= 1;

subject to transit_flow {r in Demands, n1 in ((NODES diff ORIGIN[r]) diff DESTINATION[r]), t in Time}:
		sum{n2 in ADJ_NODES[n1]} node_flow[r,n1,n2,t] = sum{n2 in ADJ_NODES[n1]} node_flow[r,n2,n1,t];
		# the incoming flow and the outgoing flow must be equal

subject to origin_flow{r in Demands, n1 in ORIGIN[r], n2 in ADJ_NODES[n1], i in (NODE_SPANS[n1] inter NODE_SPANS[n2]), t in Time}:
		node_flow[r,n1,n2,t] = span_flow[r,i,t];
		# If a span is connected to the origin, then the flow of traffic must be away from the origin node

subject to destination_flow{r in Demands, n1 in DESTINATION[r], n2 in ADJ_NODES[n1], i in (NODE_SPANS[n1] inter NODE_SPANS[n2]), t in Time}:
		node_flow[r,n2,n1,t] = span_flow[r,i,t];
		# if a span is connected to the destination then the traffic must be toward the destination

subject to one_way_traffic{r in Demands, n1 in NODES, n2 in ADJ_NODES[n1], i in (NODE_SPANS[n1] inter NODE_SPANS[n2]), t in Time}:
		node_flow[r,n1,n2,t] + node_flow[r,n2,n1,t] = span_flow[r,i,t];
		# traffic must only flow in one direction. This is only figurative, since once a link is established
		# from origin to destination, it is assumed that a link in the opposite direction is included.

subject to limit_span_flow{r in Demands, i in SPANS, t in Time}:
   span_flow[r,i,t] <= DemandVolume[r,t];

#subject to limited_path_length{d in Demands, t in Time}:
#		sum{n1 in NODES, n2 in ADJ_NODES[n1]}(node_direction[d,n1,n2,t]) <= card(SPANS);
		# the total number of spans used is less than the total number of spans in the network
```

```
#subject to mminimum_path_length{d in Demands, t in Time}:
#	sum{n1 in NODES, n2 in ADJ_NODES[n1]}(node_direction[d,n1,n2,t]) >= MinCycle[d];
	# the total number of spans used is less than the total number of spans in the network


#subject to APS {r in Demands, n1 in (ORIGIN[r]), t in Time}:
#	sum{n2 in ADJ_NODES[n1]} node_direction[r,n1,n2,t] = 2;
```

b. Shared Backup Path Protection with Multi-Period Network Augmentation Integer Linear Program

```
set Time ordered;
# The Time periods for which to model

set SPANS;
# Set of all spans and potential spans in the network

set NODES;
# set of all nodes in the network

set Demands;
#set of all Demands that exist


# *******************************************
# Parameters
# *******************************************

param ImpTechCost{t in Time} integer;
# the cost of implementing a certain technology on a given span (dependent on span length)

param CapTechCost{t in Time} integer;
# the cost of adding capacity to a given span (dependent on span length and capacity)

param SpanLength{s in SPANS};
# This is a multiplier to get the cost of adding capacity to a span.

param DemandVolume {r in Demands, t in Time};
# the amount of demand between a node pair r for Time t

#param Restorability{r in Demands, t in Time};
# the fraction of demand volume for demand r at Time t that requires single failure restorability

param initSpanCapacity{s in SPANS};
# the initial capacity in each span

param MaxFlow := sum {r in Demands, t in Time} DemandVolume[r,t];
# Used for upper bounds on flow and capacity variables
```

```
param MinCycle{d in Demands} integer;

# *******************************************
# Variables
# *******************************************

var spanCapacity{s in SPANS, t in Time} >= 0, <=MaxFlow;
# The total capacity allocated to span s at Time t

var spanChange{s in SPANS, t in Time} >=0, <=1, integer;
# 1 if technology has changed to technology i in Time t on span s, 0 otherwise

var techCost{s in SPANS, t in Time} >=0 integer;
# represents the cost of changing technologies for span s in Time t

var capCost{s in SPANS, t in Time} >=0;
# the capacity cost on span s at Time t


# *****************************************************************
# DSP
# *****************************************************************

# ************************
# TOPOLOGY DEFINITION
# ************************

set NODE_SPANS{n in NODES} within SPANS;
# Set of all spans attached to a given node

set ADJ_NODES{n in NODES} within NODES;

# ************************
# DESCRIPTION OF Demands AND THEIR ROUTING
# ************************

set ORIGIN{Demands} within NODES;
# The origin for demand d

set DESTINATION{Demands} within NODES;
# the destination for demand d



# ************************
# VARIABLES
# ************************

var span_flow{r in Demands, i in SPANS, t in Time} >=0, <=MaxFlow integer;
```

```
# The amount of lightpaths assigned to span that starts at i and finishes at j.

var node_flow{r in Demands, i in NODES, j in ADJ_NODES[i], t in Time} >=0, <=MaxFlow integer;
# Whether or not span i,j has any traffic routed on it

var totalwork >=0;

var node_direction {r in Demands, i in NODES, j in ADJ_NODES[i], t in Time} >=0, <=1 integer;
# ensures traffic is not assigned in both directions on a span

# ****************************************************************
# objective function
# ****************************************************************

minimize TotalCost:
	sum{s in SPANS, t in Time}(capCost[s,t] + techCost[s,t]);

# ****************************************************************
# Multi period constraints
# ****************************************************************


subject to EoSClassChangeDetection{s in SPANS, t in Time}:
	sum{tc in Time: ord(tc,Time) <= ord(t,Time)}(spanCapacity[s,tc]- MaxFlow *
(spanChange[s,tc]))<= 0;

subject to implementationCosts{s in SPANS, t in Time}:
	techCost[s,t] >= spanChange[s,t] * ImpTechCost[t] * SpanLength[s];

#if techSelected > 0 then capcost - differential cost >= 0
subject to capacityCosts{s in SPANS, t in Time}:
	capCost[s,t] = (if t = first(Time) then (CapTechCost[t]*SpanLength[s]*(spanCapacity[s,t] -
initSpanCapacity[s])) else (CapTechCost[t] * SpanLength[s] * (spanCapacity[s,t] -
spanCapacity[s,prev(t,Time)])) );

subject to no_capacity_reduction{s in SPANS, t in Time: ord(t, Time) > 1}:
	spanCapacity[s,t] - spanCapacity[s,prev(t,Time)] >= 0;

subject to minimum_span_count{t in Time}:
	sum{s in SPANS, tc in Time: ord(tc,Time) <= ord(t,Time)}(spanChange[s,tc]) >=
card(NODES);
	#if spans form a loop, this is the minimum requirement for the graph to be bi-connected

subject to bi_connected_nodes{n in NODES, t in Time}:
	sum{s in NODE_SPANS[n], tc in Time: ord(tc,Time) <= ord(t,Time)}(spanChange[s,tc]) >= 2;

# ****************************************************************
# Survivability model constraints
# ****************************************************************
```

```
# ***********************
# CONSTRAINTS
# ***********************

subject to calculate_span_capacity {s in SPANS, t in Time}:
	spanCapacity[s,t] = sum{r in Demands} span_flow[r,s,t];

subject to single_failure_restorability {r in Demands, n in (ORIGIN[r]), j in NODE_SPANS[n], t in Time}:
	sum{k in NODE_SPANS[n]: k<>j}span_flow[r,k,t] >= DemandVolume[r,t];
	# If a single path fails (starting with span i,j), the rest of the paths must have enough capacity to route traffic on it.

subject to translate_node_direction {r in Demands, n1 in NODES, n2 in ADJ_NODES[n1], t in Time}:
   DemandVolume[r,t] * node_direction[r, n1, n2,t] >= node_flow[r,n1,n2,t];

subject to node_flow_not_both_ways {r in Demands, n1 in NODES, n2 in ADJ_NODES[n1], t in Time}:
   node_direction[r, n1, n2,t] + node_direction[r, n2, n1,t] <= 1;

subject to disjoint_paths_1 {r in Demands, n1 in ((NODES diff ORIGIN[r]) diff DESTINATION[r]), t in Time}:
   sum{n2 in ADJ_NODES[n1]} node_direction[r,n1,n2,t] <= 1;

subject to disjoint_paths_2 {r in Demands, n1 in ((NODES diff ORIGIN[r]) diff DESTINATION[r]), t in Time}:
   sum{n2 in ADJ_NODES[n1]} node_direction[r,n2,n1,t] <= 1;

subject to transit_flow {r in Demands, n1 in ((NODES diff ORIGIN[r]) diff DESTINATION[r]), t in Time}:
	sum{n2 in ADJ_NODES[n1]} node_flow[r,n1,n2,t] = sum{n2 in ADJ_NODES[n1]} node_flow[r,n2,n1,t];
	# the incoming flow and the outgoing flow must be equal

subject to origin_flow{r in Demands, n1 in ORIGIN[r], n2 in ADJ_NODES[n1], i in (NODE_SPANS[n1] inter NODE_SPANS[n2]), t in Time}:
	node_flow[r,n1,n2,t] = span_flow[r,i,t];
	# If a span is connected to the origin, then the flow of traffic must be away from the origin node

subject to destination_flow{r in Demands, n1 in DESTINATION[r], n2 in ADJ_NODES[n1], i in (NODE_SPANS[n1] inter NODE_SPANS[n2]), t in Time}:
	node_flow[r,n2,n1,t] = span_flow[r,i,t];
	# if a span is connected to the destination then the traffic must be toward the destination

subject to one_way_traffic{r in Demands, n1 in NODES, n2 in ADJ_NODES[n1], i in (NODE_SPANS[n1] inter NODE_SPANS[n2]), t in Time}:
	node_flow[r,n1,n2,t] + node_flow[r,n2,n1,t] = span_flow[r,i,t];
	# traffic must only flow in one direction. This is only figurative, since once a link is established
	# from origin to destination, it is assumed that a link in the opposite direction is included.
```

```
subject to limit_span_flow{r in Demands, i in SPANS, t in Time}:
   span_flow[r,i,t] <= DemandVolume[r,t];

#subject to limited_path_length{d in Demands, t in Time}:
#          sum{n1 in NODES, n2 in ADJ_NODES[n1]}(node_direction[d,n1,n2,t]) <= card(SPANS);
           # the total number of spans used is less than the total number of spans in the network

#subject to mminimum_path_length{d in Demands, t in Time}:
#          sum{n1 in NODES, n2 in ADJ_NODES[n1]}(node_direction[d,n1,n2,t]) >= MinCycle[d];
           # the total number of spans used is less than the total number of spans in the network


#subject to APS {r in Demands, n1 in (ORIGIN[r]), t in Time}:
#          sum{n2 in ADJ_NODES[n1]} node_direction[r,n1,n2,t] = 2;
```

Appendix BAppendix C – Genetic Algorithm for Techno-Economic Network Design

```
# coding: utf-8

# In[1]:

import random
import numpy as np
import array
import multiprocessing
import itertools

import deap
import timeit

import networkx as nx
import matplotlib.pyplot as plt
import multiprocessing


# In[2]:


from deap import base
from deap import creator
from deap import tools
from deap import algorithms

from collections import Sequence
from itertools import repeat


# In[3]:

import csv


# In[4]:

TIME=['p1','p2','p3','p4','p5','p6','p7','p8','p9','p10','p11','p12','p13','p14','p15','p16','p17','p18','p19'
,'p20']
SPANS=['S1','S2','S3','S4','S5','S6','S7','S8','S9','S10','S11','S12','S13','S14','S15','S16','S17','S18','S19','
S20','S21','S22','S23','S24','S25','S26','S27','S28','S29','S30','S31','S32','S33','S34','S35','S36','S37','S3
8','S39','S40','S41','S42','S43','S44','S45','S46','S47','S48','S49','S50','S51','S52','S53','S54','S55','S56','
S57','S58','S59','S60','S61','S62','S63','S64','S65','S66','S67','S68','S69','S70','S71']
NODES=['N01','N02','N03','N04','N05','N06','N07','N08','N09','N10','N11','N12','N13','N14','N15']
DEMANDS=['D1','D2','D3','D4','D5','D6','D7','D8','D9','D10','D11','D12','D13','D14','D15','D16','D17','
D18','D19','D20','D21','D22','D23','D24','D25','D26','D27','D28','D29','D30','D31','D32','D33','D34','D
```

35','D36','D37','D38','D39','D40','D41','D42','D43','D44','D45','D46','D47','D48','D49','D50','D51','D5
2','D53','D54','D55','D56','D57','D58','D59','D60','D61','D62','D63','D64','D65','D66','D67','D68','D69'
,'D70','D71','D72','D73','D74','D75','D76','D77','D78','D79','D80','D81','D82','D83','D84','D85','D86','
D87','D88','D89','D90','D91','D92','D93','D94','D95','D96','D97','D98','D99','D100','D101','D102','D1
03','D104','D105']
FAILURE=['f0','S1','S2','S3','S4','S5','S6','S7','S8','S9','S10','S11','S12','S13','S14','S15','S16','S17','S18',
'S19','S20','S21','S22','S23','S24','S25','S26','S27','S28','S29','S30','S31','S32','S33','S34','S35','S36','S3
7','S38','S39','S40','S41','S42','S43','S44','S45','S46','S47','S48','S49','S50','S51','S52','S53','S54','S55','
S56','S57','S58','S59','S60','S61','S62','S63','S64','S65','S66','S67','S68','S69','S70','S71']
SpanLength =
{'S1':143,'S2':166,'S3':189,'S4':228,'S5':289,'S6':291,'S7':324,'S8':381,'S9':179,'S10':180,'S11':260,'S
12':270,'S13':292,'S14':380,'S15':381,'S16':129,'S17':142,'S18':207,'S19':210,'S20':224,'S21':294,'S
22':149,'S23':253,'S24':272,'S25':350,'S26':379,'S27':391,'S28':97,'S29':219,'S30':247,'S31':248,'S3
2':250,'S33':167,'S34':270,'S35':317,'S36':322,'S37':369,'S38':150,'S39':151,'S40':250,'S41':254,'S4
2':172,'S43':280,'S44':303,'S45':352,'S46':360,'S47':378,'S48':113,'S49':180,'S50':194,'S51':216,'S5
2':152,'S53':161,'S54':191,'S55':231,'S56':115,'S57':175,'S58':275,'S59':300,'S60':330,'S61':365,'S6
2':109,'S63':250,'S64':251,'S65':142,'S66':215,'S67':149,'S68':306,'S69':355,'S70':301,'S71':356}
DemandVolume =
{('D1','p1'):8,('D2','p1'):4,('D3','p1'):5,('D4','p1'):6,('D5','p1'):10,('D6','p1'):3,('D7','p1'):9,('D8','p1'):
9,('D9','p1'):5,('D10','p1'):10,('D11','p1'):1,('D12','p1'):1,('D13','p1'):4,('D14','p1'):4,('D15','p1'):8,('
D16','p1'):2,('D17','p1'):5,('D18','p1'):1,('D19','p1'):6,('D20','p1'):1,('D21','p1'):1,('D22','p1'):3,('D23
','p1'):9,('D24','p1'):6,('D25','p1'):5,('D26','p1'):5,('D27','p1'):7,('D28','p1'):2,('D29','p1'):9,('D30','p1
'):2,('D31','p1'):3,('D32','p1'):6,('D33','p1'):5,('D34','p1'):9,('D35','p1'):9,('D36','p1'):6,('D37','p1'):1,
('D38','p1'):4,('D39','p1'):2,('D40','p1'):3,('D41','p1'):2,('D42','p1'):4,('D43','p1'):7,('D44','p1'):9,('D
45','p1'):3,('D46','p1'):5,('D47','p1'):6,('D48','p1'):4,('D49','p1'):2,('D50','p1'):1,('D51','p1'):10,('D52
','p1'):7,('D53','p1'):9,('D54','p1'):6,('D55','p1'):3,('D56','p1'):3,('D57','p1'):1,('D58','p1'):1,('D59','p1
'):7,('D60','p1'):3,('D61','p1'):4,('D62','p1'):6,('D63','p1'):4,('D64','p1'):5,('D65','p1'):2,('D66','p1'):5,
('D67','p1'):8,('D68','p1'):9,('D69','p1'):3,('D70','p1'):5,('D71','p1'):5,('D72','p1'):7,('D73','p1'):7,('D
74','p1'):1,('D75','p1'):9,('D76','p1'):4,('D77','p1'):7,('D78','p1'):2,('D79','p1'):3,('D80','p1'):6,('D81','
p1'):8,('D82','p1'):8,('D83','p1'):2,('D84','p1'):10,('D85','p1'):8,('D86','p1'):6,('D87','p1'):8,('D88','p1
'):10,('D89','p1'):1,('D90','p1'):1,('D91','p1'):10,('D92','p1'):2,('D93','p1'):1,('D94','p1'):7,('D95','p1')
:1,('D96','p1'):5,('D97','p1'):7,('D98','p1'):1,('D99','p1'):3,('D100','p1'):3,('D101','p1'):5,('D102','p1')
:1,('D103','p1'):10,('D104','p1'):3,('D105','p1'):3,('D1','p2'):16,('D2','p2'):10,('D3','p2'):14,('D4','p2')
:11,('D5','p2'):10,('D6','p2'):5,('D7','p2'):21,('D8','p2'):15,('D9','p2'):10,('D10','p2'):13,('D11','p2'):2,
('D12','p2'):2,('D13','p2'):7,('D14','p2'):8,('D15','p2'):15,('D16','p2'):4,('D17','p2'):10,('D18','p2'):2,('
D19','p2'):10,('D20','p2'):2,('D21','p2'):2,('D22','p2'):7,('D23','p2'):22,('D24','p2'):16,('D25','p2'):11,
('D26','p2'):9,('D27','p2'):10,('D28','p2'):4,('D29','p2'):19,('D30','p2'):4,('D31','p2'):5,('D32','p2'):12,
('D33','p2'):6,('D34','p2'):23,('D35','p2'):14,('D36','p2'):11,('D37','p2'):1,('D38','p2'):8,('D39','p2'):4,
('D40','p2'):6,('D41','p2'):4,('D42','p2'):6,('D43','p2'):15,('D44','p2'):25,('D45','p2'):7,('D46','p2'):8,('
D47','p2'):12,('D48','p2'):6,('D49','p2'):3,('D50','p2'):2,('D51','p2'):19,('D52','p2'):14,('D53','p2'):15,
('D54','p2'):9,('D55','p2'):7,('D56','p2'):7,('D57','p2'):2,('D58','p2'):2,('D59','p2'):8,('D60','p2'):7,('D
61','p2'):6,('D62','p2'):10,('D63','p2'):7,('D64','p2'):5,('D65','p2'):3,('D66','p2'):11,('D67','p2'):14,('D
68','p2'):9,('D69','p2'):6,('D70','p2'):10,('D71','p2'):11,('D72','p2'):12,('D73','p2'):11,('D74','p2'):3,('
D75','p2'):18,('D76','p2'):8,('D77','p2'):13,('D78','p2'):3,('D79','p2'):5,('D80','p2'):8,('D81','p2'):17,('
D82','p2'):19,('D83','p2'):2,('D84','p2'):26,('D85','p2'):12,('D86','p2'):10,('D87','p2'):15,('D88','p2'):
23,('D89','p2'):2,('D90','p2'):2,('D91','p2'):20,('D92','p2'):4,('D93','p2'):1,('D94','p2'):17,('D95','p2'):
3,('D96','p2'):11,('D97','p2'):17,('D98','p2'):2,('D99','p2'):9,('D100','p2'):6,('D101','p2'):11,('D102','
p2'):3,('D103','p2'):24,('D104','p2'):6,('D105','p2'):7}
NODE_SPANS =
{'N01':('S1','S2','S3','S4','S5','S6','S7','S8'),'N02':('S2','S9','S10','S11','S12','S13','S14','S15'),'N03':('S5'

```
,'S10','S16','S17','S18','S19','S20','S21','S46','S71'),'N04':('S11','S16','S22','S23','S24','S25','S26','S27')
,'N05':('S3','S9','S17','S24','S28','S29','S30','S31','S32','S36','S47','S61','S63','S68'),'N06':('S15','S18','S
22','S33','S34','S35','S36','S37'),'N07':('S14','S19','S23','S30','S33','S38','S39','S40','S41','S60'),'N08':(
'S27','S34','S38','S42','S43','S44','S45','S46','S47'),'N09':('S21','S26','S32','S35','S39','S42','S48','S49','
S50','S51','S69'),'N10':('S4','S12','S20','S25','S28','S37','S40','S45','S50','S52','S53','S54','S55','S58'),'
N11':('S44','S49','S56','S57','S58','S59','S60','S61'),'N12':('S8','S41','S43','S48','S53','S56','S62','S63','
S64','S70'),'N13':('S7','S31','S51','S52','S57','S62','S65','S66'),'N14':('S6','S55','S59','S64','S65','S67','S
68','S69'),'N15':('S1','S13','S29','S54','S66','S67','S70','S71')}
ORIGIN =
{'D1':'N01','D2':'N01','D3':'N01','D4':'N01','D5':'N01','D6':'N01','D7':'N01','D8':'N01','D9':'N01','D10'
:'N01','D11':'N01','D12':'N01','D13':'N01','D14':'N01','D15':'N02','D16':'N02','D17':'N02','D18':'N02',
'D19':'N02','D20':'N02','D21':'N02','D22':'N02','D23':'N02','D24':'N02','D25':'N02','D26':'N02','D27':'
N02','D28':'N03','D29':'N03','D30':'N03','D31':'N03','D32':'N03','D33':'N03','D34':'N03','D35':'N03','
D36':'N03','D37':'N03','D38':'N03','D39':'N03','D40':'N04','D41':'N04','D42':'N04','D43':'N04','D44':'
N04','D45':'N04','D46':'N04','D47':'N04','D48':'N04','D49':'N04','D50':'N04','D51':'N05','D52':'N05','
D53':'N05','D54':'N05','D55':'N05','D56':'N05','D57':'N05','D58':'N05','D59':'N05','D60':'N05','D61':'
N06','D62':'N06','D63':'N06','D64':'N06','D65':'N06','D66':'N06','D67':'N06','D68':'N06','D69':'N06','
D70':'N07','D71':'N07','D72':'N07','D73':'N07','D74':'N07','D75':'N07','D76':'N07','D77':'N07','D78':'
N08','D79':'N08','D80':'N08','D81':'N08','D82':'N08','D83':'N08','D84':'N08','D85':'N09','D86':'N09','
D87':'N09','D88':'N09','D89':'N09','D90':'N09','D91':'N10','D92':'N10','D93':'N10','D94':'N10','D95':'
N10','D96':'N11','D97':'N11','D98':'N11','D99':'N11','D100':'N12','D101':'N12','D102':'N12','D103':'
N13','D104':'N13','D105':'N14'}
DESTINATION =
{'D1':'N02','D2':'N03','D3':'N04','D4':'N05','D5':'N06','D6':'N07','D7':'N08','D8':'N09','D9':'N10','D10'
:'N11','D11':'N12','D12':'N13','D13':'N14','D14':'N15','D15':'N03','D16':'N04','D17':'N05','D18':'N06',
'D19':'N07','D20':'N08','D21':'N09','D22':'N10','D23':'N11','D24':'N12','D25':'N13','D26':'N14','D27':'
N15','D28':'N04','D29':'N05','D30':'N06','D31':'N07','D32':'N08','D33':'N09','D34':'N10','D35':'N11','
D36':'N12','D37':'N13','D38':'N14','D39':'N15','D40':'N05','D41':'N06','D42':'N07','D43':'N08','D44':'
N09','D45':'N10','D46':'N11','D47':'N12','D48':'N13','D49':'N14','D50':'N15','D51':'N06','D52':'N07','
D53':'N08','D54':'N09','D55':'N10','D56':'N11','D57':'N12','D58':'N13','D59':'N14','D60':'N15','D61':'
N07','D62':'N08','D63':'N09','D64':'N10','D65':'N11','D66':'N12','D67':'N13','D68':'N14','D69':'N15','
D70':'N08','D71':'N09','D72':'N10','D73':'N11','D74':'N12','D75':'N13','D76':'N14','D77':'N15','D78':'
N09','D79':'N10','D80':'N11','D81':'N12','D82':'N13','D83':'N14','D84':'N15','D85':'N10','D86':'N11','
D87':'N12','D88':'N13','D89':'N14','D90':'N15','D91':'N11','D92':'N12','D93':'N13','D94':'N14','D95':'
N15','D96':'N12','D97':'N13','D98':'N14','D99':'N15','D100':'N13','D101':'N14','D102':'N15','D103':'
N14','D104':'N15','D105':'N15'}


STEADY_STATE_DEMAND = 1000 # multiply this by the first period demand volume to get the total
potential adopters
p = 0.06
q = 0.3
# These are used in the Generalized Bass Model  to differentiate advertising and pricing influences.
beta_1 = 1 # for the capacity factor
beta_2 = 1 # for the user intervention factor (pricing)

ImpTechCost = {'p1':500.00,'p2':454.54,'p3':413.22}
CapTechCost = {'p1':5.0,'p2':3.5,'p3':2.5}
ARU = 1000 #average revenue per unit of demand
```

```
AddCostUnit = 1000 #cost per advertising factor (note that 1 represents no advertising)

impDiscountRate = 0.1
capDiscountRate = 0.25
DiscountRate = 0.1
att_count = 0


# In[5]:

# LOAD NETWORK
def loadNetwork(top):
  net_file = open(top, 'r')
  is_node = False
  is_span = False
  nodes = [] # formatted (name, x, y)
  spans = [] # formatted (name, n1, n2, cost, capacity)
  net = nx.Graph()

  for line in net_file:
    words = line.split()
    if len(words) > 0:
      if is_node:
        nodes.append([words[0], words[1], words[2],])
        net.add_node(words[0], pos=(int(words[1]),int(words[2])))
      if is_span:
        spans.append([words[0], words[1], words[2], words[7],0,])
        net.add_edge(words[1],words[2], name=words[0], weight=float(words[7]), cap=0)
      if words[0] == 'NODE':
        is_node = True
        is_span = False
      elif words[0] == 'SPAN':
        is_span = True
        is_node = False
    else:
      is_node = False
      is_span = False
  net_file.close()
  return net


# In[6]:

base_net = loadNetwork("../Networks/Highly-Connected-n-2/HCon-15n30s1-71s.top")

# get network for each failure scenario
fail_net = dict()
for f in FAILURE:
  if f == FAILURE[0]:
    fail_net[f] = base_net
```

```
  else:
    span_nodes = [x for x in base_net.edges(data='name') if x[2] == f]
    f_net = base_net.copy()
    f_net.remove_edge(span_nodes[0][0],span_nodes[0][1])
    fail_net[f] = f_net


# In[7]:

SPAN_NODES = dict()
SPANS_BY_NODE = dict()
for u,v,d in base_net.edges_iter(data=True):
  l = d['name']
  SPAN_NODES[l] = SPAN_NODES.get(l, [])
  SPAN_NODES[l].append((u,v))
  SPANS_BY_NODE[(u,v)] = l
  SPANS_BY_NODE[(v,u)] = l


# In[8]:

#d = 'D5'
#zzz = nx.dijkstra_path(base_net,ORIGIN[d], DESTINATION[d],weight='weight')


# In[9]:

#zzz


# In[10]:

net_list= []

for s in SPANS:
  net_list.append(s)

net_list.append("cap")
net_list.append("usr")

# capacity factor has to be greater than 0, user factor has to be greater than 1


# In[11]:

#len(spancapacity)
att_count = len(net_list)


# In[12]:
```

```
#evaluation function
def networkValue(individual):
  # Determine user growth
  encoded_results = dict(zip(net_list,individual))
  users = dict()
  capacity = dict()
  discounted_add_cost = dict()
  discounted_revenue = dict()
  previous_t = TIME[0]
  a = encoded_results["cap"]
  b = encoded_results["usr"]
  for d in DEMANDS:

    if a == 0 or b == 0:
      return 0
    x = 1 + beta_1*np.log(a) + beta_2*np.log(b)
    m = DemandVolume[(d,TIME[0])] * STEADY_STATE_DEMAND
    if m == 0:
      print('Problem with the demand %s : %i' %(d,m))
    for t in TIME:
      Current_Users = 0
      if t != TIME[0]:
        Current_Users = users[(d,previous_t)]
      previous_t = t
      users[(d,t)] = Current_Users + (p*m + (q-p)*Current_Users - q/m*Current_Users**2.0)*x
      capacity[(d,t)] = users[(d,t)] * a
      discounted_add_cost[(d,t)] = (b-1) * AddCostUnit / (1+DiscountRate)**TIME.index(t)
      discounted_revenue[(d,t)] = users[(d,t)] * ARU / (1+DiscountRate)**TIME.index(t)
  # Determine capacity
  # Determine calculated costs

  # individual is a set of path selections indexed by [t,d,f]
  spancap = dict()
  impCost = dict()
  capCost = dict()
  shortest_paths = dict()

  #for x in encoded_results:
  #  print(x + '-' + str(encoded_results[x]))

  for t in TIME:
    cap=0
    net_change = 0 # marks if the network has changed
    if t == TIME[0]:
      # Copy the network for the first time, then just keep adding spans
      temp_net = base_net.copy()
      #span_names = nx.get_edge_attributes(temp_net,'name')
      # get spans for the network
      for span in SPANS:
```

```
      if encoded_results[(span)] > TIME.index(t):
        span_nodes = [x for x in temp_net.edges(data='name') if x[2] == span]
        if len(span_nodes) > 0:
        #if span in span_names:
          temp_net.remove_edge(span_nodes[0][0],span_nodes[0][1])
          #temp_net.remove_edge(span_names[span])
          #print('remove span - ' + span)
    else:
      for span in SPANS:
      if encoded_results[(span)] == TIME.index(t):
        # add the span to the network
        span_nodes = [x for x in base_net.edges(data='name') if x[2] == span]
        temp_net.add_edge(span_nodes[0][0],span_nodes[0][1],
name=base_net[span_nodes[0][0]][span_nodes[0][1]]['name'],
weight=base_net[span_nodes[0][0]][span_nodes[0][1]]['weight'])
        # mark that the topology has changed
        net_change = 1
        #print('Span Added %s' % span)
    #print([x for x in temp_net.edges(data='name')])
    for f in FAILURE:
      # Remove failed span
      span_removed = False
      if f != FAILURE[0]:
        span_nodes = [x for x in base_net.edges(data='name') if x[2] == f]

        if temp_net.has_edge(span_nodes[0][0],span_nodes[0][1]):
          temp_net.remove_edge(span_nodes[0][0],span_nodes[0][1])
          span_removed = True

      # route all paths in a shortest path scenario
      if t == TIME[0]:
        #temp_net.edges()
        for d in DEMANDS:
          shortest_paths[(f,d)]          =          nx.dijkstra_path(temp_net,ORIGIN[d],
DESTINATION[d],weight='weight')
      elif net_change == 1:
        # this period has a span added, re-route the paths
        for d in DEMANDS:
          shortest_paths[(f,d)]          =          nx.dijkstra_path(temp_net,ORIGIN[d],
DESTINATION[d],weight='weight')

      #for each span, check if the path contains the span and if so, add it to the capacity for this failure
scenario
      for s in SPANS:
        cap=0
        for d in DEMANDS:
          path_spans = []
          #path_weight = 0
          for x in range(len(shortest_paths[(f,d)])-1):
            #pass
```

```
        #path_weight += SpanLength[SPANS_BY_NODE[(shortest_path[x],shortest_path[x+1])]]
        #path_weight += [temp_net.get_edge_data(shortest_path[x],shortest_path[x+1])['weight']]
        path_spans.append(SPANS_BY_NODE[(shortest_paths[(f,d)][x],shortest_paths[(f,d)][x+1])])
       cap += capacity[(d,t)] if s in path_spans else 0
      if (s,t) in spancap:
       if spancap[(s,t)] < cap:
        spancap[(s,t)] = cap
      else:
       spancap[(s,t)] = cap
     if f != FAILURE[0] and span_removed:
      temp_net.add_edge(span_nodes[0][0],span_nodes[0][1],
name=base_net[span_nodes[0][0]][span_nodes[0][1]]['name'],
weight=base_net[span_nodes[0][0]][span_nodes[0][1]]['weight'])
   t0 = TIME.index(t)
   t_1 = TIME[t0-1]
   #print('Time %i, %s,%s' % (t0,t,t_1))
   for s in SPANS:
    #calculate costs

    if t0 == 0:
     #all capacities in the first are implemenation costs
     if spancap[(s,t)] > 0:
      impCost[(s,t)] = SpanLength[s] * ImpTechCost['p1'] / (1+impDiscountRate)**TIME.index(t)
      capCost[(s,t)] = spancap[(s,t)] * SpanLength[s] * CapTechCost['p1'] /
(1+capDiscountRate)**TIME.index(t)
     else:
      impCost[(s,t)] = 0
      capCost[(s,t)] = 0
    else:
     t_1 = TIME[t0-1]
     #print('Span Cap: %s,%s - %i --- %s - %i' % (s,t,spancap[(s,t)],t_1,spancap[(s,t_1)]))
     if spancap[(s,t_1)] > 0:
      #no imp cost and incremental cap costs
      capCost[(s,t)] = (spancap[(s,t)] - spancap[(s,t_1)]) * SpanLength[s] * CapTechCost['p1'] /
(1+capDiscountRate)**TIME.index(t)
      impCost[(s,t)] = 0
     elif spancap[(s,t)] > 0:
      #new span
      impCost[(s,t)] = SpanLength[s] * ImpTechCost['p1'] / (1+impDiscountRate)**TIME.index(t)
      capCost[(s,t)] = spancap[(s,t)] * SpanLength[s] * CapTechCost['p1'] /
(1+capDiscountRate)**TIME.index(t)
      #print('New Span %s at time %i'% (s,t0))
     else:
      impCost[(s,t)] = 0
      capCost[(s,t)] = 0

  #calculate discounted cost
  npv_cap = sum(capCost.values())
  npv_imp = sum(impCost.values())
  npv_add = sum(discounted_add_cost.values())
```

```
  npv_rev = sum(discounted_revenue.values())

  #print (str(npv_cap) + ',' + str(npv_imp) + ',' + str(npv_cap + npv_imp))
  return (npv_rev - (npv_cap + npv_imp+npv_add),)


# In[13]:

def networkFeasible(individual):
  # Demand is routed check if the path number is below the max path for each
  encoded_results = dict(zip(net_list,individual))
  a = encoded_results["cap"]
  b = encoded_results["usr"]
  # a > 1, b > 0, p' + q' < 1
  if a == 0 or b == 0:
    return False
  x = 1 + beta_1*np.log(a) + beta_2*np.log(b)
  if (a < 0) or (x*(p+q) >= 1) or (b<1) or (x<0):
    #print('failed a/b:' + str(a) + ',' + str(b) + ',' + str(x))
    return False

  for t in TIME:
    for f in FAILURE:
      cap=0
      temp_net = fail_net[f].copy()
      #span_names = nx.get_edge_attributes(temp_net,'name')
      for span in SPANS:
        if encoded_results[(span)] > TIME.index(t):
          span_nodes = [x for x in temp_net.edges(data='name') if x[2] == span]
          if len(span_nodes) > 0:
          #if span in span_names:
            temp_net.remove_edge(span_nodes[0][0],span_nodes[0][1])
            #temp_net.remove_edge(span_names[span])
            #print('remove span' + span)
      # try to route each demand
      for d in DEMANDS:
        o_node = ORIGIN[d]
        d_node = DESTINATION[d]
        if not nx.has_path(temp_net,ORIGIN[d], DESTINATION[d]):
          #print('No path: ' + t + ',' + f)
          return False

  return True


# In[14]:

def networkDistance(individual):
  encoded_results = dict(zip(net_list,individual))
  distance = 0
```

```
  # a > 1, b > 0, p' + q' < 1
  a = encoded_results["cap"]
  b = encoded_results["usr"]
  if a == 0 or b == 0:
    distance += 1
  else:
    x = 1 + beta_1*np.log(a) + beta_2*np.log(b)
    if (a < 1) or (x*(p+q) >= 1) or (b<0) or (x<0):
      distance += 1
  for t in TIME:
    for f in FAILURE:
      cap=0
      temp_net = fail_net[f].copy()
      #span_names = nx.get_edge_attributes(temp_net,'name')
      for span in SPANS:
        if encoded_results[(span)] > TIME.index(t):
          span_nodes = [x for x in temp_net.edges(data='name') if x[2] == span]
          if len(span_nodes) > 0:
          #if span in span_names:
            temp_net.remove_edge(span_nodes[0][0],span_nodes[0][1])
            #temp_net.remove_edge(span_names[span])
            #print('remove span' + span)
      # try to route each demand
      for d in DEMANDS:
        o_node = ORIGIN[d]
        d_node = DESTINATION[d]
        if not nx.has_path(temp_net,ORIGIN[d], DESTINATION[d]):
          distance += 1
  return distance



# In[15]:

def initPop(pop):
  #get the encoding for the shortest cycle
  cycle = [0,0,21,21,0,21,21,0,21,0,0,21,21,0,21,21,0,21,21]
  temp_ind = []

  #print ([str(x) for x in temp_ind])
  #print ([str(x) for x in [y for y in pop]])
  #for each individual
  for ind in pop:
    # find the shortest cycle
    #ind = list(temp_ind)
    #for x in range(len(SPANS)):
    #  if cycle[x] > 0:
    #    ind[x] = random.randint(0,len(TIME))
    #  else:
    #    ind[x] = 0
```

```
    #set all time periods to this topology
    i = len(SPANS)
      #set the capacity investment rate
      # first determine if capacity or advertising should be randomized first:
    if random.random() >0.5:
      ind[i] = random.triangular(0.7,3,1.5)
      x = random.triangular(0,1,0.36)
      ind[i+1] = max((np.exp(x/(p+q) - 1 - np.log(ind[i])),1))
    else:
      ind[i+1] = random.triangular(1,3,1.5)
      x = random.triangular(0,1,0.36)
      ind[i] = max((np.exp(x/(p+q) - 1 - np.log(ind[i+1])),0.7))

  #print ([str(x) for x in [y for y in pop]])
  return pop


# In[16]:

def mutNetwork(individual, spanpb, growthpb):
  #span mutations:
  # randomly set a span to be part of the network
  encoded_results = dict(zip(net_list,individual))
  if random.random() < spanpb:
    s = random.randint(0,len(SPANS)-1)
    t = random.randint(0,len(TIME))
    encoded_results[(SPANS[s])] = t
    if individual[s] > t:
      # Span is being pushed to an earlier time
      individual[s] = t
    else:
      #don't disconnect the network???
      temp_net = base_net.copy()
      #span_names = nx.get_edge_attributes(temp_net,'name')
      for span in SPANS:
        if encoded_results[(span)] > t:
          span_nodes = [x for x in temp_net.edges(data='name') if x[2] == span]
          if len(span_nodes) > 0:
          #if span in span_names:
            temp_net.remove_edge(span_nodes[0][0],span_nodes[0][1])
            #print('remove span' + span)
      #print('degree: %i' % min(list(temp_net.degree().values())))
      while min(list(temp_net.degree().values())) < 2 and t > 0:
        # Try to go back in time and add an edge
        t -= 1
        span = SPANS[s]
        span_nodes = [x for x in temp_net.edges(data='name') if x[2] == span]
        temp_net.add_edge(span_nodes[0][0],span_nodes[0][1])
      individual[s] = t
  # growth mutations
```

```
 i = len(SPANS)
 if random.random() < growthpb:
   #set the capacity investment rate
   # first determine if capacity or advertising should be randomized first:
   if random.random() >0.5:
     individual[i] = random.uniform(0.7,3)
     x = random.uniform(0,1)
     individual[i+1] = np.exp(x/(p+q) - 1 - np.log(individual[i]))
   else:
     individual[i+1] = random.uniform(1,3)
     x = random.uniform(0,1)
     individual[i] = np.exp(x/(p+q) - 1 - np.log(individual[i+1]))

     #print('New growth cap: %d, add: %d' % (individual[i],individual[i+1]))
 return individual,



# In[17]:

#Multiprocessing
pool = multiprocessing.Pool(4)

#GA
creator.create("FitnessMin", base.Fitness, weights=(-1.0,))
creator.create("FitnessMax", base.Fitness, weights=(1.0,))
creator.create("Individual",list,fitness=creator.FitnessMax)
toolbox = base.Toolbox()
toolbox.register("map", pool.map)


# In[18]:

# Attributes
max_cap = 0
toolbox.register("span_cap",random.uniform,0,max_cap)


#individual
toolbox.register("individual", tools.initRepeat, creator.Individual, toolbox.span_cap, att_count)

#population
toolbox.register("population", tools.initRepeat, list, toolbox.individual)

toolbox.register("evaluate", networkValue)
toolbox.decorate("evaluate", tools.DeltaPenality(networkFeasible, -10000000000000, networkDistance))
toolbox.register("mate", tools.cxTwoPoint)
#toolbox.register("mutate", tools.mutUniformInt, low=0, up=max_cap, indpb=0.1)
```

```
#toolbox.register("mutate", tools.mutFlipBit, indpb=0.15)
#toolbox.register("mutate", tools.mutUniformInt, low=0, up=1, indpb=0.25)
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.3)
toolbox.register("select", tools.selTournament, tournsize=3)


# In[19]:

def main():
  NGEN = 100
  MU = 50
  LAMBDA = 20
  CXPB = 0.3
  MUTPB = 0.6

  pop = toolbox.population(n=MU)
  initPop(pop)
  hof = tools.ParetoFront()
  stats = tools.Statistics(lambda ind: ind.fitness.values)
  stats.register("avg", np.mean, axis=0)
  stats.register("std", np.std, axis=0)
  stats.register("min", np.min, axis=0)
  stats.register("max", np.max, axis=0)

  algorithms.eaMuPlusLambda(pop, toolbox, MU, LAMBDA, CXPB, MUTPB, NGEN, stats,
            halloffame=hof)

  return pop, stats, hof


# In[20]:

def restart(pop):
  NGEN = 100
  MU = 50
  LAMBDA = 20
  CXPB = 0.3
  MUTPB = 0.2


  hof = tools.ParetoFront()
  stats = tools.Statistics(lambda ind: ind.fitness.values)
  stats.register("avg", np.mean, axis=0)
  stats.register("std", np.std, axis=0)
  stats.register("min", np.min, axis=0)
  stats.register("max", np.max, axis=0)

  algorithms.eaMuPlusLambda(pop, toolbox, MU, LAMBDA, CXPB, MUTPB, NGEN, stats,
            halloffame=hof)
```

```
  return pop, stats, hof


# In[21]:

#evaluation function
def write_results(individual, file_name):
  # individual is a set of path selections indexed by [t,d,f]
  output = open(file_name,'w')
  file_writer = csv.writer(output)

# Determine user growth
  encoded_results = dict(zip(net_list,individual))
  users = dict()
  capacity = dict()
  discounted_add_cost = dict()
  discounted_revenue = dict()
  previous_t = TIME[0]
  a = encoded_results["cap"]
  b = encoded_results["usr"]
  for d in DEMANDS:

    if a == 0 or b == 0:
      return 0
    x = 1 + beta_1*np.log(a) + beta_2*np.log(b)
    m = DemandVolume[(d,TIME[0])] * STEADY_STATE_DEMAND
    if m == 0:
      print('Problem with the demand %s : %i' %(d,m))
    for t in TIME:
      Current_Users = 0
      if t != TIME[0]:
        Current_Users = users[(d,previous_t)]
      previous_t = t
      users[(d,t)] = Current_Users + (p*m + (q-p)*Current_Users - q/m*Current_Users**2.0)*x
      capacity[(d,t)] = users[(d,t)] * a
      discounted_add_cost[(d,t)] = (b-1) * AddCostUnit / (1+DiscountRate)**TIME.index(t)
      discounted_revenue[(d,t)] = users[(d,t)] * ARU / (1+DiscountRate)**TIME.index(t)
  # Determine capacity
  # Determine calculated costs

  # individual is a set of path selections indexed by [t,d,f]
  spancap = dict()
  impCost = dict()
  capCost = dict()
  shortest_paths = dict()

  #for x in encoded_results:
  #  print(x + '-' + str(encoded_results[x]))

  for t in TIME:
```

```
    cap=0
    net_change = 0 # marks if the network has changed
    if t == TIME[0]:
      # Copy the network for the first time, then just keep adding spans
      temp_net = base_net.copy()
      #span_names = nx.get_edge_attributes(temp_net,'name')
      # get spans for the network
      for span in SPANS:
        if encoded_results[(span)] > TIME.index(t):
          span_nodes = [x for x in temp_net.edges(data='name') if x[2] == span]
          if len(span_nodes) > 0:
          #if span in span_names:
            temp_net.remove_edge(span_nodes[0][0],span_nodes[0][1])
            #temp_net.remove_edge(span_names[span])
            #print('remove span - ' + span)
    else:
      for span in SPANS:
        if encoded_results[(span)] == TIME.index(t):
          # add the span to the network
          span_nodes = [x for x in base_net.edges(data='name') if x[2] == span]
          temp_net.add_edge(span_nodes[0][0],span_nodes[0][1],
name=base_net[span_nodes[0][0]][span_nodes[0][1]]['name'],
weight=base_net[span_nodes[0][0]][span_nodes[0][1]]['weight'])
          # mark that the topology has changed
          net_change = 1
          #print('Span Added %s' % span)
    #print([x for x in temp_net.edges(data='name')])
    for f in FAILURE:
      # Remove failed span
      span_removed = False
      if f != FAILURE[0]:
        span_nodes = [x for x in base_net.edges(data='name') if x[2] == f]

        if temp_net.has_edge(span_nodes[0][0],span_nodes[0][1]):
          temp_net.remove_edge(span_nodes[0][0],span_nodes[0][1])
          span_removed = True

      # route all paths in a shortest path scenario
      if t == TIME[0]:
        #temp_net.edges()
        for d in DEMANDS:
          shortest_paths[(f,d)]          =          nx.dijkstra_path(temp_net,ORIGIN[d],
DESTINATION[d],weight='weight')
      elif net_change == 1:
        # this period has a span added, re-route the paths
        for d in DEMANDS:
          shortest_paths[(f,d)]          =          nx.dijkstra_path(temp_net,ORIGIN[d],
DESTINATION[d],weight='weight')
```

```
    #for each span, check if the path contains the span and if so, add it to the capacity for this failure
scenario
    for s in SPANS:
      cap=0
      for d in DEMANDS:
        path_spans = []
        #path_weight = 0
        for x in range(len(shortest_paths[(f,d)])-1):
          #pass
          #path_weight += SpanLength[SPANS_BY_NODE[(shortest_path[x],shortest_path[x+1])]]
          #path_weight += [temp_net.get_edge_data(shortest_path[x],shortest_path[x+1])['weight']]
          path_spans.append(SPANS_BY_NODE[(shortest_paths[(f,d)][x],shortest_paths[(f,d)][x+1])])
        cap += capacity[(d,t)] if s in path_spans else 0
      if (s,t) in spancap:
        if spancap[(s,t)] < cap:
          spancap[(s,t)] = cap
      else:
        spancap[(s,t)] = cap
    if f != FAILURE[0] and span_removed:
      temp_net.add_edge(span_nodes[0][0],span_nodes[0][1],
name=base_net[span_nodes[0][0]][span_nodes[0][1]]['name'],
weight=base_net[span_nodes[0][0]][span_nodes[0][1]]['weight'])
  t0 = TIME.index(t)
  t_1 = TIME[t0-1]
  #print('Time %i, %s,%s' % (t0,t,t_1))
  for s in SPANS:
    #calculate costs

    if t0 == 0:
      #all capacities in the first are implemenation costs
      if spancap[(s,t)] > 0:
        impCost[(s,t)] = SpanLength[s] * ImpTechCost['p1'] / (1+impDiscountRate)**TIME.index(t)
        capCost[(s,t)] = spancap[(s,t)] * SpanLength[s] * CapTechCost['p1'] /
(1+capDiscountRate)**TIME.index(t)
      else:
        impCost[(s,t)] = 0
        capCost[(s,t)] = 0
    else:
      t_1 = TIME[t0-1]
      #print('Span Cap: %s,%s - %i --- %s - %i' % (s,t,spancap[(s,t)],t_1,spancap[(s,t_1)]))
      if spancap[(s,t_1)] > 0:
        #no imp cost and incremental cap costs
        capCost[(s,t)] = (spancap[(s,t)] - spancap[(s,t_1)]) * SpanLength[s] * CapTechCost['p1'] /
(1+capDiscountRate)**TIME.index(t)
        impCost[(s,t)] = 0
      elif spancap[(s,t)] > 0:
        #new span
        impCost[(s,t)] = SpanLength[s] * ImpTechCost['p1'] / (1+impDiscountRate)**TIME.index(t)
        capCost[(s,t)] = spancap[(s,t)] * SpanLength[s] * CapTechCost['p1'] /
(1+capDiscountRate)**TIME.index(t)
```

```
        #print('New Span %s at time %i'% (s,t0))
      else:
        impCost[(s,t)] = 0
        capCost[(s,t)] = 0


  #calculate discounted cost
  npv_cap = sum(capCost.values())
  npv_imp = sum(impCost.values())
  npv_add = sum(discounted_add_cost.values())
  npv_rev = sum(discounted_revenue.values())

  temp = spancap
  file_writer.writerow([s+'-'+t for t in TIME for s in SPANS])
  file_writer.writerow([spancap[(s,t)] for t in TIME for s in SPANS])
  file_writer.writerow([impCost[(s,t)] for t in TIME for s in SPANS])
  file_writer.writerow([capCost[(s,t)] for t in TIME for s in SPANS])

  file_writer.writerow([d for d in DEMANDS for t in TIME])
  file_writer.writerow([t for d in DEMANDS for t in TIME])
  file_writer.writerow([users[(d,t)] for d in DEMANDS for t in TIME])
  file_writer.writerow([capacity[(d,t)] for d in DEMANDS for t in TIME])
  file_writer.writerow([discounted_add_cost[(d,t)] for d in DEMANDS for t in TIME])


  file_writer.writerow([encoded_results['cap']])
  file_writer.writerow([encoded_results['usr']])

  output.write(str(npv_cap))
  output.write(',')
  output.write(str(npv_imp))
  output.write(',')
  output.write(str(npv_add))
  output.write(',')
  output.write(str(npv_rev))
  output.write(',')
  output.write(str(npv_rev - (npv_cap + npv_imp+npv_add)))



  #print (str(npv_cap) + ',' + str(npv_imp) + ',' + str(npv_cap + npv_imp))
  output.close()

  return spancap #(npv_cap + npv_imp,)



# In[ ]:

ImpTechCost = {'p1':50.000,'p2':45.454,'p3':41.322}
```

```
CapTechCost = {'p1':5.0,'p2':3.5,'p3':2.5}
ARU = 100 #average revenue per unit of demand
AddCostUnit = 1000000 #cost per advertising factor (note that 1 represents no advertising)
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.2)
results = main()
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.0)
results = restart(results[0])
toolbox.register("mutate", mutNetwork, spanpb = 0.0, growthpb = 0.20)
results = restart(results[0])

temp = dict()
temp = write_results(results[2][0],'15n GA_Top_growth_model_scaled_path-imp-mod_mut 10 - rev 100 - uniform2.csv')


# In[ ]:

# ARU of 1000
ImpTechCost = {'p1':500.00,'p2':454.54,'p3':413.22}
CapTechCost = {'p1':5.0,'p2':3.5,'p3':2.5}
ARU = 1000 #average revenue per unit of demand
AddCostUnit = 1000000 #cost per advertising factor (note that 1 represents no advertising)
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.2)
results = main()
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.0)
results = restart(results[0])
toolbox.register("mutate", mutNetwork, spanpb = 0.0, growthpb = 0.20)
results = restart(results[0])

temp = dict()
temp = write_results(results[2][0],'15n GA_Top_growth_model_scaled_path-imp-mod_mut 100 - rev 1000 - uniform.csv')

ImpTechCost = {'p1':50.000,'p2':45.454,'p3':41.322}
CapTechCost = {'p1':5.0,'p2':3.5,'p3':2.5}
ARU = 1000 #average revenue per unit of demand
AddCostUnit = 1000000 #cost per advertising factor (note that 1 represents no advertising)
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.2)
results = main()
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.0)
results = restart(results[0])
toolbox.register("mutate", mutNetwork, spanpb = 0.0, growthpb = 0.20)
results = restart(results[0])

temp = dict()
temp = write_results(results[2][0],'15n GA_Top_growth_model_scaled_path-imp-mod_mut 10 - rev 1000 - uniform.csv')

ImpTechCost = {'p1':5000.0,'p2':4545.4,'p3':4132.2}
CapTechCost = {'p1':5.0,'p2':3.5,'p3':2.5}
```

```
ARU = 1000 #average revenue per unit of demand
AddCostUnit = 1000000 #cost per advertising factor (note that 1 represents no advertising)
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.2)
results = main()
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.0)
results = restart(results[0])
toolbox.register("mutate", mutNetwork, spanpb = 0.0, growthpb = 0.20)
results = restart(results[0])

temp = dict()
temp = write_results(results[2][0],'15n GA_Top_growth_model_scaled_path-imp-mod_mut 1000 - rev 1000 - uniform.csv')

ImpTechCost = {'p1':50000.0,'p2':4545.4,'p3':4132.2}
CapTechCost = {'p1':5.0,'p2':3.5,'p3':2.5}
ARU = 1000 #average revenue per unit of demand
AddCostUnit = 1000000 #cost per advertising factor (note that 1 represents no advertising)
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.2)
results = main()
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.0)
results = restart(results[0])
toolbox.register("mutate", mutNetwork, spanpb = 0.0, growthpb = 0.20)
results = restart(results[0])

temp = dict()
temp = write_results(results[2][0],'15n GA_Top_growth_model_scaled_path-imp-mod_mut 10000 - rev 1000 - uniform.csv')


# In[22]:

# ARU of 100
ImpTechCost = {'p1':500.00,'p2':454.54,'p3':413.22}
CapTechCost = {'p1':5.0,'p2':3.5,'p3':2.5}
ARU = 100 #average revenue per unit of demand
AddCostUnit = 1000000 #cost per advertising factor (note that 1 represents no advertising)
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.2)
results = main()
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.0)
results = restart(results[0])
toolbox.register("mutate", mutNetwork, spanpb = 0.0, growthpb = 0.20)
results = restart(results[0])

temp = dict()
temp = write_results(results[2][0],'15n GA_Top_growth_model_scaled_path-imp-mod_mut 100 - rev 100 - uniform.csv')

ImpTechCost = {'p1':50.000,'p2':45.454,'p3':41.322}
CapTechCost = {'p1':5.0,'p2':3.5,'p3':2.5}
ARU = 100 #average revenue per unit of demand
```

```
AddCostUnit = 1000000 #cost per advertising factor (note that 1 represents no advertising)
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.2)
results = main()
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.0)
results = restart(results[0])
toolbox.register("mutate", mutNetwork, spanpb = 0.0, growthpb = 0.20)
results = restart(results[0])

temp = dict()
temp = write_results(results[2][0],'15n GA_Top_growth_model_scaled_path-imp-mod_mut 10 - rev 100 - uniform.csv')

ImpTechCost = {'p1':5000.0,'p2':4545.4,'p3':4132.2}
CapTechCost = {'p1':5.0,'p2':3.5,'p3':2.5}
ARU = 100 #average revenue per unit of demand
AddCostUnit = 1000000 #cost per advertising factor (note that 1 represents no advertising)
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.2)
results = main()
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.0)
results = restart(results[0])
toolbox.register("mutate", mutNetwork, spanpb = 0.0, growthpb = 0.20)
results = restart(results[0])

temp = dict()
temp = write_results(results[2][0],'15n GA_Top_growth_model_scaled_path-imp-mod_mut 1000 - rev 100 - uniform.csv')

ImpTechCost = {'p1':50000.0,'p2':4545.4,'p3':4132.2}
CapTechCost = {'p1':5.0,'p2':3.5,'p3':2.5}
ARU = 100 #average revenue per unit of demand
AddCostUnit = 1000000 #cost per advertising factor (note that 1 represents no advertising)
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.2)
results = main()
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.0)
results = restart(results[0])
toolbox.register("mutate", mutNetwork, spanpb = 0.0, growthpb = 0.20)
results = restart(results[0])

temp = dict()
temp = write_results(results[2][0],'15n GA_Top_growth_model_scaled_path-imp-mod_mut 10000 - rev 100 - uniform.csv')


# In[22]:

# ARU of 10 000
ImpTechCost = {'p1':500.00,'p2':454.54,'p3':413.22}
CapTechCost = {'p1':5.0,'p2':3.5,'p3':2.5}
ARU = 10000 #average revenue per unit of demand
AddCostUnit = 1000000 #cost per advertising factor (note that 1 represents no advertising)
```

```
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.2)
results = main()
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.0)
results = restart(results[0])
toolbox.register("mutate", mutNetwork, spanpb = 0.0, growthpb = 0.20)
results = restart(results[0])

temp = dict()
temp = write_results(results[2][0],'15n GA_Top_growth_model_scaled_path-imp-mod_mut 100 - rev 10000 - uniform.csv')

ImpTechCost = {'p1':50.000,'p2':45.454,'p3':41.322}
CapTechCost = {'p1':5.0,'p2':3.5,'p3':2.5}
ARU = 10000 #average revenue per unit of demand
AddCostUnit = 1000000 #cost per advertising factor (note that 1 represents no advertising)
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.2)
results = main()
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.0)
results = restart(results[0])
toolbox.register("mutate", mutNetwork, spanpb = 0.0, growthpb = 0.20)
results = restart(results[0])

temp = dict()
temp = write_results(results[2][0],'15n GA_Top_growth_model_scaled_path-imp-mod_mut 10 - rev 10000 - uniform.csv')

ImpTechCost = {'p1':5000.0,'p2':4545.4,'p3':4132.2}
CapTechCost = {'p1':5.0,'p2':3.5,'p3':2.5}
ARU = 10000 #average revenue per unit of demand
AddCostUnit = 1000000 #cost per advertising factor (note that 1 represents no advertising)
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.2)
results = main()
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.0)
results = restart(results[0])
toolbox.register("mutate", mutNetwork, spanpb = 0.0, growthpb = 0.20)
results = restart(results[0])

temp = dict()
temp = write_results(results[2][0],'15n GA_Top_growth_model_scaled_path-imp-mod_mut 1000 - rev 10000 - uniform.csv')

ImpTechCost = {'p1':50000.0,'p2':4545.4,'p3':4132.2}
CapTechCost = {'p1':5.0,'p2':3.5,'p3':2.5}
ARU = 10000 #average revenue per unit of demand
AddCostUnit = 1000000 #cost per advertising factor (note that 1 represents no advertising)
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.2)
results = main()
toolbox.register("mutate", mutNetwork, spanpb = 0.3, growthpb = 0.0)
results = restart(results[0])
toolbox.register("mutate", mutNetwork, spanpb = 0.0, growthpb = 0.20)
```

```
results = restart(results[0])

temp = dict()
temp = write_results(results[2][0],'15n GA_Top_growth_model_scaled_path-imp-mod_mut 10000 -
rev 10000 - uniform.csv')
```