ON THE ECONOMIC VIABILITY OF NETWORK SYSTEMS AND ARCHITECTURES

Soumya Sen

A DISSERTATION

in

Electrical and Systems Engineering

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2011

Supervisor of Dissertation

____

Dr. Roch Guérin, Alfred Fitler Moore Professor of Telecommunications Networks, Electrical & Systems Engineering

Graduate Group Chairperson

____

Dr. Roch Guérin, Professor, Electrical & Systems Engineering

Dissertation Committee Members:
Dr. Roch Guérin, Alfred Fitler Moore Professor of Telecommunications Networks, ESE, University of Pennsylvania
Dr. Kartik Hosanagar, Associate Professor of Operations and Information Management of The Wharton School, University of Pennsylvania
Dr. Saleem Kassam, Solomon and Sylvia Charp Professor of Electrical Engineering, University of Pennsylvania
Dr. Zhi-Li Zhang, Qwest Chair Professor of Computer Science and Engineering, University of Minnesota

ON THE ECONOMIC VIABILITY OF

NETWORK SYSTEMS AND ARCHITECTURES

To my parents,

P. K. Sen and G. Sen

Acknowledgments

I am deeply indebted to my advisor, Dr. Roch Guérin, for his guidance, support, and regular feedback on my research work. His advice, candid assessment, and an eye for details have been immensely helpful in keeping me on track towards completing this thesis. I consider myself very fortunate to have an outstanding and generous researcher like him as my doctoral advisor. I am also very grateful to Dr. Kartik Hosanagar who was equally helpful in guiding me with his insightful comments and detailed feedback. He introduced me to new research topics in network economics and operations research, and throughout the length of our collaboration he provided me with encouragement and support. Their devotion to teaching and mentoring are inspiring and they set an example that I hope to follow.

I owe a special note of gratitude to Dr. Zhi-Li Zhang and Dr. Andrew Odlyzko of University of Minnesota, with whom I had the great pleasure of collaborating on various research topics. I am thankful to them for their comments, suggestions, and professional acquaintance, all of which helped to improve the quality of the work. I also want to acknowledge the help I received from my collaborator, Dr. Youngmi Jin, with whom I had several interesting technical discussions in the early years of this research initiative.

I would like to thank Dr. Saleem Kassam for kindly agreeing to serve as the chair of my dissertation committee. He shared with me his wide range of knowledge and experience for which I am very grateful.

I am deeply indebted to Penn faculty, friends, and my family for their encouragement and support through all the years leading up to this dissertation. The friends I met at the University of Pennsylvania, and particularly those at the Multimedia & Networking Lab, have made these five years fun and memorable. Finally, I would like to dedicate this thesis to my parents, without whose love, support, and sacrifice, it would not have been possible to undertake this work.

ABSTRACT

ON THE ECONOMIC VIABILITY OF

NETWORK SYSTEMS AND ARCHITECTURES

Soumya Sen

Supervisor: Roch Guérin

Understanding the relationship between technology and economics is fundamental to making judicious policy and design decisions. Many technologies that are successful in meeting their technical goals often fail to get adopted due to economic factors. This holds true even for networked systems, *e.g.,* the Internet, which witnessed failures in the adoption of QoS solutions, IPv6 migration etc., due to factors such as high costs, lack of demand, and weight of incumbency. To gain better insights into these issues, researchers need access to analytical frameworks that account for both technological and economic factors and provide useful design guidelines. This dissertation was motivated primarily by the need to undertake such a holistic, multidisciplinary approach towards creating such analytical frameworks.

We focus on three important aspect related to deployment, adoption, and design of network systems and architectures. The Internet has been one of the most successful network technologies, serving both as a shared platform for easy deployment of new services and a driver for their adoption. But recent trends in convergence of voice, video, and data services, along with advances in virtualization technologies, raise ques-

tions as to whether deploying heterogeneous services on a shared network is right or not, especially given the operational complexity and costs involved. We develop a model to investigate the trade-offs between shared and dedicated infrastructures and identify the operational metrics that influence which infrastructure choice benefits more from resource reprovisioning. Closely related to the issue of network service deployment is that of its successful adoption, which serves as the second topic of this dissertation. An entrant's success hinges not only on technical superiority but also on other factors, including its ability to win over an incumbent's installed base by using gateways. Our model for adoption of competing technologies reveals several interesting behaviors, including the possibility for converters to reduce overall market penetration across both technologies and to prevent the convergence of the adoption process to a stable state. Lastly, we consider the issue of network platform design. The emergence and adoption of new technologies depend on the functional capabilities provided by the underlying network platform. Answering whether the minimalist design of Internet is still relevant as it evolves into an ecosystem of software services, require a cost-benefit analysis of choosing between a functionality-rich and a minimalist design. We develop a two-sided market model to show how this design choice crucially depends on the relationship between the cost of adding features to the platform and the benefits that application developers derive from them. The frameworks developed in this dissertation have the potential for application in many different network settings, and can spur further research on various topics in network economics.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Network systems and architectures have a ubiquitous presence in today's world. The Internet, electrical power grids, facilities management networks etc., are just a few examples of such systems that surround us today. The potential for both economic and technological growth offered by these networked systems have attracted the attention of researchers and entrepreneurs, and allowed for convergence of ideas from various fields like computer science, economics, and operations research. The impact of these advances on different network systems are easily observable, none more so than in the case of the Internet. Over the years, the Internet has seen a huge growth both in its user base and the number of innovative services (*e.g.,* P2P, Social networks, IPTV) being offered on it [84]. For the providers of these new services, the Internet served both as a shared platform for easy and inexpensive deployment as well as a driver for their growth.

But in recent times, the convergence of voice, video, and data services has led to questioning whether using a shared platform for services with disparate requirements is recommended or not. This is because although sharing helps to save on infrastructural expenses, deploying multiple services on a shared network comes at the cost of increased complexity in operation, manageability, security, and troubleshooting. With such shortcomings of shared network solutions becoming more evident, many providers are now introducing dedicated platforms with built-in functionalities for their own services. Therefore, a natural question that arises is whether deploying a new service on dedicated infrastructure is better than a shared solution, and to determine when and why this is the case or not. The question has become even more relevant with the advent of new technologies such as virtualization [69, 83], which can further facilitate the deployment of new network "slices" dedicated to an individual new service. Conversely, even in the absence of new technologies recent instances of service deployments point to a complex decision process. For example, in 2006, AT&T introduced its U-Verse infrastructure as a dedicated solution to better control and manage its resources in delivering high quality video for its data and IPTV services [2]. In contrast, one of its competitors, Verizon, chose to share a common fiber optic network [15] for its own voice, video, and data services. These examples point to a rather complex, or perhaps an ad-hoc decision process that is being followed presently due to the lack of a coherent framework to estimate the underlying trade-offs between the network choices.

Such trends are not limited to the Internet. As networking and communication technology continue to improve and new service sectors get network-enabled, *e.g.,* healthcare [46], infrastructure monitoring [10], surveillance, etc., the question of whether to create a shared or a dedicated network becomes important. For instance, the emergence of green buildings results in a facilities management infrastructure that relies upon networked sensors and actuators to monitor and control building operation [73, 35]. This can be realized either by piggy-backing the existing IT infrastructure of a building [10], or by creating a dedicated facilities management network [51], and neither shared nor dedicated network choice emerge as an obvious winner.

The complexity of such decisions is further enhanced by technological factors like virtualization. While virtualization allows better resource sharing on a platform by minimizing interactions among the deployed services, it also allows for easier on-demand reprovisioning of network resources. The impact of the latter ability on the network choice is rather unclear, and as we shall see later in this dissertation, this factor alone can play an important role in the network choice. The rapid advancements made in virtualization technologies are now making network infrastructures akin to services, *e.g.,* IaaS, SaaS. Major corporations, such as IBM, Google, AT&T etc., are using these technologies to invest in the creation of large scale data-centers that provide on-demand computational, storage, and infrastructure resources. To get a better understanding of how the availability of these new technologies influence the way services are deployed and how network infrastructures evolve in the future, we need analytical frameworks and general principles that revise the current ad-hoc practices. The first part of this dissertation (Chapter 2) aims to provide such a framework that will help in understanding these issues and in identifying operational metrics that influence the choice of shared or dedicated infrastructures.

Closely related to the issue of new network service and technology deployment is the question of their successful adoption. Network technologies (*e.g.,* services, platform architectures, protocols *etc.*) often have to compete against formidable incumbents for adoption. An entrant's success hinges not only on technical superiority but also on several other factors. In the past, we have witnessed many Internet technologies, *e.g.,* IP multicast, QoS solutions, IPv6 etc., which were successful in meeting their technical goals but failed to get wide adoption due to factors like high costs, lack of demand, weight of incumbency etc. Understanding how these different factors can potentially impact the adoption of competing technologies is of much interest to the research community. With networking researchers exploring new experimental architectures that can eventually replace the existing Internet architecture and address many of its shortcomings [30, 40, 68], this discussion is very relevant today. The second part of this dissertation (Chapter 3) introduces a modeling framework for the adoption of competing network technologies, and in particular, reveals several interesting insights into the role of converters (or gateways) compatibility on the adoption process.

The emergence and adoption of new technologies and services also depend on the functional capabilities provided by the underlying platforms on which they are deployed. The Internet started out with a minimalist design but has since evolved from a physical infrastructure into a broader ecosystem of software and web services that serve as a platform between two market segments, application developers and consumers. The realization of services like the Amazon web services, Google App Engine, social network platforms etc, bear witness to the progress made in the creation of that ecosystem. A question that therefore arise is whether the minimalist design principle of Internet is still relevant and suitable in today's world of complex interconnected systems and infrastructures. Answering what is the right level of functionalities that a platform should offer calls for evaluating the cost-benefit trade-offs between choosing a functionality-rich and a minimalist design. The third part of the dissertation (Chapter 4) addresses this topic and identifies how various factors influence the choice between functionality-rich versus minimalist design for network platforms.

Each of these topics are introduced in greater details in Section 1.1-1.3. Each of these sections has three parts: positioning the specific research challenge addressed, providing an outline for the model by illustrating the technological and economic factors considered, and presenting a brief outline of the quantitative insights obtained using the proposed frameworks in this dissertation. The results presented here have policy implications as well as significance in the context of design and deployment of network technologies. The frameworks developed as a part of this dissertation have the potential for application in many different network settings, and can spur further research on various topics in the general area of network economics.

1.1 Network Infrastructure Choice

1.1.1 Problem Description

The ubiquity and capabilities of the Internet have led to a rapid growth in networked services and applications. This extends well beyond the migration of voice and video onto the Internet, and has the potential to reach areas either traditionally not networked or accessible only through dedicated networks, *e.g.,* health-care, infrastructure monitoring, surveillance, etc. The introduction of such new technologies and services require the network providers to identify the right architecture for their deployment, that is, whether multiple services should share resources on a common network infrastructure, or should each service be offered on a dedicated network of its own. The benefits of a shared infrastructure notwithstanding, combining services with disparate requirements onto a single common network comes at the cost of increased complexity. It often calls for upgrading the network with features required by the new services. This cost scales with overall network size, *i.e.,* is borne by services with no need for the features. It can also introduce complex interactions and the need for tracking and trouble-shooting problems of previously little consequences. Therefore, assessing the relative benefits of shared and dedicated networks calls for understanding the trade-off between the economies of scale and scope that sharing allows, and the diseconomies of scope [66] it gives rise to.

A model that allows providers to analyze these trade-offs must capture all the different network deployment and operational cost components [21], and how these costs are affected by the needs of the services. Additionally, it also has to account for the fact that the actual demand of a new service is initially uncertain, and so the provider has to allocate capacity (resources) in anticipation of the demand. But networks are not the first to face these issues. There is a long tradition of investigating the trade-offs between flexible and dedicated resources and their allocation decisions in the manufacturing systems literature. For example, the Manufacturing Process Flexibility literature has focused on efficient-plant product assignments [42, 32], the effect of process flexibility in handling demand variability [9], and the optimal resource planning and allocation in presence of demand uncertainty [25, 60]. Although the network provider's problem of choosing between shared and separate networks parallels selecting flexible or dedicated manufacturing plants, and making the right capacity allocations, there are key important differences. First, rather than explore the benefits of a flexible (shared) plant (network) in dealing with uncertain demand, our focus is on investigating the impact of various economies and diseconomies of scope in the cost components. A second and more significant difference is that these traditional manufacturing plant models assume that due to large time-lag in building new plants, production cannot be ramped-up rapidly in response to higher than expected demand, whereas in many networks it is quite feasible to increase the capacity on a relatively short time scale, and hence accommodate a portion of the excess demand. This ability to dynamically adjust the network's capacity through resource '*reprovisioning*' is becoming increasingly easy with technological advancements, such as virtualization [83, 69]. The emergence of virtualization technology has made the question of whether to add a new service on an existing network or on a network 'slice' a more practical one. This capability is also very common in distributed database, cloud computing etc. The earlier manufacturing system models are therefore no longer applicable in this new environment; they need to be extended to incorporate the impact of such dynamic provisioning ability, which can not only affect the optimal resource allocation decisions but also the choice of network infrastructure itself.

1.1.2 Model Description

The most basic setting in which the question of network sharing arise is the case of two network services. The model we develop considers the case where a network provider has an existing service that has already been deployed and runs on its existing network. The second service that the provider wants to introduce is a new service with an uncertain demand. We assume that the provider only knows the demand distribution but not the actual demand that will be realized once the service is made available. The provider has to decide whether to deploy this new service alongside the first service on the existing network, or to create another dedicated network for it. Additionally, the provider also needs to decide how much capacity (resources) has to be allocated for the new service given its choice of network architecture. These decisions have to be made prior to the actual realization of the new service's demand. Once the actual demand is known, the provider can increase its capacity if excess demand is realized. In order to account for different levels of reprovisioning ability, we introduce a *reprovisioning coefficient* that captures the extent to which the provider can recoup excess demand, *i.e.,* it provides a measure for the penalty cost of underprovisioning resources.

The first step of the model involves identifying the various deployment and operational cost components. We group these costs under the categories of fixed costs, variable costs (which grow with the realized demand), and capacity costs (which grows with allocated resources), and use them to formulate the revenue functions for shared and dedicated network infrastructure choice. The model we develop is generic enough to capture various levels of economies or diseconomies of scope in the cost and revenue parameters. We introduce a three stage sequential decision process to solve the provider's decision problems. In the first stage, the provider chooses the network infrastructure. In the second stage, the provider decides on how much capacity to allocate for the new service. The third stage is one in which the new service's actual demand finally gets realized and the provider reprovisions resources, if necessary, to accommodate excess demand. We solve this model by working backwards through the three stages, *i.e.,* given the possible realizations of demand; the corresponding profits are computed in the third stage. Using these demand-profit relationships, the net estimated profit for a given allocated capacity can be computed for each network infrastructure choice. This is done in the second stage, where the 'optimal' capacity that generates the highest estimated profit for each of the two architectural options is calculated. Using these optimal capacity decisions, the provider can compare the resulting profits for the two infrastructures in the first stage, and thus decide to choose the one with a higher profitability.

1.1.3 Summary of Key Results

In this work, we developed an analytical model that addresses the fundamental issue of network architecture selection. It creates a reasoning framework that includes not only factors like the economic relationships among various cost and revenue components, but also incorporates the impact of demand uncertainties and resource reprovisioning into the decision process. We use this model to show that the extent of (dis)economies in various cost or revenue components can impact the network choice. The results illustrate the impact that reprovisioning can have on the choice of network solution, and validates the need for models that incorporate such a feature [77]. Our analysis also identifies two key operational metrics, namely gross profit margin and return on capacity, which influence when, why, and to what extent the ability to reprovisioning can impact the choice of shared versus dedicated network infrastructure [81]. The insights gained from our model have particular significance in the context of software ecosystems and cloud computing environments where virtualization technologies are facilitating shared deployment and faster reprovisioining of resources. It can also be used to explain many trends in network choices of the past and the present, and can also be useful in developing guidelines for providers to choose between shared and dedicated infrastructures in many network settings.

1.2 Adoption of Competing Network Technologies

1.2.1 Problem Description

Networks, like other technologies, constantly improve over time as newer and better solutions become available. These entrant technologies compete against the incumbents they seek to eventually replace. The Internet itself is an example of such a technology that competed against other alternative packet data networks to finally displace the traditional phone network as the de facto communication infrastructure. But the adoption of a new network technology is fraught with challenges [65, 47, 82]. A successful migration from an incumbent technology to an entrant depends not only on technical superiority but also on economic factors [19, 48], and the ability to win over the incumbent's installed base.

The traditional networking approach to facilitate technology migration has been the introduction of converters or gateways. Converters can help a technology to increase its network externality benefits by allowing its users to connect with users of the another technology. However, developing, deploying, and operating converters come at a cost, one that often grows as a function of the converter's quality. Further, converters can play a directionally ambiguous role. On one hand, converters can help the entrant overcome the advantage of the incumbent's large installed base by allowing connectivity to it. But on the other hand, they also help the incumbent technology by mitigating the impact of its users migrating to the newer technology.

To help network technology providers better understand how the deployment of converters and other economic factors influence network technology migration, we develop an economic model which incorporates these factors (*e.g.,* quality, externality, price). We consider the utility derived from network technologies by individual heterogeneous users, and use it to build an aggregate model for technology adoption that is consistent with individual rational decision-making.

Modeling the adoption of new products and technologies has a long tradition in marketing. Fourt and Woodlock [27] proposed a product diffusion model in which a fixed fraction of consumers who have not yet bought the product did so at every period; this is also known as constant hazard rate model. Bass [6] extended it to incorporate word-of-mouth communication between current adopters and potential buyers. These earlier single technology adoption models were later extended to study the joint diffusion of successive generations of technologies by Norton & Bass [63]. However, the focus of all these works is on the aggregate adoption dynamics as opposed to modeling the individual user's decision-making process. As a result, these models fail to develop an understanding of how the consumer decision process affects adoption dynamics and how various economic factors impact adoption decisions. Only a few models have focused on individual-level adoption [11], which provide much greater insights into the mechanism through which rational individual decisions result in aggregate system dynamics. Due the complexity of such models, their use has been limited to the case of a single technology. We have extended these models to a two technology settings -an essential step towards making them suitable for studying migration from an incumbent to an entrant network technology and for investigating the role that converters play in that process. Other prior works that have considered converters in the adoption of incompatible technologies include Farrell & Saloner [24, 23], Katz [47], Choi [14], Joseph *et al.* [43]. The key findings of these works have been that network externalities can often lead to multiple equilibria and that converters can have significant impact on equilibrium adoption levels. However, these works only consider static models, that is, they do not incorporate how heterogeneous user decisions lead to the adoption dynamics, and hence, do not study the dynamics of technology adoption or identify which one of the several possible equilibria gets realized.

1.2.2 Model Description

The process of migration from an incumbent to an entrant network technology is governed by the user's adoption decisions. An individual user joins the network technology that offers a higher 'value' in terms of the technology's quality, externality benefits and price. We account for these factors and their effect on technology adoption through a utility function. For each of the two competing technologies, the utility function increases with the technology's intrinsic (stand-alone) quality and the number of other users reachable using it (externality), while it decreases with price. The utility function accounts for user heterogeneity in their evaluation of a technology's intrinsic quality. The network externality benefits that users enjoy from a technology grows in proportional to the number of users that are using the same technology as well as those who are reachable through gateways or converters that their technology deploys. The model we develop considers 'technology-level' converters (gateways), *i.e.,* these converters, once deployed, are available to all users of the technology. The price (fees) that users pay for subscription to a technology is assumed to be recurrent because of the service nature of most network technologies.

A user adopts a technology when it provides a utility that is both positive (*i.e.,* satisfies individual rationality constraint) and higher than that of the other technology (*i.e.,* satisfies incentive compatibility constraint). The users continuously re-evaluate their technology choices, and can switch from one technology to another. Since changes in the adoption decision of one user affects the externality benefits of other users as well, they all revisit their adoption decision over time, resulting in the dynamics of technology migration. This process is commonly captured through continuous time models, as in [39]. We study the process of technology diffusion using a similar continuous time model whose solution provides us with the characterization of equilibrium outcomes, adoption trajectories, and system stability.

1.2.3 Summary of Key Results

Our analysis reveals a number of interesting behaviors, as reported in our published works [41, 80, 79]. Some of the main findings are described here. Firstly, we show that the adoption process can exhibit multiple steady state equilibrium outcomes; each with a specific range of initial adoption levels of the two technologies. We also find that this behavior may arise both in the presence and absence of converters. Secondly, we find that converters can help a technology improve its own standing in the market, and even ensure its dominance while it would have entirely disappeared in the absence of converters. For example, a low-quality but low-cost technology may thwart the success of a better but more expensive competitor by preserving the ability of its users to access adopters of the costlier technology, whose usage would then be limited to a few 'techno-buffs'. Thirdly, we observe a non-intuitive behavior that improving a converter's efficiency can at times be harmful; they can result in lower market share for an individual technology or for both of them. For instance, high market penetration may depend on the combination of a cheap but low-end technology with a high-end but more expensive one to adequately serve the full spectrum of user preferences. A situation where converters allow the better technology to gain market share at the expense of the lesser technology may result in low-end users of that technology dropping out altogether; thereby contributing to a lower overall market penetration. Fourthly, we show that while in the absence of converters, technology migration always converges to stable steady-state equilibrium; this need not be so when technologies deploy converters for compatibility. The presence of converters can create ‘boom-and-bust’ cycles in which users switch back-and-forth between the two technologies.

The identification of these behaviors allow network providers to realize the potential impact of various economic and technical factors, and thus help devise competition strategies for network migration. Additionally, the knowledge of possible adverse impact of converters that our model reveals is also useful in deciding on policy interventions by regulators.

We have also verified using numerical evaluations that the results obtained from the model are robust to inclusion of switching and learning costs as well as a broad range of variations in the structure of the user utility functions [79].

1.3 Network Platform Design

1.3.1 Problem Description

The transformation of the Internet from a physical network to an ecosystem of software and services is leading many to question its original minimalist design. Answering whether or not this design remains appropriate is, however, a complex issue. Although we do not have a definitive answer to this multi-faceted question, our work presented here will offer an initial quantitative step towards exploring it. We formulate the question using a two-sided market model [74]. The network is the platform and seeks to “connect” users to services. The Internet and Android offer two recent examples of (network and operating system) platforms whose success largely comes from their ability to connect users and service/application developers.

Services are offered by developers that rely on the platform and its features. A feature-rich platform facilitates service development, which yields more services. This in turn attracts more users and benefits the platform. It, however, comes at a greater implementation cost for the platform. The focus of this work is to explore the decision problem faced by a monopolist platform provider seeking to select the level of functionality the platform should offer.

1.3.2 Model Description

A platform provider attracts developers and consumers by creating value that entices them to join the platform. This 'value' depends on a number of factors, such as the subscription fees to join it, the cost of developing applications for it, and *externalities* that affect the value that either developers or customers derive from joining the platform. In a two-sided market, both sides of the market derives cross-externality benefits from the presence of the other, *i.e.,* consumers benefit from more applications offered by developers, and conversely developers benefit from being able to target their applications to more consumers. Our model measures this value for the two market sides through utility functions that incorporate the different factors discussed above, in addition to the heterogeneity among the participants on each market side. Similarly, the impact of the decisions that the platform provider makes, *i.e.,* pricing and selection of the platform's functionality, are also reflected through the platform provider's utility function.

Two important assumptions that our model makes are: (i) applications all make use of the *same set of platform features*, and (ii) the functionality embedded in these features can be built by *either the platform or the developers,* possibly with different costs but no difference in quality. Assumption (i) implies homogeneous development needs across applications (services). In other words, they rely on the same platform application programming interfaces (APIs) or independent features created by developers. They can still be differentiated, but this clearly limits the range of their differences. Assumption (ii) calls for the platform provider to know application development needs ahead of time, and for application developers to be able to independently develop features that the platform decides not to incorporate. This is reasonable for many software products and services, where platform and applications share a common technology. Implicit in assumption (ii) is that the development quality (and cost) of a feature by either the platform or the developers, is fixed and not a decision variable. In general, these assumptions limit the model's applicability to platforms that are software ecosystems, *e.g.,* cloud computing, web services, OSes, etc.

The model is solved using a three stage sequential decision process for the platform to select the level of functionality to offer. In the first stage, the platform provider chooses the number of features to build into the platform. Given this choice for number of feature, participation prices (fees) for the two market sides are chosen in the second stage. Equilibrium adoption levels of consumers and developers are simultaneously realized in the third stage. This sequential decision process is then solved in the reverse order. Equilibrium adoption levels for users and developers are first computed for a given choice of participation prices and number of built-in features. Next, given a choice for the number of built-in features, 'optimal' participation prices are computed based on the equilibrium adoption levels of the previous step. The results characterize the platform's profit for any given number of built-in features. This is then used to find the 'optimal' number of features that maximizes the platform's profit.

1.3.3 Summary of Key Results

The analytical model we develop can be used to explore the trade-offs between functionality-rich and minimalist design from a monopolist platform provider's perspective. The solution reveals that the answer is strongly dependent on how additional features affect the rate of change of the respective costs of the platform and service developers. It also offers a somewhat "negative" result in that minor changes in these costs can produce drastically different choices for the platform. We provided illustrative examples to show that there can be multiple values for the number of platform features that maximize the platform's profit, and this can make the platform provider's decision far from obvious. In other words, minimalist or functionality-rich platforms can both arise across a wide range of scenarios.

1.4 Research Publications

The research work presented here has been published in several leading journals and conferences. An initial version of our work on the choice of shared versus dedicated networks, titled "Shared versus Separate Networks: The Impact of Re-provisioning" (coauthored with Roch Guerin and Kartik Hosanagar) [77], was first published in the proceedings of ReArch'09, CoNEXT, in Rome, Italy, on December 1, 2009. A more matured version of the work was published as "The Impact of Re-provisioning on the Choice of Network Infrastructures" (coauthored with Kristin Yamauchi, Roch Guerin and Kartik Hosanagar) [81] in the proceedings of the Ninth Workshop on E-Business (WEB'10), held in St. Louis, Missouri, USA, on December 11, 2010. A journal version of the work is currently under submission in one of the leading information systems journal.

The work on network technology adoption was first published as "Dynamics of Competition between Incumbent and Emerging Network Technologies" (coauthored with Youngmi Jin, Roch Guerin, Kartik Hosanagar, and Zhi-Li Zhang) [41] in the proceedings of NetEcon, SIGCOMM'08, held in Seattle, USA, between August 17-22, 2008. The work was further extended to consider the impact of converters, and was published as a journal version, titled "Modeling the Dynamics of Network Technology Adoption and the Role of Converters" (coauthored with Youngmi Jin, Roch Guerin, and Kartik Hosanagar) [80] in IEEE/ACM Transactions on Networking, vol. 18(6), 2010, along with a detailed technical report in 2009 [79], available at the University of

Pennsylvania repository.

Our results on the last topic is available in the paper titled, “Functionality-rich versus Minimalist Platforms: A Two-sided Market Analysis” (coauthored with Roch Guerin and Kartik Hosanagar) [78] in 2011, is under submission at a networking conference and plans for a journal version of the work is under consideration.

Chapter 2

Network Infrastructure Choice:

Shared Versus Dedicated Networks

2.1 Introduction

Advances in network technologies have resulted in the Internet evolving from a simple data network to a global communication infrastructure that carries a multiplicity of services. This integration has many obvious advantages, but combining services with disparate requirements onto a shared network can also have a cost. It often calls for the entire network to be "upgraded" with features required by only a handful of services, and at a cost that is borne by all of them. Resource sharing can also introduce complex interactions among services and call for tracking and trouble-shooting problems of previously little consequence, *e.g.,* minor routing instabilities don't affect most data services but can severely degrade voice or video quality.

Hence, while sharing a network across many services is often advantageous, it need not always be, and it is of interest to determine when and why this is the case or not. The question has become even more relevant with the advent of new technologies such as virtualization [69, 83], which can further facilitate the deployment of new network "slices" dedicated to an individual new service. Conversely, even in the absence of new technologies recent instances of service deployments point to a complex decision process. For example, in deploying its new U-verse TV service AT&T chose to create a dedicated network. This was in part to ensure it could be managed more easily for better reliability and for delivering higher quality video [2]. In contrast, one of its competitors, Verizon, chose to share a common fiber optic network [15] for its own voice, video, and data services. As networking and communication technology continues to improve and more services are network-enabled, *e.g.,* health-care, infrastructure monitoring, surveillance, etc., the question of whether to offer this access over shared or dedicated networks will become even more important.

For instance, the emergence of green buildings results in a facilities management infrastructure that relies upon networked sensors and actuators to monitor and control building operation. This can be realized either by piggy-backing the existing IT infrastructure of a building [10], or by creating a dedicated facilities management network [51]. In particular, [10] provides an example of greener schools in New York that are using a shared IT and facilities management infrastructure to reduce their peak electrical usage by enabling real-time monitoring over the web. In contrast, it also cites a second example, that of a sportswear retailer, which created a dedicated ethernet backbone for its facilities management traffic because of concerns over costs, throughput and security. Thus, neither shared nor dedicated network choice emerge as an obvious winner.

Making these decisions calls for a framework that systematically examines the trade-off between shared and dedicated network infrastructures. Although the issue has received some attention, particularly from the business press, it has seen little formal analysis. Developing a framework to evaluate the underlying trade-off is the primary motivation for this work.

In this chapter, we propose a model for offering two network services, an existing service with a known demand and a new one with uncertain demand that can either be deployed on the same network as the existing service or on its own dedicated network. The model allows for economies or diseconomies of scope in network resources, and also accounts for the ability to adjust network resources (reprovision) in response to a higher than anticipated demand for the new service. The main contribution of this study is in offering a framework for service providers to evaluate network infrastructure options, and in particular to decide whether it is profitable to deploy a new service on an existing network infrastructure. The model also establishes that the extent to which reprovisioning is feasible can by itself affect which infrastructure, shared or dedicated, is more effective. In particular, two operational metrics, the gross profit margin and the return on capacity play a major role in determining which infrastructure benefits more from reprovisioning.

The rest of the chapter is structured as follows. Section 2.2 reviews prior work from the manufacturing flexibility literature, and highlights its relevance to our research question. Section 2.3 introduces the model and its parameters. Section 2.4 presents the analysis. Section 2.6 summarizes the work's findings and concludes the study.

2.2 Literature Review

Recall that investigating whether to use shared or dedicated infrastructures involves two main aspects, provisioning and reprovisioning capacity in response to demand uncertainty and assessing the economies and diseconomies of scope that arise from sharing resources across services. Both of those issues have been the subject of a number of past investigations, and we now briefly review the most relevant ones.

Capacity Planning under Uncertain Demand: The key question in our work is to choose between shared and dedicated infrastructures, while accounting for demand uncertainty and economies or diseconomies of scope in various costs. For either infrastructure choice, we need to find the optimal capacity allocation and compare the expected profits. The question of capacity sizing for a given infrastructure choice in our setting is analogous to that of the classical news-vendor problem. A number of papers in Operations Management have studied capacity sizing under stochastic demand. The most relevant papers are those on the news-vendor problem [50, 54]. The classical single-period single-product news-vendor problem is to select an inventory/order level for a product under uncertain demand so as to maximize the expected profit in a single period. Both over-provisioning and under-provisioning have associated costs and the inventory level cannot be readjusted if demand exceeds capacity. A rich literature has extended the study of the classical news-vendor to allow for multiple periods [70, 52], multiple products [88, 1, 55, 22], and multi-product multi-period decision problems [61]. The capacity sizing decision in our model is similar to these classical news-vendor models, but with an additional feature that the service provider can reprovision resources to accommodate some excess demand. Although this feature of resource reprovisioning is present in a some works like [5, 31], their main focus is on accounting, product costing, and pricing. In contrast, our central question relates to deciding between shared and dedicated infrastructures under optimal capacity allocation. This also differentiates our model from papers on studied multi-product news-vendors in that they focus on finding the optimal production quantities of each product under capacity or budget constraints and do not delve into the tradeoffs associated with servicing the demands for the two products on dedicated versus shared infrastructure.

Manufacturing Flexibility: The manufacturing flexibility literature investigates the trade-off between using flexible resources to manufacture multiple products versus using dedicated resources for each product. Flexible plants capable of producing different types of products are more expensive to build, but have benefits in dealing with uncertain demand. There is, therefore, a trade-off that needs to be investigated to determine how much capacity to build into flexible and dedicated plants. In all these models, investment decisions in manufacturing plants have to be made before the actual demand for products are realized. Fine and Freund [25] develop a *two-stage model* to analyze this trade-off. Plant capacity decisions are made in the first stage, when demand is still uncertain. Production decisions are implemented in the second stage after demand is realized. The authors set up an optimization problem to establish the firm's optimal investments in flexible and/or dedicated resources and the optimal production levels using these resources. A similar setting is considered by Van Mieghem [60], with an emphasis on the role of price margin and cost mix differentials. The author shows that an investment in flexible resources can be beneficial even with perfectly positively correlated product demands because a flexible plant can shift production towards the product with a higher profit margin.

Our decision problem also shares basic properties with these works. Choosing between shared and dedicated networks parallels selecting flexible or dedicated manufacturing plants, as does the need to decide how to provision the network in the face of demand uncertainty. There are, however, several differences between our setup and these earlier works. Unlike manufacturing plants where production usually cannot be rapidly ramped-up in response to higher than expected demand, "upgrading" network capacity on a relatively short time-scale is becoming increasingly feasible[1]. As a result,

[1]As mentioned in the previous section, the advent of virtualization technology will contribute further

even if some excess demand is ultimately lost during reprovisioning, network services can recover from suboptimal upfront capacity decisions. We show that this affects not only the optimal capacity levels chosen by providers, but can also impact the decision to go with a shared or dedicated infrastructure. In addition, the manufacturing flexibility literature focuses on the benefits from pooling uncertain demand for two or more products but does not consider the impact of economies and diseconomies of scope in the underlying cost parameters, which is a key aspect of our investigation.

The above discussion reveals two themes. The literature on news-vendor problems investigates how to size capacity when demand for a product is uncertain. Recent extensions consider multi-product problems but do not delve into the benefits of infrastructure sharing between these products. The manufacturing flexibility literature considers the benefits of resource sharing but focuses primarily on manufacturing settings in which reprovisioning of capacity is often too slow or completely infeasible. This is a major limitation when considering network services because virtualization and other technologies facilitate the reprovisioning of network resources in short order. Our study builds on these streams of work to study the deployment of network services in shared versus dedicated infrastructure, while allowing for the reprovisioning of network resources in response to realized demand.

to this ability.

2.3 Model Formulation

We consider the most basic setting in which to explore whether to share a network across services, or instead deploy them on separate dedicated networks[2]. Specifically, one service has already been deployed and has a stable demand, and the service provider is introducing a second one. There is uncertainty in the demand for the second service, and possible economies or diseconomies of scope when adding it to the same network as the existing service. Our goal is to develop a simple model that accounts for these economies and diseconomies of scope in determining the optimal infrastructure choice. For analytical tractability, we ignore any economies of scale that may arise when combining services on a shared infrastructure. The magnitude of such economies of scale are typically limited in networks, *e.g.,* [53](ref. Fig. 5), [71] show almost linear growth of both access and backbone bandwidth and router costs, and furthermore Section 2.5 shows that they do not qualitatively affect the model's outcome.

The provider's objective is to maximize its total profit from the two services. This decision problem can be modeled as a three stage sequential process, as shown in Figure 2.1.

In the first stage, the provider makes an infrastructure choice, namely a shared or a dedicated network. At this stage, the provider does not know the profit from Service 2 since its demand is uncertain. Given an infrastructure choice, in the second stage the provider provisions capacity for the yet unknown demand for Service 2. Demand for

[2] We use the words network and infrastructure interchangeably throughout this chapter.

Service 2 is realized in the third stage, where the provider now has the opportunity to reprovision the network if this demand exceeds the capacity provisioned upfront. A penalty for under-provisioning is incurred, and only a fraction of the excess demand can be captured through reprovisioning. Conversely, when the realized demand is lower than the existing capacity, the provider takes no further action[3]. The three stages of the decision process are referred to as *Infrastructure Decision Stage*, *Capacity Allocation Stage* and *Reprovisioning Stage*, respectively.

The above sequential decision problem is solved in the reverse order. We first solve for the provider's decision in the Reprovisioning Stage, *i.e.,* we evaluate whether the provider must reprovision resources after demand is realized, conditional on both the capacity provisioned upfront and the infrastructure choice. Next, we evaluate the provider's expected profit as a function of its capacity sizing decision in stage 2 when demand is uncertain. This is used to compute the optimal capacity to be provisioned upfront. Based on these results for Capacity Allocation Stage and Reprovisioning Stage, we finally evaluate the provider's total expected profit for each infrastructure choice, and select the one that yields the higher expected profit. These three steps are discussed in greater details in Section 2.3.2 after introducing the model parameters.

[3]Contractual obligations are assumed to preclude downward adjustment of resources.

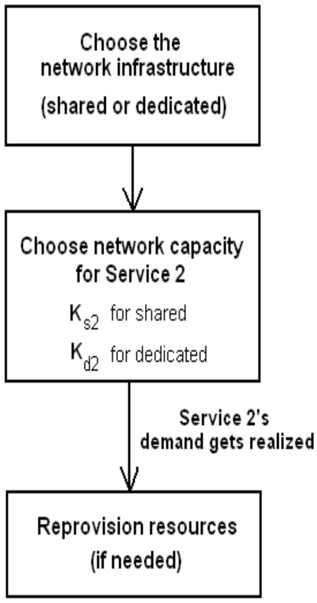

Figure 2.1: The three-stage sequential decision process

2.3.1 Model Parameters

Given that Service 1 is an existing service with a mature demand, we assume for simplicity that it operates at full capacity, *i.e.,* its provisioned capacity matches its realized demand, X_1. The new service, Service 2, has uncertainty in its demand that is denoted by a random variable x_2 with known distribution, f_{x_2}. We use the notation X_2 to indicate a realization of this demand. The provisioned capacity (number of users the network can handle) for Service 2 is a decision variable denoted by K_{s2} and K_{d2} for shared and dedicated networks, respectively[4]. If the demand for Service 2 exceeds the provisioned level ($X_2 > K_i,\ i = \{s2, d2\}$), network resources can be adjusted to accommodate a fraction α of the excess demand, *i.e.,* resources are increased to $K_i + \alpha(X_2 - K_i)$. The

[4] A shared network is, therefore, provisioned to handle $X_1 + K_{s2}$ users.

parameter α or *reprovisioning coefficient*, represents the fraction of excess demand that reprovisioning can capture. This fraction is assumed to be independent of the magnitude of the reprovisioning effort. In other words, the feasibility of and latency in securing additional capacity are the same regardless of the amount of capacity requested (at least within some bounds), with the latter possibly affecting the network's ability to retain all the excess demand that was present. This is consistent with the provisioning process of most computing and communications facilities. When $\alpha = 0$, reprovisioning is unable, *e.g.,* too slow, to capture any excess demand, while $\alpha = 1$ corresponds to a scenario where reprovisioning succeeds in accommodating the entire excess demand. In other words, when $\alpha = 1$, a "provisioning phase," is unnecessary as resources can be secured on-the-fly. Different levels of provisioning flexibility, *e.g.,* as afforded by different types of virtualization technology, can be accounted for by varying α. Of interest, as discussed in Section 2.4, is the fact that changing α can also change the outcome of the provider's decision process, *i.e.,* which infrastructure yields the highest profit. Note that α is not a decision variable for the provider; it is an exogenous system parameter whose value depends on the reprovisioning technology available to the service provider.

An additional point which bears mention is that a linear adjustment to α can transform our model to an alternative one in which a service provider can always accommodate the entire excess demand, but pays a higher per unit capacity cost in doing so. This equivalence between the two models is shown in Appendix A.3.

Next, we describe and contrast revenue and cost components of shared and dedicated networks. To facilitate comparisons, we follow the standard practice, *e.g.,* [25, 60], and consider only the present value of all future revenues and costs.

Services generate revenues from subscription fees paid by users. These fees are assumed set based on exogeneous market factors. Offering a service also incurs a per user connection cost, *e.g.,* cost of enabling last-mile connectivity, installing end-user access equipments, operational costs of billing, etc. We denote by p_{s1} and p_{s2} the per user *contribution margins*- price less the variable costs- for Services 1 and 2 respectively in a shared network. Similarly, p_{d1} and p_{d2} denote contribution margins for the two services in dedicated networks. We note that p_{s1} and p_{d1} can differ from each other. For example, support for voice service in a FiOS network[5] (a shared network used to carry voice, data and video) calls for network termination equipment that is significantly more complex than that used in a traditional voice network, *e.g.,* the FiOS equipment needs to come with a battery pack to handle power outages. This then translates into $p_{s1} > p_{d1}$. We also note that implicit in the definition of individual contribution margins for each service is the assumption that they are incurred independently for each service, *i.e.,* there are no economies of scope associated with users subscribing to *both* services. This assumes that there is no bundling discount for subscribing to both services, and that per user connection costs are additive across services. This is reasonable in settings such as the FiOS example, where the bulk of the connection costs are in termination equipment specific to each service.

[5] See http://en.wikipedia.org/wiki/Verizon_FiOS for an informal description.

In addition to per-user connection costs, offering network services also involves *fixed* and *capacity* costs. Upfront fixed costs are independent of demand and capacity levels, *e.g.,* they include facility rent, research & development expenses. These costs are denoted by c_s for a shared network, and by c_{d1} and c_{d2} when each service is deployed on a dedicated network. Capacity costs grow with network resources; they are incurred upfront because of provisioning and may also be incurred subsequently during reprovisioning. Unit capacity costs for Service 1 are denoted by a_{s1} and a_{d1} in shared and dedicated networks, respectively, and by a_{s2} and a_{d2} for Service 2. We use the term *return on capacity* to refer to the ratio of contribution margin to capacity cost, $\frac{p_i}{a_i}$, $i = \{s2, d2\}$ for Service 2, with a parallel definition for Service 1.

The values that the above parameters take in shared and dedicated networks are obviously related to each other. These relationships can exhibit different levels of economies and diseconomies of scope. We illustrate this through the example of overlay and integrated networks, which represent two possible options for realizing a shared network.

An overlay involves limited use of an existing infrastructure to deploy a new network service. For example, early versions of the Internet were deployed as an overlay on the existing phone network. End-systems connected using modems to transmit data over existing phone lines, and early routers were interconnected using available telephony transmission facilities such as T1 and T3 links. Control functions of the nascent Internet were, however, kept separate from those of the phone network, *e.g.,* the Internet relied on its own routing protocols and did not use the phone network signalling system (SS7).

In general, when a new service is deployed by way of an overlay, the networks of the two services share a common infrastructure (that of Service 1), but remain largely decoupled from each other. This limits the diseconomies of scope that could arise from complex interactions between them, but it also precludes significant economies of scope.

In contrast, an integrated network solution will operate both services on a truly common network infrastructure. For example, many cable providers with "triple-play" offerings, have upgraded their infrastructure (backbone network and cable access network) so that it can carry the voice, data, and video traffic from those three services. This required upgrading backbone and access routers to allow differentiation (and prioritization) of different traffic types, but allowed reuse of the same router platforms and transmission facilities for all three services. In other words, an integrated network solution offers opportunities for greater economies of scope, but often mandates more expensive equipment to handle the individual requirements of each service. This in turn can translate into higher diseconomies of scope. The model of Section 2.3.2 can be configured to reflect any combination of economies and diseconomies of scope between shared and dedicated networks.

2.3.2 Model Setup and Solution

We describe next solving the Three Stage model of Figure 2.1 to obtain the expected profits associated with shared and dedicated networks. As alluded to earlier, the solution proceeds in the reverse order of the decision process of Figure 2.1, *i.e.,* the reprovisioning stage is first solved, followed by capacity allocation stage, and finally the infrastructure choice stage. Because the solution method is essentially identical for dedicated and shared networks, we present it for the former and simply provide final expressions for the latter. Furthermore, we note that because Service 1 is a mature service with a stable, known demand, it is only relevant in the final infrastructure choice stage.

2.3.2.1 Reprovisioning Stage

As mentioned earlier, reprovisioning takes place after the demand for Service 2 has been realized. In the presence of excess demand, *i.e.,* the realized demand exceeded the originally provisioned capacity, the provider secures additional capacity to capture a fraction α of the excess demand. In the absence of excess demand, no reprovisioning takes place.

We present next an expression for the gross profit from Service 2 after the reprovisioning phase in a dedicated network. As defined in Sub-section 2.3.1, the contribution margin for Service 2 is p_{d2}, and the variable cost of provisioning capacity is a_{d2}. If the realized demand X_2 exceeds the provisioned capacity K_{d2}, the capacity is adjusted to accommodate a fraction α of the excess demand, *i.e.,* capacity increases to $K_{d2} + \alpha(X_2 - K_{d2})$. The gross profit for Service 2 is then given by

$$R_{d2}(X_2 > K_{d2}) = (p_{d2} - a_{d2})(K_{d2} + \alpha(X_2 - K_{d2})) \tag{2.1}$$

Conversely, when the realized demand is less than the provisioned capacity K_{d2}, the gross profit for Service 2 is

$$R_{d2}(X_2 \leq K_{d2}) = p_{d2}X_2 - a_{d2}K_{d2} \tag{2.2}$$

Similar expressions can be obtained in the case of a shared network, and Equations (2.3) and (2.4) parallel the expressions of Equations (2.1) and (2.2).

$$R_{s2}(X_2 > K_{s2}) = (p_{s2} - a_{s2})(K_{s2} + \alpha(X_2 - K_{s2})) \tag{2.3}$$

$$R_{s2}(X_2 \leq K_{s2}) = p_{s2}X_2 - a_{s2}K_{s2} \tag{2.4}$$

Next, we use these expressions to compute the optimal upfront capacity for Service 2 in the capacity allocation stage.

2.3.2.2 Capacity Allocation Stage

Assuming a known distribution f_{x_2} for the demand of Service 2, the expected gross profit R_{d2} *given* the capacity provisioned upfront K_{d2} in a dedicated network can be expressed as

$$\mathbf{E}(R_{d2})_{[K_{d2}]} = \int_0^{K_{d2}} R_{d2}(X_2 \leq K_{d2}) f'_{x_2} d(x_2) + \int_{K_{d2}}^{X_2^{\max}} R_{d2}(X_2 > K_{d2}) f'_{x_2} d(x_2), \tag{2.5}$$

where $R_{d2}(X_2 > K_{d2})$ and $R_{d2}(X_2 \leq K_{d2})$ are given in Equations (2.1) and (2.2).

For analytical tractability, we assume that f_{x_2} is uniformly distributed[6] in $[0, X_2^{\max}]$.

[6]Results typically extend [8, 28] to other distributions that share with the uniform distribution the important property of a non-decreasing hazard-rate function $F'(\theta)/(1 - F(\theta))$.

Assuming a uniform distribution for X_2, Equation (2.5) becomes

$$E(R_{d2})_{[K_{d2}]} = \frac{1}{X_2^{max}} \left[\left(\frac{(\alpha-1)}{2} p_{d2} - \frac{\alpha}{2} a_{d2}\right) {K_{d2}^{*}}^2 + (1-\alpha)(p_{d2}-a_{d2}) X_2^{max} K_{d2}^{*} + \left(\frac{\alpha(p_{d2}-a_{d2})}{2} (X_2^{max})^2 - c_{d2} X_2^{max}\right) \right] \quad (2.6)$$

Using Equation (2.6), we can compute the optimal capacity K_{d2}^{*} such that $\frac{\partial \mathbf{E}(R_{d2})_{[K_{d2}]}}{\partial K_{d2}} = 0$:

$$K_{d2}^{*} = \left(\frac{(1-\alpha)(p_{d2}-a_{d2})}{(1-\alpha)p_{d2} + \alpha a_{d2}} \right) X_2^{\max} \quad (2.7)$$

The expression for K_{d2}^{*} from Equation (2.7) can be used to show that

$$a_{d2} P(X_2 \leq K_{d2}^{*}) = (1-\alpha)(p_{d2}-a_{d2}) P(X_2 > K_{d2}^{*}) \quad (2.8)$$

Equation (2.8) shows that the optimal capacity K_{d2}^{*} is one at which the cost incurred from each unit of over-provisioning (*i.e.,* a_i for $P(X_2 \leq K_{d2}^{*})$) is balanced against the loss from each unit of under-provisioning (*i.e.,* $(1-\alpha)(p_i - a_i)$ for $P(X_2 > K_{d2}^{*})$).

Also notice that the optimal capacity in Equation (2.7) is analogous to the notion of 'critical fractile'[7] in single period news-vendor problems. As expected, Equation (2.7) yields $K_{d2}^{*} = 0$ when $\alpha = 1$, *i.e.,* the ability to reprovision without penalty obviates the need for provisioning upfront.

[7] Single-period news-vendor problems with selling price p and inventory purchase price c use the notion of 'critical fractile', given by the ratio $(p-c)/p$, to determine the optimal inventory level q^{*}, where $q* = F^{-1}\left(\frac{p-c}{c}\right)$ and F^{-1} is the inverse cumulative distribution function of demand. Intuitively, the critical fractile balances the cost of being understocked (a lost sale worth $(p-c)$) and the total costs of being either overstocked or understocked (where the cost of being overstocked is the inventory cost, c, thus giving a total cost of simply $p = (p-c)+c$).

Substituting the expression for K^*_{d2} from Equation (2.7) in Equation (2.6), we get

$$\mathbf{E}(R_{d2})_{[K^*_{d2}]} = \frac{(p_{d2}-a_{d2})X_2^{\max}}{2}\left(1-\frac{(1-\alpha)a_{d2}}{(1-\alpha)p_{d2}+\alpha a_{d2}}\right). \tag{2.9}$$

Similar expressions can be obtained for a shared network as shown below

$$K^*_{s2} = \left(\frac{(1-\alpha)(p_{s2}-a_{s2})}{(1-\alpha)p_{s2}+\alpha a_{s2}}\right)X_2^{\max}. \tag{2.10}$$

$$\mathbf{E}(R_{s2})_{[K^*_{s2}]} = \frac{(p_{s2}-a_{s2})X_2^{\max}}{2}\left(1-\frac{(1-\alpha)a_{s2}}{(1-\alpha)p_{s2}+\alpha a_{s2}}\right). \tag{2.11}$$

Next we proceed to use the results of Equations (2.9) and (2.11) to compute profits from shared and dedicated networks and finalize a choice of infrastructure.

2.3.2.3 Infrastructure Choice Stage

In this last stage, the overall profit of the two network options, shared or dedicated, are evaluated to select the one with the higher profit. We consider dedicated and shared networks in turn.

Dedicated Networks

The gross profit for Service 1 in a dedicated network is of the form $R_{d1} = X_1(p_{d1} - a_{d1})$, where p_{d1} and a_{d1} are as defined earlier. As a result, the profit Π_{d1} of Service 1 deployed on a dedicated network is of the form

$$\Pi_{d1} = X_1\left(p_{d1}-a_{d1}\right)-c_{d1} \tag{2.12}$$

The expected profit under optimal provisioning for Service 2 is given by subtracting the fixed cost from Equation (2.9):

$$\Pi_{d2} = \frac{(p_{d2} - a_{d2})X_2^{\max}}{2}\left(1 - \frac{(1-\alpha)a_{d2}}{(1-\alpha)p_{d2} + \alpha a_{d2}}\right) - c_{d2} \tag{2.13}$$

The total profit from both services, $\Pi_d = \Pi_{d1} + \Pi_{d2}$, is therefore

$$\Pi_d = \frac{(p_{d2} - a_{d2})X_2^{\max}}{2}\left(1 - \frac{(1-\alpha)a_{d2}}{(1-\alpha)p_{d2} + \alpha a_{d2}}\right) - c_{d2} + X_1\left(p_{d1} - a_{d1}\right) - c_{d1} \tag{2.14}$$

Shared Networks

The gross profit R_{s1} for Service 1 is of the form

$$R_{s1} = X_1\left(p_{s1} - a_{s1}\right), \tag{2.15}$$

while the gross profit for Service 2 is given by Equation (2.11). Hence, the total profit from both services is given by

$$\Pi_s = \frac{(p_{s2} - a_{s2})X_2^{\max}}{2}\left(1 - \frac{(1-\alpha)a_{s2}}{(1-\alpha)p_{s2} + \alpha a_{s2}}\right) + X_1\left(p_{s1} - a_{s1}\right) - c_s \tag{2.16}$$

The optimal network infrastructure choice is the one yielding the highest overall profit. In the next section, we explore how this choice is affected by the model parameters. Before proceeding, we first derive a number of basic properties that are used later in the analysis (the proofs of the corresponding lemmas are available in Appendix A.1.

Lemma 2.3.1 *The network infrastructure with a higher return on capacity has a higher optimal capacity,* i.e., *if* $\frac{p_{d2}}{a_{d2}} > \frac{p_{s2}}{a_{s2}}$, *then* $K_{d2}^* \geq K_{s2}^*$, *and vice-versa*[8].

[8]The equality $K_{d2}^* = K_{s2}^*$ can hold only at $\alpha = 1$, where both K_{d2}^* and K_{s2}^* are zero.

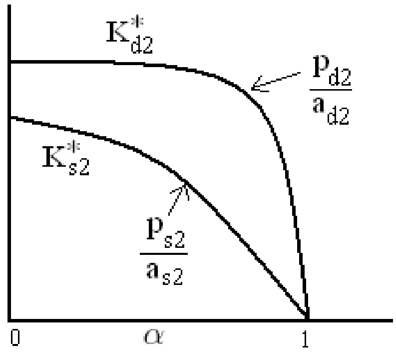

Figure 2.2: Network Resource Allocation for $p_{d2}/a_{d2} > p_{s2}/a_{s2}$

For a given demand distribution, investing in greater capacity increases the likelihood of having excess capacity. Thus the network with greater capacity has greater expected cost of unused capacity. But if the network also has higher return on capacity then that is offset by the greater potential upside, namely higher returns if the capacity is utilized. Thus, the network with the greater return on capacity is the one that can invest in additional capacity.

Lemma 2.3.2 *For both shared and dedicated infrastructures, the optimal capacity decreases with* α*.*

An increase in alpha allows a provider to recover more of the excess demand and therefore reduces the provider's cost of under-provisioning resources. With the cost of under-provisioning going down while the cost of over-provisioning remaining the same, the provider reduces the capacity it provisions upfront. In particular, when $\alpha = 1$, *i.e.,* the entire excess demand is captured, no provisioning is needed *i.e.,* $K^*_{d2} = K^*_{s2} = 0$.

Lemma 2.3.3 *When the return on capacity* (i.e., $\frac{p_i}{a_i}$, $i = \{d2, s2\}$) *is large, the optimal capacity decreases slowly with* α *when* α *is small and decreases rapidly with* α *as* $\alpha \to 1$.

Lemmas 2.3.1 to 2.3.3 are illustrated in Figure 2.2 (the ratio p_{d2}/a_{d2} is assumed large in this figure). In the figure, the optimal capacity provisioned upfront for the dedicated network is always greater than or equal to that for the shared network on account of its greater return on capacity. The optimal capacities for both network choices are decreasing with α. And finally, the decrease is relatively small at low α for the dedicated network and very rapidly as $\alpha \to 1$.

2.4 Analysis

In this section, we use the analytical results obtained in Section 2.3 to study the impact of various system parameters on the choice of network infrastructure. This will be done in two phases. First, in Section 2.4.1, we consider the impact of the different cost and revenue parameters (c_{d1}, c_{d2}, c_s, a_i, p_i; $i = \{d1, s1, d2, s2\}$). All these parameters exhibit similar behavior, *i.e.,* economies of scope in costs favor the creation of shared networks while diseconomies of scope favor dedicated networks. Second, in Section 2.4.2, we focus on the impact of the reprovisioning coefficient (α) on the network choice, and show that it can produce more subtle and interesting behaviors.

2.4.1 Impact of Cost/Revenue parameters

The preferred network infrastructure is found by comparing Π_s and Π_d (Equations (2.14) and (2.16)) of the shared and dedicated networks respectively, and choosing the one that yields a higher profit. These profits depend on the various cost and revenue parameters, all of which have a similar kind of impact on the network choice. We provide an illustrative example by evaluating the impact of p_{s2} (*i.e.,* Service 2's contribution margin in a shared network). Detailed analyses for other cost parameters are provided in Appendix A.4.

Impact of p_{s2}

Using Equations (2.14) and (2.16), the condition for choosing shared over dedicated ($\Pi_s > \Pi_d$) can be written as:

$$\frac{-1}{(1-\alpha)\frac{p_{s2}}{a_{s2}}+\alpha} < \frac{p_{s2}}{a_{s2}} - \gamma_0 \tag{2.17}$$

where $\gamma_0 = -\frac{2}{a_{s2}X_2^{max}}\left[\Pi_d - X_1(p_{s1}-a_{s1}) + c_s\right]$ is independent of p_{s2}.

Suppose p_{s2} is small and Equation (2.17) is invalid, *i.e.,* dedicated networks are preferred. When p_{s2} increases sufficiently, it is possible for Equation (2.17) to be satisfied. That is, a shared network becomes more profitable, as expected. Additionally, it can be shown that there can be at most one such transition from dedicated to shared network as p_{s2} increases. In Appendix A.4, we show that other system cost/revenue parameters give rise to similar behaviors.

2.4.2 Impact of Reprovisioning

To study the impact of reprovisioning coefficient (α), we substitute the expressions for K_{d2}^* and K_{s2}^* from Equations (2.7) and (2.10) into the condition $\Pi_s > \Pi_d$ to obtain:

$$a_{d2}K_{d2}^*(\alpha) - a_{s2}K_{s2}^*(\alpha) > 2\gamma, \tag{2.18}$$

where γ is independent of α and is given by

$$\gamma = \big((p_{d2} - a_{d2}) - (p_{s2} - a_{s2})\big)\frac{X_2^{\max}}{2} + (R_{d1} - R_{s1}) - (c_{d1} + c_{d2} - c_s) \tag{2.19}$$

As seen in Equation (2.19), γ captures the difference in expected profits between the dedicated and shared networks conditioned on capacity exactly meeting the realized demand (as would for example be the case when $\alpha = 1$). The left hand side of Equation (2.18) captures the difference in capacity costs under optimal provisioning between the dedicated and shared network infrastructures as a function of α. For ease of exposition, we introduce the notation $h(\alpha) = a_{d2}K_{d2}^*(\alpha) - a_{s2}K_{s2}^*(\alpha)$ to denote this difference. Note that $h(1) = 0$ as the required up-front capacity is zero for both shared and dedicated infrastructures when $\alpha = 1$. In contrast, $h(0)$ can be positive or negative depending on whether the dedicated or shared network incurs a higher capacity cost in the absence of reprovisioning.

As specified in Equation (2.18), the network infrastructure choice at any value of α depends on the value of $h(\alpha)$ relative to the constant baseline of 2γ. At each value of α where $h(\alpha)$ intersects with 2γ, a switch occurs from preferring one network choice to another. Understanding how reprovisioning affects network choice therefore calls for understanding how the capacity cost difference, $h(\alpha)$, varies with α. This is the topic of Subsection 2.4.2.1. In Subsection 2.4.2.2, we enumerate the possible intersection(s) of $h(\alpha)$ with 2γ and their implications on network choice.

2.4.2.1 Analyzing the effect of α on capacity cost difference

The cost of the capacity that needs to be provisioned up-front is decreasing with α for both shared and dedicated networks. This follows from Lemma 2.3.2 which shows that an increase in α benefits both shared and dedicated networks by helping them reduce their up-front capacity requirements, and hence their corresponding capacity costs ($a_iK_i^*,\ i = \{d2, s2\}$). But the difference in these costs, as captured by $h(\alpha)$, may increase or decrease as α varies in $[0,1]$. Proposition 1 specifies the conditions under which $h(\alpha)$ is increasing (shared benefits more) or decreasing (dedicated benefit more).

Proposition 1 *Increasing α benefits both shared and dedicated networks by reducing their optimal capacity costs. Additionally,*

(i) if $h'(0) \geq 0$ and $h'(1) \geq 0$, an increase in α benefits a shared network more than a dedicated network $\forall \alpha \in [0,1]$.

(ii) if $h'(0) < 0$ and $h'(1) < 0$, an increase in α benefits a dedicated network more than a shared network $\forall \alpha \in [0,1]$.

(iii) if $h'(0) \geq 0$ and $h'(1) < 0$, an increase in α benefits a shared network more at low α and a dedicated network at high α.

(iv) if $h'(0) < 0$ and $h'(1) \geq 0$, an increase in α benefits a dedicated network more at

low α and a shared network at high α.

Proposition 1 establishes that the signs of $h'(0)$ and $h'(1)$ fully characterize the behavior of $h(\alpha)$. Implicit in this characterization is that $h'(\alpha)$ can change its sign at most once for $\alpha \in [0,1]$. The proof that there can be at most one value of α at which $h'(\alpha) = 0$ for $\alpha \in [0,1]$ is given in Appendix A.1.

Proposition 1 is useful for two reasons. First, it helps identify key operational metrics that determine whether a dedicated or a shared network benefits more from improvements in reprovisioning. Second, it provides a useful graphical aid to understand the factors driving the optimal network choice. We elaborate on both these points below.

Since the sign of $h'(\alpha)$ at $\alpha = 0$ and $\alpha = 1$ determines which network benefits more from reprovisioning, substituting, we focus on the relations $h'(0) = 0$ and $h'(1) = 0$:

$$h'(0) = 0: \qquad \frac{p_{d2} - a_{d2}}{(\frac{p_{d2}}{a_{d2}})^2} = \frac{p_{s2} - a_{s2}}{(\frac{p_{s2}}{a_{s2}})^2} \tag{2.20}$$

$$h'(1) = 0: \qquad p_{d2} - a_{d2} = p_{s2} - a_{s2} \tag{2.21}$$

From Equations (2.20) and (2.21), we observe that two operational metrics, the *return on capacity* ($\frac{p_i}{a_i}$; $i = \{s2, d2\}$) and the *gross profit margin* for each unit of used capacity ($p_i - a_i$; $i = \{s2, d2\}$), determine which network choice benefits more from increases in α. In Figure 2.3, we identify the regions in the $\frac{p_{s2}}{a_{s2}}$ (y-axis) and $p_{s2} - a_{s2}$ (x-axis) plane associated with the four conditions from Proposition 1 ($\frac{p_{d2}}{a_{d2}}$ and $p_{d2} - a_{d2}$ are held constant). Note that the y-axis, $\frac{p_{s2}}{a_{s2}}$, only takes values greater than 1 since

$\Pi_2 > 0$. A similar plot can be obtained for the $(\frac{p_{d2}}{a_{d2}}, p_{d2} - a_{d2})$ plane, by holding $\frac{p_{s2}}{a_{s2}}$ and $p_{s2} - a_{s2}$ constants.

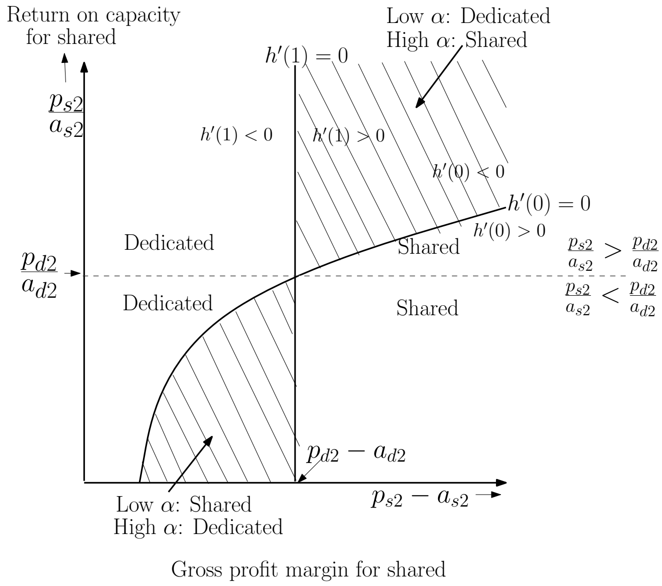

Figure 2.3: Partition of parameter space into regions corresponding to cases of Proposition 1 for the $(p_{s2} - a_{s2}, \frac{p_{s2}}{a_{s2}})$ plane

We observe from Figure 2.3 that the line $p_{d2} - a_{d2} = p_{s2} - a_{s2}$ ($h'(1) = 0$) partitions the plane into two regions such that at high α a dedicated network always benefits more on one side and a shared network on the other. This observation is formalized in Proposition 2.

Proposition 2 *A dedicated network benefits more from reprovisioning at high* α *(*i.e., $\alpha \to 1$*) iff* $p_{d2} - a_{d2} > p_{s2} - a_{s2}$, i.e., *if it has a higher gross profit margin.*

For α close to 1, the network choice with a higher gross profit margin has the higher capacity cost[9]. When $\alpha = 1$, those costs become zero (for both network choices), *i.e.,* $K_i^*(\alpha = 1) = 0$; $i = \{s2, d2\}$. As a result, the network with the greater gross profit margin that started with a higher capacity cost experiences a more significant decrease in its up-front capacity cost as α approaches 1. Given that the only impact of reprovisioning is on up-front capacity needs, this network clearly benefits more from reprovisioning.

Proposition 2 focused on scenarios with high α, and we now turn to scenarios with low α, *i.e.,* very limited reprovisioning. We observe from Figure 2.3 that $h'(0) = 0$ partitions the plane into two regions. More formally,

Proposition 3 *A dedicated network benefits more from reprovisioning at low* α *(*i.e., $\alpha \to 0$*) if* $\frac{p_{d2}-a_{d2}}{\left(\frac{p_{d2}}{a_{d2}}\right)^2} > \frac{p_{s2}-a_{s2}}{\left(\frac{p_{s2}}{a_{s2}}\right)^2}$*, and a shared network benefits benefits more at low* α *if* $\frac{p_{d2}-a_{d2}}{\left(\frac{p_{d2}}{a_{d2}}\right)^2} < \frac{p_{s2}-a_{s2}}{\left(\frac{p_{s2}}{a_{s2}}\right)^2}$.

Proposition 3 indicates that in addition to $p_i - a_i; a_i = \{s2, d2\}$, another metric, return on capacity ($\frac{p_i}{a_i}$, $i = \{d2, s2\}$), affects which network choice benefits more from increase in reprovisioning at low α.

Our analysis thus far identifies for given values of the metrics, $\frac{p_i}{a_i}$ and $p_i - a_i$, $i = \{d2, s2\}$, which region of Figure 2.3 we operate in. Next, we further examine how these metrics influence how each network choice benefits from reprovisioning.

---

[9] $a_i K_i^* = \frac{(p_i - a_i)}{\frac{p_i}{a_i} + \frac{\alpha}{1-\alpha}}$; $i = \{s2, d2\}$. Hence, for $\alpha \approx 1$, the numerator $p_i - a_i$ determines which network has a higher $a_i K_i^*$.

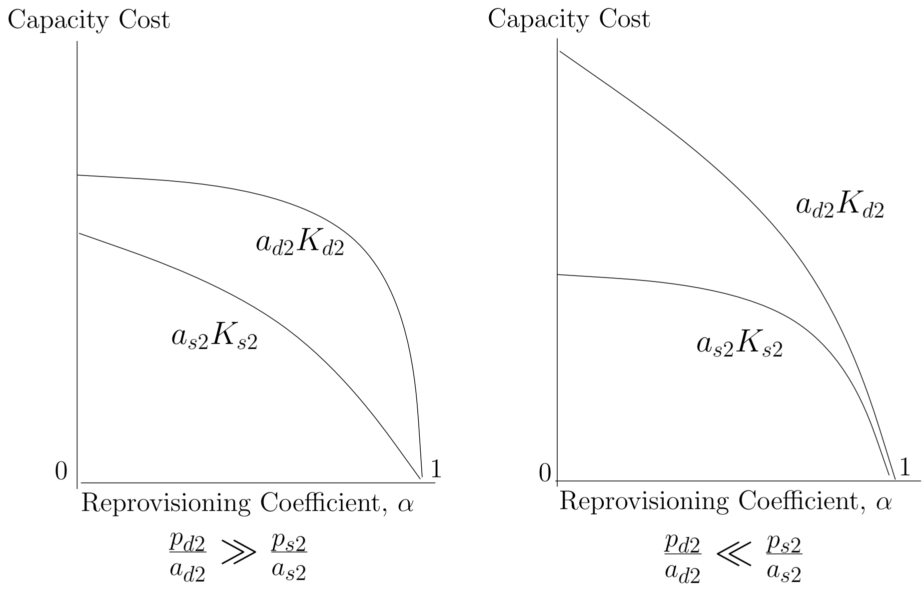

Figure 2.4: (A) Shared benefits more at low α, dedicated benefits more at high α

(B) Dedicated benefits more $\forall \alpha \in [0,1]$

Consider the case in which the dedicated network enjoys a higher gross profit margin than the shared, *i.e.,* $p_{d2} - a_{d2} > p_{s2} - a_{s2}$. Suppose first that $\frac{p_{d2}}{a_{d2}} \gg \frac{p_{s2}}{a_{s2}}$, then a scenario similar to the one shown in Figure 2.4(A) may arise. Since $\frac{p_{d2}}{a_{d2}}$ is high, we know from Lemma 2.3.3 that the up-front capacity remains almost unaffected for the dedicated network for low α. Consequently, the drop in the capacity provisioning cost for the dedicated network is not very significant at low α. On the other hand, since the shared network has a much lower return on capacity, its capacity and the associated provisioning cost decreases faster with increase in α at low α. But when α is high the dedicated network starts to benefit more because the capacity requirements drop from a relatively high value to zero. Therefore, a shared network benefits more from an increase in reprovisioning at low α while a dedicated network does at high α. This observation is consistent with Figure 2.3, where the relationships $p_{d2} - a_{d2} > p_{s2} - a_{s2}$ and $\frac{p_{d2}}{a_{d2}} \gg \frac{p_{s2}}{a_{s2}}$ result in a point in the shaded region in the lower left hand side.

Next consider the case in which $\frac{p_{d2}}{a_{d2}} \ll \frac{p_{s2}}{a_{s2}}$. In Figure 2.4(B), since $\frac{p_{s2}}{a_{s2}}$ is high, the capacity cost of the shared network remains unaffected at low α in accordance with Lemma 2.3.3. But the dedicated network, which has a lower return on capacity, drops its capacity faster at low α. Moreover, in accordance with Proposition 2, the dedicated network continues to benefit more from increase in α even as $\alpha \to 1$ since $p_{d2} - a_{d2} > p_{s2} - a_{s2}$ (*i.e.,* $h'(1) < 0$). Thus, in this scenario, a dedicated network benefits more for all $\alpha \in [0, 1]$. Once again, this is consistent with Figure 2.3(A), where the conditions $p_{d2} - a_{d2} > p_{s2} - a_{s2}$ and $\frac{p_{d2}}{a_{d2}} \ll \frac{p_{s2}}{a_{s2}}$ correspond to a point in the upper left hand side region.

Similar explanations can be given for other parameter values, *e.g.,* $p_{d2} - a_{d2} < p_{s2} - a_{s2}$, and how these map onto the regions of Figure 2.3.

2.4.2.2 Optimal Network Choice

The analysis in Section 2.4.2.1 characterized which network choice benefits more from reprovisioning. However, the provider's optimal network choice depends on how these relative benefits compare to the other cost and revenue parameters. As specified in Equation (2.18), this choice depends on the value of $h(\alpha)$ with respect to the baseline of 2γ, and each intersection between them marks a switch in network choice. In this section, we evaluate the provider's optimal network choice.

As specified in Proposition 1 (see also Figure 2.3), there are four possible behaviors associated with an increase in reprovisioning coefficient, α.

First, consider the region in which a shared network always benefits more from increases in α. Now, if shared network is already the preferred choice at $\alpha = 0$, then it obviously remains the provider's optimal network choice irrespective of reprovisioning ability. This requires $\gamma < 0$ (because $h(\alpha) > 2\gamma$, $\forall\alpha$ and $h(1) = 0$), which can arise if the shared network enjoys significantly lower fixed costs (*i.e.,* $c_s \ll c_{d1} + c_{d2}$) or variable costs ($p_{s2} - a_{s2} \gg p_{d2} - a_{d2}$ and/or $R_{s1} \gg R_{d1}$). A numerical example is shown in Figure 2.5(A). On the other hand, if a dedicated network is initially preferred and if the benefits that the shared network receives from reprovisioning are never sufficient to overcome the impact of other parameters (*i.e.,* $h(\alpha) < 2\gamma$, $\forall\alpha \in [0,1]$), then a dedicated network remains preferred irrespective of α, as shown in Figure 2.5(B) (where a low fixed cost, $c_{d2} = 0.5$ favors a dedicated network). A more interesting outcome arises when a dedicated network is the preferred choice for $\alpha = 0$, but as α increases, the benefits that the shared network receives are sufficiently high to overcome the impact of diseconomies of scope in other costs (*i.e.,* $h(\alpha)$ and 2γ intersect). As a result, the optimal network choice switches to a shared network at high α. A numerical example is shown in Figure 2.5(C), in which a dedicated network is preferred for $\alpha \lesssim 0.6$ and shared for higher values.

Second, consider the region in which a shared network benefits more from increases in α at low α and a dedicated network at high α. A numerical example for this sce-

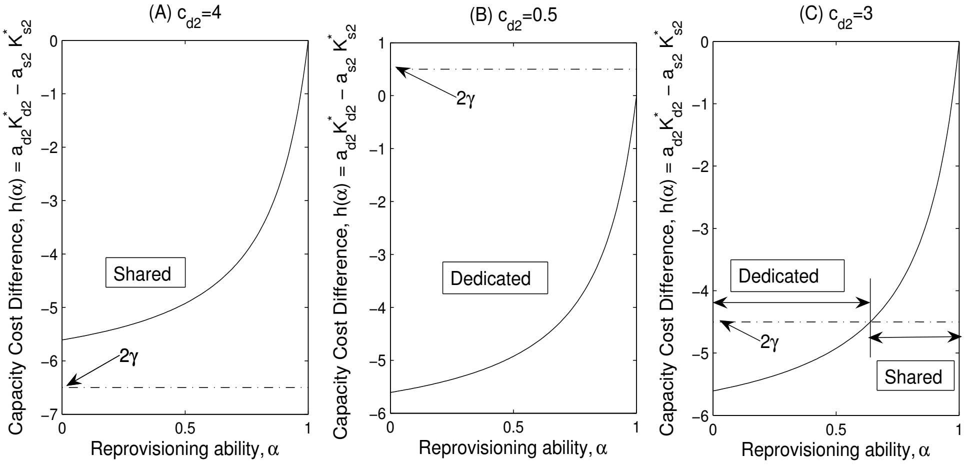

Figure 2.5: (A) Shared network $\forall\alpha \in [0,1]$, (B) Dedicated network $\forall\alpha \in [0,1]$, (C) Dedicated at low α, shared network at high α (Parameter values: $p_{d2} - a_{d2} = 9 < 16.7 = p_{s2} - a_{s2}$, $\frac{p_{d2}}{a_{d2}} = 10 > 8.2609 = \frac{p_{s2}}{a_{s2}}$, $R_{d1} = 30, R_{s1} = 25, c_s = 20, c_{d1} = 5, c_{d2} = 4$).

nario is shown in Figure 2.6, which shows that there are four possible network choice outcomes depending on 2γ: (i) a dedicated network is preferred irrespective of α, (ii) a shared network is preferred irrespective of α, (iii) a shared network is preferred at low α and a dedicated network at high α, (iv) a dedicated network is preferred at both low and high α, and a shared network for intermediate values. A dedicated (shared) network is chosen irrespective of α if there are significant diseconomies (economies) of scope as shown in Figure 2.6(A) (Figure 2.6(B)). In both cases, the impact of reprovisioning is negligible relative to the impact of other cost and revenue parameters. For values of cost parameters such that $h(\alpha)$ and 2γ intersect, the optimal network choice switches. As shown in Figure 2.6(C) and (D), there can be one or two such switches in the optimal network choice.

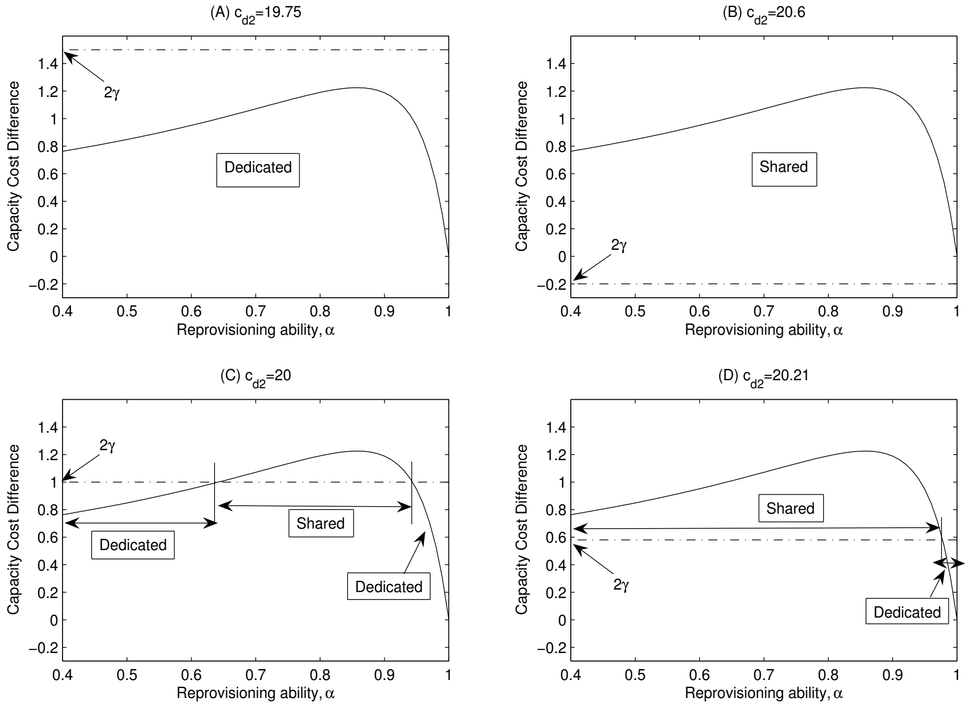

Figure 2.6: (A) Dedicated network $\forall\alpha \in [0,1]$, (B) Shared network $\forall\alpha \in [0,1]$, (C) Dedicated at both high and low α, shared for intermediate α, (D) Shared at low α, dedicated network at high α (Parameter values: $p_{d2} - a_{d2} = 18 > 11 = p_{s2} - a_{s2}, \frac{p_{d2}}{a_{d2}} = 10 > 6.5 = \frac{p_{s2}}{a_{s2}}, R_{d1} = 25, R_{s1} = 25, c_s = 14, c_{d1} = 11$).

Next, we consider the region in which a dedicated network always benefits more than a shared network $\forall\alpha \in [0,1]$. In this scenario, if the diseconomies (economies) of scope in the costs are very large, a dedicated (shared) network is preferred irrespective of α, else the network choice switches from shared to dedicated as α increases. Lastly, when a dedicated network benefits more from increase in α at low α and a shared network at high α, there can be four possible outcomes. The analyses for these last two cases are analogous to the previous ones in which the shared network was benefiting more.

2.4.3 Discussion

In this section, we illustrate the paper's findings through two examples that help contrast the different possible outcomes the model predicts when the ability to reprovision resources improves (α increases).

Recall that shared infrastructures benefit from reusing equipment across services, but that these benefits can all but disappear when the sharing is poorly controlled and produces diseconomies rather than economies of scope. Technologies that control sharing have, therefore, played an important role in the emergence of shared solutions, *e.g.,* witness the impact of "virtualization" on the growing popularity of both cloud computing and virtual networks. However, the same technologies that enable better control of shared resources, often also facilitate more dynamic provisioning of those same resources. As seen in the previous section, better reprovisioning abilities and greater economies of scope, as measured by improvements in *gross profit margin* and *return on capacity,* need not always combine to favor shared solutions. We illustrate this next through two examples[10].

Consider the task of providing computing services in the late eighties, early nineties. There were two major competing options for delivering such services. Systems such as IBM mainframes were representative of shared solutions that would support multi-

[10]The examples draw from computation as opposed to network services, but as mentioned earlier, the model is applicable to different types of services infrastructures.

ple services (and users). In contrast, DEC mini-computers and later on a wide range of "workstations" were the pillars of dedicated solutions, with individual machines assigned to specific tasks or users. The cost of equipment and therefore computational capacity was substantially lower for dedicated solutions than it was for IBM mainframe shared solutions, *i.e.,* $a_d < a_s$. As a result, and even if services based on IBM mainframes often carried a premium ($p_s \gtrsim p_d$), this is a scenario that maps to the lower left quadrant of Figure 2.3, *i.e.,* shared solutions boast lower gross profit margins and return on capacity than dedicated ones. Furthermore, in the context of this example, both solutions can be argued to have remained within competitive range of each other. Hence, they belong to the "shaded area" of the lower left quadrant of Figure 2.3; an area where better reprovisioning abilities (α) can favor either shared or dedicated solutions. In particular, improving α in the low-α region benefits shared solutions more than dedicated ones. This can be claimed to capture the small improvements in reprovisioning that technology advances afforded both mainframes and mini-computers, *e.g.,* through processor and memory upgrades or even additional processors cards. As shown in Figure 2.3, these improvements would have then favored (shared) IBM mainframes more than (dedicated) mini-computer based solutions. This factor, and obviously many others, may have enabled mainframes to survive in spite of the emergence of cheaper distributed solutions.

Contrast the previous situation with the current environment for computing services, where dedicated and shared solutions both rely on the same type of equipment, *i.e.,* a stand-alone blade server fulfills the needs of an individual computation service, while racks of the same blade servers can be shared across services. In addition, previously mentioned technologies such as virtualization offer tight control of resources sharing across services, which enables shared solutions to take full advantage of the economies of scope they afford. The similar equipment costs $(a_s \approx a_d)$ and the ability to fully leverage the economies of scope of shared solutions $(p_s \geq p_d)$ imply that we are now operating in the upper-right quadrant of Figure 2.3, *i.e.,* shared solutions display both higher gross profit margins and return on capacity than their dedicated counterparts. Furthermore, the same technology that is behind stackable blade servers and virtualization makes highly dynamic reprovisioning a reality, *i.e.,* idle CPUs can be rapidly allocated to individual services, and adding new blades to an existing system can be done with little turn-around time. In other words, we are now in a high α environment. Hence, throughout the upper-right quadrant of Figure 2.3, improvements in reprovisioning abilities only further the advantage of shared solutions. In other words, unlike the "mainframe vs. workstation" scenario where improving reprovisioning tilted the balance back towards the less competitive mainframe solution, it now further strengthens the solution of choice, shared systems, which augurs well for the continued growth of large-scale cloud computing systems.

2.5 Robustness to Alternative Models

In this section, we provide numerical examples to show that the different possible outcomes reported in Section 2.4 about the impact of reprovisioning coefficient ($\alpha \in [0,1]$) on the network infrastructure choice can arise even in the presence of economies of scale and non-uniform demand distribution. Of particular interest is the observation that even the more complex outcome that involves choosing one type of network for both high and low values of α and the other for intermediate values is also present. Some of these results are presented below.

2.5.1 Economies of Scale and Non-uniform Demand Distribution

The motivation for these numerical investigations is to show that while economies of scale may provide an additional advantage to a shared network, and a non-uniform demand distribution can affect the benefits of reprovisioning for the two network choices differently, their presence (or absence) alone is not enough to strongly favor a shared solution over dedicated ones (or vice-versa).

We use $a_i K_i^{\gamma}, i = \{s2, d2\}, \gamma < 1$ to account for the economies of scale in capacity costs. In addition to economies of scale, the two examples provided below consider positively and negatively skewed beta distributions, respectively, for Service 2's demand.

In Figure 2.7 ($R_{s1} = 30.5, R_{d1} = 10, c_{d1} = 12.03, c_{d2} = 14, c_s = 6, p_{s2} = 6.8, a_{s2} = 2, p_{d2} = 16.5, a_{d2} = 2$), the economies of scale in capacity grow as $a_i K_i^{0.8}, i = \{s2, d2\}$

and Service 2's demand distribution follows a positively skewed beta distribution between $[0, X_2^{\max}]$ with parameters (1.5, 2), *i.e.,* skewness of 0.2226. The example shows an instance of infrastructure choice where dedicated networks are preferred at both high and low α, while a shared network is preferred in the intermediate range of α.

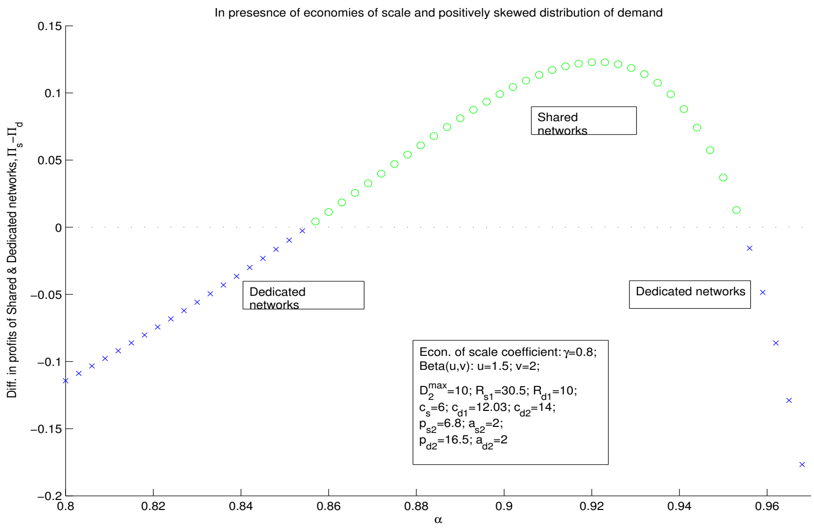

Figure 2.7: Impact of α for positively skewed beta distribution & economies of scale

In Figure 2.8 ($R_{s1} = 9.68, R_{d1} = 20, c_{d1} = 10, c_{d2} = 13, c_s = 16.1, p_{s2} = 5.2, a_{s2} = 2.1, p_{d2} = 4.5, a_{d2} = 2$), we have $a_i K_i^{0.85}, i = \{s2, d2\}$ (*i.e.,* $\gamma = 0.85$) and the demand distribution of Service 2 follows a negatively skewed beta distribution between $[0, X_2^{\max}]$ with parameters (1.5,1), *i.e.,* skewness of -0.2223. In this scenario, a shared network is preferred at both high and low α, while dedicated networks are preferred for intermediate values.

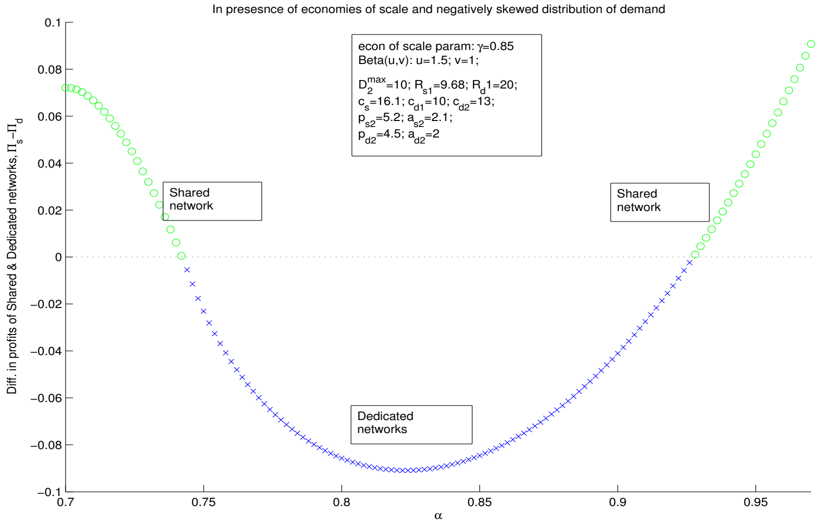

Figure 2.8: Impact of α for negatively skewed beta distribution & economies of scale

2.6 Conclusions

This work introduces an analytical framework to investigate when shared or dedicated infrastructures offer a more cost-effective solution in the deployment of new services. The choice of an infrastructure is influenced by many factors such as fixed and variable costs, capacity costs, and the ability to dynamically reprovision resources as made possible by technologies such as virtualization. The results demonstrate that although strong economies or diseconomies of scope in the cost components can, as one would expect, favor a shared or a dedicated solution, the ability to dynamically provision resources can have a similar influence. Reprovisioning improves the profits of both shared and dedicated solutions, but can do so differently as a function of their respective gross profit margins and returns on capacity. The selection of a preferred infrastructure is, therefore, influenced not only by economies and diseconomies of scope, but also by how the infrastructure is affected by reprovisioning. The work identifies the presence of complex interactions between these factors, and characterizes when they give rise to different outcomes.

Although the model demonstrates the impact of reprovisioning and identifies operational metrics that influence network choice, it relied on a number of assumptions that we briefly review. First, the model focused on economies of scope, and ignored the possible impact of economies of scale. As discussed in Section 2.3, this is a realistic assumption in many settings, including networks. Furthermore, the numerical investigation of a model that incorporates economies of scale in Section 2.5 demonstrated qualitatively similar results. Second, for analytical tractability the model assumed a uniform distribution for the demand of Service 2. As shown again in Section 2.5, the results remain qualitatively valid even under non-uniform demand distributions. The model also assumed that reprovisioning was equally available in shared and dedicated networks. The use of different reprovisioning parameters for shared and dedicated solutions, *i.e.,* α_s and α_d, can be readily incorporated in the model, albeit at the cost of greater notational complexity. Finally, the model also assumed that reprovisioning was invoked only in the presence of excess demand, *i.e.,* provisioned capacity could not be relinquished when demand was insufficient. Allowing symmetric reprovisioning represents an interesting extension to the model, although it is intuitively not expected to yield drastically different results.

Chapter 3

Adoption of Competing Network Technologies

3.1 Introduction

Advances in technology often see newer and better solutions replacing older ones. Networking is no exception. For example, the Internet competed against alternative packet data technologies before finally displacing the phone network as the de facto communication infrastructure. Recently, there have been calls for new architectures to succeed it, and these will face a formidable incumbent in the Internet. Their eventual success in replacing it will likely depend not just on technical superiority, but also on economic factors, and on their ability to win over the Internet's installed base.

A large installed base can give an incumbent an edge even if a new (entrant) technology is technically superior. The traditional networking approach to this problem has been converters (a.k.a. gateways) to ease migration from one technology to another. This is not unique to networks, but converters are particularly important in network settings where "communication" is the primary function, and its benefits grow with the number of users that can be reached, *e.g.,* as in Metcalfe's Law. Since converters allow users of one technology to connect with users of another, they are an important tool in the adoption of network technologies. However, developing, deploying, and operating converters comes at a cost, one that often grows as a function of the converter's quality. Further, converters can play a directionally ambiguous role. On one hand, a converter can help the entrant overcome the advantage of the incumbent's large installed base by allowing connectivity to it. On the other hand, the converter also helps the incumbent technology by mitigating the impact of its users migrating to the newer technology. Understanding the impact of converters on network technology adoption is, therefore, a topic that deserves further scrutiny.

In this work, we develop a modeling framework to study adoption dynamics of entrant and incumbent technologies in the presence of network externalities. Specifically, we introduce a model for the utility derived by an individual user from a communication network, and use it to build an aggregate model for technology adoption that is consistent with individual rational decision-making. We apply the model to study the role that converters can play in the adoption of network technologies. Our main findings are:

- The adoption process can exhibit multiple steady state outcomes (equilibria); each associated with a specific region of initial adoption levels for the two technologies. For example, an entrant technology may succeed only if the incumbent is not already well entrenched.
- Converters can help a technology improve its own standing, *i.e.,* market share, and even ensure its dominance while it would have entirely disappeared in the absence of converters. For example, a low-quality but low-cost technology may thwart the success of a better but more expensive competitor by preserving the ability of its users to access adopters of the pricier technology, whose usage would then be limited to a few "techno-buffs."
- Improving converters' efficiency can at times be harmful. They can result in lower market shares for an individual technology or for *both*. For instance, high market penetration may depend on the combination of a cheap but low-end technology with a high-end but more expensive one to adequately serve the full spectrum of user preferences. A situation where converters allow the better technology to gain market share at the expense of the lesser technology may result in low-end users of that technology dropping out altogether; thereby contributing to a lower overall market penetration.
- While in the absence of converters technology adoption always converges to a stable steady-state equilibrium, this need not be so when converters are present. Boom-and-bust cycles in which users switch back-and-forth between technologies

can arise when technologies are asymmetric in the externality benefits they offer.

The rest of this chapter is organized as follows: Section 3.2 introduces our model and problem formulation. Section 3.3 characterizes technology adoption trajectories and equilibrium adoption levels. Section 3.4 explores the role of converters in influencing adoption outcomes. Section 3.6 reviews prior work and positions the work in the literature. We discuss the limitations of this study and conclude the chapter with remarks on future work in Section 3.7. The proofs of all propositions can be found in Appendix B.

3.2 Technology Adoption Model

3.2.1 Technology Valuation

As in most competitive situations, the choice of one technology over another depends on the "value" they provide. Value is a somewhat elusive concept that depends in part on the quality and functionality of the technology and its cost. In the context of network technologies whose main purpose is to enable communication among users, the *number* of users[11] accessible through it is another important component, often termed network effect or externality. As commonly done, we account for these factors and their effect on technology adoption through a *utility function.* For two competing network

[11] Users can be individuals or organizations, and include resources and content.

technologies, 1 and 2, the respective utility functions are given by eqs. (3.1) and (3.2).

$$U_1 = \theta q_1 + (x_1 + \alpha_1 \beta x_2) - p_1 \tag{3.1}$$

$$U_2 = \theta q_2 + (\beta x_2 + \alpha_2 x_1) - p_2 \tag{3.2}$$

Eqs. (3.1) and (3.2) consist of three distinct terms. Focusing on, say, Technology 1, the first term, θq_1 represents the stand-alone benefits from the technology, with q_1 representing the intrinsic quality of the technology, and θ a random variable accounting for heterogeneity in how users value technology. The quantity q_1 incorporates aspects of functionality, reliability, performance, etc., for the technology. In the rest of the chapter, we assume $q_2 > q_1$, *i.e.,* Technology 2 is superior to Technology 1 and correspondingly can be viewed as the entrant with Technology 1 playing the role of the incumbent. The model, however, does not mandate such an assignment of roles, *e.g.,* it can be used to study settings where Technology 1 is the entrant and offers, say, a lower quality but cheaper alternative than the incumbent Technology 2. The random variable $\theta \in [0,1]$ determines the relative weight a user places on the intrinsic quality of a technology. It is private information, but we assume that the distribution of θ across users is known. We make the common assumption [11] that θ is uniformly distributed in the interval $[0,1]$. This choice affects the magnitude of equilibrium adoption levels, but does not qualitatively affect findings regarding technology adoption dynamics and outcomes as demonstrated in Section 3.5.1.

The second component of the user's utility is the network externality (or network effect), which refers to benefits derived from the ability to connect with other users. Network externalities are chosen to be proportional to the number of users to which each technology gives access. This linear dependency of network benefits on the number of adopters is consistent with Metcalfe's Law and commonly used in the literature [24]. In Sections 3.5.2 and 3.5.3, we investigate other models and demonstrate the robustness of our findings across different functional forms for network externality, including non-linear ones. Denoting as x_1 and x_2 the fractions of adopters of each technology out of a large population of size N, the externality benefits for Technology 1 consist of x_1, the fraction of Technology 1 users, plus $\alpha_1\beta x_2$, a term that includes the fraction of Technology 2 users weighed by two additional factors. The first, $0 \leq \alpha_1 \leq 1$, captures the availability of converters offering connectivity *from* Technology 1 *to* Technology 2 ($\alpha_1 = 0$ corresponds to no converter and $\alpha_1 = 1$ to "perfect" converters). The second parameter, β, allows different externality benefits for the two technologies[12]. We note that converters, once deployed, are available to *all* users of the technology. This corresponds to what we term "technology-level" converters, *i.e.,* their development and deployment are decisions made by the providers of network technologies.

Converters can be characterized as either *duplex* or *simplex*, *symmetric* or *asymmetric*, and *constrained* or *unconstrained.* Duplex converters provide bi-directional connectivity between technologies, while simplex converters are present in only one direction (most network technologies involve duplex converters, but the model does not mandate them). Asymmetric converters simply refer to the fact that converter ef-

[12]Eqs. (3.1) and (3.2) implicitly express utility in units of Technology 1 externality benefits, *i.e.,* Technology 1 externality benefits are equal to 1 when its penetration level is 100% ($x_1 = 1$).

ficiency can be different in each direction *i.e.,* $\alpha_1 \neq \alpha_2$. The notion of constrained vs. unconstrained converters arises in the presence of technologies that exhibit different externalities, *i.e.,* $\beta \neq 1$. For example, when $\beta > 1$, it captures whether converters allow users of Technology 1 access to the greater externality benefits of Technology 2 when connecting to its users. A converter is unconstrained if this is permitted, *i.e.,* $\alpha_1\beta > 1$. We discuss an example where this can arise at the end of the section.

The last element of eqs. (3.1) and (3.2) is the price, $p_i, i \in \{1,2\}$. Because of our focus on networks and connectivity that is typically offered as a *service* rather than a good or product, price is recurrent. In other words, maintaining connectivity through a particular network technology incurs new charges at regular intervals. As a result, users continuously reevaluate their technology choices, and can switch from one technology to another and possibly back. For analytical tractability, the model assumes that switching costs are negligible. This represents a reasonable assumption in many settings. For example, the very high customer churn (reported to range from 72% all the way up to 98% per year [17, 59]) that prevails in the ISP market points to little or no switching costs in that market. On the other hand, non-zero switching costs, in the form of contract breaking penalties and learning costs, are the norm in many settings. A natural question is, therefore, whether behaviors observed in the absence of switching costs hold when they are present. Extending the analytical model to incorporate switching costs is challenging, especially if one is to consider the many possible types of switching costs, but numerical investigations are feasible. Section 3.5.4 reports on these investigations, which demonstrate a relative insensitivity to switching costs. Obviously, as switching costs increase, they eventually eliminate all adoption dynamics, but the different behaviors that the (simple) model of this work helped reveal persist over a non-trivial range.

We note that the model parameters, *i.e.,* q_i, p_i, α_i, β, are static and exogenously specified. An obvious extension is to make them time-varying, *e.g.,* technology gets better and/or cheaper as time goes by, and the outcome of strategic decision-making. Developing game-theoretic models that incorporate such effects is clearly of interest, especially in the context of competitive scenarios where firms may offer introductory pricing or seed the market to gain an initial foothold. Exploring those issues, however, requires that we first understand the basic tenets of technology adoption and dynamics in the simpler setting considered in this work.

Another important question is how to assign actual values to the model's parameters. This is a topic that goes well beyond this work, and we only point to a possible approach. A common method to estimate utility weights is conjoint analysis, a technique that has been widely adopted by marketing researchers and practitioners (see [34, 33] for a detailed review). It relies on surveys offering users different combinations of functionality and attributes to extract a relative ordering among them, and ultimately produce individual weight assignments.

3.2.2 User Decisions

Given current adoption levels, x_1 and x_2, the utility functions of eqs. (3.1) and (3.2) identify how a user values each technology, which in turn determines her technology selection decisions. Specifically, a user chooses Technology i whenever it provides a surplus that is both positive (Individual Rationality constraint) and higher than that of the other technology (Incentive Compatibility constraint). In other words, a user chooses

$$\begin{cases} \text{no technology} & \text{if} \quad U_i < 0 \text{ for all } i, \\ \text{Technology 1} & \text{if} \quad U_1 > 0 \text{ and } U_1 > U_2, \\ \text{Technology 2} & \text{if} \quad U_2 > 0 \text{ and } U_2 > U_1. \end{cases}$$

Note that the model assumes an exclusive choice of technology by users, *i.e.,* they select Technology 1, or 2, or neither, but not both. This translates into the constraint $0 \leq x_1 + x_2 \leq 1$. The dynamics of technology adoption arise from the dependency of the U_i's on the x_i's that change with users' adoption decisions. Capturing these dynamics, therefore, calls for specifying when users become aware of changes in the x_i's and update their adoption decisions. Knowledge of changes in adoption levels is likely to diffuse through the user population and users' reactions are often heterogeneous, *i.e.,* some switch quickly, while others defer. An approach, commonly used in individual-level diffusion models [39] and that captures these aspects is a continuous time approximation. Specifically, assume that at time t the "current" technology adoption levels, $\underline{x}(t) = (x_1(t), x_2(t))$, are known to all users. With this information, users can compute their utility for each technology and make adoption decisions. Let $H_i(\underline{x}(t)),\ i \in \{1,2\}$ denote the fraction of users for whom Technology i provides the highest (and positive) utility[13]. The quantity $H_i(\underline{x}(t)) - x_i(t)$ corresponds to the fraction of users that would normally proceed to adopt (disadopt) Technology i at time t. To capture a progressive adoption process, we assume that the rate of change in users' technology choices is proportional to this quantity, namely,

$$\frac{dx_i(t)}{dt} = \gamma\big(H_i(\underline{x}(t)) - x_i(t)\big),\ i \in \{1,2\}, \tag{3.3}$$

The quantity $\gamma < 1$ is analogous to the hazard rate in diffusion models, and can be viewed as the expected conditional probability that an individual who has not yet adopted technology i will do so at time t. In our analysis, we assume that the propensity of individuals to adopt does not change with time, *i.e.,* γ is constant.

Two aspects of this diffusion process need further clarification. First, users are myopic. At any instant, the adoption decisions are driven by the number of adopters at that time ($x_i(t)$) and users are not able to anticipate likely adoption levels in the future. This is analogous to a best response dynamic. Second, the model identifies the rate of technology adoption across users, but not *which* users are making the change. To preserve consistency with user preferences, θ, we assume that the first users to adopt Technology i are those that stand to benefit most from the action. This ensures that at all times the sets of users having adopted either technology correspond to blocks of users with contiguous technology preferences.

[13]We discuss the derivation of $H_i(\underline{x}(t))$ in Section 3.3.1.

The diffusion dynamics governed by eq. (3.3) can converge to a steady-state equilibrium $\underline{x}^*$ characterized by:

$$\left.\frac{dx_i(t)}{dt}\right|_{\underline{x}(t)=\underline{x}^*} = 0 \quad \Leftrightarrow x_i^* = H_i(\underline{x}^*) \text{ for } i \in \{1,2\}. \tag{3.4}$$

In other words, at equilibrium, the fraction of users for whom it is individually rational and incentive compatible to choose Technology i equals the current fraction of adopters of Technology i. Based on this formulation, our goal is to characterize, as a function of the exogenous system parameters $\beta, p_i, q_i, \alpha_i$ for $i \in \{1,2\}$, the equilibrium adoption levels, *i.e.,* the fixed points of eq. (3.4), and the dynamics leading to them.

Before exploring the dynamics and equilibria of technology adoption that the model gives rise to, we pause to briefly introduce a couple of examples that illustrate the model's parameters and applicability.

IPv4 vs. IPv6

The impending exhaustion of IPv4 addresses, *e.g.,* http:/www.potaroo.net/tools/ipv4 for a daily countdown, implies that service providers signing up new Internet customers will have to start using IPv6 addresses or charge more users who insist on an IPv4 address, *i.e.,* $p_{\text{IPv4}} = p_1 > p_2 = p_{\text{IPv6}}$. As technologies, although IPv4 and IPv6 are *incompatible,* they are largely similar so that for the purpose of our model one can reasonably assume $q_1 \lesssim q_2$ and $\beta = 1$. Because of their incompatibility, converters (gateways), *e.g.,* see [18] for a representative recent proposal, are needed for IPv6 users to access the IPv4 content that is the bulk of today's Internet content and unlikely to become natively accessible over IPv6 any time soon[14]. Conversely, those converters also enable the reverse flow from IPv4 to IPv6, *i.e.,* they are duplex converters, albeit not necessarily delivering the same performance in both directions, *i.e.,* they can be asymmetric, so that both α_1 and α_2 are non-zero but not always equal.

Users then decide between subscribing to an IPv4 or IPv6 service on the basis of price (p_i), the level of content they are able to access (x_i), and the quality of that access (α_i).

Low Def. vs. High Def. Video

The previous example illustrated a common adoption scenario with two mostly equivalent technologies and duplex, asymmetric converters. Because of the similarity of the two technologies $(\beta = 1)$, converters were by default constrained $(\alpha_1\beta \leq 1)$. However, when technologies exhibit significant differences in externality benefits, *e.g.,* $\beta > 1$, converters can be unconstrained $(\alpha_1\beta > 1)$ and we present next a possible example.

Consider a provider that offers its customers a video-conferencing service with the associated equipment. The service comes in two versions, high-definition (HD) and standard-definition (SD), *i.e.,* HD equipment generates a high-quality (q_2) video signal while SD equipment produces a lower resolution $(q_1 < q_2)$. Users derive value from video-conferencing with one another, with $\beta > 1$ reflecting the higher quality of an HD signal. The two services are priced accordingly $(p_2 > p_1)$. However, because video is a

[14]Although the servers hosting most web sites can typically get an IPv6 address, very few have bothered registering one with DNS, *e.g.,* see http://bgp.he.net/ipv6-progress-report.cgi.

highly asymmetric technology (encoding is hard but decoding is comparatively easy), it is possible for the provider to enable the *decoding* of HD signal on SD equipment (and obviously conversely). This conversion can introduce quality degradations ($\alpha_1 < 1$), but more importantly it allows SD users access to the externality benefits associated with receiving HD signals. Assuming HD$\leftrightarrow$SD conversion is available in both direction, this is an instance of a duplex, possibly asymmetric ($\alpha_1 \neq \alpha_2$), and unconstrained ($\alpha_1\beta > 1$) converter.

Many users may then opt for the SD service because of its lower price and the ability to still enjoy the higher benefits of viewing HD signals. On the other hand, if all users were to select the SD service, those externality benefits would disappear. In general, users with high technology valuation (θ close to 1) may still opt for the HD service, but the decision depends on choices made by others.

3.3 Trajectories and Equilibria

Solving the evolution of technology adoption decisions over time described in eq. (3.3) calls for first computing expressions for $H_i(\underline{x}(t)),\ i = \{1,2\}$ as functions of known parameters.

3.3.1 Characterizing Potential Adoption Levels

For notational convenience we omit dependency on time and write $\underline{x}(t)$ simply as $\underline{x}$. Recall that $H_i(\underline{x}),\ i \in \{1,2\}$, corresponds to the fraction of users for whom it is rational to adopt Technology i, *given* the current adoption levels, $\underline{x}$. To determine the fraction of adopters of each technology, we introduce the notion of *indifference points,* which identify thresholds in users technology valuation (θ) corresponding to changes in technology preference. Specifically, $\theta_i^0(\underline{x}), i \in \{1,2\}$ identify the θ value separating users with a negative utility for Technology i from those with a positive utility. In other words, for technology penetration levels $\underline{x}$, $\theta_i^0(\underline{x})$ is such that $U_i(\theta_i^0,\underline{x}) = 0$, and $U_i(\theta,\underline{x})$ is positive (negative) for θ values larger (smaller) than θ_i^0.

From eqs. (3.1) and (3.2), $U_i(\theta_i^0,\underline{x}) = 0$ gives

$$\theta_1^0(\underline{x}) = \frac{p_1 - (x_1 + \alpha_1 \beta x_2)}{q_1} \tag{3.5}$$

$$\theta_2^0(\underline{x}) = \frac{p_2 - (\beta x_2 + \alpha_2 x_1)}{q_2} \tag{3.6}$$

Similarly, $\theta_2^1(\underline{x})$ corresponds to the θ value separating users preferring Technology 1 from those preferring Technology 2, *i.e.,* $U_1(\theta_2^1,\underline{x}) = U_2(\theta_2^1,\underline{x})$ and users with $\theta > \theta_2^1(\underline{x})$ derive greater utility from Technology 2 than Technology 1 (recall that $q_2 > q_1$). Setting, $U_1(\theta_2^1,\underline{x}) = U_2(\theta_2^1,\underline{x})$ gives

$$\theta_2^1(\underline{x}) = \frac{(1-\alpha_2)x_1 - \beta(1-\alpha_1)x_2 + p_2 - p_1}{q_2 - q_1} \tag{3.7}$$

Combining eqs. (3.5)-(3.7) gives

$$\theta_2^1(\underline{x}) - \theta_1^0(\underline{x}) = \frac{q_2}{q_2 - q_1}(\theta_2^0(\underline{x}) - \theta_1^0(\underline{x})), \tag{3.8}$$

$$\theta_2^1(\underline{x}) - \theta_2^0(\underline{x}) = \frac{q_1}{q_2 - q_1}(\theta_2^0(\underline{x}) - \theta_1^0(\underline{x})), \tag{3.9}$$

from which the following Proposition can be derived.

Proposition 4

If $\theta_1^0(\underline{x}) < \theta_2^0(\underline{x})$, *then* $\theta_2^1(\underline{x}) > \theta_2^0(\underline{x}) > \theta_1^0(\underline{x})$.

If $\theta_1^0(\underline{x}) \geq \theta_2^0(\underline{x})$, *then* $\theta_2^1(\underline{x}) \leq \theta_2^0(\underline{x}) \leq \theta_1^0(\underline{x})$.

Proposition 4 constrains the possible orderings of the indifference points given by eqs. (3.5)-(3.7), so that $H_i(\underline{x}), i \in \{1,2\}$ can be characterized in a compact manner.

$$H_1(\underline{x}) = \begin{cases} [\theta_2^1(\underline{x})]_{[0,1]} - [\theta_1^0(\underline{x})]_{[0,1]} & \text{if } \theta_1^0(\underline{x}) < \theta_2^0(\underline{x}) \\ 0 & \text{otherwise} \end{cases}$$
$$H_2(\underline{x}) = \begin{cases} 1 - [\theta_2^1(\underline{x})]_{[0,1]} & \text{if } \theta_1^0(\underline{x}) < \theta_2^0(\underline{x}) \\ 1 - [\theta_2^0(\underline{x})]_{[0,1]} & \text{otherwise} \end{cases} \tag{3.10}$$

where $y_{[a,b]}$ is the 'projection[15]' of y into $[a,b]$.

The expressions for $H_1(\underline{x})$ and $H_2(\underline{x})$ determine the trajectory as well as the equilibrium outcome of the adoption process as per eqs. (3.3) and (3.4) respectively. Eq. (3.10) characterizes $H_i(\underline{x}),\ i = \{1,2\}$ through multiple possible expressions that depend on the relative ordering of $\theta_1^0(\underline{x}), \theta_2^0(\underline{x})$ and $\theta_2^1(\underline{x})$, and the outcome of their projections on $[0,1]$. Identifying the different combinations that eq. (3.10) gives rise to calls for partitioning the (x_1, x_2)-plane into *regions*, each corresponding to unique expressions for the $(H_1(\underline{x}), H_2(\underline{x}))$ pair. A method for constructing such a partition is described next.

First, consider all values of $\underline{x}$ which satisfy $\theta_1^0(\underline{x}) \geq \theta_2^0(\underline{x})$. In the corresponding half-plane of the (x_1, x_2)-plane, $H_1(\underline{x})$ is always 0, but the value of $H_2(\underline{x})$ depends on the projection of $\theta_2^0(\underline{x})$ on $[0,1]$ (*i.e.,* whether $\theta_2^0(\underline{x}) \leq 0$, $0 < \theta_2^0(\underline{x}) < 1$, or $1 \leq \theta_2^0(\underline{x})$). This creates three regions in the (x_1, x_2)-plane, each with a different expression for the

[15] *i.e.,* its value is y for $y \in [a,b]$, a for $y < a$, and b for $y > b$.

$(H_1(\underline{x}), H_2(\underline{x}))$ pair. These three regions, labeled R_1, R_2 and R_3, and the corresponding conditions on $\theta_2^0(\underline{x})$ appear in the left side of Table 3.1. The expressions for $H_i(\underline{x}),\ i = \{1,2\}$ in each region are provided in Table 3.2.

Conversely, for values of $\underline{x}$ which satisfy $\theta_1^0(\underline{x}) < \theta_2^0(\underline{x})$ (the other half-plane), eq. (3.10) indicates that expressions for $H_1(\underline{x})$ and $H_2(\underline{x})$ depend on the positions of both $\theta_2^1(\underline{x})$ and $\theta_1^0(\underline{x})$ relative to 0 and 1 (*i.e.,* whether $\theta_2^1(\underline{x}) \leq 0$, $0 < \theta_2^1(\underline{x}) < 1$ or $1 \leq \theta_2^1(\underline{x})$, and similarly for $\theta_1^0(\underline{x})$). This yields nine possible combinations. The number of feasible combinations can, however, be reduced to six using Proposition 4, which constrains $\theta_1^0(\underline{x}) < \theta_2^0(\underline{x}) < \theta_2^1(\underline{x})$. For example, when $\theta_1^0(\underline{x}) < \theta_2^0(\underline{x})$, $\theta_2^1(\underline{x}) \leq 0$ and $1 \leq \theta_1^0(\underline{x})$ is infeasible per Proposition 4. Thus in the half-plane $\theta_1^0(\underline{x}) < \theta_2^0(\underline{x})$, there are six possible expressions for $H_i(\underline{x}),\ i = \{1,2\}$. These expressions are reported on the right side of Table 3.1, with the corresponding regions labeled R_4 to R_9. Combining the two half-planes gives a total of nine regions, R_1 to R_9, where in each region the $(H_1(\underline{x}), H_2(\underline{x}))$ pair has a unique expression as specified in Table 3.2.

This partitioning in nine regions has a graphical representation, as shown in Figure 3.1. The line $\theta_1^0(\underline{x}) = \theta_2^0(\underline{x})$ splits the (x_1, x_2)-plane in the two previously mentioned half-planes. The two lines corresponding to $\theta_2^0(\underline{x}) = 0$ and $\theta_2^0(\underline{x}) = 1$ are parallel, and define the three regions R_1, R_2 and R_3 in the half-plane $\theta_1^0(\underline{x}) \geq \theta_2^0(\underline{x})$. Similarly, the lines $\theta_1^0(\underline{x}) = 0$, $\theta_1^0(\underline{x}) = 1$, $\theta_2^1(\underline{x}) = 0$ and $\theta_2^1 = 1$, divide the second half-plane into the six regions, R_4 to R_9. Figure 3.1 also illustrates that the lines $\theta_2^0(\underline{x}) = 0, \theta_1^0(\underline{x}) = 0$ and $\theta_2^1(\underline{x}) = 0$ always intersect at a point denoted as P, and the lines $\theta_2^0(\underline{x}) = 1, \theta_1^0(\underline{x}) = 1$

and $\theta_2^1(\underline{x}) = 1$ always intersect at a point denoted as Q, with both P and Q[16] lying on the line $\theta_1^0 = \theta_2^0$. The points P and Q can be shown to act as "anchors" that ensure that the (x_1, x_2)-plane is always partitioned into exactly nine regions with fixed relative positions.

It should also be noted that all nine regions need not always be feasible. Whether or not they are feasible depends on their relative position in the solution space, $S = \{(x_1, x_2)\, s.t.\, 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, x_1 + x_2 \leq 1\}$. The number of regions that lie inside S (and hence are relevant to the analysis) is a function of the system parameters $(q_i, p_i, \beta, \alpha_i; i = \{1,2\})$. Last but not least, as shown in Appendix B.1, the partitioning of the solution space into nine regions actually holds for more generic (monotonic) externality functions, *i.e.,* it is not an artifact of the simplified linear externality function.

Finally, we pause to briefly interpret the conditions that define each region, and their implications for solutions. We do so by way of an example, focusing on Region R_8. Region R_8 is defined as the set of adoption levels, $\underline{x} = (x_1, x_2)$, for which $1 \leq \theta_2^1(\underline{x})$ and $0 \leq \theta_1^0(\underline{x}) < 1$. The condition $1 \leq \theta_2^1(\underline{x})$ implies that in Region 8 all users prefer Technology 1 over 2. Hence any existing Technology 2 adopter will disadopt. Thus, in R_8, users can either be non-adopters of both technologies ($0 < \theta < \theta_1^0(\underline{x})$) or adopters of Technology 1 (if $\theta > \theta_1^0(\underline{x})$). Similar interpretations are possible for other regions.

[16]The coordinates of the points P and Q can be readily found by solving simple systems of linear equations given by eqs.(3.5)-(3.7).

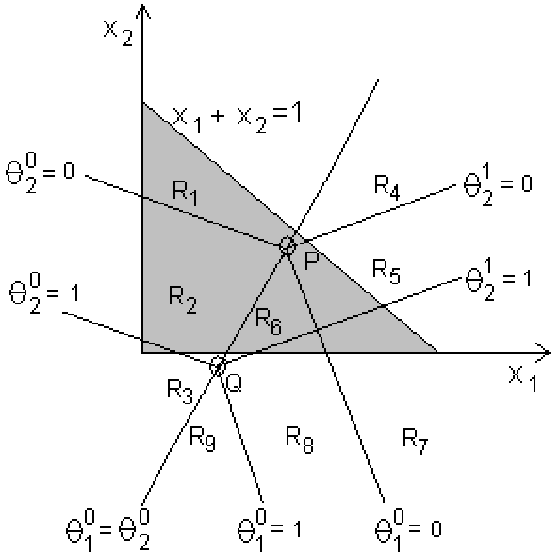

Figure 3.1: Region Partitions

Table 3.1: Partitions characterizing $H_i(\underline{x})$

$\theta_1^0(\underline{x}) \geq \theta_2^0(\underline{x})$		$\theta_1^0(\underline{x}) < \theta_2^0(\underline{x})$		
Region	condition	Region	condition	
R_1	$\theta_2^0(\underline{x}) \leq 0$	R_4	$\theta_2^1(\underline{x}) \leq 0$,	$0 \leq \theta_1^0(\underline{x})$
R_2	$0 < \theta_2^0(\underline{x}) < 1$	R_5	$0 < \theta_2^1(\underline{x}) < 1$,	$\theta_1^0(\underline{x}) \leq 0$
R_3	$1 \leq \theta_2^0(\underline{x})$	R_6	$0 < \theta_2^1(\underline{x}) < 1$,	$0 < \theta_1^0(\underline{x}) < 1$
		R_7	$1 \leq \theta_2^1(\underline{x})$,	$\theta_1^0(\underline{x}) \leq 0$
		R_8	$1 \leq \theta_2^1(\underline{x})$,	$0 < \theta_1^0(\underline{x}) < 1$
		R_9	$1 \leq \theta_2^1(\underline{x})$,	$1 \leq \theta_1^0(\underline{x})$

Table 3.2: Expressions for $H_i(\underline{x})$

R_1		$H_2(\underline{x}) = 1$
R_2	$H_1(\underline{x}) = 0$	$H_2(\underline{x}) = 1 - \frac{p_2-(\beta x_2+\alpha_2 x_1)}{q_2}$
R_3		$H_2(\underline{x}) = 0$
R_4	$H_1(\underline{x}) = 0$	$H_2(\underline{x}) = 1$
R_5	$H_1(\underline{x}) = \frac{(1-\alpha_2)x_1-\beta(1-\alpha_1)x_2+p_2-p_1}{q_2-q_1}$	$H_2(\underline{x}) = 1 - \frac{(1-\alpha_2)x_1-\beta(1-\alpha_1)x_2+p_2-p_1}{q_2-q_1}$
R_6	$H_1(\underline{x}) = \frac{(1-\alpha_2)x_1-\beta(1-\alpha_1)x_2+p_2-p_1}{q_2-q_1} - \frac{p_1-(x_1+\beta\alpha_1 x_2)}{q_1}$	$H_2(\underline{x}) = 1 - \frac{(1-\alpha_2)x_1-\beta(1-\alpha_1)x_2+p_2-p_1}{q_2-q_1}$
R_7	$H_1(\underline{x}) = 1$	
R_8	$H_1(\underline{x}) = 1 - \frac{p_1-(x_1+\beta\alpha_1 x_2)}{q_1}$	$H_2(\underline{x}) = 0$
R_9	$H_1(\underline{x}) = 0$	

3.3.2 Characterizing Adoption Trajectories

By combining eqs. (3.5) to (3.7) with eq. (3.10), explicit expressions can be obtained for $H_i(\underline{x})$ in each of the nine regions (as listed in Table 3.2). Using these expressions, it is now possible to solve eq. (3.3) and characterize the trajectory of technology adoption in each region. The trajectories have the following general form:

$$x_i(t) = a_i + b_i e^{\lambda_1 t} + c_i e^{\lambda_2 t}, \quad i \in \{1,2\} \tag{3.11}$$

where λ_1 and λ_2 can be positive, negative, or complex depending on the region. Individual solutions for each region are listed in Table B.2 of the Appendix.

The full trajectory of technology adoption starting at some initial adoption levels $\underline{x}(0)$ within a given region, can then be obtained by "stitching" together trajectories in individual regions as region boundaries are crossed. The next question is to determine whether and where these trajectories may eventually converge as $t \to \infty$. We tackle this issue next.

3.3.3 Computing Steady-state Equilibria

From eq. (3.11), we see that a technology adoption trajectory in, say, region R_k, converges to a stable equilibrium $x_i(\infty) = a_i, i \in \{1,2\}$, if λ_1 and λ_2 are both negative (equilibrium is locally stable), *and* $(a_1,a_2) \in S \cap R_k$ (the equilibrium is valid, *i.e.,* in the region associated with the trajectory). In other words, solutions to eq. (3.4) $(H_i(\underline{x}^*) = x_i^*, i \in \{1,2\})$, must satisfy *stability* and *validity* conditions to be valid steady-state outcomes of the technology adoption process[17]. The simple nature of eq. (3.4) makes characterizing valid and stable solutions relatively straightforward, albeit tedious. The results are listed in Tables B.3 and B.4 in the Appendix. Table B.3 of the Appendix gives the stability conditions associated with each equilibrium, along with the joint validity and stability conditions (they are inter-dependent) in the last column.

The derivations are mechanical in nature, but we review the implications and properties of their solutions.

First, possible equilibria include instances where one technology wipes out the other while achieving either full $(x_i^* = 1)$ or partial $(0 \leq x_i^* < 1)$ market penetration, and instances where both technologies coexist, again at either full $(x_1^* + x_2^* = 1)$ or partial market penetration $(0 \leq x_1^* + x_2^* < 1)$. Instances where both technologies die-out, *i.e.,*

[17]Our model is well-behaved and instances of boundary fixed points do not arise

$\underline{x}^* = (0,0)$, while possible (the equilibrium lies in regions R_3 or R_9), are absent from Table B.3 (see Appendix), as we restrict our focus to scenarios where Technology 1 survives in the absence of the Technology 2's introduction. This precludes a $(0,0)$ outcome.

Second, although not explicitly indicated in Table B.3 (see Appendix), configurations can be found for which the validity and stability conditions of multiple equilibria are simultaneously satisfied. In other words, depending on the initial conditions $\underline{x}(0)$, technology adoption converges to different outcomes. The following proposition identifies the configurations of multiple equilibria that can simultaneously arise for a given set of parameter values.

Proposition 5 *The only combination of multiple valid and stable equilibria that can coexist are:*

1. $(1,0)$ *and* $(0,1)$
2. $(x^*_{1_{R_8}},0)$ *and* $(0,1)$
3. $(x^*_{1_{R_8}},0)$ *and* $(0,x^*_{2_{R_2}})$
4. $(1,0)$ *and* $(0,x^*_{2_{R_2}})$
5. $(x^*_{1_{R_5}},1-x^*_{1_{R_5}})$ *and* $(0,x^*_{2_{R_2}})$
6. $(x^*_{1_{R_6}},x^*_{2_{R_6}})$ *and* $(0,1)$
7. $(x^*_{1_{R_6}},x^*_{2_{R_6}})$ *and* $(1,0)$

Additionally, no combination of three or more equilibria can coexist as valid and *stable equilibria.*

The proof of the above proposition is available in Appendix B.4. When multiple equilibria arise, the initial market penetration determines the equilibrium to which the adoption process converges. Therefore it is useful to identify the set of all initial market levels, $\underline{x}(0)$, for which the adoption trajectory converges to a particular stable equilibrium. This set is known as the 'Basin of Attraction' of that stable equilibrium. If the stable equilibrium is the only stable equilibrium in the system *i.e.,* globally stable, then the entire region S is its basin of attraction. That is, all starting points lead to the equilibrium. But whenever a pair of stable equilibria coexist, a 'separatrix', demarcating the basins of attraction of the two stable equilibria can be computed. Section 3.3.4 provides the expressions for these separatrices and illustrates the methodology for computing them.

Figure 3.2 provides an illustrative example. The figure, called a phase diagram, shows the path of the diffusion process in the (t, x_1, x_2) space projected onto the (x_1, x_2) plane. In other words, it plots $x_1(t)$ versus $x_2(t)$ and is what one would see if one stood high on the time axis and looked down into the (x_1, x_2) plane, sometimes referred to as the phase plane. We observe that there are two stable steady-state equilibria (of the form $(0, x_2^*)$ and $(x_1^*, 0)$) and an unstable equilibrium in R_6. A separatrix passes through this unstable equilibrium, separating the basins of attraction of the stable equilibria.

The framework developed here can be used in a wide range of situations to model the dynamics of adoption. As an illustration of the useful insights that such a model can

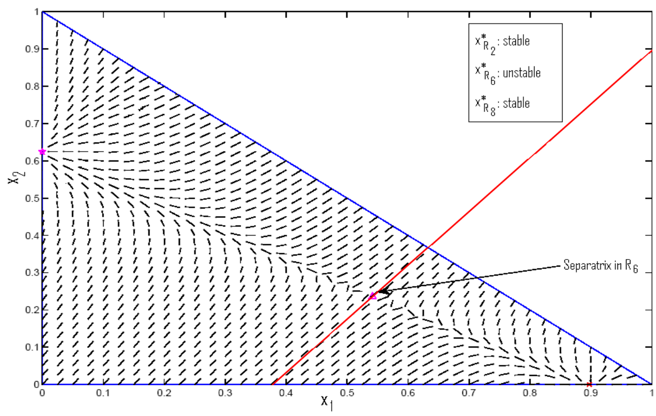

Figure 3.2: Separatix and the Basins of Attraction

$(p_1 = 1.2, q_1 = 2.95, p_2 = 2.54, q_2 = 5.1, \alpha_2 = \alpha_1 = 0.01, \beta = 1)$

offer, we apply our model to studying the role of converters in the adoption of incompatible technologies. We see from Tables B.3 and B.4 of the Appendix that converters can influence (through the parameters α_i) both the validity and the stability of equilibria. In other words, converters may lead technology adoption to an entirely different equilibrium. A rapid inspection of Table B.2 (see Appendix) shows that a similar conclusion holds for trajectories. In particular, converters can affect the values of λ_1 and λ_2 of eq. (3.11). Investigating if and when such changes can happen, is the topic of Section 3.4.

3.3.4 Computing Separatrices

Table 3.3 provides the characterization of the separatrices in each region when the unstable equilibrium that it passes through lies in that region. The regions R_1, R_3, R_7, R_9 are not included in the table as no separatrix can arise in them because by definition these equilibria are always stable whenever they are valid.

Table 3.3: Candidates for local separatrix

Region	Candidate of local separatrix
R_2	$x_2 = -\frac{\alpha_2}{\beta}x_1 + \frac{p_2-q_2}{\beta-q_2}$
R_5	$x_2 = \frac{1-\alpha_2}{\beta(1-\alpha_1)}(x_1 - x^*_{1R_6}) + x^*_{2R_5}$
R_6	$x_2 = \frac{2(1-\alpha_2)(x_1-x^*_{1R_6})}{\alpha_2+\beta(1-\alpha_1)-\frac{q_2}{q_1}+(q_2-q_1)\sqrt{A^2-4B}} + x^*_{2R_6}$
R_8	$x_2 = -\frac{1}{\beta\alpha_2}x_1 + \frac{p_1-q_1}{\beta\alpha_2(1-q_1)}$

We briefly illustrate how the expressions for the separatrices may be derived. Consider the separatrix of Region R_2 passing through the unstable equilibrium $(0, x^*_{2R_2})$. Note that the equilibrium is valid but unstable, and thus $q_2 < \beta$ from Table B.2. From the expressions of trajectory in R_2 it is clear that if $q_2 < \beta$, $x_2(t)$ increases if $c_2 > 0$ while it decreases if $c_2 < 0$. Depending on the sign of c_2, the trajectory diverges in opposite directions. Thus $c_2 = 0$ separates the trajectories with different convergence behaviors and is the candidate for the local separatrix in R_2, which gives

$$x_2 = \frac{p_2 - q_2}{\beta - q_2} - \frac{\alpha_2}{\beta}x_1$$

as the expression for the boundary separating different basins of attraction.

3.4 The Impact of Converters

As we shall see, converters are capable not just of shifting equilibria around; they can also *eliminate* or *create* equilibria. An exhaustive investigation of the full influence of converters, while possible, results in a situation where it is difficult to "see the forest

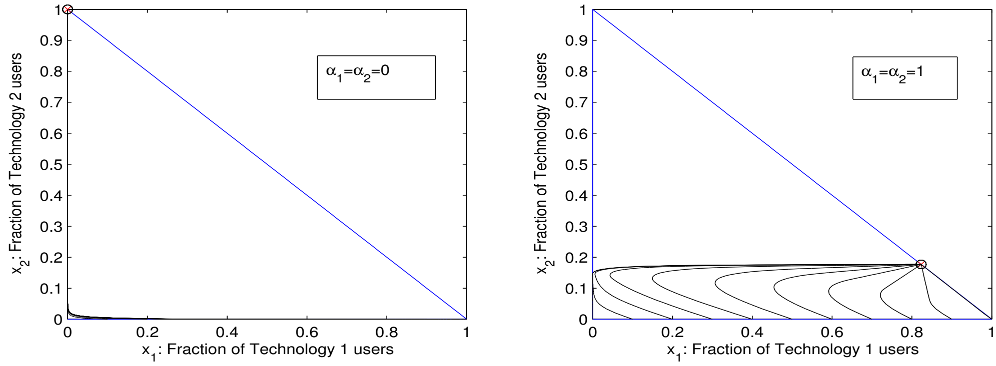

Figure 3.3: On the effect of converters on the existence of equilibria. $(p_1 = 1.01, q_1 = 0.7, p_2 = 2.5, q_2 = 2.51, \beta = 3)$

for the trees." As a result, we focus on what we believe are some of the more revealing and significant effects of converters. We identify the reasons behind these effects, and provide conditions under which they can arise.

The investigation proceeds along the following thrusts: (i) Can converters help a network technology improve its market standing and in particular avoid elimination? (ii) Can improving the efficiency of one's converter hurt a technology? (iii) Can improving the efficiency of one's converter hurt the overall market? and (iv) Can the introduction of converters affect overall market stability? Note that when referring to converters of a particular technology, we mean converters developed by that technology provider to let its users communicate with users of the other technology. This distinction is moot when using symmetric converters, but worth highlighting as the model allows it.

3.4.1 Impact on Adoption Levels

We begin our investigation with a simple numerical example that illustrates how converters can induce drastic changes in the adoption of network technologies. Specifically, consider the scenario of Figure 3.3 that shows two adoption outcomes for the same two network technologies ($p_1 = 1.01, q_1 = 0.7, p_2 = 2.5, q_2 = 2.51, \beta = 3$), with and without converters.

The plot on the left corresponds to a scenario without converters ($\alpha_1 = \alpha_2 = 0$) and in which Technology 2 eventually eliminates Technology 1 and achieves full market penetration[18]. This corresponds to a single, stable equilibrium $(0,1)$. The right hand plot shows how the use of perfect converters results in the elimination of the original $(0,1)$ equilibrium, so that the only possible outcome of technology adoption is now one where both technologies co-exist.

Figure 3.3 answers our question regarding a technology's ability to avoid elimination through the introduction of converters, and thus leading to a new equilibrium adoption outcome. We now state it more formally in the following proposition.

Proposition 6 *Converters can help a technology alter market equilibrium from a scenario where it has been eliminated to one where it coexists with the other technology, or even succeeds in nearly eliminating it.*

The proofs of Proposition 6 and subsequent propositions can all be found in Ap-

[18]Note that this is a scenario in which Technology 1 is marginally competitive, *i.e.,* if left alone it would achieve a relatively low market penetration.

pendix B.4.

As discussed above, Figure 3.3 provides a sample configuration illustrating Proposition 6, *i.e.,* Technology 1 goes from elimination to dominating Technology 2 simply by introducing an efficient converter. Table B.3 (see Appendix) identifies that the equilibrium $(0,1)$ becomes invalid when the converter efficiency of Technology 1 verifies $\alpha_1 > 1 - \frac{p_2 - p_1}{\beta}$. Note that since $0 \leq \alpha_1 \leq 1$, this requires $p_1 < p_2$. Assuming this is the case, the difference between the maximum intra-network benefits of Technology 2 and the maximum cross-networks (through the converter) benefits that the users of Technology 1 derive, becomes equal to the price differential between the two technologies. As a result, low-end users (with small θ values) become indifferent to choosing either technology *i.e.,* $\theta_2^1 = 0$, and any further increase in α_1 leads them to switching to Technology 1. Depending on the values of the other system parameters, it is possible that further increases in α_1 can allow it to nearly eliminate Technology 2. Note that while Technology 1 may succeed in nearly eliminating Technology 2, a small number of users of Technology 2 must remain present to contribute externality benefits to the users of Technology 1. Note also that Figure 3.3 considers symmetric converters and thus the outcome is not one that can be changed by the other technology deploying its own converters. This is a general phenomenon, and most if not all of the results in this section also hold under the constraint of symmetric converters (we will explicitly highlight those that don't).

A similar set of results hold for Technology 2 that, under some conditions, can enjoy the same benefits from converters. The symmetric condition that allows Technology 2 to overcome elimination ($(1,0)$ is now the initial equilibrium), is to introduce a converter whose efficiency α_2 exceeds $\alpha_2 \geq 1 + (p_2 - p_1) - (q_2 - q_1)$. In other words, Technology 2 needs to develop a converter whose efficiency compensates for both the maximum intra-network benefits of Technology 1 and the difference between the price and quality differentials of the two technologies[19]. At that point, $\theta_2^1 = 1$ so that with any further improvement in its converter efficiency, Technology 2 will start attracting some high-end users (large θ values) and eventually re-emerge. As with Technology 1, further improvements in its converter efficiency can in some cases allow Technology 2 to nearly wipe out Technology 1, although again not entirely.

Similar results can also be obtained from Table B.3 for $(x_1^*, 0)$ and $(0, x_2^*)$, *i.e.,* instances when the elimination of a technology does not coincide with full market penetration for the other.

The behavior highlighted by Proposition 6 is relatively common. Consider our earlier example of IPv4 and IPv6. The large IPv4 installed base (or conversely, the small amount of content natively accessible over IPv6) mandates converters (gateways) that allow IPv6-only users[20] to access IPv4 content. Without such converters, IPv6 is unlikely to ever take-off. Conversely, once such converters are in place, it is possible for IPv6 to eventually fully replace IPv4.

[19]The price differential must be lower than the quality differential, *i.e.,* $p_2 - p_1 < q_2 - q_1$, for this to be possible.

[20]Those that have only an IPv6 address once IPv4 addresses have been exhausted.

Proposition 6 focused on a scenario where converters help a technology avoid elimination. Next, we explore whether it is possible for an increase in converter efficiency to actually harm a technology, *i.e.,* lower its market penetration.

Proposition 7 *Technology 1 can hurt its market penetration by introducing a converter and/or improving its efficiency if Technology 2 offers higher externality benefits (*$\beta > 1$*) and the users of Technology 1 are able to access these benefits (*$\alpha_1\beta > 1$*). Furthermore, whenever Technology 1 hurts its own market penetration, it also reduces the overall market penetration. In contrast, Technology 2 can never hurt itself while improving its own converter efficiency.*

Note that the proposition implicitly assumes asymmetric converters, *i.e.,* explores the effect of unidirectional converter introduction or improvement.

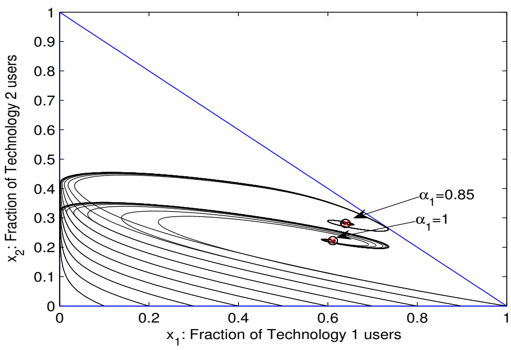

Figure 3.4: Better converters harm Technology 1 and the overall market when α_1 is increased from 0.85 to 1, $(p_1 = 1.3, q_1 = 0.8, p_2 = 2.3, q_2 = 2.4, \alpha_2 = 0.6, \beta = 2.5)$

The following discussion tries to shed light on when and why the outcome of Proposition 7 arises. Intuitively, the original impetus for Technology 1 to improve the efficiency of its converters, is to make itself more attractive to potential users by allowing them to better tap into the (higher) externality benefits of Technology 2. It may then attract new users, either from among those that had not previously adopted any technology or among users of Technology 2 who decide to switch to Technology 1. It is the acquisition of the latter type of users that can prove harmful to Technology 1. Specifically, because $\alpha_1\beta > 1$, the switching of users from Technology 2 to Technology 1 negatively affects the externality benefits of all Technology 1 users. When β is high, the decrease in externality benefit can be significant. As illustrated in Figure 3.5, the result of this decrease can be that some low-end (small θ) users decide to leave Technology 1 and exit the market. When the influx of new users is less than the outflow, the overall penetration of Technology 1 decreases. Figure 3.4 shows an instance of such a decrease. Additionally, the same reasoning shows that this also results in a decrease in overall market penetration (both x_1 and x_2 decrease).

This behavior can arise in the earlier example of competing HD and SD video services, as it satisfies the requirement that $\alpha_1\beta > 1$. Specifically, although SD users are limited to generating SD quality videos, through converters they can receive and enjoy the higher-quality of HD videos. As a result, they will be negatively affected by any move of HD users back to SD. This can in turn lead some SD users to disadopt the service altogether. Hence, lowering their own user base and the overall market penetration of both services.

When $\beta \geq 1$, it is easy to see that the argument used for Technology 1 does not hold

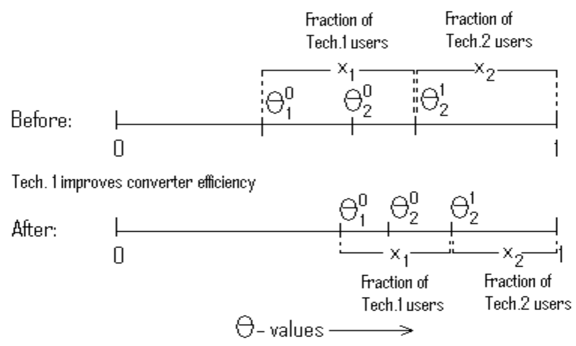

Figure 3.5: Technology 1 hurts itself as well as the overall market penetration.

for Technology 2, *i.e.,* acquiring a customer from Technology 1 will never decrease the externality benefits of Technology 2 users, so that it cannot experience such a reversal when improving its own converter. A proof that this property actually holds for all values of β, *i.e.,* even when $\beta \leq 1$ is provided in Appendix B.4.

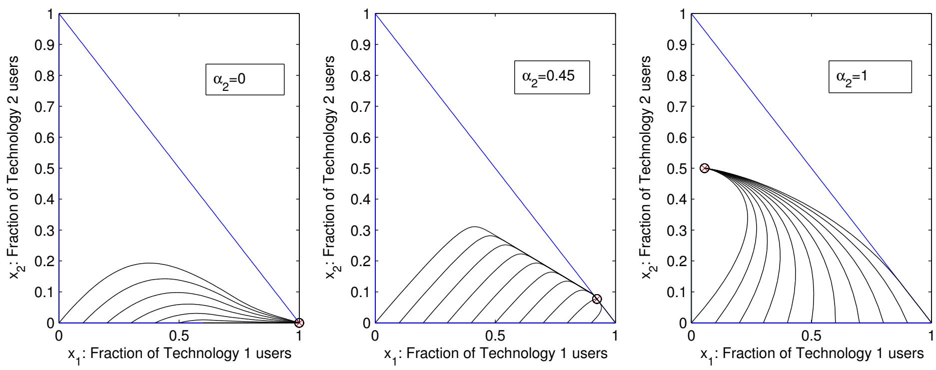

Figure 3.6: Greedy Technology 2 harms overall market penetration. $(p_1 = 0.9, q_1 = 1.9, p_2 = 2.7, q_2 = 4.3, \alpha_1 = 0, \beta = 1.2)$

Proposition 7 indicated that Technology 1 could not only hurt itself through better converters, but also the overall market penetration. The next proposition investigates the negative impact of converters on overall market penetration, and formally identifies conditions under which this takes place.

Proposition 8 *Both technologies can hurt overall market penetration through better converters. Technology 2 can have such an effect only when* $\alpha_1\beta < 1$, i.e., *Technology 1 users derive lower externality benefits from connecting to Technology 2 users than to their peers. Conversely, Technology 1 demonstrates this behavior only when* $\alpha_1\beta > 1$, i.e., *its users derive greater externality benefits from connecting to Technology 2 users than to their peers.*

As the discussion of Proposition 7 already highlighted how this could occur with Technology 1, we focus instead on Technology 2. The motivation for better converters remains the same, namely, allow users of Technology 2 to derive higher externality benefits by connecting to users of Technology 1. This improvement in the externality benefits of Technology 2 leads some users (those close to the θ_2^1 boundary) to switch. When $\alpha_1\beta < 1$, the migration of those users from Technology 1 to Technology 2 translates into a net drop in the overall utility Technology 1 offers its remaining users (the externality benefits contributed by every user that migrates goes down from a relative weight of 1 to one of $\alpha_1\beta < 1$). This decrease in Technology 1 value then leads some low valuation users (small θ) to drop out altogether. Technology 1 fails to provide them with enough externality benefits to justify even its low cost, while Technology 2 remains too expensive for them. This brings the overall market penetration down.

It is interesting to note that the use of converters by Technology 2 can effectively force low valuation users to leave the market. This may not be desirable and suggests the possibility of policy interventions- regulations and/or market mechanisms- to offer low valuation users alternatives that allow them to stay in the market. This can include increasing the attractiveness of Technology 1 (*e.g.,* subsidizing improvements in its converter efficiency), or by asking Technology 2 to provide a low-tier, low-cost version of its service that caters to low valuation users.

For a real world instance where the conditions of Proposition 8 could be satisfied, consider again the IPv4 vs. IPv6 scenario for which $\beta \approx 1$, and assume that IPv6 has taken off but that providers serving low valuation customers have not bothered with converting them to IPv6. If the converters that allow these legacy IPv4 users to access the now increasingly IPv6-only content are of low quality, it is possible that some of them will, if not drop their IPv4 service, at least significantly reduce their usage.

Figure 3.6 provides an example. In this configuration, in the absence of converters, Technology 1 had reached full market penetration. When Technology 2 introduces a converter of efficiency $\alpha_2 = 0.45$, it emerges and both technologies coexist at equilibrium, while still achieving full market penetration. If the efficiency of Technology 2 converter further improves, it still sees a rise in its own market penetration, but the overall market penetration now decreases to $\approx 55\%$, as low valuation users drop out.

3.4.2 Impact on Adoption Dynamics

The previous sub-section explored the effect that converters can have on equilibria. In this sub-section we extend the investigation to both trajectories and equilibria. In particular, we concentrate on an unexpected effect of converters, one that can be shown not to be possible in their absence, namely, the possibility that the introduction of converters can render the process of technology adoption unstable. In the next proposition, we specify the conditions under which it can arise.

Proposition 9 *The introduction of converters can create "boom and bust" cycles in the technology adoption process. This behavior arises only when Technology 2 exhibits higher externality benefits (*$\beta > 1$*) than Technology 1 and the users of Technology 1 are unconstrained in their ability to access these benefits (*$\alpha_1\beta > 1$*).*

Conversely, the next corollary establishes that this never occurs in the absence of converters. The proofs are again in Appendix B.4.

Corollary 3.4.1 *In the absence of converters, technology adoption trajectories always converge to a stable equilibrium.*

Before trying to offer some insight into the emergence of instabilities when converters are introduced, we offer an example to illustrate the type of outcomes that can arise.

Figure 3.7 provides a sample scenario of converters affecting the stability of technology adoption, and in particular introducing cycles in the adoption trajectories. The left-

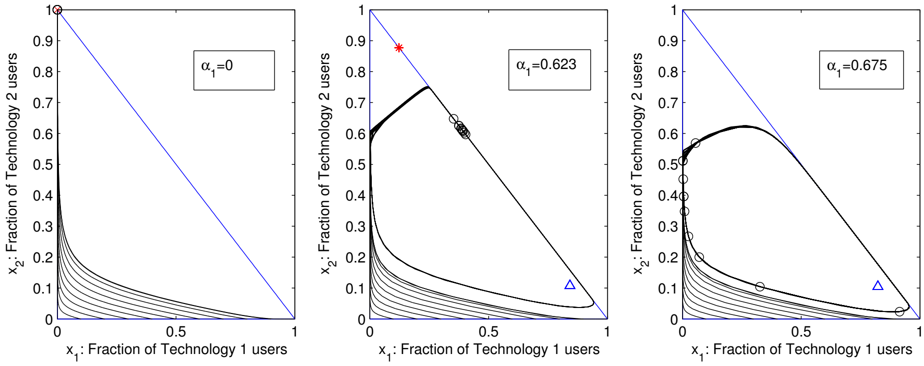

Figure 3.7: Effect of converters on adoption stability $(p_1 = 1.05, q_1 = 0.4, p_2 = 2.1, q_2 = 2.11, \alpha_2 = 0.3, \alpha_1 = 0.675, \beta = 2.8)$

hand-side of the figure shows how in the absence of converters, Technology 2 displaces Technology 1 and achieves full market penetration. The introduction of a reasonably efficient converter $(\alpha_1 \approx 0.623)$ by Technology 1, however, drastically changes the situation by introducing two new equilibria; both of them unstable (middle diagram). As a result, while the original equilibrium of $(0,1)$ remains valid, its basin of attraction has now shrunk considerably . Instead, under most initial conditions, a cyclical pattern of adoption decisions emerges. In other words, users repeatedly switch back and forth between the two network technologies. Matters only become worse if the efficiency of the converter of Technology 1 continues improving[21], and with a perfect converter the original equilibrium of $(0,1)$ has all but disappeared and only one, unstable equilibrium remains around which adoption decisions keep circling.

The intuition behind the emergence of such a situation is somewhat similar to that of a technology harming itself and/or the overall market through the introduction of

[21] As mentioned before, similar situations arise under symmetric converters.

better converters. Specifically, consider an instance where Technology 2 offers higher externality benefits that users of Technology 1 can tap into if a converter is available. When converters are absent, users that value the higher quality of Technology 2 adopt it (when it offers a higher overall utility), eventually leading to full adoption as shown on the left most part of Figure 3.7. However, once a converter is introduced, users have the option to remain with Technology 1 (and enjoy its lower price) without forfeiting all the benefits of Technology 2, and in particular its externality benefits. As a result, while Technology 2 will initially still gain market share by attracting high technology valuation users away from Technology 1, this now happens with Technology 1 also gaining new customers (low technology valuation customers are now adopting because of the externality benefits accessible through the new users who joined Technology 2). This combined effects results in a steady increase in overall market share until a limit is reached. This limit corresponds to a point where Technology 2 has tapped out all the high technology valuation users it could attract. As Technology 2 growth tapers off, Technology 1 continues growing as it still attracts new low technology valuation customers. Continued growth in Technology 1 customer base eventually makes it attractive to some mid-range technology valuation customers that start switching back to it. This fuels an accelerated growth in the user base of Technology 1 that now acquires customers from both Technology 2 and non-adopters. This continues until the user base of Technology 2 becomes so small that it starts affecting the ability of Technology 1 to grow. At this point, both technologies start loosing customers. This ends when the customer base of Technology 1 is small enough to allow Technology 2 to again start attracting customers (its own customer base had by then all but disappeared), and the process repeats anew.

To illustrate this behavior in a less abstract setting, we return to the example of HD and SD video conference services of Section 3.2. The higher quality HD service when introduced attracts high-valuation users, who switch over from the existing SD service. This eventually results in a new market equilibrium. If the SD service responds to this competition by introducing its own converter, it will entice some (low-valuation) non-adopters to adopt, as they now have access to the higher benefits of viewing other users in HD quality. As the number of SD users grows, the technology attracts back some of the lesser valuation HD adopters, because of its lower cost and increased externality benefits (from its larger user base and access to HD users). This results in an increase in SD adoption level and a corresponding drop in HD's. However, as the switching from HD to SD continues, the drop in the number of HD users lowers the overall externality benefits available to SD users. Consequently, the lowest valuation SD users begin to disadopt. This decrease in the number of SD users, and therefore the externality benefits that the SD service affords, makes the higher valuation SD users switch back to the HD service. This creates a situation where SD adoption drops, while the HD service grows. As before, the growth of the HD service eventually draws low-valuation users (non-adopters) to the lower-priced SD service. The two services then grow until SD's user population has once again grown large enough to attract the lesser valuation HD users.

At which point the cycle repeats anew.

3.5 Robustness to Alternative Models

In the work we identified several interesting behaviors that arise in presence of converters for a model where users are heterogeneous in their evaluation of the technology's quality and benefit from linear externality. However, many of the behaviors identified in this work will be present in a wide range of alternative models.

To show this robustness of our findings we first show that quantitatively similar behaviors arise for more generic distribution of heterogeneous user preferences. We do so by considering Beta-distribution of the user preference with positive and negative skewness. Following this, we also consider some different types of network externality benefits, namely, sub-linear, super-linear and logarithmic network benefits and provide examples that demonstrate similar behaviors as well. We also consider the case where user heterogeneity is extended to network benefits in addition to the technology's quality. Again for all these scenarios we present illustrative examples for qualitatively similar behaviors of interest.

3.5.1 Non-Uniform Distribution of User Heterogeneity

We consider the same user utility functions as those in eqs.(3.1) and (3.2), but the heterogeneous user preferences are assumed to follow a Beta-distribution in $[0, 1]$ as opposed to an Uniform distribution. The density of Beta-distribution is given by

$\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}$ where $B(a,b)$ is the beta function with parameters a and b. Its skewness is given by $\frac{2(b-a)\sqrt{a+b+1}}{(a+b+2)\sqrt{ab}}$. We show qualitatively similar behaviors for both beta distributions with positive and negative skewness in the following cases.

(i) Positively skewed Beta-distributions

Figure 3.8 shows that instabilities can arise even with such alternative distributions where $a = 1.45, b = 2$ with a positive skewness of 0.25.

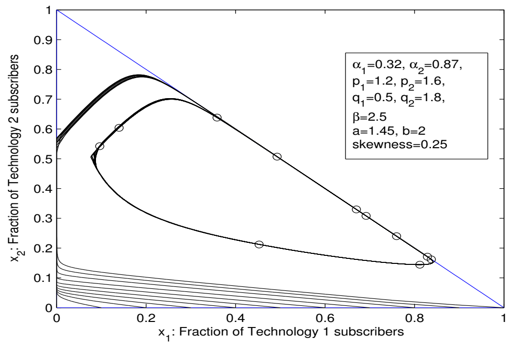

Figure 3.8: Positively skewed Beta-distribution showing Instability.

Figure 3.9 shows that Technology 1 can hurt itself as well as the overall market penetration as it improves its converter efficiency from 0.85 to 1. For the scenario shown in the figure, the beta distribution has $a = 0.65, b = 1$ and a positive skewness of 0.3872.

(ii) Negatively skewed Beta-distributions

Figure 3.10 shows that instabilities can arise even for negatively skewed (-0.3205) beta-distributions where $a = 2.2, b = 1.45$.

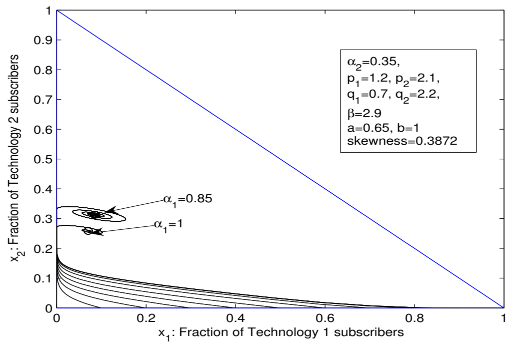

Figure 3.9: Technology 1 hurts itself and overall market for a positively skewed Beta-distribution.

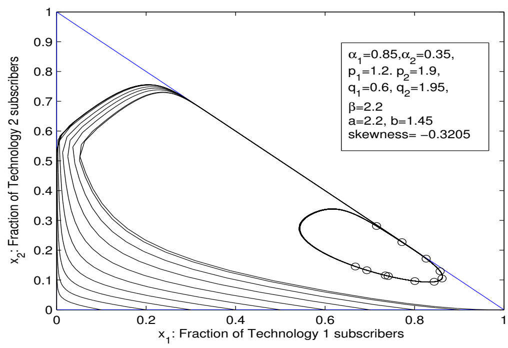

Figure 3.10: Negatively skewed Beta-distribution showing Instability.

Figure 3.11 shows that Technology 1 can hurt itself as well as the overall market penetration as it improves its converter efficiency from 0.85 to 1 as shown in the figure with a beta distribution for parameters $a = 2, b = 1.5$ and a negative skewness of -0.2227.

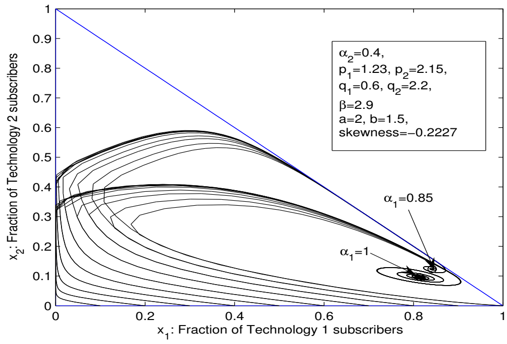

Figure 3.11: Technology 1 hurts itself and overall market for a negatively skewed Beta-distribution.

3.5.2 Non-linear Network Externality Effects

$$U_1 = \theta q_1 + ({x_1}^{\rho} + \alpha_1 \beta {x_2}^{\rho}) - p_1 \tag{3.12}$$

$$U_2 = \theta q_2 + (\beta {x_2}^{\rho} + \alpha_2 {x_1}^{\rho}) - p_2 \tag{3.13}$$

(i) Sub-linear network benefits

The plot on the left in Figure 3.12 shows an example of instability in the adoption process when the externality benefits are of the form ${x_i}^{0.7}, \ i = \{1, 2\}$. The plot on the right provides an example where Technology 1 on improving its converter efficiency from 0.85 to 0.9 hurts its own market as well as the overall market levels across the two technologies.

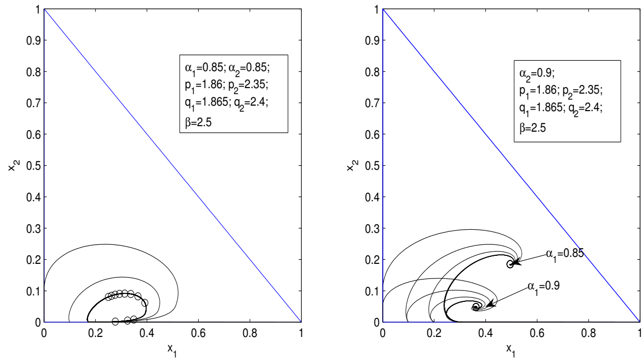

Figure 3.12: Effects of sublinear network benefits.

(ii) Super-linear network benefits

The plot on the left in Figure 3.13 shows an example of instability in the adoption process when the externality benefits are of the form ${x_i}^{1.4},\ i = \{1,2\}$. Once again, even for the superlinear externalities, we find that the plot on the right shows that Technology 1 can potentially harm itself as well as the overall market. In this figure, the Technology 1 improves its converter efficiency from 0.87 to 0.91 leading to a drop in its own market by about 0.16.

(iii) Logarithmic network benefits

We also considered the case where the externality benefits are of the form $\log_2(x_i +$

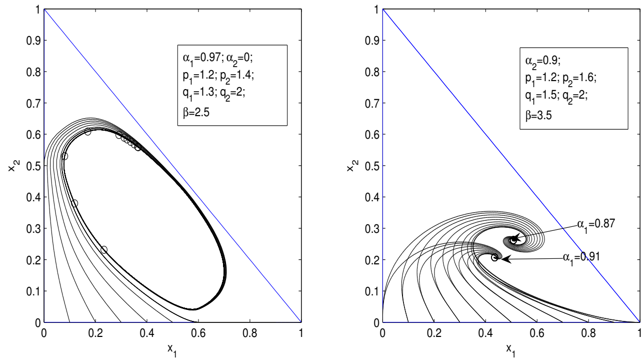

Figure 3.13: Effects of superlinear network benefits.

1), $i = \{1,2\}$.

$$U_1 = \theta q_1 + (\log_2(x_1+1) + \alpha_1\beta\log_2(x_2+1)) - p_1 \tag{3.14}$$

$$U_2 = \theta q_2 + (\beta\log_2(x_2+1) + \alpha_2\log_2(x_1+1)) - p_2 \tag{3.15}$$

The plot on the left in Figure 3.14 shows an example of the instabilities that arise in presence of converters in case of a logarithmic externality function. Technology 1 may again hurt itself as well as the overall market while improving its converter efficiency as shown on the plot on the right. Such a behavior is shown in this plot as Technology 1 improves its converter efficiency from 0.9 to 0.99.

3.5.3 User Heterogeneity in Intrinsic Quality and Externality Benefits

If the users have similar heterogeneous preferences over both the intrinsic (stand-alone) quality of the technology and the network externality benefit, then their utility from the

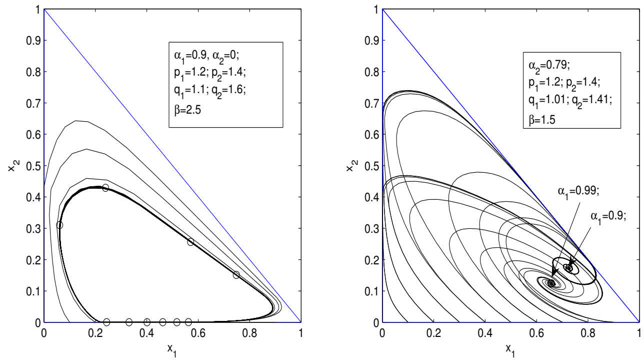

Figure 3.14: Instability for logarithmic network benefits.

two alternative technologies are:

$$U_1 = \theta(q_1 + x_1 + \alpha_1 \beta x_2) - p_1 \tag{3.16}$$

$$U_2 = \theta(q_2 + \beta x_2 + \alpha_2 x_1) - p_2 \tag{3.17}$$

For this utility form, we again identify that deployment and improvement in converters can lead to behaviors like drop in overall market penetration and adoption instability.

(i) Drop in overall market penetration

The plot on the right in Figure 3.15 considers a case where $p_1 = 0.6, p_2 = 3.9, q_1 = 0.5, q_2 = 4.2, \beta = 7, \alpha_2 = 0$ *i.e.,* Technology 1 is cheaper and lower in quality than Technology 2, which also provides larger network benefits. In this plot, when Technology 1 introduces a converter of 0.45 efficiency then the overall market penetration at equilibrium is about 0.54. However if the first technology introduces a converter, it improves its own market but the overall penetration drops to 0.3737. Therefore even for this utility form, the overall market penetration across the two technologies can be hurt by the converters.

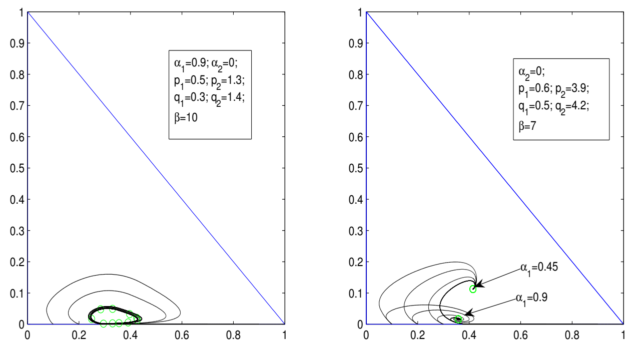

Figure 3.15: Drop in overall market penetration.

(ii) Creation of instability in adoption process

We also find that instabilities may be arise in the adoption process when converters are present. The plot on the left in Figure 3.15 shows such a scenario for $\alpha_1 = 0.9, \alpha_2 = 0, p_1 = 0.5, p_2 = 1.3, q_1 = 0.3, q_2 = 1.4, \beta = 10$. Thus the behavior for instability in adoption dynamics may also arise for this alternative utility form.

(iii) Special Case: User's only value network benefits

$$U_1 = \theta q_1(x_1 + \alpha_1 \beta x_2) - p_1 \tag{3.18}$$

$$U_2 = \theta q_2(\beta x_2 + \alpha_2 x_1) - p_2 \tag{3.19}$$

In this case as well we find instances where deployment of converters can hurt the overall market. Figure 3.16 illustrates such a scenario.

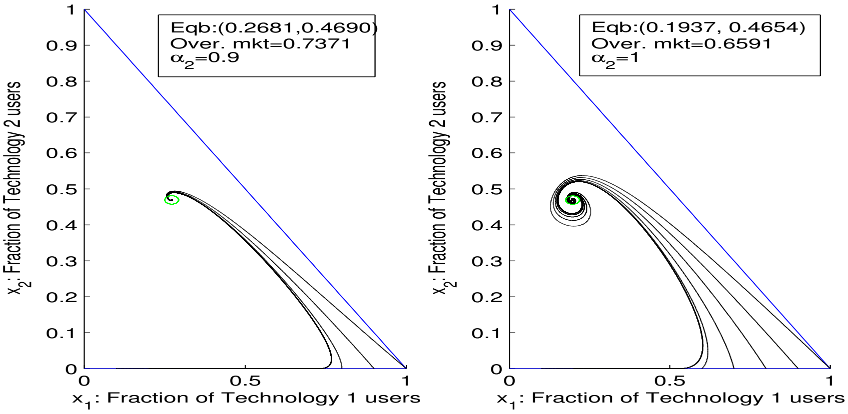

Figure 3.16: Drop in overall market penetration. ($\alpha_1 = 0.185, p_1 = 0.4, p_2 = 1.8, q_1 = 4.65, q_2 = 7.3, \beta = 0.7$)

3.5.4 Presence of Switching Costs

Network technologies often try to introduce switching costs by implementing contracts with penalty and by developing 'lock-in' strategies. For example, ISPs such as AOL practice 'lock-in' by restricting their users to send instant messages only to other fellow subscribers, thus preventing a user who switches to another ISP from messaging to his/her previous network. Another such strategy is the lowering of provider specific email addresses (e.g., Comcast, AOL). In spite of such efforts, the annual customer turnover in the ISP market remain very high at above 72%, suggesting that consumers in the ISP market still have sufficiently low switching costs [17, 59]. In other online markets, the use of 'lock-in' strategies based on proprietary IT are also on the decline due to the emergence of web browsers and technologies like XML that lower switching costs by allowing interoperability between disparate systems [17]. However, some amount of switching costs may indeed continue to exist through contractual commitments, learning curve, specialized formats and customer loyalty programs.

Our model may be extended to include such switching costs, but it will introduce non-trivial complexity in the modeling effort because of the “memory” it adds to the individual users behavior. Additionally, there are different possible types of switching cost configurations one may need to consider, each requiring different utility forms for user decision. For example, the learning costs may significantly affect non-adopters when they join a technology, while contract- breaking cost affect only the users who disadopt or switch technologies. Also, the lack of a clear answer as to when and how many times such costs can affect switching behavior, adds complexity from a modeling standpoint. Therefore, a general analytical solution with switching costs quickly becomes intractable. However, it is possible in some cases to formulate generalized expressions for the indifference thresholds introduced in this work and derive results in the way outlined next.

Assume that when a user switches from one technology to the other, or becomes a non-adopter, he/she incurs a symmetric switching cost of S due to prior contractual commitments, and that the learning costs for all users are negligible ($L = 0$). Then utilities U_1 and U_2 for the current non-adopters remain the same as before and so does the corresponding expressions for the indifference points (cite Eqns.). The utilities of the current adopters of Technology 1 become

$U_0 = -S$

$U_1 = \theta q_1 + (x_1 + \alpha_1 \beta x_2) - p_1$

$U_2 = \theta q_2 + (\beta x_2 + \alpha_2 x_1) - (p_2 + S)$

which give the indifference points as

$\theta_1^0(1) = \frac{p_1 - S - (x_1 + \alpha_1 \beta x_2)}{q_1}$

$\theta_2^0(1) = \frac{p_2 - (\beta x_2 + \alpha_2 x_1)}{q_2}$

$\theta_2^1(1) = \frac{(1-\alpha_2)x_1 - \beta(1-\alpha_1)x_2 + p_2 - p_1 + S}{q_2 - q_1}$

Similarly, the utilities of the current adopters of Technology 2 will be

$U_0 = -S$

$U_1 = \theta q_1 + (x_1 + \alpha_1 \beta x_2) - (p_1 + S)$

$U_2 = \theta q_2 + (\beta x_2 + \alpha_2 x_1) - p_2$

which give the indifference points as

$\theta_1^0(2) = \frac{p_1 - (x_1 + \alpha_1 \beta x_2)}{q_1}$

$\theta_2^0(2) = \frac{p_2 - S - (\beta x_2 + \alpha_2 x_1)}{q_2}$

$\theta_2^1(2) = \frac{(1-\alpha_2)x_1 - \beta(1-\alpha_1)x_2 + p_2 - p_1 - S}{q_2 - q_1}$

Note that the indifference points will now need to be represented as as $\theta_j^i(k)$, where the additional index k, $k = \{0, 1, 2\}$ will be used to represent the user category. Additionally, for any values of $x_1(t)$ and $x_2(t)$ and $S > 0$, the set of indifference points for the above three categories of users must satisfy the following relationships:

$$\theta_1^0(1) < \theta_1^0(0) = \theta_1^0(2)$$

$$\theta_2^0(2) < \theta_2^0(0) = \theta_2^0(1)$$

$$\theta_2^1(2) < \theta_2^1(0) < \theta_2^1(1) \quad (3.20)$$

Therefore if at time t, the set of indifference points are represented by $\theta_j^i(t)$, the new values at $t+1$ *i.e.,* $\theta_j^i(t+1)$ will have to calculated based on the relative positions of $\theta_j^i(k)(t+1)$ with respect to $\theta_j^i(t)$. In other words,

$$\theta_j^i(t+1) = F(\theta_j^i(k)(t), \theta_j^i(t)),\ for\ k = \{0,1,2\},\ i,j = \{0,1,2\},\ i < j,\ i \neq j.$$

where the function F needs to be carefully determined by considering the possible arrangements of these indifference points. For example, in the above case if we consider that the initial arrangement of the indifference point followed the order $\theta_1^0(t) < \theta_2^0(t) < \theta_2^1(t)$, then the function F for the new position of the indifference point $\theta_2^1(t+1)$ (based on relationships in eqn.(3.20)) will be given by:

$$\begin{aligned} \theta_2^1(t+1) &= \theta_2^1(2)(t+1)\ if\ \theta_2^1(2)(t+1) \geq \theta_2^1(t) \\ &= \theta_2^1(1)(t+1)\ if\ \theta_2^1(1)(t+1) \leq \theta_2^1(t) \\ &= \theta_2^1(t) \qquad if\ \theta_2^1(2)(t+1) < \theta_2^1(t) < \theta_2^1(1)(t+1) \end{aligned}$$

Given the complexity of the cases, solutions that account for switching costs need to resort to either numerical solutions (when it is possible to generalize equations for indifference thresholds) or simulations. We have investigated using both approaches to demonstrate that the results that our analytically tractable simplified model allow us to explicitly identify, namely the possible presence of instability in technology adoption and that better converters can hurt the incumbent as well as the overall market level etc., remain present across different switching cost configurations. Figure 3.17 and Figure 3.18 show these behaviors using the numerical solution. For clarity of the plots only the adoption paths for initial penetration levels of $x_1 = 0.5, x_2 = 0$ in Fig. 3.17 and $x_1 = 1, x_2 = 0$ in Fig. 3.18 are shown.

In the next section we provide more evidence of these behaviors under different cost configurations (learning cost, contract-termination cost etc.) through simulation results to establish that the results are robust to the introduction of a broad range of switching costs.

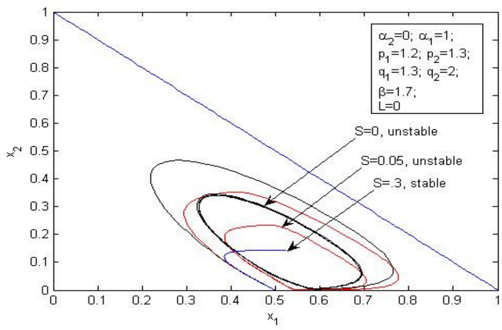

Figure 3.17: Instability in presence of Switching Costs

3.5.4.1 Simulation Results

We consider three types of configurations for the purpose of our simulation to show the robustness of the observed behavior under different switching cost configurations. The simulations consider a population size of $N = 500$, each with a type value θ that is

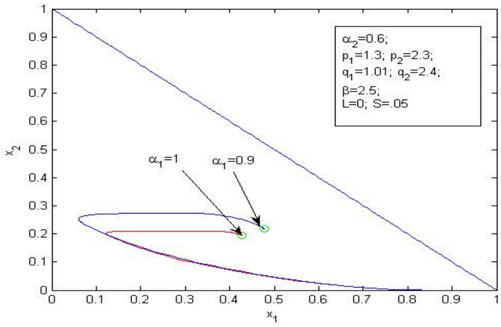

Figure 3.18: Market Level drops in presence of Switching Costs

uniformly distributed between 0 and 1. The plots for that (i) presence of instability and (ii) incumbent's converter hurting itself as well as the overall market, are shown only for initial penetration levels of $x_1 = 0.3, x_2 = 0$ and $x_1 = 1, x_2 = 0$ respectively for the purpose of clarity.

Case (A): Switching Cost due to Contract breaking

In this scenario we consider that a user of either technology who decides to become a non-adopter or switches to the other technology incurs a certain cost as penalty for breaking a contract. It is assumed in this case that there is no learning cost for the users. The instability plot in Fig 3.19 shows the sample diffusion trajectories for switching cost of $S = 0.05$. But as the switching cost increases, the outcome stabilizes (*e.g.,* , $S = 0.3$) since the high switching cost makes it difficult for users to infinitely switch back and forth between the two technologies. Fig 3.20 shows the drop in the overall market and incumbent's market as its converter efficiency is increased from 0.9 to 1.

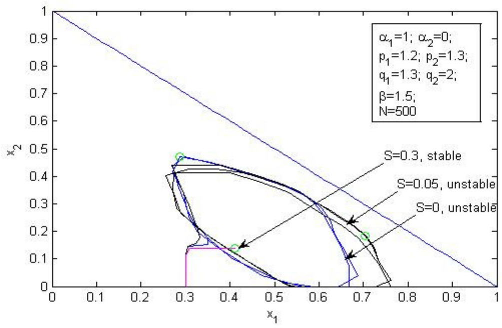

Figure 3.19: Instability in presence of Switching Costs

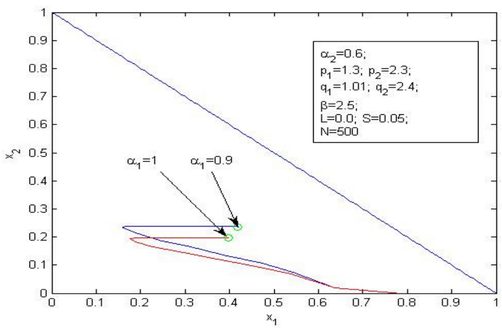

Figure 3.20: Market Level drops in presence of Switching Costs

Case (B): Switching Cost due to Contract breaking & Learning Costs for new adopters

This scenario considers both the presence of learning and switching costs. A non-adopter incurs a learning cost of *L* on joining either technology while an existing user of a technology incurs a switching cost of *S* for either becoming a non-adopter or switching over to the competing technology. As before, Figures 3.21 and 3.22 demonstrate the presence of the interesting behaviors.

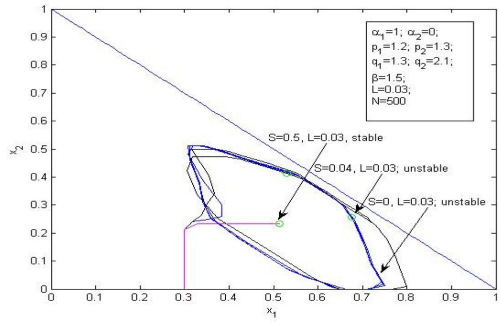

Figure 3.21: Instability in presence of Learning & Switching Costs

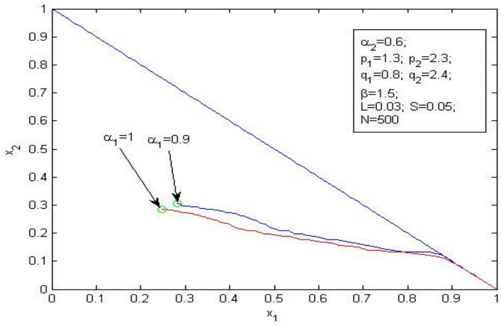

Figure 3.22: Market Level drops in presence of Learning & Switching Costs

Case (C): Switching Cost due to 'Lock-In' but no contract breaking costs

In this scenario we consider a case where a user only incurs a switching cost when he/she has to move from one technology to its competitor but not if he/she becomes a non-adopter. This situation arises mainly when the switching cost is not in the form of a contract but due to 'lock-in' strategies. For example, if a user of a online music service with customization options decides to migrate to a competing site, he/she incurs a switching cost due to 'lock-in', but however if the person gets disinterested in the technology and becomes a non-adopter he/she does not incur this cost. Again, Figure 3.23 and 3.24 provide examples of the noticed behaviors for this scenario.

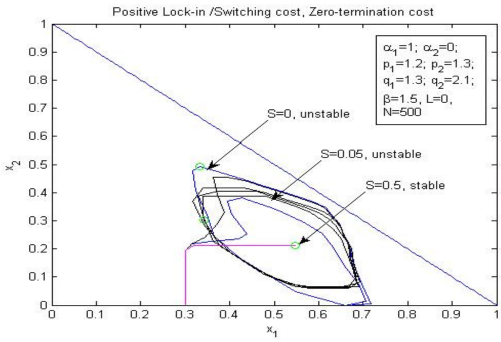

Figure 3.23: Instability in presence of only 'Lock-in' Costs

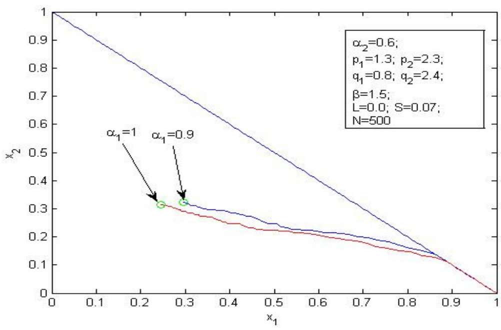

Figure 3.24: Market Level drop in presence of only 'Lock-in' Costs

These simulation results therefore demonstrate that the results we identified with the help of our simplified model are quite robust, and they provide insights into the possible interesting adoption behaviors that can arise in a wide variety of switching cost configurations.

3.6 Related work

Modeling the diffusion of new products and technologies has a long tradition in marketing. Fourt and Woodlock [27] first proposed a product diffusion model in which a fixed fraction of consumers who have not yet bought the product do so at every period; this is known as the constant hazard rate model. Bass [6] proposed an extension that additionally incorporates word-of-mouth communication between current adopters and potential adopters. A large body of work has since built on these earlier models (see [56] for an overview of this literature). Although most of the literature deals with single-product settings, Norton and Bass [63] study the joint diffusion of successive generations of technologies. Their model belongs to a class of substitution models that assume that the newer generation eventually replaces the earlier generation and thus their interest is only in the time it takes for this to occur. Significantly, both single-product and multiple-generation diffusion models focus on aggregate adoption dynamics without explicitly modeling individual decision-making processes. The advantage of such an approach is that it results in relatively simple diffusion models that can, in turn, be used to study dynamic policies (*e.g.,* dynamic pricing). Unfortunately, these aggregate models do not shed sufficient light on the decision processes that lead to the emergent system dynamics or the exact mechanism through which various decision variables (pricing, quality, advertising, etc.) impact adoption decisions.

A few models have focused on individual-level adoption (*e.g.,* [39]). These models provide far greater insight into the mechanism through which rational individual decision-making results in aggregate system dynamics. Given the complexity of these models, much of the progress to date has been in settings with a single technology. In contrast, the adoption of new network technologies is often influenced by incumbents. Moreover, all of the above models and indeed much of the literature refers to generic durables, *e.g.,* washing machines. Such models do not account for the unique features of network technologies, including network externalities and the role of converters.

A recent stream of work in economics has studied the role of network externalities on equilibrium adoption of standards and technologies. Cabral [11] develops a model of individual decision-making in the presence of network externalities and characterizes the aggregate adoption dynamics. He shows that network externalities are potential drivers of S-shaped diffusion curves. We build on Cabral's model but differ in our focus by considering a two-technology setting. Put another way, we are interested in the diffusion of a new network technology in the presence of an incumbent. A related paper by Farrell and Saloner [23] evaluates the impact of an installed base on the transition to a new standard. They show that the installed base can cause "excess inertia" which prevents the transition to the new standard. At the same time, the adoption of the new standard by a few users can create “excess momentum” as well. In their model, users are homogeneous except for the time of their arrival into the system. As a result, they observe a bandwagon effect in which the adoption of a standard by one set of users makes the same choice more attractive to all other users. Thus, one standard always wins and coexistence is not feasible. Choi [14] extends the model by Farrell and Saloner [23] to include converters and shows that converters can in some instances blockade the transition by weakening the threat of being stranded for users of the incumbent technology. In a more recent study, Joseph et al. [44] also show that increase in efficiency of a converter can hinder the adoption of a new network architecture.

An important distinction of our work relative to these papers is that we incorporate heterogeneity in user preferences. We show that this gives rise to equilibria in which the technologies may coexist, *i.e.,* neither network technology fully captures the market. Further, very little attention is paid to the adoption path in these papers because all users make the same decision. In contrast, we show that the heterogeneity across users can result in interesting adoption dynamics including non-monotonic evolution of the market shares of the technologies. Additionally, these papers focus on environments in which users make the decisions to adopt the converters. This is meaningful in environments in which the converter functionality and its deployment resides with individual users, *e.g.,* converters for two incompatible software applications that a user decides to download. In contrast, our interest is in environments in which converters are usually deployed by the technology providers upon incurring high fixed costs, and in the process made available to all its users.

3.7 Conclusion and Extensions

The work provides a framework to study the adoption and diffusion of a new network technology in the presence of an incumbent and offers insight into the role of converters. Our model accounts for both externalities and user heterogeneity, and helps reveal several unexpected behaviors. Of note are that the presence of converters can hurt overall market penetration, and that under certain conditions they can preclude the adoption process from ever converging. In Section 3.5, we showed that these behaviors remain present across a wide range of utility models that differ from the one used for analytical tractability in this paper. These robustness tests consider nonlinear externality functions, non-uniform distribution of user preferences, user heterogeneity in both standalone and network benefits and switching costs.

As the first step of our investigation in the dynamics of technology adoption in the presence of converters, the model clearly has limitations that can be addressed in the future. As mentioned earlier, allowing some of the system parameters to be time-varying is of obvious interest. Similarly, letting prices be endogenous and/or dynamic variables that are chosen by strategic service providers is another direction we have started investigating. This work represents an initial framework towards understanding adoption dynamics of network technologies. Further research building on this work would likely provide additional insight.

Chapter 4

Network Platforms: Functionality-rich Versus Minimalist Design

4.1 Introduction

Platforms provide foundations on which services or products can be developed that deliver value to both their users and the providers that develop them. In general, platforms succeed based on their ability to "connect" *consumers* of applications and services to the *developers* of those applications and services. The Internet and Android offer two recent examples of (network and operating system) platforms whose success largely comes from their ability to connect users and service/application providers. Platforms are, therefore, commonly studied within the framework of *two-sided markets* [74]. The platform is the 'market' and customers and applications/services developers are the 'two sides' of the market. This work follows in this tradition.

The specific issue this work is concerned with is the level of *functionality* that the platform should offer. A platform typically provides a set of capabilities through built-in APIs, modules, tools, etc., which make it easier for developers to innovate new applications and services of interest/value to consumers. This, however, comes at a cost to the platform, and this cost grows with the number of features offered. The question for the platform provider is then to determine the number of features that maximizes profits. A minimalist platform has a low cost but makes developing services and applications more complex, which limits the number of application developed for the platform. This makes the platform less attractive to consumers and lowers revenues. Conversely, a functionality-rich platform is expensive to build, but this cost may be offset by facilitating the development of more applications, therefore attracting more consumers.

This trade-off arises in many environments and properly assessing it can have far-reaching consequences. For example, many attribute the Internet initial success to its minimalist design principles. However, as it matures and transforms from a "physical" network platform to a broader ecosystem of software and web services, the question of whether or not to abandon this minimalist principle is increasingly being raised [58, 13, 85]. The focus of this work is to explore the decision problem faced by a platform provider[22] seeking to select the level of functionality the platform should offer.

[22]The work assumes a monopoly setting with a single platform.

The investigation identifies the ratio of the *rate of change* in the cost (to the platform) of adding new features and the cost of developing applications given a number of platform features, as a key factor in determining the optimal (for the platform) number of features to offer. This optimal choice is, however, highly sensitive to small relative changes in these two costs, with minor differences producing drastically different outcomes, *i.e.,* shifting the optimal operating point from a minimalist to a functionality-rich platform. This negative result not-withstanding, the model provides a framework for reasoning about the impact of introducing more features to a platform. In addition, in cases where the costs of developing new features and their benefits in lowering application development costs can be estimated, the model offers quantitative tools that can assist decision makers.

The chapter is organized as follows: Section 4.2 introduces our two-sided market model. Section 4.3 outlines the solution methodology for this work. Section 4.4 presents the analysis and explores the impact of different factors on the level of platform functionalities. Section 4.5 reviews prior works and positions this work in the literature. Section 4.6 concludes the chapter with remarks on future work.

4.2 Model Formulation

A platform provider attracts developers and consumers by creating value that entices them to join the platform. This 'value' depends on a number of factors, such as the platform's intrinsic value, the subscription fees to join it, the cost of developing applications for it, and *externalities* that affect the value that either developers or customers derive from joining the platform. When modeling a platform as a two-sided market, externalities are usually classified as *same-side* externalities and *cross-side* externalities. Same-side externalities arise in each side of the market from the presence of other users and can be positive or negative [47, 20, 87]. For example, the users of Android devices can derive positive benefits from using the same applications as other Android users (*e.g.,* for applications like Groupon), while developers who produce similar applications face negative externalities from each other through increased competition. Cross-side externalities measure benefits that one side of the market derives from the other. These are usually positive, *i.e.,* consumers benefit from more applications offered by developers, and conversely developers benefit from being able to target their applications to more consumers.

The adoption of the platform by either developers or consumers depends on the overall value they derive from it. As commonly done, we measure this value through a *utility function* that incorporates the different factors that contribute to it. Similarly, the impact of the decisions that the platform provider makes, *i.e.,* pricing and selection of the platform's functionality, are also reflected through the platform provider's utility function. The utility functions for the platform, the developers, and the consumers are described in Subsections 4.2.2, 4.2.3, and 4.2.4, respectively. However, before introducing these utility functions, we briefly review a number of assumptions we make in the model and their implications on its applicability.

4.2.1 Assumptions and Implications

Although both same-side and cross-side externalities can arise, the latter are usually more important in how they affect adoption decisions. As a result and following [4, 20, 67], the model targets cross-side externalities only[23]. In addition, it relies on several further assumptions that we now review.

We assume that developers generate revenue from advertisements and not from consumers purchasing the applications, *i.e.,* application downloads are free and transaction costs for both developers and consumers are negligible. This is reasonable in many settings where applications are offered for free and the bulk of the developer's revenue comes from location based and personalized advertising [57, 76], a trend that is only expected to grow in the future [26]. The advertising revenue generated by an application is also assumed to be linear in the number of users of the applications[24].

Two other important assumptions, both of which affect the model's applicability are that (i) applications all make use of the *same set of platform features*, and (ii) the functionality embedded in these features can be built by *either the platform or the developers themselves,* albeit at possibly different costs. We pause briefly to expand on these two assumptions.

The first implies that applications (services) are homogeneous in their development requirements. In other words, they make use of the same set of platform application

[23] Appendix C.2.1 establishes that the work's main results are qualitatively unaffected by the introduction of same-side externalities.

[24] A non-linear relation changes the results quantitatively but not qualitatively

programming interfaces (APIs) and/or set of independent features created by the developers. Applications can still be differentiated in their offering, but this clearly limits the range of those differences. The second assumption requires that application development needs be known ahead of time to the platform provider, and that application developers be able to independently develop features that the platform decides not to incorporate. This is reasonable when dealing with many software products and services, where both the platform and applications share a common technology. However, this excludes from the platform decision process hardware features whose presence or absence determines the feasibility (or not) of certain applications, *e.g.,* a graphic co-processor is mandatory to enable certain rendering effects.

In general, the last two assumptions limit the model's applicability to platforms that are software ecosystems, such as cloud computing platforms, web services, operating systems, etc. We also assume that application users are oblivious to whether the features that their applications use are provided natively by the platform or independently by application developers. In other words, there is no performance or quality penalty with either choice. In the next sub-sections, we review the utility functions that drive the decisions of the platform provider, application developers, and users.

4.2.2 Platform Utility

The goal of the platform provider is to maximize its profit, which depends on the revenue it generates from the two sides of the market and the cost of the features it has decided to embed in the platform.

We use x_c and n_d to denote the fraction of a large population of N_c and N_d consumers and developers, respectively, who join the platform. As in [4, 20], the platform charges flat fees of p_c and b_d to the consumers and to the developers, respectively[25]. These fees may be incurred as a monthly membership fee for consumers and as a licensing or certification fee for application developers.

The revenue for the platform is, therefore, $p_c x_c + b_d n_d$.[26]

The set of platform features of potential benefits to application developers is assumed known to the platform provider. Embedding more features in the platform incurs a greater cost, and we denote as $C(F)$ the cumulative cost of incorporating F features. We assume that the set of possible features is large. Hence, when mapped on to an interval $[0, F^{\max}]$, they result in a differentiable, monotonically increasing function of $C(F)$ for $F \in [0, F^{\max}]$. An integrality constraint on F is, therefore, not considered explicitly.

In Sub-section 4.2.5, we discuss specific, real-world examples that illustrate different possible behaviors for $C(F)$, *i.e.,* concave or convex. The profit (utility) of a platform with F built-in features and fees of p_c and b_d is given by

$$U_p = p_c x_c + b_d n_d - C(F) \tag{4.1}$$

[25]The model does not require either p_c or b_d to be positive; a negative value is akin a subsidy by the platform to the corresponding market side.

[26]See AppendixC.1 for relabeling of parameters to account for consumer and developer population sizes.

As discussed in Section 4.3, Equation (4.1) together with similar expressions for the utility of consumers and application developers will guide the decisions of how many features to embed in the platform and how to price it.

4.2.3 Developer Utility

Developing applications incurs a certain cost, which depends on the level of support provided by the platform (number of features and associated functionality). A feature-poor platform will usually have lower subscription fees, which may partly offset its correspondingly higher development costs. The revenues generated by an application depend on the number of users the platform has managed to attract, and grow in proportion to that number. Equation (4.2) captures the combined effect of these factors on the developers' utility.

$$U_d = \alpha x_c - b_d - (K(F) + \tau\phi) \tag{4.2}$$

The first component of Equation (4.2) represents the application revenues generated from the x_c consumers that joined the platform (the factor α is a normalization constant that can also be interpreted as the marginal value that a consumer generates for the developer). Those revenues are in the form of advertising revenues, as is commonly assumed in many two-sided markets. For example, the online service iLike has developed a free application for Facebook (the platform) that based on a user profile allows her to play clips of music she may like, and collect revenues from both advertising and referrals to iTunes or Ticketmaster [38].

The second component of Equation (4.2) is the flat-fee, b_d, a developer pays the platform, *e.g.,* to be certified. This charging model is used by many platforms, *e.g.,* Android charges a $25 market developer fee, while Apple's iPhone offers an annual $99 licensing fee to distribute applications and $299 for its iOS developer program.

The last component of Equation (4.2) reflects development costs, which as alluded to earlier depends on the number F of features embedded in the platform. This is captured by the function $K(F)$ that is a differentiable, monotonically decreasing function of $F \in [0, F^{\max}]$. As for $C(F)$, Section 4.2.5 reviews illustrative, real-world examples associated with different behaviors for $K(F)$. In particular, $K(F)$ can be convex or concave depending on whether the marginal reduction in development costs is increasing or decreasing as the platforms adds more features. For a given F, $K(F)$ is the same for all developers. Hence, it can be interpreted as the base cost of developing applications when the platform includes F built-in features. This assumes that developers have similar expertise in developing applications, *e.g.,* software engineers draw from a similar skill base. Developers can, however, be expected to exhibit heterogeneity in their overall development costs, *e.g.,* because of different fixed costs, overhead, benefit levels, etc. This is captured in the factor $\tau\phi$ of Equation (4.2), where following [20, 87] ϕ is uniformly distributed on a unit interval. The value of ϕ for an individual developer is private information, but the distribution is known to all.

4.2.4 Consumer Utility

The value that consumers derive from joining a platform depends on the subscription fees charged, and the number of applications and services accessible through the platform. Consumers are typically heterogeneous, and this heterogeneity can manifest itself in how they value the platform, access to applications and services (cross-side externality), or both. For simplicity and analytical tractability, we focus on a model where heterogeneity is present only in how users value access to applications. Appendix C.2.2 presents an alternative utility function and its analysis, where consumers are instead heterogeneous in how they value the platform. The results under both utility functions are qualitatively similar.

The consumer utility function is then of the form:

$$U_c = \theta\beta n_d - p_c \tag{4.3}$$

The first component, $\theta\beta n_d$, captures the cross-side externality benefits that consumers enjoy from accessing applications available on the platform. These benefits grow with the number of developers, n_d, creating applications for the platform, *e.g.,* the many iPhone developers are responsible for the large number of applications available on it, which contributes to its attractiveness. The assumption of linear growth in n_d is consistent with past literature [87, 4, 80]. The factor β denotes the marginal externality benefit associated with each developer. The term $\theta \in [0,1]$ is a random variable that accounts for heterogeneity in how users value these externality benefits. A user who uses many applications will have a higher θ value, and thus derive higher externality benefits for a given number of application developers. The value of θ for individual users is private information, but its distribution across users is known. We make the common assumption [20, 80, 4] that θ is uniformly distributed[27] in $[0,1]$. The last element of Equation (4.3) is the price p_c, which is a flat membership fee paid to the platform provider.

4.2.5 Representative Examples

Before presenting how the three utility functions just introduced combine in the platform provider decision process, we pause to introduce examples that illustrate possible combinations of the cost functions $C(F)$ and $K(F)$. In all examples, there is an inherent ordering of the features the platform provider is considering offering, *i.e.,* from basic features to more advanced ones, with the latter building on the former. The examples differ in the relative cost of more advanced features compared to basic ones, and in how useful each additional feature is to application developers. Fig. 4.1 illustrates the main combinations of interest between the costs of platform features and their benefits to application developers.

1. Amazon Web Services Platform: Amazon Web Services (AWS) is a cloud computing platform that offers functionalities which third-party developers can use to create services for clients (consumers). These functionalities or features include Amazon EC2 (computation), SimpleDB (database), Amazon S3 (storage), FPS (flexible payment),

[27] Results typically extend [8, 28] to other distributions that share with the uniform distribution the important property of a non-decreasing hazard-rate function $F'(\theta)/(1-F(\theta))$.

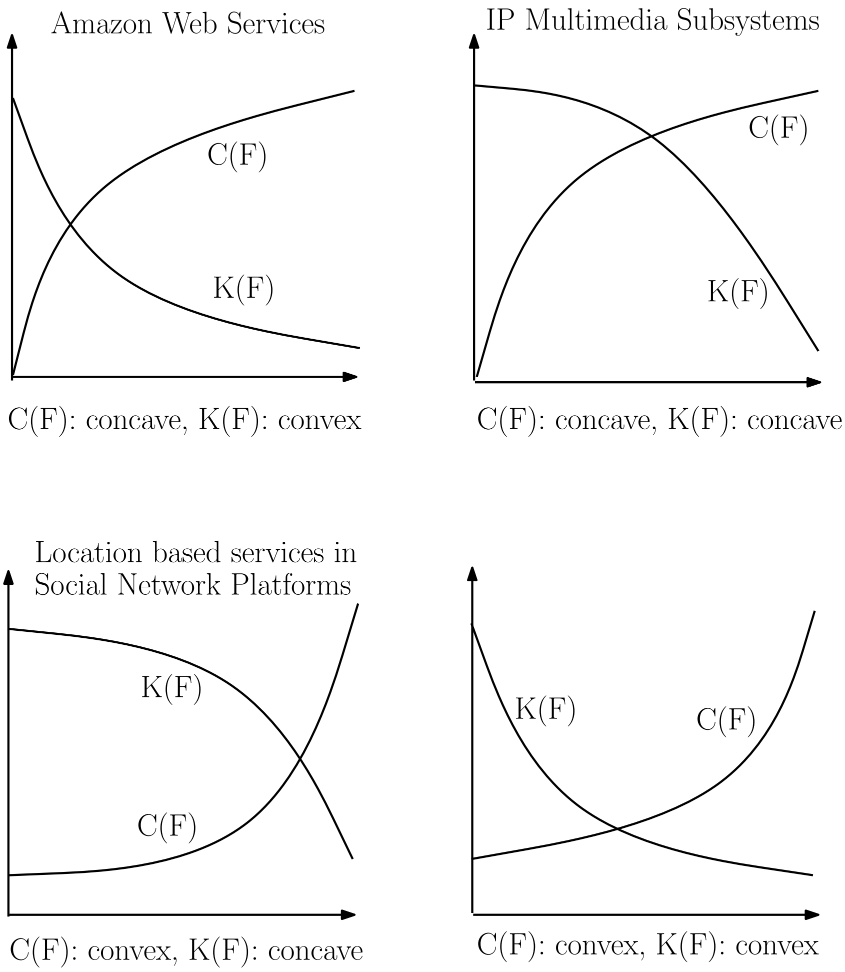

Figure 4.1: Examples for possible shapes of $C(F)$ and $K(F)$

CloudFront (content delivery), MTurk (Internet marketplace), etc. Consumers and developers of services and applications on the AWS platform enjoy cross-side externality benefits from joining the platform, for which they pay subscription fees[28].

The introduction of features on the AWS platform proceeded in two steps. Between

[28]Although AWS allows usage based pricing, large customers have the option to make a low, one-time payment for each reserved instance and in turn receive a significant discount on the hourly usage charged for that instance.

2006-2007, Amazon introduced a set of core features (EC2, FPS, SimpleDB, etc.) that offered basic capabilities such as computation, database, and payment for its AWS platform. Additional features (*e.g.,* SQS, SNS, DevPay, etc.) that built on these core capabilities were subsequently introduced.

Adding each feature to the AWS platform came at a cost. Using API complexity[29] as a proxy for the platform's development cost together with data from [29], it can be observed that capabilities such as EC2, FPS, and SimpleDB came at a higher cost than that of follow-on enhancements such as SNS and DevPay. From this data, one can infer that the AWS platform has a feature development cost function, $C(F)$, that is a *concave increasing* function of F.

Conversely, the benefits of each feature can be estimated based on its "popularity" among developers, *i.e.,* presuming that more useful features are more likely to be used by developers. Using again [29], we see that most core features are significantly more popular than subsequent enhancement features[30]. In other words, the features that were the most costly to develop and incorporate in the platform were also the most useful to developers; at least based on how often developers took advantage of them. As a result, one can conclude that while the development cost function $C(F)$ of the AWS platform is a concave increasing function, the benefits that developers derive from those features, as captured by the function $K(F)$, is a *convex decreasing* function,

[29]The development cost of a feature can be approximated through the complexity of its API. [29] measured API complexity based on the number of operations that the feature supports, as captured in the data required in the specification of its 'Web Services Description Language' (WSDL).

[30]This is most noticeable when comparing EC2, SimpleDB and FPS to SQS, SNS and DevPay.

i.e., the more expensive initial features are also the most useful.

2. IP Multimedia Subsystems (IMS) Platform: The IMS platform is meant to facilitate the development of new integrated multimedia applications and services. Both applications developers and subscribers (consumers of services) pay a fee to the IMS platform. The platform offers a number of built-in capabilities such as a registration mechanism, co-location of multiple IMS services, quality of service, etc. These capabilities are exposed to developers through APIs using Java specifications (JSRs). There are multiple "layers" of JSRs [45, 64], from low-level JSRs such as JSR-180, to high-level JSRs such as JSR-186/187, to more developer friendly APIs for Communication Services such as JSR-281+. Each layer builds on those below, with the base layer that implements the core capabilities of the platform the most expensive to develop. Additional layers are typically "wrappers" meant to hide low-level details from developers, and therefore typically easier for the platform to implement. The development cost function $C(F)$ of the IMS platform is, therefore, a *concave increasing* function of the number F of features (JSRs) it offers.

On the developer side, application development costs are high when only low-level APIs are available. This is mainly because of the greater technical knowledge and programming consistency they require from developers. As APIs that hide many of the platform's low-level intricacies are made available, development costs decrease rapidly. In other words, the function $K(F)$ that captures development costs as a function of the number F of features (APIs) that the platform offers is a *concave decreasing* function, *i.e.,* low-level APIs have little effect on developers costs, while higher-level ones deliver significant benefits.

3. Social Network Platform with location-based services (LBS) support: A social network platform such as Facebook provides application developers access to basic capabilities, *e.g.,* APIs to access the users' social graph, database of user interests, affiliations, etc. However, it also offers more sophisticated functionalities such as real time updates and location-based services (LBS). These have enabled the rapid growth of applications that offer personalized services, *e.g.,* Facebook's Recommendation and Places [72].

Adding this level of sophistication to the platform is, however, technically challenging. It calls for integrating capabilities such as spatial database management, location tracking, real time generation of cryptographic data [7], all of which are significantly more complex than the basic functionalities at the core of a social network platform, *e.g.,* access to the underlying social graph or to a user database. In other words, the function $C(F)$ that captures the cost of adding new (sophisticated) capabilities to a social network platform such as Facebook is a *convex increasing* function of F.

On the other hand, the benefits to application developers of those advanced features can be very high. For example, in the absence of LBS support from the platform, developers would need to build this capability into their application, *e.g.,* by interfacing to the GPS service built into the user's mobile device, when available. That those development costs are high is readily seen from the growth in the number of applications that rely on location information once LBS became available. Specifically, in spite of the large revenue potential of location-based services [72], there were relatively few applications that used location information before LBS became readily accessible to application developers[31] In other words, the substantial decrease in development costs that this produced, enabled many more developers to offer such applications. As a result, it can be argued that a social network platform such as Facebook is an environment where while sophisticated features are expensive to add, they are the ones that deliver the most benefits (reduction in development costs) to application developers. This means that the corresponding function $K(F)$ is a *concave decreasing* function of F.

In the next section, we introduce the methodology used by the platform provider to decide on the "optimal" number of features to incorporate in the platform.

4.3 Solution Methodology

The objective of the platform provider is to select the number of features to include in the platform, and the fees to charge to consumers and developers so as to maximize its

[31] The CEO of Skyhook Wireless noted that in 2010 the number of applications that location information have jumped from 10K to 50K in one year [86].

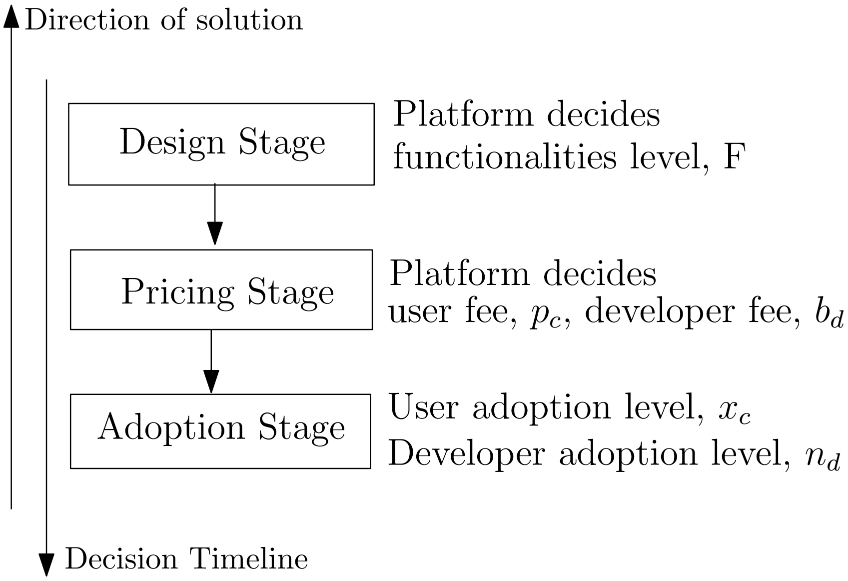

Figure 4.2: Timeline for the sequential decision process

profit. This objective can be realized by using a three-stage sequential process as shown in Fig. 4.2. In the first stage, the platform provider chooses the number F of features to build into the platform. Given a choice for F, participation prices (fees) for the two market sides are chosen in the second stage. Equilibrium adoption levels of consumers and developers are simultaneously realized in the third stage. The three stages are referred to as the *Design Stage*, *Pricing Stage*, and *Adoption Stage,* respectively.

This sequential decision process can then be solved in the reverse order. Equilibrium adoption levels for users and developers are first computed for a given choice of participation prices and number of built-in features. Next, given a choice for the number of built-in features, 'optimal' participation prices are computed based on the equilibrium adoption levels of the previous step. The results characterize the platform's profit for any given number of built-in features. This can then be used to find the 'optimal' number of features, F^*, that maximizes the platform's profit. These steps are detailed next.

4.3.1 Adoption Stage

Both consumers and application developers are assumed to make incentive-compatible and rational decisions. Therefore consumers and developers who join the platform derive positive utility from doing so. Let x_c^* and n_d^* be the expected fraction of consumers and developers joining the platform at equilibrium.

Given p_c, b_d, and F, the value $\widehat{\theta}$ of the marginal consumer who is indifferent between joining the platform or not is

$$\widehat{\theta} = 1 - x_c = \frac{p_c}{\beta n_d^*}\,. \tag{4.4}$$

Conversely, the value $\widehat{\phi}$ of the marginal developer who is indifferent between joining the platform or not is

$$\widehat{\phi} = n_d = \alpha x_c^* - b_d - K(F)\,. \tag{4.5}$$

Note that the system parameters were normalized with respect to the maximum fixed costs that application developers incur, *i.e.,* we set $\tau = 1$. At equilibrium, $x_c^* = x_c$ and $n_d^* = n_d$. Thus, equilibrium adoption levels satisfy

$$p_c = (1 - x_c^*)\beta n_d^* \tag{4.6}$$

$$b_d = \alpha x_c^* - n_d^* - K(F) \tag{4.7}$$

4.3.2 Pricing Stage

For a given number of features F, the platform provider's decision problem is to select fees p_c and b_d that maximize its profit, subject to constraints on the fractions of consumers and developers joining the platform. This yields

$$\begin{aligned} \max_{x_c^*, n_d^*} U_p = \quad & p_c x_c^* + b_d n_d^* - C(F) \\ s.t. \quad & 0 \leq x_c^* \leq 1 \\ & 0 \leq n_d^* \leq 1 \end{aligned} \tag{4.8}$$

Using Equations (4.6) and (4.7) in Equation (4.8), 'optimal' fees and corresponding adoption levels can be derived as described next. The derivation considers separately the cases of interior $\left(0 < x_c^*, n_d^* < 1\right)$ and boundary ($x_C^* = 0, 1$ or $n_d^* = 0, 1$) solutions.

4.3.2.1 Interior Solutions

Interior solutions for optimal prices can be obtained by solving $\frac{\partial U_p}{\partial x_c^*} = 0$ and $\frac{\partial U_p}{\partial n_d^*} = 0$ simultaneously. Details of the derivations can be found in Appendix C.3, and the results are summarized in Proposition 10.

Proposition 10 *Optimal price levels (p_c^*, b_d^*) and optimal adoption levels (x_c^*, n_d^*) that* *maximize the platform provider's profit are given by*

$$p_c^* = \frac{(\beta-\alpha)((\alpha+\beta)^2 - 4\beta K(F))}{16\beta} \tag{4.9}$$

$$b_d^* = \frac{(3\alpha-\beta)(\alpha+\beta) - 4\beta K(F)}{8\beta} \tag{4.10}$$

$$x_c^* = \frac{\alpha+\beta}{2\beta} \tag{4.11}$$

$$n_d^* = \frac{(\alpha+\beta)^2 - 4\beta K(F)}{8\beta} \tag{4.12}$$

The conditions for an interior solution, *i.e.,* $0 < x_c^*, n_d^* < 1$, are $\alpha < \beta$ and $4\beta K(F) < (\alpha+\beta)^2 < 4\beta(2-K(F))$. The second order conditions of the Hessian are then also found to hold.

Proposition 10 reveals properties that are consistent with prior works in two-sided markets [4, 87, 12]. In particular, optimal pricing is typically asymmetric, *i.e.,* different prices are levied on the two sides of the market, and in some cases one market side may be subsidized, *i.e.,* $b_d^* < 0$ while $p_c^* > 0$.

4.3.2.2 Boundary Solutions

Boundary solutions arise when the fraction of customers joining the platform on either market side is either 0 or 1. Such an outcomes are typically associated with less interesting configurations, but are provided here for completeness.

Equations (4.6) and (4.7) are easily seen to imply that solutions of the form $(0, n_d^* > 0)$ or $(x_c^* > 0, 0)$ are not feasible. This is because $n_d^* = 0$ forces $p_c^* = 0$, which results in a negative profit for the platform. Similarly, when $x_c^* = 0$, the platform needs to subsidize developers $b_d^* < 0$, which again translates into a negative profit. So $(0,0)$ is the only feasible equilibrium in such cases, *i.e.,* the system parameters are such that the platform cannot be profitable.

Other possible boundary solutions arise when one side of the market reaches full penetration. There are three sub-cases to consider:

i. Both market sides are at full penetration $(x_c^* = 1, n_d^* = 1)$*:* In this scenario, Equations (4.6) and (4.7) give $p_c^* = \beta$ and $b_d^* = \alpha - K(F) - 1$.

ii. Only the consumers side of the market is at full penetration $(x_c^* = 1, 0 < n_d^* < 1)$*:* The adoption level on the developers side of the market is then given by $n_d^* = (\alpha + \beta - K(F))/2$, and the prices for the two sides are $b_d^* = (\alpha - \beta - K(F))/2$ and $p_c^* = \beta(\alpha + \beta - K(F))/2$, respectively. The constraints $0 < n_d^* < 1$ imply $K(F) < \alpha + \beta < K(F) + 2$, *i.e.,* cross-side benefits, as measured by $\alpha + \beta$, cannot be either too large or too small compared to development costs $K(F)$. When they are large, $(1,1)$ is the equilibrium.

iii. Only the developers side of the market is at full penetration $(0 < x_c^* < 1, n_d^* = 1)$*:* The adoption level on the consumers side of the market is then given by $x_c^* = (\alpha + \beta)/2\beta$ (which requires $\alpha < \beta$), and the prices for the two sides are $b_d^* = (\alpha(\alpha + \beta) - 2\beta(K(F) - 1))/2\beta$ and $p_c^* = (\beta - \alpha)/2$.

As in the case of interior solutions, pricing is typically asymmetric and instances where the platform subsidizes one side and charges the other also occur. Boundary solutions arise mostly when cross-side externalities dominate other system parameters.

In the remaining of the chapter, the focus is on (the more interesting) scenarios where neither market-side has achieved full market adoption.

4.3.3 Design Stage

Using the results of Proposition 10 in Equation (4.8), the platform provider can determine the 'optimal' number F^* of features to include in the platform to maximize profits. Solving for $\frac{\partial U_p}{\partial F} = 0$, F^* can be shown to verify the conditions of Proposition 11.

Proposition 11 *The optimal number F^* of features that should be built into the platform to maximize profit satisfies*

$$\frac{C'(F^*)}{K'(F^*)} = \frac{K(F^*)}{2} - \frac{(\alpha+\beta)^2}{8\beta} \tag{4.13}$$

$$\Rightarrow \quad \frac{C'(F^*)}{K'(F^*)} = -n_d^*(F^*) \tag{4.14}$$

$$\text{and} \quad C''(F^*) > -n_d^*(F^*)K''(F^*) + \frac{1}{2}[K'(F^*)]^2 \tag{4.15}$$

where Equation (4.14) is obtained by using Equation (4.12) in Equation (4.13).

The condition $C'(F^*)/K'(F^*) = -n_d^*(F^*)$ of Equation (4.14) implies that at F^*, the marginal increase in the cost to the platform of adding more features equals the marginal decrease in development costs across all developers subscribed to the platform[32].

Note though that selecting an optimal number of platform features still calls for evaluating the platform profits at *all* F^* values that satisfy Equations (4.13) and (4.15) (see Section 4.4), *and* at the boundaries $F = 0$ and $F = F^{max}$.

[32]Equation (4.14) remains valid for boundary cases where either market sides is at full market penetration.

4.4 Analysis

In this section, we use Proposition 11 to explore generic properties of F^* and the influence of system parameters. We begin with Equations (4.13) and (4.15), which we use to explore how changes in (cross-side) externality benefits affect F^*. These are summarized Proposition 12.

Proposition 12 *for any interior Solution (*i.e., $0 < x_c^*, n_d^* < 1$*), increases in the cross-side externality benefits* α *and* β*, favor adding more functionalities to the platform. In other words,* $\frac{\partial F^*}{\partial \alpha} > 0$ *and* $\frac{\partial F^*}{\partial \beta} > 0$.

Proof 4.4.1 *The proof relies on applying the conjugate pair theorem [16], which gives*

$$\begin{aligned} sign\frac{\partial F^*}{\partial \alpha} &= sign\frac{\partial^2 U_p}{\partial F \partial \alpha} > 0 \\ sign\frac{\partial F^*}{\partial \beta} &= sign\frac{\partial^2 U_p}{\partial F \partial \beta} > 0 \end{aligned}$$

Figures 4.3 and 4.4[33] provide representative examples which demonstrate that an increase in the cross-externality benefits (α or β) increases the optimal level of functionalities in the platform, F^*.

Figures 4.3 shows that F^* increases as developer's cross-side benefits increase from $\alpha = 0.65$ to $\alpha = 0.67$. We note that Proposition 12 is consistent with arguments in favor of expanding the Internet's functionality at a time where the services it offers become more valuable, and the providers of those services derive more value than previously

[33] The parameters in all the figures of this section are assumed to be normalized with respect to populations of size $N_c = N_d = 10^3$, and maximum fixed cost for developers, $\tau = \$10^3$.

feasible.

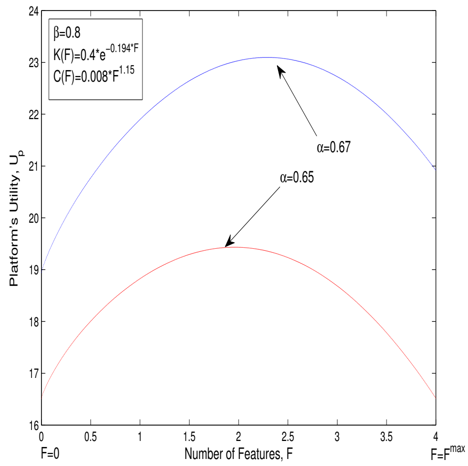

Figure 4.3: Increases in α increases the optimal number of features

Next, we investigate how F^* is affected by changes in the relationship between the cost of adding new features to the platform and the benefits that application developers derive from them. The platform development costs $C(F)$ increase with F, while application development costs $K(F)$ decrease. The relative rates of these increases and decreases ultimately determine F^* and the associated optimal prices, $p_c^*(F)$ and $b_d^*(F)$. The dependency of F^* on the relative rate of change of $C(F)$ and $K(F)$ is captured in Equation (4.13). Note that the platform utility function of Equation (4.1) includes product terms of the form $p_c x_c$ and $b_d n_d$, which imply complex dependencies on the functions $C(F)$ and $K(F)$. Hence, the function $U_p(F)$ that the platform provider seeks to maximize can exhibit a wide range of variations, *e.g.,* multiple maxima and/or minima, even when the functions $C(F)$ and $K(F)$ are themselves "well-behaved," *e.g.,* concave or convex. Clearly, the possibility of more than one value of F satisfying Equa-

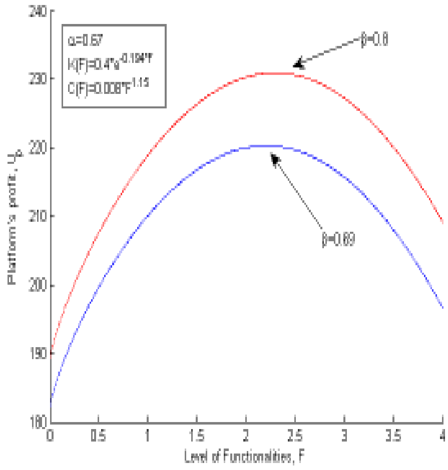

Figure 4.4: Increases in β increases the optimal number of features

tions (4.13) and (4.15) complicates the platform provider's decision process. Furthermore, the optimal decision also depends on how profits at these local maxima compare to "boundary" profits, *i.e.,* for $F = 0$ and $F = F^{max}$.

Numerical investigations demonstrate that multiple local maxima of the platform's utility as a function of F easily arise, *i.e.,* for various combinations of $C(F)$ and $K(F)$. In general small adjustments in the relative rate of increase and decrease of $C(F)$ and $K(F)$ are sufficient to yield drastic shifts in outcome. This is to a large extent borne by Equation (4.13), which shows that small changes in either $C(F)$ or $K(F)$ can substantially vary the value of the ratio $\frac{C'(F)}{K'(F)}$. The implications are that determining how much to invest in a platform and deciding whether or not to develop new features cannot be easily predicted from general properties of $C(F)$ and $K(F)$, *e.g.,* concavity or convexity. Instead, it calls for a fine-grain comparison of the costs of developing features and their benefits to application developers.

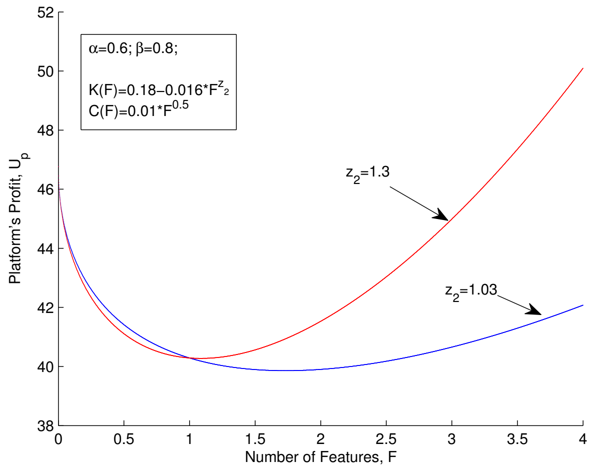

Figure 4.5: Provider provisions either at $F = 0$ or $F = F^{max}$ when both K(F) and C(F) are concave

The one instance for which it is possible to somewhat narrow down the range of possible outcomes is when both $C(F)$ and $K(F)$ are concave, *e.g.,* the IMS platform example. In this case, the optimal number of features can be shown to always be at one of the two boundaries, *i.e.,* $F = 0$ or $F = F^{max}$ (see Appendix C.3 for a proof). This still does not finalize a decision between a minimalist or a functionality-rich platform, but by eliminating intermediate values, it considerably reduces the number of options to consider. In particular, it calls for only evaluating "extreme" scenarios. We illustrate this with a numerical example in Fig. 4.5. The figure shows the platform's profit as a function of the number of features it offers for two different configurations. In both configurations, the platform cost grows like $\sqrt{F}$, while the cost decrease that application developers experience is super-linear in the number of features (the parameter z_2 of the legend). The figure shows that when development costs decrease nearly linearly in F ($z_2 = 1.03$), a minimalist platform is more efficient, while when the decrease if steeper ($z_2 = 1.3$) a functionality-rich platform is preferred. This is obviously intuitive, but the model offers a quantitative framework in which to carry out such an assessment.

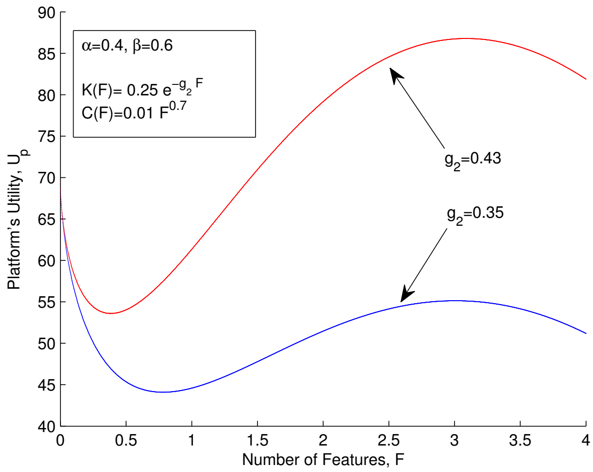

Figure 4.6: Presence of multiple local maxima for convex K(F) and concave C(F)

For all other combinations of convex or concave $C(F)$ and $K(F)$, more complex outcomes can arise, including instances where the optimal outcome involves selecting an intermediate number of features. We illustrate this with a numerical example in Fig. 4.6, which corresponds to a scenario where the platform's cost function $C(F)$ is concave increasing and the developer's cost function $K(F)$ is convex decreasing in F, much like the AWS example of Section 4.2.5. Once again, the plot shows the platform's (profit) utility U_p (y-axis) as a function of the number of features F (x-axis), where $F \in [0,4]$. The figure displays two utility curves corresponding to two different developer cost functions $K(F)$ (convex decreasing), while keeping the functional expression for the platform provider's cost $C(F) = 0.01F^{0.7}$ (concave increasing) identical in both cases. The two application developer cost functions, given by $K(F) = 0.25e^{-g_2 F}$; $g_2 = \{0.35, 0.43\}$, differ in their rate of decrease with the number of platform features F. The figure reveals the following three interesting behaviors.

First, it shows that there are two local maxima for the platform provider's utility for the given system configuration; one corresponding to a minimalist choice ($F = 0$), and the other to an intermediate value F^* that satisfies Equations (4.13) and (4.15). Selecting the globally optimal solution calls not only for computing F^*, but also for comparing profits at $F = 0$ and $F = F^*$.

Second, it shows that a relatively small changes in the rate of decrease in the developers' cost $K(F)$ can result in drastically different choices for the platform provider. In the case of $K(F) = 0.25e^{-0.35F}$, the provider's optimal decision is to choose a minimalist design (*i.e.,* $F = 0$) , while for $K(F) = 0.25e^{-0.43F}$, the provider should create a platform with a large set of built-in features (*i.e.,* $F = F^* \approx 3$). This illustrates the dependency of the decision process in the rate at which development costs decrease as the number of features increases. A similar outcome could have been obtained by keeping $K(F)$ constant, and changing the rate of increase in the platform provider's cost (*i.e.,* $C'(F)$).

Third, the figure illustrates a behavior that at first sight may seem counter-intuitive. Consider the two developers cost functions $K(F) = 0.25e^{-0.35F}$ and $K(F) = 0.25e^{-0.43F}$. The rate of decrease of $K(F)$ is higher in the second case, so that most of the benefits are realized early on as the first few features are added. In contrast, the slower decrease in the first case implies that more features need to be added before a similar decrease is realized. This would seem to suggest that a larger number of features would be preferable in that case than in the other one. The figure shows that the opposite is actually true, *i.e.,* a minimalist choice ($F = 0$) is preferred when $K(F) = 0.25e^{-0.35F}$, while $K(F) = 0.25e^{-0.43F}$ calls for investing in a relatively large number of features in the platform. The reason is that when $K(F) = 0.25e^{-0.43F}$ and costs drop fast, adding features ultimately yields a lower *absolute* value of $K(F)$, which encourages more developers to join and ultimately produces a higher profit. In contrast, the slower cost decrease of $K(F) = 0.25e^{-0.35F}$ is such that the smaller number of developers that join is not sufficient to produce a higher profit than when $F = 0$.

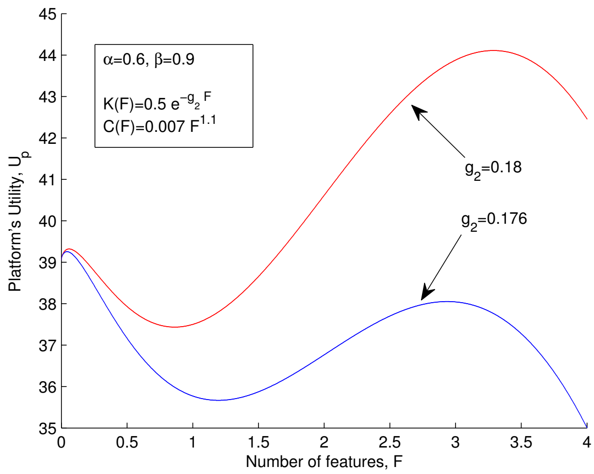

Figure 4.7: Presence of multiple local maxima for convex K(F) and convex C(F)

Finally, we should point out that while the scenario of Fig. 4.6 showed only one interior maximum, it is possible to have more than one. This is illustrated in Fig. 4.7, which involves a convex decreasing $K(F)$ function and a convex increasing $C(F)$ function. As in the previous example and for essentially the same reasons, the choice of which maximum yields the highest overall profit depends on the relative rates of change of the two cost functions.

4.5 Related Literature

This work is related to two streams of information systems and network economics literature, namely, (i) Two-sided markets and (ii) Platform intermediaries in e-commerce markets.

Our model shares a key structural element with the two-sided markets literature in that we consider a platform provider who facilitates interaction between two interdependent customer groups (*i.e.,* app developers and consumers). We also address the topics of platform pricing and customer adoption decisions which have received considerable attention in many of the earlier works.

The second stream of literature mentioned above focus on the question of how can a platform invest in impacting the cross-side network effects in an e-commerce market so as to increase its profits. A key difference from these earlier works is that the platform provider in our model does not directly alter the network effects, instead it invests in adding platform functionalities that reduce the development costs for application developers. Such a scenario is typical in many software ecosystems where the trade-offs lie in the costs borne by the platform in adding functionalities and the benefits these bring to the developers. It also opens up an interesting question that we investigate, namely, what factors influence whether a platform is functionality-rich or minimalist in its design. Thus, we contribute to the existing e-commerce platform literature by investigating the trade-offs in the platform's investments in built-in functionalities. Next, we discuss the related works in these two areas of research.

Two-sided markets: Two-sided markets are made of two interdependent groups of customers (*e.g.,* sellers and buyers) who benefit from each other's participation. A platform intermediary facilitates interactions between these two customer groups and generates its revenue by charging them a price for joining the platform. The adoption of the platform by the two customers groups and the volume of interaction between them depend not only the prices set by the platform provider but also on the price structure [74]. [37, 74] provide definition and examples of two-sided platforms along with extensive literature survey on the topic. Two-sided models have also been used recently to analyze net-neutrality issues [20, 62]. Many works have also focused on pricing strategies for two-sided platform [67, 36, 87]. While our work builds on the existing literature, we use the two-sided platform model to consider issues that were not the focus of these earlier works. We consider software based platforms that bring application developers and consumers together, and focus on the question of functionality design. This environment is particularly relevant in the context of today's Internet as it evolves from a physical infrastructure to a software ecosystem. The success of the Amazon Web Services platform, Facebook platform etc, bear witness to the progress made in that direction. Although the Internet started out with a minimalist design, its gradual evolution raises new questions as to whether a functionality-rich or a minimalist design is desirable in this new environment. The model of the two-sided platform we develop is aimed towards creating an analytical framework that explores such questions.

Platform intermediaries in e-commerce markets: The literature in electronic intermediaries is also related to this work. Electronic market intermediaries lower search costs for buyers and increases price competition among sellers [3]. While much of the work in this area have analyzed the role of electronic intermediaries, some [4, 87] have focused on the impact of infrastructural investments of the intermediary platform on its cross-network externalities. [4] shows that it is optimal for an intermediary to invest in network externalities asymmetrically to maximize the network benefits for one market side. Our work differs from the above in that we consider scenarios where the platform provider does not have the means to directly impact the cross-externalities. Instead, the platform can invest in functionalities that make application development easier for developers, and thus indirectly influence the customer adoption levels and pricing. Such scenarios are typical for most web services and social network platforms where the level of functionality investments determine how the costs and benefits are shared by the platform and its developers. Also, our focus is different from the previous works since we are interested in analyzing the factors that influence the platform's decision regarding a functionality-rich or a minimalist design. The results of our work will therefore contribute to the growing literature on e-commerce intermediary investments and platform design.

4.6 Conclusions

This chapter develops a model to explore when a platform should seek minimalist or functionality-rich designs. The question is formulated using a two-sided market model in which the platform is the market and service developers and consumers are the two sides of the market. Consumers, developers and the platform provider have utility functions that account for externality benefits, prices, and costs. The platform provider seeks to identify how many features to include in the platform to maximize its profit. This is formulated as a three stage sequential decision process, for which the solution is characterized.

The solution reveals the impact of cross-externalities and confirms the benefits of asymmetric pricing. More importantly, it shows that the platform choice is highly sensitive to how additional features affect the costs of the platform and service developers. This unfortunately establishes that minor changes in either costs can yield drastically different solutions. This is illustrated through numerical examples, which point to the significant challenge of answering the question of optimal platform design in practice. In spite of the limitations of its results, the work provides initial insight and a possible methodology for tackling the complex question of (software) platform design.

There are many directions in which the work's initial results can be extended. Empirical validations are obviously at the forefront, and exploring if this can be done for one of the examples of Section 4.2.5 is of interest. Some modeling extensions are also worth pursuing. One involves allowing developers to use different subsets (bundles) of features, each assigned a different price. Another is to introduce multiple platforms and allow competition, *e.g.,* between minimalist and functionality-rich platforms. Both are topics that can be pursued by building on the model presented in the chapter.

Chapter 5

Conclusions

In this dissertation, we focused on understanding the interplay between technological and economic factors in the context of network systems and architectures, and investigated in depth three fundamental issues. These issues deal with the deployment of network services, the adoption of competing network technologies, and the design of network platforms. Our contribution involves developing robust analytical frameworks that study these issues by accounting for relevant technological and economic factors, identifying behaviors of interest, and demonstrating their impact on design decisions.

The first part of this dissertation considered the question of network choice for deployment of new services, namely whether heterogeneous services should share a common network infrastructure or have dedicated networks of their own. We introduced an analytical framework to explore the trade-offs involved in choosing between shared and dedicated infrastructures. This required identifying various economic factors of relevance, *e.g.,* deployment and operational costs, capacity costs, fixed costs, etc., and accounting for possible economies and/or diseconomies of scope that these costs exhibit in a shared infrastructure. We then developed on the traditional news-vendor models to find the optimal capacity allocation for the new service in the presence of demand uncertainty and the ability to reprovision resources in response to excess demand. Resource reprovisioning is becoming increasingly feasible with progress in virtualization technologies, and therefore, its impact on the network choice is a key area of study. We showed that the extent to which reprovisioning is feasible can by itself affect which infrastructure, shared or dedicated, is more effective. Moreover, we demonstrated that the impact of reprovisioning can be quite non-intuitive, thereby motivating the need for network providers to consider its impact while making network design choices. We also identified two operational metrics, the gross profit margin and the return on capacity, that play a major role in determining which infrastructure benefits more from reprovisioning, and are thus useful in making managerial decisions. Thus, the first part of the thesis contributes to offering a framework for service providers to evaluate network infrastructure options, and in particular to decide whether it is profitable to deploy a new service on an existing network infrastructure while accounting for both technological capabilities and economic factors.

In the second topic of this dissertation, we dealt with the issue of network technology adoption. As new network technologies and services emerge, questions about their successful adoption arise, as they often need to compete against formidable incumbents with large installed base. Developing an understanding of the competition between network technologies, and identifying the extent to which different factors, in particular converters (gateways), affect the outcome was the focus of our work. To this end, we proposed and solved a model for adoption of competing network technologies by individual heterogeneous users. We identified a number of interesting and at times unexpected behaviors, including the possibility for converters to hurt the adoption of its own technology, reduce overall market penetration across both the technologies, and prevent convergence to a stable state; something that never arises in their absence. The findings were tested for robustness, *e.g.,* different utility functions and adoption models, and found to remain valid across a broad range of scenarios. The key contribution of this work is in providing a framework to study the adoption and diffusion of a new network technology in the presence of an incumbent and in offering insight into the role of converters.

Related to the questions of new service deployment and adoption is the issue of the right design for the underlying network platform (*e.g.,* Internet). Platforms generate value by bringing together two interdependent groups of customers, the developers and the consumers, and provide functionalities to foster service innovation and user adoption. But there is natural trade-offs in the cost of creating a functionality-rich versus a minimalist platform. The last part of the dissertation is concerned with this issue. In particular, we develop a model to explore the decision problem faced by a monopolist platform provider seeking to select the level of functionality that the platform should offer. In doing so, we considered a framework of a two-sided market, with the platform as the market and the developers and consumers as the two sides of this market. We introduced utility functions for the two sides and the platform and incorporated in them the impact of network externalities, prices, and functionality development costs. The investigation revealed a number of interesting properties depending on how the cost of features to the platform and the benefits application developers derive from them relate to each other. In particular, it showed that the ratio of the marginal increase in platform provider's cost to the marginal decrease in application developer's cost from an increase in the platform's functionalities play an important role in determining the optimal functionality level. Besides contributing to the growing economics literature on two-sided markets, this work improves our understanding of functionality-rich versus minimalist platform design by providing an analytical framework to assess the underlying design trade-offs.

Next, we discuss some areas for future investigation and explore the potential for improvements to the frameworks we have proposed in this dissertation.

5.1 Potential Extensions

The broad area of network economics is ripe for investigation along several directions. In this work, we dealt with three key aspects related to technology adoption, deployment, and design, each of which have potential for several interesting extensions as discussed below.

In the context of new service deployment, one related topic that can be investigated is the issue of choosing between private and public clouds. Companies, such as Google, are promoting the idea of a shared public cloud in which many different companies can utilize a common pool of network resources and infrastructure for purpose of storage, computation, etc. In contrast, companies like IBM are pushing for the adoption of private or dedicated cloud which can be provided on their mainframe infrastructure. An investigation that aims to understand the trade-offs in this tussle between the proposed ideas of private and public cloud can be of considerable interest. The framework for such a study can use ideas from the model we developed for understanding the trade-offs between shared and dedicated networks, albeit with the additional feature that users can potentially impose negative externality on each other.

A potential extension of our work on the role of converters in technology adoption is to consider a game-theoretic framework in which the two competing technology providers choose converter efficiencies and/or pricing in a strategic manner. Additionally, allowing some of the system parameters, *e.g.,* technology's quality, price, to be time-varying is also of obvious interest. These issues may be investigated for additional insights in by building on the framework laid out in our work.

Lastly, the issue of platform design too provides many opportunities for further studies. As discussed in the work on the choice of functionality-rich versus minimalist platform design, our model is tailored to investigate platforms that are software ecosystems, such as cloud computing platforms, web services, etc. This is because we assume that the applications make use of the same set of features and that these can be developed either by the platform or the developers themselves. One natural extension would be to extend this work and allow different applications to use different subset of functionalities, and then explore the impact of bundling subsets of functionalities on the platform's profitability.

These three pieces of work that form the core of this dissertation provide the initial impetus for investigations into a wide range of interesting questions about the economic viability of network solutions. They present a reasoning framework that can be used to systematically account for key economic and technological factors and to study their impact in a variety of network system settings.

Appendix A

Appendix: Network Infrastructure Deployment

A.1 Proofs of Lemmas and Proposition

Proof of Lemma 2.3.1: It can be easily verified from Equations (2.7) and (2.10) that if $\frac{p_{d2}}{a_{d2}} > \frac{p_{s2}}{a_{d2}}$ ($\frac{p_{s2}}{a_{s2}} > \frac{p_{d2}}{a_{d2}}$) then $\forall\alpha\ K^*_{d2} \geq K^*_{s2}$ ($\forall\alpha\ K^*_{s2} \geq K^*_{d2}$). Moreover, this relationship is independent of α. □

Proof of Lemma 2.3.2: $\frac{\partial K^*_{d2}}{\partial\alpha} = \frac{-a_{d2}{}^2(p_{d2}-a_{d2})X_2^{\max}}{[(1-\alpha)p_{d2}+\alpha a_{d2}]^2} < 0$ and $\frac{\partial K^*_{s2}}{\partial\alpha} = \frac{-a_{s2}{}^2(p_{s2}-a_{s2})X_2^{\max}}{[(1-\alpha)p_{s2}+\alpha a_{s2}]^2} < 0$ □

Proof of Lemma 2.3.3: K^*_{d2}, K^*_{s2} from Eq.(2) ($i = \{s2, d2\}$) of Chapter 2 can be written as $K_i^* = \frac{(1-\frac{a_i}{p_i})X_2^{\max}}{1+(\frac{\alpha}{1-\alpha})\frac{a_i}{p_i}}$, $i = \{d2, s2\}$, We see that the dependency of K_i^* on α is through the ratio $\frac{\alpha}{1-\alpha}$ that multiplies the term $\frac{a_i}{p_i}$ (the inverse of the return on capacity). When return on capacity is large, $\frac{a_i}{p_i}$ is small, and the term $\left(\frac{\alpha}{1-\alpha}\right)\frac{a_i}{p_i}$ remains small as α increases from an initial value of 0. As a result, the denominator of K_i^* remains essentially constant for small values of α. This only changes when α starts approaching 1, so that the ratio $\frac{\alpha}{1-\alpha}$ becomes large, which eventually translates in a steep decrease of K_i^* down to a value of 0 when $\alpha = 1$. □

Proof related to Proposition 1: Show that $h'(\alpha)$ ***can change its sign at most once for*** $\alpha \in [0,1]$**.**

Proof: We need to show that the equation $h'(\alpha) = 0$ can have at most one solution for $\alpha \in [0,1]$. $\frac{\partial h(\alpha)}{\partial \alpha} = 0$ can be rewritten as: $\frac{\beta_1}{(\gamma_1-\alpha)^2} = \frac{\beta_2}{(\gamma_2-\alpha)^2}$, where $\beta_1 = \frac{{a_{d2}}^2}{(p_{d2}-a_{d2})} > 0$, $\beta_2 = \frac{{a_{s2}}^2}{(p_{s2}-a_{s2})} > 0$, $\gamma_1 = \frac{p_{d2}}{p_{d2}-a_{d2}} > 1$, and $\gamma_2 = \frac{p_{s2}}{p_{s2}-a_{s2}} > 1$. The solutions to the above equation are obtained either from $(\gamma_1 - \alpha) = \sqrt{\frac{\beta_1}{\beta_2}}(\gamma_2 - \alpha)$ or $(\gamma_1 - \alpha) = -\sqrt{\frac{\beta_1}{\beta_2}}(\gamma_2 - \alpha)$.

Note that $\gamma_1 > 1$, $\gamma_2 > 1$ and $0 \leq \alpha \leq 1$ together imply that $(\gamma_1 - \alpha) > 0$ and $(\gamma_2 - \alpha) > 0$. Moreover, since $\frac{\beta_1}{\beta_2} > 0$, a valid value of $\alpha \in [0,1]$ may only be obtained from solving $(\gamma_1 - \alpha) = \sqrt{\frac{\beta_1}{\beta_2}}(\gamma_2 - \alpha)$. Hence, there can be at most one solution to $h'(\alpha) = 0$ for $\alpha \in [0,1]$. □

A.2 Optimal Resource Allocation for Generalized Demand

This section provides the results for the optimal resource allocation for generic demand distributions. Let $i = \{s2, d2\}$ be the index for the cases of Service 2 deployment on shared and dedicated networks respectively. The contribution margin for Service 2 is p_i, and the capacity cost is a_i. If the realized demand X_2 exceeds the provisioned capacity K_i, the capacity is adjusted to accommodate a fraction α of the excess demand, *i.e.,* capacity is increased to $K_i + \alpha(X_2 - K_i)$. The gross profit for Service 2 is then given by

$$R_i(X_2 > K_i) = (p_i - a_i)(K_i + \alpha(X_2 - K_i)) \tag{A.1}$$

Conversely, when the realized demand is less than the provisioned capacity K_i, the gross profit for Service 2 is

$$R_i(X_2 \leq K_i) = p_i X_2 - a_i K_i \tag{A.2}$$

Assuming a known distribution F_{x_2} with a support in $[0, X_2^{max}]$[34] for the demand of Service 2, the expected gross profit R_i *given* the capacity provisioned upfront K_i in a dedicated network can be expressed as

$$\mathbf{E}(R_i)_{[K_i]} = \int_0^{K_i} R_i(X_2 \leq K_i) f_{x_2} d(x_2) + \int_{K_i}^{X_2^{\max}} R_i(X_2 > K_i) f_{x_2} d(x_2)\,, \tag{A.3}$$

where $R_i(X_2 > K_i)$ and $R_i(X_2 \leq K_i)$ are given in Equations (A.1) and (A.2).

[34] $F(0) = 0, F(X_2^{max}) = 1$, and to simplify the analysis we assume that $F(.)$ is strictly increasing and continuous, with a valid inverse function $F^{-1}(.)$.

$$\mathbf{E}(R_i)_{[K_i]} = \quad p_i \int_0^{K_i} x_2 f_{x_2} d(x_2) + \alpha(p_i - a_i) \int_{K_i}^{X_2^{\max}} x_2 f_{x_2} d(x_2)$$
$$+(1-\alpha)(p_i - a_i)K_i - [(1-\alpha)p_i + \alpha a_i]K_i F(K_i) \qquad \text{(A.4)}$$

Taking $\frac{d\mathbf{E}(R_i)_{[K_i]}}{dK_i} = 0$, we get K_i^*:

$$F(K_i^*) = \frac{(1-\alpha)(p_i - a_i)}{(1-\alpha)p_i + \alpha a_i}$$

If $F^{-1}(.)$ is the inverse distribution function, then the optimal capacity provisioned is

$$K_i^* = F^{-1}\left(\frac{(1-\alpha)(p_i - a_i)}{(1-\alpha)p_i + \alpha a_i}\right) \qquad \text{(A.5)}$$

Notice that this formulation for K_i^* is analogous to the notion of 'critical fractile' for the newsvendor problem, with the extension that in this case we allow for capacity relaxation when excess demand is realized, albeit with a penalty. The optimal capacity decreases with increase in reprovisioning ability, α. Moreover, as with the case of uniform demand distribution, the optimal capacity, K_i^*, depends on $p_i - a_i$ and p_i/a_i values. The capacity cost will be $a_i K_i^* = a_i F^{-1}\left(\frac{(1-\alpha)(p_i - a_i)}{(1-\alpha)p_i + \alpha a_i}\right)$, which suggests that these two operational metrics can play an important role, as with the uniform demand distribution.

$\mathbf{E}(R_i)_{[K_i^*]}$ for optimal provisioning is given by

$$\mathbf{E}(R_i)_{[K_i^*]} = \quad p_i \int_0^{K_i^*} x_2 f_{x_2} d(x_2) + \alpha(p_i - a_i) \int_{K_i^*}^{X_2^{\max}} x_2 f_{x_2} d(x_2)$$
$$+\left[((1-\alpha)p_i + \alpha a_i)(1 - F(K_i^*)) - a_i\right] K_i^* \qquad \text{(A.6)}$$

Note that when we substitute the expression for K_i^*, the component $\left[((1-\alpha)p_i +\right.$

$\alpha a_i)(1-F(K_i^*))-a_i\big]K_i^*=0$. This is no coincidence, and it must be satisfied. Notice that it can be rearranged to write:

$$a_iK_i^*F(K_i^*)=(1-\alpha)(p_i-a_i)K_i^*(1-F(K_i^*)) \quad \text{(A.7)}$$

$$\Rightarrow a_iP(X_2\leq K_i^*)=(1-\alpha)(p_i-a_i)P(X_2>K_i^*) \quad \text{(A.8)}$$

The above equation means that at the optimal capacity allocation, the cost incurred from an unit of over-provisioning (*i.e.,* $P(X_2\leq K_i^*)$) given by the left hand side must be balanced against the penalty cost from an unit of under-provisioning (*i.e.,* $P(X_2>K_i^*)$) given by the right hand side. This is the principle what governs the expression we obtained earlier, *i.e.,* $K_i^*=F^{-1}\big(\frac{(1-\alpha)(p_i-a_i)}{(1-\alpha)p_i+\alpha a_i}\big)$, which can also be obtained from Equation (A.7).

A.3 Capacity Cost increase as a Reprovisioning Penalty

In our model we considered that the reprovisioning coefficient, α, accounts for the amount of excess demand that can be recaptured through resource reprovisioning. In other words, α governed the extent of penalty paid by the service provider for under-provisioning network resources. Consider that the realized demand for Service 2 is X_2 and the provisioned capacity is K_i, where $i=\{s2,d2\}$ is the index for the cases when Service 2 is deployed on a shared and a dedicated network respectively. Let the contribution margin for Service 2 be p_i, and the unit capacity cost be a_i for these cases. Thus, when an excess demand is realized (*i.e.,* $X_2 > K_i$), the revenue function is given by

$$R_i(X_2 > K_i) = (p_i - a_i)(K_i + \alpha(D_2 - K_i)) \tag{A.9}$$

An alternative way of modeling the penalty of underprovisioning is to consider that whenever an excess demand is realized, the provider can accommodate it entirely, but at a higher per unit capacity cost for each unit of excess demand accommodated. As before, let p_i be the profit margin for unit demand and a_i be the unit capacity cost. But now, the provider incurs an additional cost of $a_i + \delta$ for each unit of excess demand[35]. Therefore, when an excess demand is realized (*i.e.,* $X_2 > K_i$), the revenue function for this model is given by

$$R_i(X_2 > K_i) = p_i X_2 - a_i K_i - (a_i + \delta_i)(X_2 - K_i) \tag{A.10}$$

We show that these two models are equivalent *i.e.,* there is a one-to-one mapping between the parameters α and δ which account for the penalty of underprovisioning in the two respective models.

$$(p_i - a_i)(K_i + \alpha(X_2 - K_i)) = p_i X_2 - a_i K_i - (a_i + \delta_i)(X_2 - K_i) \tag{A.11}$$

$$\alpha = 1 - \frac{\delta_i}{p_i - a_i}, \quad 0 \le \delta_i \le p_i - a_i \tag{A.12}$$

This establishes that the findings of our model can also be used for this alternative scenario where a provider can accommodate the entire excess demand, but at a penalty of a larger unit capacity cost.

[35]Notice that $\delta < p_i - a_i$ because the profit per user needs to be positive.

A.4 Impact of cost parameters on Network Choice

In this section we consider the impact of various cost and revenue components on the network choice. As stated in Section 2.4, a shared network is preferred if $\Pi_s > \Pi_d$. Economies of scope in costs favor the creation of shared networks while large diseconomies of scope favor dedicated networks. This behavior is easily observed for certain cost components, such as fixed costs (c_{d1}, c_{d2}, c_s) and Service 1 profit components $(p_i, a_i; i = \{d1, s1\})$, by analyzing the impact of these parameters on the expressions for Π_d and Π_s (ref. Equations (2.14) and (2.16) of Chapter 2). But, the impact of the other cost and revenue parameters, *e.g.,* the contribution margins and capacity costs, p_i and a_i (i={d2, s2}), are less obvious. In particular, we would like to know if (dis)economies of scope in these parameters can alter the network choice from shared to dedicated (or vice versa), and whether there can be more than one threshold for altering back and forth between the network choices as these (dis)economies increase.

To investigate the impact of the cost parameters, we first identify the different deployment and operational cost components that these parameters represent. This identification of cost components is important in order to identify which of these parameters may vary independent of the others, and which parameters will vary in lock-step if increase in (dis)economies of scope were to be found in these cost components.

Economies (diseconomies) in contribution margins can be captured by considering the presence of diseconomies (economies) in the variable costs, v_{s1}, v_{d1} and v_{s2}, v_{d2} for the two services. The contribution margins can be written as $p_{d1} = p_1 - v_{d1}$, $p_{s1} =$

$p_1 - v_{s1}$, $p_{d2} = p_2 - v_{d2}$ and $p_{s2} = p_2 - v_{s2}$, where p_1, p_2 are the prices of Service 1 and 2 respectively. These variable cost for Service 2 consists of both the deployment costs for new equipments (e, e_{s2}) and operational costs (η, η_s). Service 1, which is an existing service incurs an operational cost of η and η_s in the dedicated and shared networks. Therefore, the total variable costs for Service 1 and Service 2 in dedicated and shared networks are captured by the following parameters:

$$\begin{aligned} &Service\ 1: \ v_{d1} = \eta, v_{s1} = \eta_s \\ &Service\ 2: \ v_{d2} = e + \eta, v_{s2} = e_{s2} + \eta_s \end{aligned} \tag{A.13}$$

Note that the parameter v_{d2} can increase/decrease alone if the equipment cost e increases/decreases in the dedicated network. This corresponds to the case where the contribution margin p_{d2} changes. On the other hand, we could have a scenario where both v_{s1} and v_{s2} change simultaneously, which happens when the operational cost, η_s, is increased/decreased. This corresponds to scenarios where the corresponding contribution margins p_{s2} and p_{s1} both changes in lock-step. We will analyze this case later in the discussion. The effect of changing other parameters is also similar because decreasing costs in a shared network (greater economies of scope) has the same effect as that of increasing costs in dedicated networks, and vice-versa.

The capacity costs are captured by introducing parameters for cost of bandwidth and operational costs. The parameters b and b_s capture the unit bandwidth costs for dedicated and shared networks. We assume that Service 2 requires λ units of more bandwidth than Service 1. The operational costs are captured by the parameters q_{d1}, q_{d2} and

q_{s1}, q_{s2} for dedicated and shared networks respectively. Therefore, the total capacity costs for Service 1 and Service 2 in dedicated and shared networks are captured by the following parameters:

$$Service\ 1: \ a_{d1} = \beta + q_{d1}, a_{s1} = \beta_s + q_{s1} \tag{A.14}$$

$$Service\ 2: \ a_{d2} = \lambda\beta + q_{d2}, a_{s2} = \lambda\beta_s + q_{s2}$$

We will study the impact of the various cost parameters by considering the following scenarios: (1) a_{s1} is varied (*i.e.,* q_{s1} varies), (2) a_{s2} is varied (*i.e.,* q_{s2} varies) , and (3) both a_{s1} and a_{s2} are varies (*i.e.,* b_s varies). The effect of changing the other parameters is also similar because decreasing costs in a shared network (greater economies of scope) has the same effect as that of increasing costs in dedicated networks, and vice-versa. For example, increasing/decreasing b will have the same kind of impact as that of decreasing/increasing b_s.

Next, we provide the analysis for the scenarios discussed previously, starting with the impact of variable costs, followed by that of capacity costs.

A.4.1 Varying v_{s1} and v_{s2} together

We consider the case where v_{s1} and v_{s2} vary together. If we substitute in the values for K_2^*, K_{s2}^*, and γ into $\Pi_s > \Pi_d$, we can rearrange the terms to obtain the inequality.

$$v_{s2} + \frac{2v_{s1}X_1}{X_2^{max}} - \frac{a_{s2}^2}{(1-\alpha)(p_2 - v_{s2}) + \alpha a_{s2}} < \gamma_1 \tag{A.15}$$

Where

$$
\begin{aligned}
\gamma_1 &= a_{d2} - \frac{1}{X_2^{max}}[a_{s2}(X_2 + 2X_1) + 2(c_s - c_{d1} - c_{d2}) - 2a_{d1}X_1] \\
&\quad - a_{s2} + \frac{2v_{d1}X_1}{X_2^{max}} + v_{d2} - \frac{a_{d2}^2}{(1-\alpha)p_2 + \alpha a_{d2} - (1-\alpha)v_{d2}}
\end{aligned} \tag{A.16}
$$

Note that the cost parameters, $v_{s2} = e_{s2} + \eta_s$ and $v_{s1} = \eta_s$, both vary when η_s is varying. We will therefore study the impact of changing η_s on the validity of $\Pi_s > \Pi_d$. Introducing the above parameters in (A.15), we obtain:

$$
\eta_s\left(1 + \frac{2X_1}{X_2^{max}}\right) - \frac{a_{s2}^2}{(1-\alpha)(p_2 - e_{s2}) + \alpha a_{s2} - (1-\alpha)\eta_s} < \gamma_1 - e_{s2} \tag{A.17}
$$

Let $\beta_{\eta_s} = (1-\alpha)(p_2 - e_{s2}) + \alpha a_{s2}$ and $\gamma_2 = \gamma_1 - e_{s2}$ then:

$$
\eta_s\left(1 + \frac{2X_1}{X_2^{max}}\right) - \frac{a_{s2}^2}{\beta_{\eta_s} - (1-\alpha)\eta_s} < \gamma_2 \tag{A.18}
$$

Notice that η_s can vary only between 0 and $\frac{\beta_{\eta_s} - a_{s2}}{1-\alpha}$ (because $p_2 - e_{s2} - a_{s2} > \eta_s$ for $\Pi > 0$).

A single intersection between $\eta_s\left(1 + \frac{2X_1}{X_2^{max}}\right) - \frac{a_{s2}^2}{\beta_{\eta_s} - (1-\alpha)\eta_s}$ and γ_2 will mark a switch from shared to dedicated (or vice versa). Next, we prove that two intersections can never arise *i.e.,* multiple switching between shared and separate does not occur when v_{s1} and v_{s2} vary together. We proceed to prove by the method of contradiction.

If two crossings were to occur, the baseline of γ_2 must intersect the left hand side function $\eta_s\left(1 + \frac{2X_1}{X_2}\right) - \frac{a_{s2}^2}{\beta_{\eta_s} - (1-\alpha)\eta_s}$ twice. This function is concave between $\eta_s = 0$ and $\eta_s = \frac{\beta_{\eta_s} - a_{s2}}{1-\alpha}$ (and the second derivative w.r.t to η_s is $\frac{-2a_{s2}^2(1-\alpha)^2}{(\beta_{\eta_s} - (1-\alpha)\eta_s)^3} < 0$). Therefore, for γ_2 to cut the line twice, the slope of the function at $\eta_s = \frac{\beta_{\eta_s} - a_{s2}}{1-\alpha}$ must be negative, which requires,

$$\left(1+\frac{2X_1}{X_2^{max}}\right)-\frac{a_{s2}^2(1-\alpha)}{[\beta\eta_s-(1-\alpha)\eta_s]^2}\bigg|_{\eta_s=\frac{\beta\eta_s-a_{s2}}{1-\alpha}}<0 \tag{A.19}$$

However, we find that this is not possible because (A.19) simplifies to $\frac{2X_1}{X_2^{max}}+\alpha<0$, which can never occur because X_1, X_2^{max}, and α are always positive. Therefore, it is never possible that the network choice will be switch twice as the costs increase/decrease.

A.4.2 Impact of varying a_{s1}

Consider the case where a_{s1} varies. Substituting in K_2^*, K_{s2}^* in $\Pi_s>\Pi_d$ and rewriting the expression in terms of a_{s1}, we get:

$$\frac{2X_1}{X_2^{max}}a_{s1}>\gamma_4 \tag{A.20}$$

where

$$\begin{aligned}\gamma_4 = \quad & 2a_{d2}-\frac{a_{d2}^2}{(1-\alpha)(p_2-v_{d2})+\alpha a_{d2}}-2a_{s2}\\ & -\frac{a_{s2}^2}{(1-\alpha)(p_2-v_{s2})+\alpha a_{s2}}-\frac{2}{X_2^{max}}\left[\left(\frac{v_{s2}X_2^{max}}{2}+v_{s1}X_1+c_s\right)\right.\\ & \left.-\left(\frac{v_{d2}D_2^{max}}{2}+v_{d1}X_1+c_{d1}+c_{d2}\right)\right]+\frac{2X_1}{X_2^{max}}\end{aligned} \tag{A.21}$$

The left hand side of A.20 is linear in a_{s1}. Therefore, there can be at most one valid crossing (dedicated to shared as a_{s1} increases). However, in order for a crossing to occur, $p_1-v_{s1}-a_{s1}>0$ must be satisfied. This results in the following constraint for one crossing.

$$0 < \gamma_4 \frac{X_2^{max}}{2X_1} < p_1 - v_{s1} \tag{A.22}$$

If this constraint is not satisfied, then no crossing occurs *i.e.,* one type of network is always chosen. The constraint shows that when capacity cost, a_{s1}, is low (*i.e.,* economies of scope) a shared network is preferred but if a_{s1} is high, dedicated networks are preferred.

A.4.3 Varying a_{s2}

Now, let us consider the case where we vary a_{s2}. If we substitute K_2^*, K_{s2}^* in $\Pi_s > \Pi_d$ and rewrite it in terms of a_{s2} on one side, we get:

$$\gamma_5 = 2a_{s2} - \frac{a_{s2}^2}{\beta_{a_{s2}} + \alpha a_{s2}} \tag{A.23}$$

where $\beta_{a_{s2}} = (1-\alpha)(p_2 - v_{s2})$ and

$$\begin{aligned}\gamma_5 = \quad & 2a_{d2} - \frac{a_{d2}^2}{\beta_{a_{s2}} + \alpha a_{d2}} - \frac{2X_1}{d_2^{max}}(a_{s1} - a_{d1}) - \frac{2}{X_2^{max}}\left[\left(\frac{v_{s2}X_2^{max}}{2}\right.\right. \\ & \left.\left. + v_{s1}X_1 + c_s\right) - \left(\frac{v_{d2}X_2^{max}}{2} + v_{d1}X_1 + c_{d1} + c_{d2}\right)\right]\end{aligned}$$

As before, for multiple switches between shared and dedicated to occur, γ_5 must intersect the function $2a_{s2} - \frac{a_{s2}^2}{\beta_{a_{s2}}+\alpha a_{s2}}$ twice. Note that under the constraint from $\Pi_2 > 0$, a_{s2} can vary from 0 to $\frac{\beta_{a_{s2}}}{1-\alpha}$, and the function $2a_{s2} - \frac{a_{s2}^2}{\beta_{a_{s2}}+\alpha a_{s2}}$ is monotonically increasing in the domain of a_{s2}. This is because its slope w.r.t. $\beta_{a_{s2}}$ is:

$$2 - \frac{a_{s2}}{\beta_{a_{s2}} + \alpha a_{s2}}\left(1 + \frac{\beta_{a_{s2}}}{\beta_{a_{s2}} + \alpha a_{s2}}\right) \tag{A.24}$$

The above expression for the slope is always positive since both $\frac{a_{s2}}{\beta_{a_{s2}}+\alpha a_{s2}} < 1$ and $\frac{\beta_{a_{s2}}}{\beta_{a_{s2}}+\alpha a_{s2}} < 1$. Since the function is monotonically increasing, it can not have more than one intersection with the constant baseline of γ_5. Hence, increasing a_{s2} can only create at most one transition from dedicated to shared and vice-versa.

A.4.4 Varying a_{s1} and a_{s2} together

Lastly, consider the case when a_{s1} and a_{s2} vary together. As before, if we substitute K_2^*, K_{s2}^* in $\Pi_s > \Pi_d$ and rearrange the terms with a_{s1} and a_{s2} on one side of the inequality, we get:

$$\gamma_6 < 2a_{s2} - \frac{a_{s2}^2}{(1-\alpha)(p_2 - v_{s2}) + \alpha a_{s2}} + 2a_{s1}\frac{X_1}{X_2^{max}} \tag{A.25}$$

where

$$\begin{aligned}\gamma_6 = \quad & 2a_{d2} - \frac{a_{d2}^2}{\beta_{a_{s2}} + \alpha a_{d2}} + \frac{2X_1}{X_2^{max}}a_{d1} - \frac{2}{X_2^{max}}\left[\left(\frac{v_{s2}X_2^{max}}{2}\right.\right. \\ & \left.\left. + v_{s1}X_1 + c_s\right) - \left(\frac{v_{d2}X_2^{max}}{2} + v_{d1}X_1 + c_{d1} + c_{d2}\right)\right]\end{aligned} \tag{A.26}$$

Substituting $a_{s2} = \lambda b_s + q_{s2}$ and $a_{s1} = b_s + q_{s1}$ leads to the following inequality:

$$\gamma_7 < \psi_1 b_s - \frac{(\lambda b_s + q_{s2})^2}{\beta_{a_{s1}a_{s2}} + \alpha\lambda b_s} \tag{A.27}$$

where $\gamma_7 = \gamma_6 - 2q_{s2} - 2q_{s1}\frac{X_1}{X_2^{max}}$, $\beta_{a_{s1}a_{s2}} = (1-\alpha)(p_2 - v_{s2}) + \alpha q_{s2}$ and $\psi_1 = 2\lambda + \frac{2X_1}{X_2^{max}}$.

We will study the impact of changing the parameter b_s which varies both a_{s1} and a_{s2}. Specifically, we need to show that increasing b_s can not create situations where the preferred network choice switches between shared and dedicated twice. The proof that two switches does not occur requires proving that γ_7 cannot intersect the left hand side function $\psi_1 b_s - \frac{(\lambda b_s + q_{s2})^2}{\beta_{a_{s1}a_{s2}} + \alpha\lambda b_s}$ twice. b_s can take values between 0 and $\frac{\beta_{a_{s1}a_{s2}} - q_{s2}}{(1-\alpha)\lambda}$ (from the profit constraint $p_2 - v_{s2} - a_{s2} > 0$). This function is monotonically increasing in the domain of b_s because its derivative w.r.t b_s *i.e.,* $2\lambda(1 - (\frac{\lambda b_s + q_{s2}}{\beta_{a_{s1}a_{s2}}) + \frac{2X_1}{X_2^{max}} + \alpha\lambda b_s}) + \alpha\lambda(\frac{\lambda b_s + q_{s2}}{\beta_{a_{s2}a_{s2}} + \alpha\lambda b_s})^2$, is always positive. This is because $(\frac{\lambda b_s + q_{s2}}{\beta_{a_{s1}a_{s2}} + \alpha\lambda b_s})$ is a fraction (from the upper bound on $b_s = \frac{\beta_{a_{s1}a_{s2}} - q_{s2}}{(1-\alpha)\lambda}$), and all the other parameters are positive. But since γ_7 is a constant and the function is monotonically increasing, there can be at most one crossing. Once again, we find that the network choice cannot alter back and forth between shared and dedicated more than once.

The analysis shows that the diseconomies/economies of scope in cost components can trigger at most one switch in the network choice *i.e.,* from shared to dedicated or vice-versa. These translate to choosing one type of network when the parameter value is small (as would be the case if economies of scope were to be realized in some costs) and the other when it is large (*e.g.,* diseconomies of scope). Moreover, we showed that there cannot be switches back and forth between shared and dedicated networks as the costs increase/decrease. In other words, outcomes where one type of network is preferred at both high and low costs, and the other for intermediate values do not arise. This result is in contrast with the behavior observed previously for the parameter α (reprovisioning coefficient), and hence provides additional support for models that incorporate reprovisioning ability in its analysis of network choice.

Appendix B

Appendix: Network Technology Adoption

B.1 Generalized Region Partition

The analysis of solution regions in the above subsection was based on the fact that for linear externality functions, the (x_1, x_2)-plane can be partitioned into nine regions, each representing a unique ordering of the indifference points, and therefore different expressions for $H_i(\underline{x})$ and diffusion trajectory. The unique points P and Q on the line $\theta_1^0(\underline{x}) = \theta_2^0(\underline{x})$ acted as 'pivots' for the partition of the plane into nine regions. Although one may expect that for arbitrary externality functions, the lines denoting the region boundaries will intersect in arbitrary ways, we show here that even for more generic monotonically increasing network externality functions, the two 'pivot' points

P and Q remain unique. This uniqueness of P and Q, along with constraints on how the boundary lines can intersect as imposed by the monotonic property of externality functions, result in the partitioning of the plane into "nine" regions.

Let the network externality of the two technologies, Technology 1 and 2 be given by positive increasing externality functions $g_1(x_1)$ and $g_2(x_2)$ for the respective adoption levels of x_1 and x_2 (*i.e.*, $g_i(x_i) \geq 0,\ g_i'(x_i) > 0,\ i = \{1,2\}$).

The end user's utility from using Technologies 1 and 2 is given by:

$$U_1 = \theta q_1 + (g_1(x_1) + \alpha_1 \beta g_2(x_2)) - p_1 \tag{B.1}$$

$$U_2 = \theta q_2 + (\beta g_2(x_2) + \alpha_2 g_1(x_1)) - p_2 \tag{B.2}$$

Setting $U_i(\theta, \underline{x}) = 0$, we get

$$\theta_1^0(\underline{x}) = \frac{p_1 - (g_1(x_1) + \alpha_1 \beta g_2(x_2))}{q_1} \tag{B.3}$$

$$\theta_2^0(\underline{x}) = \frac{p_2 - (\beta g_2(x_2) + \alpha_2 g_1(x_1))}{q_2} \tag{B.4}$$

Similarly, setting $U_1(\theta, \underline{x}) = U_2(\theta, \underline{x})$ gives

$$\theta_2^1(\underline{x}) = \frac{(1-\alpha_2) g_1(x_1) - \beta(1-\alpha_1) g_2(x_2) + p_2 - p_1}{q_2 - q_1} \tag{B.5}$$

To simplify notation, we use from now on θ_i^0 and θ_2^1 instead of $\theta_i^0(\underline{x})$ and $\theta_2^1(\underline{x})$. After simple manipulations, we get

$$\theta_2^1 - \theta_1^0 = \frac{q_2}{q_2 - q_1}(\theta_2^0 - \theta_1^0), \tag{B.6}$$

$$\theta_2^1 - \theta_2^0 = \frac{q_1}{q_2 - q_1}(\theta_2^0 - \theta_1^0) \tag{B.7}$$

Given that Technology 2, the entrant, is technically superior (*i.e.,* $q_2 > q_1$), from the above relation we establish the following Proposition.

Proposition 13 *If* $\theta_1^0 < \theta_2^0$*, then* $\theta_2^1 > \theta_2^0 > \theta_1^0$*. If* $\theta_1^0 \geq \theta_2^0$*, then* $\theta_2^1 \leq \theta_2^0 \leq \theta_1^0$.

$$H_1(\underline{x}) = \begin{cases} [\theta_2^1]_{[0,1]} - [\theta_1^0]_{[0,1]} & \text{if } \theta_1^0 < \theta_2^0 \\ 0 & \text{otherwise} \end{cases}$$
$$H_2(\underline{x}) = \begin{cases} 1 - [\theta_2^1]_{[0,1]} & \text{if } \theta_1^0 < \theta_2^0 \\ 1 - [\theta_2^0]_{[0,1]} & \text{otherwise} \end{cases} \tag{B.8}$$

where $x_{[a,b]}$ is the projection of x into the interval $[a,b]$, *i.e.,* is equal to x for $x \in [a,b]$, a for $x < a$, and b for $x > b$.

As the preference levels θ of all users lie in $[0,1]$, Equation (B.8) fully determine $H_i(\underline{x})$, albeit with possibly different expressions depending on the outcome of the projections of the indifference thresholds on $[0,1]$. Hence, we partition the (x_1,x_2) plane into regions where $H_i(\underline{x})$ has a unique expression. This can be achieved by combining Equation (B.3) to (B.5) with Equation (B.8).

The line $\theta_1^0(\underline{x}) = \theta_2^0(\underline{x})$ separates regions that require different expressions of $H_i(\underline{x}), i = \{1,2\}$. For points above that line $(\theta_1^0(\underline{x}) > \theta_2^0(\underline{x}))$, the expression of $H_2(\underline{x})$ depends on the projection of $\theta_2^0(\underline{x})$ on $[0,1]$. Therefore the lines $\theta_2^0(\underline{x}) = 0$ and $\theta_2^0(\underline{x}) = 1$ delineate regions associated with different $H_2(\underline{x})$. Similarly, for points below that line, the lines $\theta_1^0(\underline{x}) = 0$, $\theta_1^0(\underline{x}) = 1$, $\theta_2^1(\underline{x}) = 0$ and $\theta_2^1(\underline{x}) = 1$ introduce additional region boundaries for $H_i(\underline{x})$, $i = \{1,2\}$. Expressions for all the lines can be obtained from eqs. (B.3) to (B.5).

Next we show that irrespective of the choice of system parameters, the lines $\theta_2^0(\underline{x}) = 0, \theta_1^0(\underline{x}) = 0$ and $\theta_2^1(\underline{x}) = 0$ always intersect at a point P, and the lines $\theta_2^0(\underline{x}) = 1, \theta_1^0(\underline{x}) = 1$ and $\theta_2^1(\underline{x}) = 1$ always intersect at a point Q, with both P and Q lying on the line $\theta_1^0(\underline{x}) = \theta_2^0(\underline{x})$. The points P and Q act as "anchors" of the partition of the solution space.

We denote the lines $\theta_1^0(\underline{x}) = 0$ and $\theta_2^0(\underline{x}) = 0$ by functions $f_1(x_1,x_2) = 0$ and $f_2(x_1,x_2) = 0$. Let (x_1^*,x_2^*) denote the co-ordinates in (x_1,x_2)-plane where these lines intersect (*i.e.*, $f_i(x_1^*,x_2^*) = 0, i = \{1,2\}$). Note that the lines $\theta_2^1(\underline{x}) = 0$ and $\theta_1^0(\underline{x}) = \theta_2^0(\underline{x})$, which can then be represented as $f_2(x_1,x_2) - f_1(x_1,x_2) = 0$ and $(1/q_2)f_2(x_1,x_2) - (1/q_1)f_1(x_1,x_2) = 0$ respectively, also must pass through (x_1^*,x_2^*). This point of intersection of all these lines can be labeled as P. Additionally, it can be seen that if any two of these lines intersect at some point, all the other curves must also pass through that point. Similarly, we obtain the other 'pivot' point, Q, at which the lines $\theta_1^0(\underline{x}) = \theta_2^0(\underline{x}), \theta_2^0(\underline{x}) = 1, \theta_1^0(\underline{x}) = 1$ and $\theta_2^1(\underline{x}) = 1$ must intersect.

Proof of Uniqueness of P ***and*** Q

Let us consider the intersection of the lines $\theta_1^0(\underline{x}) = 0$, $\theta_2^0(\underline{x}) = 0$, $\theta_2^1(\underline{x}) = 0$ and $\theta_1^0(\underline{x}) = \theta_2^0(\underline{x})$ at some point P as shown in Figure B.1. Assume that there exist another

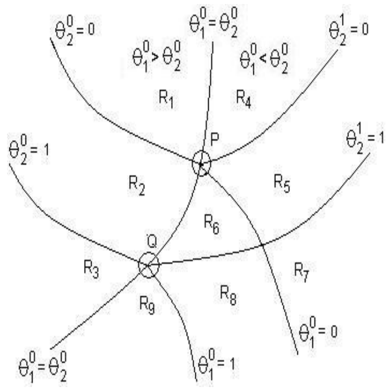

Figure B.1: Generalized Region Partitions

such point P' where all the lines again intersect (because we showed that whenever any two of these lines intersect, the other lines should also intersect). However, using eqn.(3.5-3.7) and $g_i'(x_i) \geq 0,\ i = \{1,2\}$, we see that the line $\theta_2^1 = 0$ is always increasing in x_1 and x_2, while the line $\theta_0^1 = 0$ is decreasing in x_2 for increase in x_1. Therefore in the entire region $\theta_1^0 < \theta_2^0$ the lines can only intersect once, and therefore the point P must be unique. A similar argument holds for point Q as well.

Thus the (x_1, x_2)-plane can only be partitioned into the nine regions shown in Figure B.1, and the relative positions of these regions in the plane remain fixed. Moreover, each of these regions is a connected set. As shown in the figure, each region corresponds to a different arrangement of the indifference points with respect to the $0,1$ boundary under the two feasible orderings (from Proposition 13). The classification of these nine regions based on the different orderings is provided in Table B.1. A brief explanation of the meaning of the regions is provided next.

Meaning of Regions

Table B.1: Partitions characterizing $H_i(\underline{x})$

$\theta_1^0 \geq \theta_2^0$		$\theta_1^0 < \theta_2^0$		
Region	condition	Region	condition	
R_1	$\theta_2^0 \leq 0$	R_4	$\theta_2^1 \leq 0,$	$0 \leq \theta_1^0$
R_2	$0 < \theta_2^0 < 1$	R_5	$0 < \theta_2^1 < 1,$	$\theta_1^0 \leq 0$
R_3	$1 \leq \theta_2^0$	R_6	$0 < \theta_2^1 < 1,$	$0 < \theta_1^0 < 1$
		R_7	$1 \leq \theta_2^1,$	$\theta_1^0 \leq 0$
		R_8	$1 \leq \theta_2^1,$	$0 < \theta_1^0 < 1$
		R_9	$1 \leq \theta_2^1,$	$1 \leq \theta_1^0$

Each region correspond to a particular ordering of the indifference points, which in turn maps to unique expressions for $H_i(\underline{x})$ in eqn.(B.8). For example, consider the region R_8, which is the set of all (x_1, x_2) penetration levels for which $\theta_1^0 < \theta_2^0$, $1 \leq \theta_1^2$ and $0 \leq \theta_1^0 < 1$. In this region, because $\theta_1^2 > 1$ for any current (x_1, x_2) adoption levels, no user has a preference for choosing Technology 2 over 1. But since $0 < \theta_1^0 < 1$, some users whose preference $\theta_1^0 < \theta$ derive positive utility from Technology 1, will be willing to adopt it. Therefore if at any instant t, the system reaches adoption levels (x_1, x_2) in Region R_8, the diffusion from that point on in the inside of Region R_8 will proceed with a decrease in the value of x_2 *i.e.,* users leave Technology 2 as $\theta_2^1 > 1$. This fact is also reflected in the exponentially decreasing value of the $x_2(t)$ co-ordinate of the diffusion trajectory in R_8 (as given in Table B.2). Each region can be interpreted in a similar manner, and Table B.1 essentially connects this abstract notion of each region to the corresponding ordering of the indifference thresholds that define it.

B.2 Trajectories and Equilibria (validity & stability)

The expressions for equilibria in R_5 and R_6, $\underline{x}^*_{R_5}$ and $\underline{x}^*_{R_6}$ of Tables B.2, B.3 and B.4 are provided separately as eqs. (B.9) and (B.10) respectively for better readability.

$$
\begin{aligned}
x^*_{1\ R_5} &= \frac{(p_2-p_1)-\beta(1-\alpha_1)}{(q_2-q_1)-[(1-\alpha_2)+\beta(1-\alpha_1)]} \\
x^*_{2\ R_5} &= 1-x^*_{1\ R_5} = \frac{(q_2-q_1)-(p_2-p_1)-(1-\alpha_2)}{(q_2-q_1)-[(1-\alpha_2)+\beta(1-\alpha_1)]}
\end{aligned} \tag{B.9}
$$

$$
\begin{aligned}
x^*_{1\ R6} &= \frac{p_1q_2-p_2q_1+\beta\alpha_1(p_2-q_2)-\beta(p_1-q_1)}{(q_1-1)(\beta-q_2)+(q_1-\alpha_1\beta)(q_1-\alpha_2)} \\
x^*_{2\ R6} &= \frac{p_2q_1-p_1q_1-p_2+p_1\alpha_2+{q_1}^2-q_1q_2+q_2-q_1\alpha_2}{(q_1-1)(\beta-q_2)+(q_1-\alpha_1\beta)(q_1-\alpha_2)}
\end{aligned} \tag{B.10}
$$

Table B.2: Technology adoption trajectories

	$x_1(t)$	$x_2(t)$
R_1	$x_1(t_0)e^{-\gamma(t-t_0)}$	$\frac{(x_2(t_0)-1)}{x_1(t_0)}e^{-\gamma t}+1$
R_2	$c_1e^{-\gamma(t-t_0)}$ $c_1 = x_1(t_0)$	$\frac{p_2-q_2}{\beta-q_2}+c_2e^{-\gamma(1-\beta/q_2)(t-t_0)}-c_1\frac{\alpha_2}{\beta}e^{-\gamma t}$ $c_2 = [x_2(t_0)+\frac{\alpha_2 x_1(t_0)}{\beta}-\frac{p_2-q_2}{\beta-q_2}]$
R_3	$x_1(t) = x_1(t_0)e^{-\gamma(t-t_0)}$	$x_2(t) = x_2(t_0)e^{-\gamma(t-t_0)}$
R_4	same as R_1	same as R_1
R_5	$x^*_{1R_5}+c_2e^{-\gamma(t-t_0)}$ $+c_1e^{(-1+\frac{(1-\alpha_2)+\beta(1-\alpha_1)}{q_2-q_1})\gamma(t-t_0)}$	$x^*_{2R_5}+c_2\frac{(1-\alpha_2)}{\beta(1-\alpha_1)}e^{-\gamma(t-t_0)}$ $-c_1e^{(-1+\frac{(1-\alpha_2)+\beta(1-\alpha_1)}{q_2-q_1})\gamma(t-t_0)}$
	$c_1 = \frac{\beta(1-\alpha_1)}{1-\alpha_2+\beta(1-\alpha_1)}(x^*_{2R_5}-x_2(t_0))$ $-\frac{1-\alpha_2}{1-\alpha_2+\beta(1-\alpha_1)}(x^*_{1R_5}-x_1(t_0))$	$c_2 = (\frac{\beta(1-\alpha_1)}{1-\alpha_2+\beta(1-\alpha_1)})[x_1(t_0)+x_2(t_0)-1]$
R_6	$x^*_{1R_6}+c_1K_1e^{\frac{A+\sqrt{A^2-4B}}{2}\gamma(t-t_0)}$ $+c_2K_2e^{\frac{A-\sqrt{A^2-4B}}{2}\gamma(t-t_0)}$	$x^*_{2R_6}+c_1e^{\frac{A+\sqrt{A^2-4B}}{2}\gamma(t-t_0)}$ $+c_2e^{\frac{A-\sqrt{A^2-4B}}{2}\gamma(t-t_0)}$
	$c_1 = \frac{(1-\alpha_2)\{x^*_{1R_6}-x_1(t_0)-K_2(x^*_{2R_6}-x_2(t_0))\}}{(q_2-q_1)\sqrt{A^2-4B}}$ $K_1 = \frac{\alpha_2+\beta(1-\alpha_1)-q_2/q_1-(q_2-q_1)\sqrt{A^2-4B}}{2(1-\alpha_2)}$ $K_2 = \frac{\alpha_2+\beta(1-\alpha_1)-q_2/q_1+(q_2-q_1)\sqrt{A^2-4B}}{2(1-\alpha_2)}$	$c_2 = \frac{(1-\alpha_2)\{-(x^*_{1R_6}-x_1(t_0))+K_1(x^*_{2R_6}-x_2(t_0))\}}{(q_2-q_1)\sqrt{A^2-4B}}$ $A = \frac{1-\alpha_2+\beta(1-\alpha_1)}{q_2-q_1}+\frac{1}{q_1}-2$ $B = (\frac{1}{q_1}-1)(\frac{\beta(1-\alpha_1)}{q_2-q_1}-1)+\frac{1-\alpha_2}{q_2-q_1}(\frac{\beta\alpha_1}{q_1}-1)$
R_7	$(x_1(t_0)-1)e^{-\gamma(t-t_0)}+1$	$x_2(t_0)e^{-\gamma(t-t_0)}$
R_8	$\frac{p_1-q_1}{1-q_1}+c_1e^{-\gamma(1-\frac{1}{q_1})(t-t_0)}-c_2\beta\alpha_1e^{-\gamma(t-t_0)}$ $c_1 = [x_1(t_0)+\beta\alpha_1x_2(t_0)-\frac{p_1-q_1}{1-q_1}]$	$c_2e^{-\gamma(t-t_0)}$ $c_2 = x_2(t_0)$
R_9	$x_1(t) = x_1(t_0)e^{-\gamma(t-t_0)}$	$x_2(t) = x_2(t_0)e^{-\gamma(t-t_0)}$

Table B.3: Conditions for stable, valid Equilibria

Region	Equilibria	Stability Conditions	Validity and Stability Conditions
R_1	$(0,1)$	always locally stable	$p_2 \leq \beta$, $\alpha_1 \leq \frac{p_1}{\beta} + \frac{q_1}{q_2}(1 - \frac{p_2}{\beta})$
R_2	$(0, \frac{p_2-q_2}{\beta-q_2})$	$\beta < q_2$	$\beta < p_2 < q_2$ $\alpha_1\beta(q_2 - p_2) \leq$ $\beta(q_1 - p_1) + p_1q_2 - p_2q_1$
R_4	$(0,1)$	always locally stable	$p_1 < \alpha_1\beta$, $\frac{p_1}{\beta} + \frac{q_1}{q_2}(1 - \frac{p_2}{\beta}) \leq \alpha_1 \leq 1 + \frac{p_1-p_2}{\beta}$
R_5	$(x^*_{1R_5}, 1 - x^*_{1R_5})$ (See Eq. (B.9))	$q_2 - q_1 >$ $1 - \alpha_2 + \beta(1 - \alpha_1)$	$p_2 - p_1 > \beta(1 - \alpha_1)$ $q_2 - q_1 - (p_2 - p_1) \geq 1 - \alpha_2$ $q_2 - q_1 > \beta(1 - \alpha_1) + 1 - \alpha_2$ $\alpha_1\beta(\alpha_2 + q_2 - q_1 - p_2) \geq$ $\beta - p_2 - p_1(\beta - \alpha_2 - (q_2 - q_1))$
R_6	$(x^*_{1R_6}, x^*_{2R_6})$ (See Eq. (B.10))	See Table B.4	$0 < x^*_{1R_6}$, $0 < x^*_{2R_6}$, $0 < x^*_{1R_6} + x^*_{2R_6} < 1$
R_7	$(1,0)$	always locally stable	$p_1 \leq 1$, $\alpha_2 \leq 1 + p_2 - p_1 - (q_2 - q_1)$
R_8	$(\frac{p_1-q_1}{1-q_1}, 0)$	$1 < q_1$	$1 < p_1 < q_1$ $\alpha_2(q_1 - p_1) \leq$ $(1 - q_1)(q_2 - p_2) + q_1(q_1 - p_1)$

Table B.4: Stability conditions for $\underline{x}^*_{R_6}$

Case	Conditions
$A^2 - 4B \geq 0$ (Ref. Table B.2 for exp. of A and B)	$A < 0 \Leftrightarrow \beta(1-\alpha_1) - \alpha_2 < 2(q_2 - q_1) - \frac{q_2}{q_1}$ $B > 0 \Leftrightarrow (q_1 - 1)(\beta - q_2) + (q_1 - \alpha_1\beta)(q_1 - \alpha_2) < 0$
$A^2 - 4B < 0$	$A < 0 \Leftrightarrow \beta(1-\alpha_1) - \alpha_2 < 2(q_2 - q_1) - \frac{q_2}{q_1}$

B.3 Conditions for valid and stable equilibria

The expressions in Table B.3 and B.4 for the validity and stability conditions for each of the equilibrium are rearranged and presented below for clarity.

Region R_1 – Equilibrium: $(0,1)$

Validity and Stability Conditions are:

$$p_2 \leq \beta \tag{B.11}$$

$$\alpha_1 \leq \frac{p_1}{\beta} + \frac{q_1}{q_2}\left(1 - \frac{p_2}{\beta}\right) \tag{B.12}$$

Region R_2 – Equilibrium: $(0, \frac{p_2 - q_2}{\beta - q_2})$

Validity and Stability Conditions:

$$\beta < p_2 < q_2 \tag{B.13}$$

$$\alpha_1 \leq \frac{\beta(q_1 - p_1) + p_1 q_2 - p_2 q_1}{\beta(q_2 - p_2)} \tag{B.14}$$

Region R_3 – Candidate equilibrium $(0,0)$

By assumption, this is not a feasible equilibrium.

Region R_4 – Equilibrium: $(0,1)$

Validity and Stability Conditions:

$$p_1 \leq \alpha_1 \beta \tag{B.15}$$

$$\frac{p_1}{\beta} + \frac{q_1}{q_2}\left(1 - \frac{p_2}{\beta}\right) \leq \alpha_1 \leq \frac{p_1}{\beta} + 1 - \frac{p_2}{\beta} \tag{B.16}$$

Region R_5 – Equilibrium: $(x^*_{1_{R5}}, x^*_{2_{R5}})$

$$\begin{aligned} x^*_{1_{R5}} &= \frac{(p_2-p_1)-\beta(1-\alpha_1)}{(q_2-q_1)-[\beta(1-\alpha_1)+(1-\alpha_2)]} \\ x^*_{2_{R5}} &= 1-x^*_{1_{R5}} = \frac{(q_2-q_1)-(p_2-p_1)-(1-\alpha_2)}{(q_2-q_1)-[\beta(1-\alpha_1)+(1-\alpha_2)]} \end{aligned}$$

Rewriting the Validity and Stability conditions given in Table B.3, we have:

$$p_2-p_1 > \beta(1-\alpha_1) \tag{B.17}$$

$$q_2-q_1-(p_2-p_1) \geq 1-\alpha_2 \tag{B.18}$$

$$q_2-q_1 > \beta(1-\alpha_1)+1-\alpha_2 \tag{B.19}$$

$$\alpha_1\beta(\alpha_2+q_2-q_1-p_2) \geq \beta-p_2-p_1(\beta-\alpha_2-(q_2-q_1)) \tag{B.20}$$

Region R_6 – Equilibrium: $(x^*_{1_{R6}}, x^*_{2_{R6}})$

$$\begin{aligned} x^*_{1_{R6}} &= \frac{p_1q_2-p_2q_1+\beta\alpha_1(p_2-q_2)-\beta(p_1-q_1)}{(q_1-1)(\beta-q_2)+(q_1-\alpha_1\beta)(q_1-\alpha_2)} \\ x^*_{2_{R6}} &= \frac{p_2q_1-p_1q_1-p_2+p_1\alpha_2+q_1^2-q_1q_2+q_2-q_1\alpha_2}{(q_1-1)(\beta-q_2)+(q_1-\alpha_1\beta)(q_1-\alpha_2)} \end{aligned}$$

The validity conditions for these equilibrium expression requires $x^*_{1_{R6}} \geq 0$, $x^*_{2_{R6}} \geq 0$ and $x^*_{1_{R6}} + x^*_{2_{R6}} < 1$. We will denote the numerators of $x^*_{1R_6}$ and $x^*_{2R_6}$ in eq. (B.10) as N_1 and N_2, respectively, and their common denominator as D. Table B.4 shows that the if $A^2-4B \geq 0$ the stability conditions require $A < 0$ and $B > 0$, while if $A^2-4B < 0$ then $A < 0$ is required (and $B > 0$ since $B > A^2/4 > 0$). Thus an equilibrium in R_6 can satisfy stability conditions only if $B > 0$ and $A < 0$. Additionally $B > 0$ implies that the denominator of the expressions for the equilibrium adoption levels (given in eqs. (B.10)) is negative (*i.e., $D < 0$*).

Therefore a valid, stable equilibrium in R_6 must have:

$$A < 0:$$

$$\beta(1-\alpha_1)-\alpha_2 < 2(q_2-q_1)-\frac{q_2}{q_1} \tag{B.21}$$

$$D < 0 \; (B > 0):$$

$$(q_1-1)(\beta-q_2)+(q_1-\alpha_1\beta)(q_1-\alpha_2) < 0 \tag{B.22}$$

$$N_1 \leq 0:$$

$$\alpha_1\beta(q_2-p_2) \geq \beta(q_1-p_1)+p_1q_2-p_2q_1 \tag{B.23}$$

$$N_2 \leq 0:$$

$$\alpha_2(q_1-p_1) \geq (1-q_1)(q_2-p_2)+q_1(q_1-p_1) \tag{B.24}$$

$$\frac{N_1+N_2}{D} < 1:$$

$$\alpha_1\beta(\alpha_2+q_2-q_1-p_2)$$

$$< \beta-p_2-p_1(\beta-\alpha_2-(q_2-q_1)) \tag{B.25}$$

Region R_7 – Equilibrium: $(1,0)$

Validity and Stability Conditions:

$$p_1 \leq 1 \tag{B.26}$$

$$\alpha_2 < 1+(p_2-p_1)-(q_2-q_1) \tag{B.27}$$

Region R_8 – Equilibrium: $(\frac{p_1-q_1}{1-q_1},0)$

Validity and Stability Conditions:

$$1 < p_1 < q_1 \tag{B.28}$$

$$\alpha_2(q_1 - p_1) \leq (1-q_1)(q_2 - p_2) + q_1(q_1 - p_1) \tag{B.29}$$

Region R_9 – Equilibrium: $(0,0)$

By assumption, $(0,0)$ is not a feasible equilibrium.

B.4 Proofs of Propositions

Proof of Proposition 5:

In this proof we will show that the following pairs of equilibria cannot coexist together as valid and stable equilibria. Consequently, it is easy to verify that the only combination of multiple equilibria that can coexist are the ones mentioned in Proposition 5.

1. $(1,0)$ and $\underline{x}^*_{R_8}$
2. $(1,0)$ and $\underline{x}^*_{R_5}$
3. $(0,1)$ and $\underline{x}^*_{R_2}$
4. $(0,1)$ and $\underline{x}^*_{R_5}$
5. $\underline{x}^*_{R_8}$ and $\underline{x}^*_{R_5}$
6. $\underline{x}^*_{R_8}$ and $\underline{x}^*_{R_6}$
7. $\underline{x}^*_{R_2}$ and $\underline{x}^*_{R_6}$

8. $\underline{x}^*_{R_5}$ and $\underline{x}^*_{R_6}$

The following analysis will use the expressions for validity and stability conditions for the different equilibria listed in the Subsection B.3 of the Appendix.

1. $(1,0)$ and $\underline{x}^*_{R_8}$

This pair cannot coexist because the equilibrium $(1,0)$ in R_7 requires $p_1 \leq 1$ (eq. (B.26)) while the equilibrium $\underline{x}^*_{R_8}$ requires $p_1 > 1$ (eq. (B.28)).

2. $(1,0)$ and $\underline{x}^*_{R_5}$

Eq.(B.27) for equilibrium $(1,0)$ in R_7 and condition in eq.(B.18) for $\underline{x}^*_{R_5}$ have contradictory requirements, and therefore these pair cannot coexist.

3. $(0,1)$ and $\underline{x}^*_{R_2}$

The equilibrium $(0,1)$ can either exist in Region R_1 or R_4. In either it requires $\beta \geq p_2$ for being a valid, stable equilibrium. While this is explicit for Region R_1 (refer to eq.(B.11)), the relation is implicitly implied by the conditions of Region R_4. Note that eq.(B.15) and eq.(B.16) can be written together as $0 \leq \alpha_1\beta - p_1 \leq \beta - p_2$. Thus $(0,1)$ equilibria requires $\beta \geq p_2$ which contradicts with the requirement in eq.(B.13) for $\underline{x}^*_{R_2}$.

4. $(0,1)$ and $\underline{x}^*_{R_5}$

The equilibrium $(0,1)$ is valid in R_4 if $\alpha_1 \leq \frac{p_1}{\beta} + 1 - \frac{p_2}{\beta}$. It is valid in R_1 if the bound is stricter *i.e.,* $\alpha_1 \leq \frac{p_1}{\beta} + \frac{q_1}{q_2}\left(1 - \frac{p_2}{\beta}\right)$ (since $\frac{q_1}{q_2} < 1$ and $\beta \geq p_2$).

However equilibrium $\underline{x}^*_{R_5}$ requires $\alpha_1 > \frac{p_1}{\beta} + 1 - \frac{p_2}{\beta}$ from eq.(B.17). Therefore the two equilibria cannot coexist.

5. $\underline{x}^*_{R_8}$ and $\underline{x}^*_{R_5}$

Equilibria $\underline{x}^*_{R_8}$ requires $q_1 > p_1$ from eq.(B.28). When this relation holds, the condition in eq.(B.29) for $\underline{x}^*_{R_8}$ and eq.(B.18) for equilibria $\underline{x}^*_{R_5}$ can be written as:

$$\begin{aligned}
\frac{(1-q_1)(q_2-p_2)+q_1(q_1-p_1)}{q_1-p_1} &\geq \alpha_2 \\
&> 1-(q_2-q_1)+p_2-p_1 \\
\Rightarrow \quad (q_2-q_1-(p_2-p_1)) & \\
&> p_1(q_2-q_1-(p_2-p_1)) \qquad \text{(B.30)}
\end{aligned}$$

Since $(q_2-q_1-(p_2-p_1)) \geq 1-\alpha_2 \geq 0$ from eq.(B.18), we must have $p_1 < 1$, which contradicts with the requirement in eq.(B.28) for the equilibrium $\underline{x}^*_{R_8}$. Therefore this pair cannot coexist.

6. $\underline{x}^*_{R_8}$ and $\underline{x}^*_{R_6}$

Condition in eqs.(B.28) and (B.29) when considered together contradicts the requirement of eq.(B.24). Therefore these equilibria cannot coexist as valid, stable equilibria.

7. $\underline{x}^*_{R_2}$ and $\underline{x}^*_{R_6}$

Eqs.(B.13) and (B.14) for $\underline{x}^*_{R_2}$ together contradict the requirement in eq.(B.23) for $\underline{x}^*_{R_6}$, and thus cannot coexist as valid, stable equilibria pair.

8. $\underline{x}^*_{R_5}$ and $\underline{x}^*_{R_6}$

The condition in eq.(B.20) for $\underline{x}^*_{R_5}$ and eq.(B.25) cannot hold together and therefore these equilibria never coexists as a pair of valid, stable equilibria.

Proof: No combination of three or more equilibria can coexist as valid, stable equilibria in the presence of converters

Given Proposition 5, all but one combination of three equilibria can be excluded from further consideration as at least a pair of equilibria in these combinations will not coexist as per the proposition. The only combination of three equilibria that can potentially coexist is $\{(0,1),(1,0),\underline{x}^*_{R_6}\}$. We will show that the validity and stability conditions of these three equilibria cannot be satisfied together. Thus since no pair of three equilibria may coexist, it will directly follow that no combination of four or more equilibria can therefore coexist, thus proving the present proposition.

The equilibrium $(0,1)$ to exist in R_4 can be shown to require $\beta \geq p_2$ and $\alpha_1\beta \leq \beta + p_1 - p_2$ from eqs.(B.15) and (B.16). For $(0,1)$ to exist in R_1 the constraint imposed by eq.(B.12) is even more stringent than $\alpha_1\beta \leq \beta + p_1 - p_2$. Therefore the validity and stability of $(0,1)$ requires at least $\alpha_1\beta \leq \beta + p_1 - p_2$ and $\beta \geq p_2$. Using this and eq.(B.27) for $(1,0)$ in R_7, we have:

$$
\begin{aligned}
(\alpha_1 - 1)\beta \quad \leq p_1 - p_2 \leq & \quad 1 - \alpha_2 - (q_2 - q_1) \\
\Rightarrow (\alpha_1 - 1)\beta \quad \leq & \quad 1 - \alpha_2 - (q_2 - q_1)
\end{aligned}
$$

Using the above inequality and eq.(B.21), we get:

$$
\begin{aligned}
\Rightarrow q_2 - q_1 - 1 &\leq \beta(1-\alpha_1) - \alpha_2 \\
&< 2(q_2 - q_1) - q_2/q_1 \\
\Rightarrow (q_2 - q_1)(q_1 - 1) &> 0 \\
\Rightarrow q_1 &> 1 \ \ (as \ \ q_2 > q_1)
\end{aligned}
$$

Eq.(B.24) gives:

$$
(\alpha_2 - q_1)(q_1 - p_1) \geq (1 - q_1)(q_2 - p_2)
$$

Since we have $1 \geq p_1$ from eq.(B.26) as a condition for the $(1,0)$ equilibrium in R_7 and we established that $q_1 > 1$, we get $q_1 > p_1$. This in addition to the relation $q_1 > 1 \geq \alpha_2$, enforces $q_2 > p_2$ for the previous inequality expression.

Eq.(B.24) could also be rearranged as:

$$(\alpha_2 - q_1 + q_2 - p_2)(q_1 - p_1) \geq (q_2 - p_2)(1 - p_1)$$

The expression on the right hand side is positive since $q_2 > p_2$ and $1 \geq p_1$ as discussed previously. Therefore the left hand side expression also needs to be positive. Using eq.(B.25) and $q_1 > p_1$, we must have $\alpha_2 - q_1 + q_2 - p_2 > 0$.

Now using eqs.(B.23) and (B.25), and the facts $q_2 > p_2$ and $\alpha_2 - q_1 + q_2 - p_2 > 0$ as established above, we can write:

$$\begin{aligned}
\frac{\beta(q_1 - p_1) + p_1 q_2 - p_2 q_1}{q_2 - p_2} &\leq \alpha_1 \beta \\
&< \frac{\beta - p_2 - p_1(\beta - \alpha_2 - (q_2 - q_1))}{\alpha_2 + q_2 - q_1 - p_2} \\
\Rightarrow \alpha_2(q_1 - p_1) &< (1 - q_1)(q_2 - p_2) + q_1(q_1 - p_1)
\end{aligned}$$

It can be easily seen that the above inequality contradicts with the condition in eq.(B.24).

Hence all the validity and stability conditions for the three equilibria $\{(0,1),(1,0),\underline{x}^*_{R_6}\}$ cannot be satisfied together.

Additionally, since no pair of three valid, stable equilibria can coexist a set of given parameter values, it follows that no combinations of four or more equilibria can coexist either, thus completing the proof.

Proof of Proposition 6:

The proposition has two parts: converters can help a technology (i) alter market equilibrium from a scenario where it has been eliminated to one where it coexist with the other technology; (ii) and even succeed in nearly eliminating it.

Condition (i) is relatively easy to establish. Consider a scenario where one of the technologies has been eliminated, *i.e.,* an equilibrium of the form $(1,0)$, $(0,1)$, $(0,x_2^*)$ or $(x_1^*,0)$. The validity conditions from Table B.3 identify the minimum converter efficiency required to invalidate that equilibrium. From the table, such an invalidation is easily seen to correspond to the re-emergence of the other technology (these are the only equilibria whose validity conditions are compatible with the invalidation of the previous equilibrium), and thus co-existence of the two technologies.

Turning to condition (ii), assume that for a given set of system parameters, $(0,1)$ is the initial equilibrium in the absence of converters. Users with the lowest technology valuation ($\theta = 0$) must, therefore, derive greater utility from Technology 2 than Technology 1 *i.e.,* $U_1(\theta = 0) < U_2(\theta = 0)$. This implies

$$\beta > p_2 - p_1 \tag{B.31}$$

Next, using perfect, symmetric converters $(\alpha_1 = \alpha_2 = 1)$, we show that is is possible to satisfy both eq. (B.31) and the validity conditions of an equilibrium of the form $(1-x_2^*, x_2^*)$, with x_2^* arbitrarily small. This identifies a configuration satisfying condition (ii).

An equilibrium of the form $(1-x_2^*, x_2^*)$ requires that users of preference $\theta = 0$ adopt Technology 1, *i.e.,* $U_1(\theta = 0) \geq 0$, thus

$$x_2^* \geq \frac{p_1 - 1}{\beta - 1} \tag{B.32}$$

and users with preference $\theta = 1 - x_2^*$ are to be indifferent to the two technologies *i.e.,* $U_1(1-x_2^*, x_2^*) = U_2(1-x_2^*, x_2^*)$. This gives

$$x_2^* = 1 - \frac{p_2 - p_1}{q_2 - q_1} \tag{B.33}$$

From eq. (B.33), for Technology 2 to nearly disappear, *i.e.,* $x_2^* \approx 0$, we need $p_2 - p_1 \lesssim q_2 - q_1$. We also need β large enough for eqs. (B.32) and (B.31) to continue holding. Combinations of system parameters that allow these conditions to be simultaneously satisfied are easily found, which establishes that the introduction of converters can take the system from an equilibrium of the form $(0,1)$ to one of the form $(1-\varepsilon, \varepsilon)$, where $\varepsilon \approx 0$.

Consider now the reverse scenario, where the equilibrium in the absence of converters is $(1,0)$ for $\alpha_1 = \alpha_2 = 0$. For this, we need $\theta_2^1 > 1$ and $\theta_0^1 < 0$, *i.e.,*

$$1 + p_2 - p_1 > q_2 - q_1 \tag{B.34}$$

$$p_1 < 1 \tag{B.35}$$

As before, we assume next perfect, symmetric converters, and establish that with them it is possible to achieve a new equilibrium of the form $(x_1^*, 1-x_1^*)$, where $x_1^* \approx 0$. The new equilibrium requires that users with preference $\theta = 0$ derive positive utility from the Technology 1, *i.e.,*

$$(\beta - 1)x_1^* < \beta - p_1, \tag{B.36}$$

and that users with preference $\theta = x_1^*$ be indifferent to the two technologies *i.e.,* $U_1(x_1^*, 1 - x_1^*) = U_2(x_1^*, 1 - x_1^*)$.

$$x_1^* = \frac{p_2 - p_1}{q_2 - q_1} \tag{B.37}$$

It is again easy to find a combination of system parameters that simultaneously satisfy eqs. (B.34) to (B.37), while ensuring $x_1^* \approx 0$.

Proof of Proposition 7:

We first consider Technology 1 hurting itself by introducing or improving a converter. Converter efficiencies affect the expressions of the adoption levels only for the equilibria in R_5 and R_6 (eq.(B.9) and (B.10)). Region R_5 is easily eliminated from consideration as its validity conditions can be shown to force a positive derivative of x_1^* w.r.t. α_1. Therefore, the remainder of the proof focuses on a stable equilibrium in R_6.

As before, the numerators of $x_{1R_6}^*$ and $x_{2R_6}^*$ in eq. (B.10) are denoted as N_1 and N_2 respectively, and with D as their common denominator. The stability of the equilibrium can be shown to imply that $D < 0$. The requirement $D < 0$ implies $N_1 < 0$ and $N_2 < 0$, which has important consequences on the impact of converter efficiency.

Specifically, better converters hurt Technology 1 if

$$\frac{\partial x_{1R_6}^*}{\partial \alpha_1} = \frac{(\beta - q_2)N_2}{D^2} < 0 \tag{B.38}$$

Since $N_2 < 0$, the derivative is negative only if $\beta > q_2$. As a result, for better converters to hurt the incumbent, the condition $\beta > q_2$ and one of the sets of stability conditions in Table B.4 must be simultaneously satisfied. We use the Mathematica symbolic manipulation software to establish that the intersection of parameter sets satisfying these combinations of conditions is non-empty. Figure 3.4 is an instance of one combination of parameters in that intersection.

To prove that $\alpha_1\beta > 1$ is a necessary condition for this behavior to arise, we will show that if $\alpha_1\beta \leq 1$ then the validity and stability conditions of R_6 and the condition $\beta > q_2$, required for this behavior, cannot hold together. The proof will proceed by considering several subcases depending on the relationships between the parameters.

(A) Case: $q_1 > 1$

From eq. (B.22) for $D < 0$ we have

$(q_1 - 1)(\beta - q_2) + (q_2 - \alpha_1\beta)(q_1 - \alpha_2) < 0$

Using the fact that $\beta > q_2$ and $q_1 > 1 > \alpha_2$, we find that the above inequality can only hold if $q_1 < \alpha_1\beta \Rightarrow \alpha_1\beta > 1$.

(B) Case: $q_1 \leq 1$

Here we will need to consider two subcases for $q_1 \geq \alpha_2$ and $q_1 < \alpha_2$.

(B.1) Subcase: $q_1 \geq \alpha_2$

Since $\beta > q_2$ and $q_1 \geq \alpha_2$, it implies $\beta > \alpha_2 + q_2 - q_1$. Note that this is also the condition for the converter of Technology 1 to hurt the overall market penetration. We show in the proof of Proposition 8 that this condition for the drop in overall penetration, can only be satisfied with the validity and stability conditions for $\bar{x}^*{}_{R_6}$ only if $\alpha_1\beta > 1$. Therefore this particular subcase will require $\alpha_1\beta > 1$ to hold.

(B.2) Subcase: $q_1 < \alpha_2$

For this subcase we will again need to consider two more subcases: (a) $q_2 \geq p_2$ and (b) $q_2 < p_2$.

(B.2.a) subcase: $q_2 \geq p_2$

If $q_2 \geq p_2$ and $q_1 < \alpha_2$ then $\alpha_2 + q_2 - q_1 - p_2 > 0$. From eqs. (B.23) and (B.25) and using the fact that $\beta > q_2 \geq p_2$, we have

$$\begin{aligned} \frac{\beta(q_1 - p_1) + p_1 q_2 - p_2 q_1}{q_2 - p_2} &\leq \alpha_1\beta \\ &< \frac{\beta - p_2 - p_1(\beta - \alpha_2 - (q_2 - q_1))}{\alpha_2 + q_2 - q_1 - p_2} \\ \Rightarrow \alpha_2(q_1 - p_1) &< (1 - q_1)(q_2 - p_2) + q_1(q_1 - p_1) \end{aligned}$$

It can be easily seen that the above inequality contradicts the condition in eq.(B.24). Therefore, the relationships considered in this subcase cannot hold together.

(B.2.b) subcase: $q_2 < p_2$

In this subcase we again need to consider two further subcases depending on the parameter relations: (i) $q_1 \geq p_1$ and (ii) $q_1 < p_1$.

When $q_1 \geq p_1$, using the facts that $\beta > q_2$, $q_1 \geq p_1$, $q_2 < p_2$ and eq. (B.23), we get $\alpha_1\beta \leq q_1$. This also implies $\alpha_1 q_2 < q_1$ since $\beta > q_2$. However, the relation, $\alpha_1 q_2 < q_1 < \alpha_2 < 1$, and eq.(B.21) together imply $q_1 > q_2$ which contradicts the requirement $q_2 > q_1$ of the model. Therefore this subcase cannot arise.

The subcase $q_1 < p_1$ also cannot arise by our assumption that $(0,0)$ is an invalid equilibrium. When both $q_1 < p_1$ and $q_2 < p_2$ then both $\theta_1^0(x_1 = 0, x_2 = 0) > 0$ and $\theta_1^0(x_1 = 0, x_2 = 0) > 0$, which makes $(0,0)$ a valid equilibrium.

Technology 2 cannot hurt itself while improving its converter efficiency, α_2.

Proof: The only equilibrium outcomes where the adoption level of Technology 2 varies as a function of α_2 are those that arise in regions R_5 and R_6 (as given by Eqs. (B.9) and (B.10)). If the equilibrium is in R_5 (*i.e.,* the technologies coexist at full market penetration) the derivative of the adoption level x_2 w.r.t. α_2 is

$$\frac{\partial x^*_{2_{R_5}}}{\partial \alpha_2} = \frac{(p_2 - p_1) - \beta(1-\alpha_2)}{[(q_2 - q_1) - (1-\alpha_2) - \beta(1-\alpha_1)]^2} \tag{B.39}$$

This expression is always positive since $(p_2 - p_1) > \beta(1-\alpha_2)$ is a required validity condition for the equilibrium in R_5. thus increasing α_2 cannot hurt Technology 2 for an equilibrium in R_5.

Next we consider the effect of α_2 on the equilibrium in R_6. In this region, the indifference points obey the relation $0 < \theta_1^0 \leq \theta_2^0 \leq \theta_2^1 < 1$. To show that Technology 2 cannot hurt itself by increasing α_2, we will consider the two cases $\alpha_1\beta \leq 1$ and $\alpha_1\beta > 1$ separately.

For $\alpha_1\beta \leq 1$, consider that $x^*_{1_{R_6}}(0)$ and $x^*_{2_{R_6}}(0)$ are the initial equilibrium adoption levels. Since

$$U_2 - U_1 = \theta(q_2 - q_1) + \beta(1-\alpha_1)x_2 - (1-\alpha_2)x_1 - (p_2 - p_1)$$

on increasing α_2, the difference of $U_2 - U_1$ is increased. Therefore a small fraction, say δ, of users of Technology 1 switch to Technology 2, thus making $x^*_{1_{R_6}}(1) = x^*_{1_{R_6}}(0) - \delta$ and $x^*_{2_{R_6}}(1) = x^*_{2_{R_6}}(0) + \delta$. The indifference point θ_2^1 shifts to the right to $\theta_2^1(1) = \theta_2^1(0) - \delta$. The second order effect of the switch-overs leads to changes in the adoption decisions of the lower-end users of Technology 1. The indifference point θ_1^0 shifts to $\theta_1^0(1) = \theta_1^0(0) - \frac{(\alpha_1\beta-1)\delta}{q_1}$. Since $\alpha_1\beta \leq 1$, if θ_1^0 shifts, it will shift to the right; thus further decreasing x_1. The new adoption level of Technology 1, therefore, becomes $x^*_{1_{R_6}}(1) = x^*_{1_{R_6}}(0) - \delta + \frac{(\alpha_1\beta-1)\delta}{q_1}$. Given these new values for $x^*_{1_{R_6}}(1)$ and $x^*_{2_{R_6}}(1)$, a new value can be computed for θ_2^1:

$\theta_2^1(1) = \theta_2^1(0) - \delta - \frac{\delta}{q_2-q_1}[\beta + (1-\alpha_1\beta)(1+\frac{1}{q_1})]$.

Since $q_2 > q_1$ and $\alpha_1\beta \leq 1$, the change in $\theta_2^1(1) - \theta_2^1(0)$ is again negative *i.e.,* θ_2^1 shifts further to the left, leading to more users switching from Technology 1 to 2. Thus, the compounding of the first and second order effects of a small increase in α_2 leads to decreases in x_1 decreases and increases in x_2. Both reinforce the initial increase in $U_2 - U_1$ after increasing α_2. As a result, as the process converges to a new equilibria after an increase in α_2, the final x_2 value exceeds the original one. Hence, improving its converter cannot hurt Technology 2.

We consider next the case $\alpha_1\beta > 1$. In this scenario, we know[36] that the overall market penetration must increase when increasing α_2. Therefore, if Technology 2's market share were to drop upon increasing α_2, then the market share of Technology 1 must increase so that the overall market share increases. We proceed to show that such a scenario is infeasible.

Assume that $\alpha_1\beta > 1$ and Technology 2 hurts itself by increasing α_2, *i.e.,* the indifference point θ_2^1 moves to the right to $\theta_2^1 + \varepsilon_1$, $(\varepsilon_1 \gtrsim 0)$, then a user with preference θ is in the range $\theta_2^1 < \theta < \theta_2^1 + \varepsilon_1$ will switch from using Technology 2 to using Technology 1. The switch-over of each such users decreases the utility U_1 of Technology 1 users by an amount $(\alpha_1\beta - 1) > 0$. This affects the lower-end users of Technology 1, *i.e.,* users with preference in the range $\theta_1^0 < \theta < \theta_1^0 + \varepsilon_2$, $(\varepsilon_2 \gtrsim 0)$, whose utility then becomes negative. These users, therefore, disadopt Technology 1. These disadoptions imply that the overall market penetrations decreases, which contradicts the fact that the overall market cannot drop when $\alpha_1\beta > 1$. This establishes that it is not possible for Technology 2 to hurt itself by increasing α_2.

[36]Using the notation form the proof of Proposition 8, note that $N_1 \leq 0$. Therefore, for $\alpha_1\beta > 1$, the derivative $\frac{\partial(x^*_{1R_6} + x^*_{2R_6})}{\partial\alpha_2} = \frac{(1-\alpha_1\beta)N_1}{D^2}$ is strictly positive for $N_1 < 0$ while it equals zero for $N_1 = 0$. However $N_1 = 0$ corresponds to the $x^*_{1_{R_6}} = 0$, *i.e.,* Technology 1 has no users, in which case an increase in α_2 can never hurt x_2. Therefore our present discussion only requires us to consider the case where the derivative is strictly increasing.

Proof of Proposition 8:

Using the same notation as in Proposition 7, decreasing the overall market penetration by increasing the converter efficiency of Technology 1 requires $\frac{\partial(x^*_{1R_6}+x^*_{2R_6})}{\partial\alpha_1} < 0$.

$$\frac{\partial(x^*_{1R_6}+x^*_{2R_6})}{\partial\alpha_1} = \frac{\beta(\beta-\alpha_2-(q_2-q_1))N_2}{D^2}$$

Using the same reasoning as in the proof of Proposition 7, a valid and stable equilibrium in R_6 implies $D < 0$, and consequently $N_2 < 0$. The above derivative is, therefore, negative only if $\beta > (q_2-q_1)+\alpha_2$. There are many combinations of parameters that simultaneously satisfy this condition and the validity and stability conditions of an equilibrium in R_6. Figure 3.4 is again one such combination. Furthermore, this condition $\beta > (q_2-q_1)+\alpha_2$ can only hold along with the validity and stability conditions for the equilibrium in R_6 only if $\alpha_1\beta > 1$. We now provide the proof for this.

Proof: $\alpha_1\beta > 1$ *is a necessary condition for the incumbent to hurt the overall market*

First consider the case when $q_1 \leq 1$. From eq. (B.21) and $\beta > (q_2-q_1)+\alpha_2$, we have

$0 < \beta-\alpha_2-(q_2-q_1) < \alpha_1\beta+q_2-q_1-\frac{q_2}{q_1}$

$\Rightarrow \alpha_1\beta-q_1 > \frac{q_2}{q_1}(1-q_1)$

$\Rightarrow \frac{\alpha_1\beta-q_1}{1-q_1} > \frac{q_2}{q_1} > 1$

Since $q_1 \leq 1$, we need $\alpha_1\beta > 1$.

Next consider the case when $q_1 > 1$. For this case, we will need to consider two subcases, corresponding to $\alpha_2 + q_2 - q_1 - p_2 > 0$ and $\alpha_2 + q_2 - q_1 - p_2 < 0$.

Subcase (1): Let $\alpha_2 + q_2 - q_1 - p_2 > 0$

When $q_1 > 1$, the above condition implies $q_2 - p_2 > q_1 - \alpha_2 > 0$ (*i.e.,* $q_2 > p_2$). However, when $\alpha_2 + q_2 - q_1 - p_2 > 0$ and $q_2 > p_2$, then eqs. (B.23) and (B.25) together result an inequality that contradicts the inequality in eq. (B.24). Since the conditions considered in this subcase cannot hold together, we do not need to consider it further.

Subcase (2): Let $\alpha_2 + q_2 - q_1 - p_2 < 0$

Let us assume that $\alpha_1\beta < 1$. We show here that the conditions for validity and stability of the R_6 equilibrium, and $\beta > (q_2 - q_1) + \alpha_2$ cannot hold together if $\alpha_1\beta < 1$. Using eq.(B.25), we have

$$\alpha_2 + q_2 - q_1 - p_2 < \beta - p_2 - p_1(\beta - \alpha_2 - (q_2 - q_1)) \Rightarrow p_1 < 1 < q_1$$

From eq.(B.22) we have

$$(q_1 - 1)(\beta - q_2) < (q_1 - \alpha_2)(\alpha_1\beta - q_1) < 0$$

which implies $\beta < q_2$ (since $q_1 > 1$).

Using the condition $\beta - \alpha_2 - (q_2 - q_1) > 0$ needed for hurting the overall market and eq.(B.24) and the condition $p_1 < 1 < q_1$ obtained previously, we get

$\beta-(q_2-q_1)>\alpha_2>\frac{(1-q_1)(q_2-p_2)+q_1(q_1-p_1)}{q_1-p_1}$

$\Rightarrow(\beta-q_2)(q_1-p_1)>(1-q_1)(q_2-p_2)$

$\Rightarrow q_2>p_2$ since $\beta<q_2$ and $q_1>1>p_1$.

Now using the condition of this subcase *i.e.,* $\alpha_2+q_2-q_1-p_2<0$ and eq. (B.24) we get $p_1>1$, which contradicts the previously obtained relation $p_1<1<q_1$. Therefore when $\alpha_1\beta<1$, all these conditions do not hold together and the so behavior will not arise for this case.

However, when $\alpha_1\beta>1$, it can be shown using Mathematica that for this subcase there exists numerical values for the various parameters for which the overall market drops. The above analysis of all the different cases establishes that $\alpha_1\beta>1$ is a necessary condition for this behavior to arise.

Similarly, when Technology 2 increases its converter efficiency, the overall market penetration will drop if

$$\frac{\partial(x^*_{1R_6}+x^*_{2R_6})}{\partial\alpha_2}=\frac{(1-\alpha_1\beta)N_1}{D^2}<0 \tag{B.40}$$

For a valid, stable equilibrium in R_6, we have $N_1<0$, and therefore the above expression is negative only if $\alpha_1\beta<1$. This establishes the second part of Proposition 8.

Proof of Proposition 9:

In Figure 3.7 we identified a scenario where instabilities in adoption dynamics arose for $\alpha_1\beta > 1$, and the parameter values satisfied $A^2 - 4B < 0$.

To prove that $\alpha_1\beta > 1$ is necessary for the formation of cyclic instability, we proceed in two steps: (i) we show that if $A^2 - 4B \geq 0$ then cycles (closed orbits) cannot arise in the adoption trajectories, and (ii) if $\alpha_1\beta \leq 1$ then $A^2 - 4B \geq 0$ always holds, and therefore there cannot be any closed orbits.

It may be noted that if any such 'closed orbit' or cycle were to arise in the adoption process, then its locus must be the equilibrium in R_6. This follows from the Index Theory which states that if J is a closed orbit[37] of a system enclosing an open set D (*i.e.,* $A = D \cup J$ is a compact set), then the set D must include an equilibrium point. Thus, every closed orbit in the plane encloses an equilibrium point. In our adoption process, a cyclic trajectory can either lie entirely inside the S-plane or may touch its boundaries. Note that when the trajectory touches or includes a portion of the boundary, it is not possible to have an equilibrium on the boundary itself because then the system would have attained stability as soon as the trajectory reaches that equilibrium. Therefore every closed trajectory must enclose an equilibrium that lie exclusively in the interior of the S-plane. The equilibrium in R_6 is the only equilibrium that satisfies this requirement (as all the others lie on the boundaries *i.e.,* x_1, x_2-axes or the line $x_1 + x_2 = 1$ by their definition). So the equilibrium in R_6 will be the focus for the proof of the two steps mentioned earlier.

---

[37] It is a trace of the trajectory of a non-trivial (i.e. not a point) periodic solution [75]

(i) Proof: If $A^2 - 4B \geq 0$ then cycles cannot arise in the adoption trajectories

To show that cycles cannot arise in the adoption trajectories when $A^2 - 4B \geq 0$, we have to consider the two possibilities that the equilibrium in R_6 is either stable or unstable.

(a) $A^2 - 4B \geq 0$ and $\underline{x}_{R_6}$ is stable.

We will prove that for this case, the entire R_6 region is the basin of attraction of the stable equilibrium $\underline{x}_{R_6}$ and therefore a closed trajectory cannot pass through (or be entirely located in) R_6 as it would have converged to that equilibrium. Furthermore, we show that it is not possible to realize any closed trajectory that has the R_6 equilibrium as its locus but does not ever pass through the Region R_6. These statements together eliminates the possibility of cycles in this case.

A stable equilibrium in R_6 requires $A < 0$ and $B > 0$. When $A^2 - 4B \geq 0$, we have $A - \sqrt{A^2 - 4B} \leq A + \sqrt{A^2 - 4B} < 0$. Therefore once a trajectory enters R_6, the exponential terms in its expression (Table B.2) decrease exponentially over time and converges to the equilibrium. In other words, the entire R_6 region is the basin of attraction of the stable equilibria located in it. Hence a closed trajectory cannot be realized if it were to pass through R_6.

Recall that in order to necessarily eliminate the possibility of $(0,0)$ being a valid equilibrium, the point P (in Figure 3.1) cannot lie in the positive quadrant of the (x_1, x_2) plane. As a result, the region R_6 will always touch either the boundary x_1 or x_2 axis.

Recall that we previously established that if a closed orbit were to arise in this system, the equilibrium in R_6 must lie in its interior. However, since the region R_6 touches at least one of the axes, it is never possible to realize a closed orbit that encircles the equilibrium in R_6 as its locus but doesn't pass through this region. Therefore, cyclic instabilities cannot be realized for this case.

(b) $A^2 - 4B \geq 0$ and $\underline{x}_{R_6}$ is unstable.

In this proof we will need two more results from the Index Theory. First, the index of a closed orbit, J, is +1 [38]. All such closed orbits (trajectories) must encircle the equilibrium in R_6. Moreover, since the equilibrium $\underline{x}_{R_6}$ is the only equilibrium point enclosed by J, the index of this equilibrium is I($\underline{x}_{R_6}$)=+1 [39]. Therefore a closed contour, C, around $\underline{x}_{R_6}$ in its neighborhood set must also have the same index as I($\underline{x}_{R_6}$) *i.e.,* +1.

However, C can have an index of +1 only if either all the trajectories are pointing radially inward(outward) towards(from) the equilibrium. But we show below that when $A^2 - 4B \geq 0$ and the equilibrium is unstable, the trajectories in R_6 take the shape of hyperbolas, which implies that all the trajectories are not directed consistently either inward or outward from the equilibrium, and thus the index computed for the closed contour, C, will not be +1. This contradiction implies that a closed orbit J cannot be present, which will therefore eliminate the possibility of cyclic instability in this case.

[38] ref. [49], pp. 299, lemma 7.1(c)

[39] ref. [75], pp.50

Using the expressions for the trajectories in R_6 provided in Table B.2 and the fact that $A^2 - 4B \geq 0$ and the equilibrium is unstable, the expressions for the trajectories around $\underline{x}_{R_6}$ can be rearranged in the following form to show their hyperbolic shapes:

$$(x_1 - x^*_{1R_6} - K_2(x_2 - x^*_{2R_6}))(x_1 - x^*_{1R_6} - K_1(x_2 - x^*_{2R_6})) \\ = -c_1 c_2 (K_1 - K_2)^2 e^{A\gamma(t-t_0)} \quad \text{(B.41)}$$

Substituting $\bar{x}_1 = x_1 - x^*_{1R_6}$, $\bar{x}_2 = x_2 - x^*_{2R_6}, a = \frac{\alpha_2+\beta(1-\alpha_1)-\frac{q_2}{q_1}}{(q_2-q_1)\sqrt{A^2-4B}}$, we have $K_1 = a - 1$, $K_2 = a + 1$.

Eqn. (B.41) can therefore be written as:

$$(\bar{x}_1 - a\bar{x}_2)^2 - \bar{x}_2{}^2 = -4b^2 c_1 c_2 e^{A\gamma(t-t_0)} \quad \text{(B.42)}$$

The above expression clearly shows that at any time t, the trajectories around the equilibrium have hyperbolic shapes.

Thus, when $A^2 - 4B \geq 0$, irrespective of whether the equilibrium in R_6 is stable or not, there cannot be any cyclic trajectories in the adoption process.

(ii) Proof: If $\alpha_1\beta \leq 1$ *then* $A^2 - 4B \geq 0$

Note that $a^2 + b^2 \geq 2ab$ for any real numbers a, b. Hence we have

$$\frac{\left(1-\alpha_2+\beta(1-\alpha_1)\right)^2}{(q_2-q_1)^2} + \frac{1}{q_1^2} \geq \frac{2(1-\alpha_2)+2\beta(1-\alpha_1)}{q_1(q_2-q_1)} > 0$$

since $0 < \alpha_1 < 1, 0 < \alpha_2 < 1, \beta > 0$.

Using the expression for $A^2 - 4B$ and the above inequality, we get:

$$A^2 - 4B \geq \frac{4(1-\alpha_1\beta)(1-\alpha_2)}{q_1(q_2-q_1)} \geq 0 \ \ if \ \alpha_1\beta \leq 1$$

Since we can only have cyclic instability in the system when $A^2 - 4B < 0$ and that this condition can only be satisfied when $\alpha_1\beta > 1$, it also becomes necessary that $\beta > 1$ (as $\alpha_1 < 1$).

Appendix C

Appendix: Network Platform Design

C.1 Platform Utility Simplification

This appendix explains briefly the parameter relabeling and normalization method used in the utility function of Equation (4.1). Suppose that the total customer population for the two market sides, developers and consumers, are N_d and N_c, respectively. Equation (4.2) and Equation (4.3) can be written as

$$U_d = \alpha x_c N_c - b_d - K(F) - \phi\tau \quad \text{(C.1)}$$

$$U_c = \theta\beta n_d N_d - p_c \quad \text{(C.2)}$$

Therefore, the platform's utility function can be written as

$$U_p = p_c x_c N_c + b_d n_d N_d - C(F) \quad \text{(C.3)}$$

Substituting $p_c = (1 - x_c)\beta n_d N_d$ and $b_d = \alpha x_c N_c - K(F) - n_d \tau$, we have

$$U_p = \left((1 - x_c)\overline{\beta} n_d x_c + (\overline{\alpha} x_c - n_d - \overline{K(F)}) n_d - \overline{C(F)} \right) \tau N_d \tag{C.4}$$

where $\overline{\beta} = (\beta N_c)/\tau$, $\overline{\alpha} = (\alpha N_c)/\tau$, $\overline{K(F)} = K(F)/\tau$ and $\overline{C(F)} = C(F)/(N_d \tau)$. Thus, with these proper normalization of the parameter values, the platform's optimization problem reduces to an equivalent maximization problem, $max_{0 \leq n_d^*, x_c^* \leq 1} U_p$, where

$$\begin{aligned} U_p &= ((1 - x_c^*)\overline{\beta} n_d^*) x_c^* + (\overline{\alpha} x_c^* - n_d^* - \overline{K(F)}) n_d^* - \overline{C(F)} & \text{(C.5)} \\ &= \overline{p_c} x_c^* + \overline{b_d} n_d^* - \overline{C(F)} & \text{(C.6)} \end{aligned}$$

where $\overline{p_c} = (1 - x_c^*)\overline{\beta} n_d^*$ and $\overline{b_d} = \overline{\alpha} x_c^* - n_d^* - \overline{K(F)}$. Note that the normalized function in Equation (C.6) is structurally identical to Equation (4.1) used in Section 4.2.2, and similarly, $\overline{p_c}$ and $\overline{b_d}$ are equivalent to p_c and b_c of Equation (4.6) and Equation (4.7) in Section 4.3.1.

C.2 Alternative Utility Functions

C.2.1 Competition among Developers: Same-side Externalities

In our model we consider that the applications that developers innovate, although differentiated in their offering and characteristics, share some degree of homogeneity in terms of the underlying functionalities they use. Consequently, developers may have to compete for same advertisement revenue sources and consumers, and thus incur negative same-side externalities from each other's presence. To account for such situations, we show that our model can be easily generalized to include competition among the developers.

In the utility function of developers given in Equation (4.2), we include a term $-\eta n_d$ where n_d is the fraction of developers on the platform and η captures the marginal externality that each developer imposes on the other. Therefore, the new utility function becomes

$$U_d = \alpha x_c - \eta n_d - b_d - (K(F) + \tau\phi) \tag{C.7}$$

In using Equation (C.7) to compute the equilibrium adoption levels (as done in Section 4.3.1), the developer who is indifferent between joining and not joining the platform ($U_d = 0$) is obtained from $\widehat{\phi} = n_d$:

$$\begin{aligned} \tau\widehat{\phi} = \tau n_d &= \alpha x_c^* - \eta n_d - b_d - K(F) \\ \Rightarrow n_d(\tau + \eta) &= \alpha x_c^* - b_d - K(F) \end{aligned} \tag{C.8}$$

If all the parameters are normalized with respect to $\tau + \eta$, the resulting expression is identical to Equation (4.5). Therefore, the qualitative outcomes observed from our earlier model do not change upon introducing competition among developers, only the quantitative values need to be adjusted with respect to the new normalization coefficient.

C.2.2 Consumer heterogeneity in Intrinsic benefits

In this section, we consider alternative consumer utility function that capture scenarios where all consumers have similar valuation for the number of available applications, but are heterogeneous in how they evaluate the platform's intrinsic qualities (*e.g.,* reliability, performance, brand name etc). This is typically true for platforms that deliver a strong core value and the availability of software applications are added bonuses. For example, consider a gaming platform (*e.g.,* Xbox 360) where almost all users value the availability of interesting games equally but are quite heterogeneous in how they perceive the platform's intrinsic quality, as determined by the screen resolution, controls, loading time etc. The utility function for the consumers in this scenario is:

$$U_c = \theta q + \beta n_d - p_c \tag{C.9}$$

The term q stands for the intrinsic value of the platform to the consumers, and it is weighted by their individual preference parameter θ. We assume θ to be uniformly distributed in $[0,1]$, and that it value is a private information for each consumer, but its distribution is known. The term βn_d captures the cross-externality benefits that the consumers enjoy from the presence of n_d developers on the platform. As before, β captures the marginal externality benefit that each developer brings to the consumers. p_c is the flat fee paid to the platform. All the parameters of the model are appropriately normalized with respect to the customer population on each side and the maximum fixed cost that developers may incur (*i.e.,* $\tau = 1$). The utility functions for the platform and the developers remain the same, as in Equations (4.1) and (4.2), respectively.

Using the solution methodology outlined in Section 4.3, we get the following outcomes in the three stages:

Adoption Stage:

For a given functionality level, F, and a set of prices, p_c and b_d, for the consumers and developers respectively, the marginal consumer who is indifferent between joining and not joining the platform is $\widehat{\theta} = 1 - x_c = \frac{p_c - \beta n_d^*}{q}$ and similarly, the marginal indifferent developer is $\widehat{\phi} = n_d = \alpha x_c^* - b_d - K(F)$ At equilibrium, we have $x_c^* = x_c$ and $n_d^* = n_d$.

Pricing Stage:

As before, we solve the platform provider's profit maximization problem to find the 'optimal' prices for the two market sides and the equilibrium adoption levels at these prices. The interesting situation is one where neither market side has reached full adoption, *i.e.,* where the outcome of the equilibrium adoption is an interior solution of the maximization problem. The results for it are provided below.

The optimal price levels (p_c^*, p_d^*) and the optimal adoption levels of consumers and developers (x_c^*, n_d^*) of the two-sided market, which maximize the platform provider's *profit are given by*

$$p^* = \frac{q(2q+(\alpha-\beta)K(F)-\alpha(\alpha+\beta))}{4q-(\alpha+\beta)^2} \tag{C.10}$$

$$b^* = \frac{(\alpha-\beta)q-(2q-\beta(\alpha+\beta))K(F)}{4q-(\alpha+\beta)^2} \tag{C.11}$$

$$x_c^* = \frac{2q-(\alpha+\beta)K(F)}{4q-(\alpha+\beta)^2} \tag{C.12}$$

$$n_d^* = \frac{(\alpha+\beta-2K(F))q}{4q-(\alpha+\beta)^2} \tag{C.13}$$

For the above outcome to be an interior solution, $0 < \{2q-(\alpha+\beta)K(F), (\alpha+\beta-2K(F))q\} < 4q-(\alpha+\beta)^2$ needs to be satisfied. The second order condition for the Hessian to be positive definite is also satisfied when the above inequality holds.

Design Stage:

The optimal level of built-in functionalities (F^*) for the platform which maximizes its profit is given by

$$\frac{C'(F^*)}{K'(F^*)} = \frac{2q}{4q-(\alpha+\beta)^2}K(F^*) - \frac{(\alpha+\beta)q}{4q-(\alpha+\beta)^2} \tag{C.14}$$

$$\Rightarrow \frac{C'(F^*)}{K'(F^*)} = -n_d^*(F^*) \tag{C.15}$$

$$C''(F^*) > -n_d^*(F^*)K''(F^*) + \frac{2q[K'(F^*)]^2}{4q-(\alpha+\beta)^2} \tag{C.16}$$

As before with the utility functions of Section 4.3, we have the following result by using the conjugate pair theorem. The level of functionality investment by the platform increases with increase in cross-externality benefits enjoyed by either customer side, *i.e., sign*$\frac{\partial F^*}{\partial \alpha} = sign\frac{\partial^2 U_p}{\partial F \partial \alpha} > 0$, *i.e.,* $\frac{\partial F^*}{\partial \alpha} > 0$ and *sign*$\frac{\partial F^*}{\partial \beta} = sign\frac{\partial^2 U_p}{\partial F \partial \beta} > 0$, *i.e.,* $\frac{\partial F^*}{\partial \beta} > 0$.

C.3 Proofs

Proof of Proposition 10:

The utility functions of the developers and consumers at the indifferent points provides the price levels, p_c and b_d, as in Equation (4.6) and Equation (4.7). Substituting these prices into the profit function of the platform provider in Equation (4.1), we get the first order condition for each market side, $\frac{\partial U_p}{\partial x_c^*} = 0$ and $\frac{\partial U_p}{\partial n_d^*} = 0$, which gives

$$\frac{\partial U_p}{\partial x_c^*} = (1-2x_c^*)\beta n_d^* + \alpha n_d^* = 0$$
$$\frac{\partial U_p}{\partial n_d^*} = (1-x_c^*)\beta x_c^* + \alpha x_c^* - 2n_d^* - K(F) = 0$$

Solving the above equations simultaneously gives x_c^* and n_d^*, which also gives p_c^* and b_d^* for the optimal two-sided pricing by the platform.

$$x_c^* = \frac{\alpha+\beta}{2\beta}$$
$$n_d^* = \frac{(\alpha+\beta)^2 - 4\beta K(F)}{8\beta}$$

The conditions for interior solution, *i.e.,* $0 < x_c^*, n_d^* < 1$, are satisfied when $\alpha < \beta$ and $4\beta K(F) < (\alpha+\beta)^2 < 4\beta(2-K(F))$. The second order maximization conditions for a negative definite Hessian (principle minors have alternate signs), $|H_1|_{(x_c^*,n_d^*)} = -2\beta n_d^* < 0$ and $|H_2|_{(x_c^*,n_d^*)} = 4\beta n_d^* > 0$, also hold true for interior solutions.

Bibliography

[1] L. Abdel-Malek, R. Montanari, and L. C. Morales. Exact, approximate, and generic iterative models for the multi-product newsboy problem with budget constraint. *International Journal of Production Economics*, 91(2):189 – 198, 2004.

[2] AT&T. What is IPTV? http://www.att.com/Common/merger/files/pdf/IPTV_background.pdf, 2009.

[3] Y. Bakos. Reducing buyer search costs: Implications for electronic marketplaces. *MIS Quarterly*, 15(3):295–310, 1991.

[4] Y. Bakos and E. Katsamakas. Design and ownership of two-sided networks: Implications for internet platforms. *Journal of Management Information Systems*, 25(2):171–202, 2008.

[5] R. D. Banker and J. S. Hughes. Product costing and pricing. *The Accounting Review*, 69(3):pp. 479–494, 1994.

[6] F. M. Bass. A new product growth for model consumer durables. *Management Science*, 15(5):215–227, 1969.

[7] R. Beaubrun, B. Moulin, and N. Jabeur. An Architecture for Delivering Location-Based Services. *International Journal of Computer Science and Network Security*, 7(7):160–166, 2007.

[8] H. K. Bhargava and V. Choudhary. Economics of an Information Intermediary with Aggregation Benefits. *Information Systems Research*, 15(1):22–36, 2004.

[9] E. K. Bish, A. Muriel, and S. Biller. Managing flexible capacity in a make-to-order environment. *Management Science*, 51(2):167–180, 2005.

[10] M. Brandel. Facilities management goes green. computerworld.com, http://www.computerworld.com/s/article/295379/Green_Buildings_Smarter_More_Energy_Efficient, 25 June 2007.

[11] L.M.B. Cabral. On the adoption of innovation with network externalities. *Mathematical Social Sciences*, 19:299–308, 1990.

[12] B. Caillaud and B. Jullien. Chicken and egg: Competing matchmakers. *Rand Journal of Economics*, 34(2):309–328, 2003.

[13] J. Catone. eBay Launches Dev Platform - Too Little, Too Late? *www.readwriteweb.com/archives/ebay_launches_ dev_platform.php*, 2008.

[14] J.P. Choi. The provision of (two-way) converters in the transition process to a new incompatible technology. *The Journal of Industrial Economics*, 45(2):139–153, 1997.

[15] T. Crosby. How FiOS Works. HowStuffWorks.com, 7 August 2008. http://communication.howstuffworks.com/fiber-optic-communications/fios.htm/printable.

[16] K. Currier. *Comparative Statics Analysis in Economics*. World Scientific Publishing Company, August 2000.

[17] D. Didem, V. S. Jacob, and S. Raghunathan. Strategic it investments: the impact of switching cost and declining it cost. *Management Science*, 53(2):208–226, 2007.

[18] A. Durand, R. Droms, B. Haberman, and J. Woodyatt. Dual-stack lite broadband deployments post IPv4 exhaustion. Internet Draft (draft-ietf-softwire-dual-stack-lite-00.txt), March 2009. (Work in progress).

[19] N. Economides. The economics of networks. *International Journal of Industrial Organization*, 14(6):673–699, 1996.

[20] N. Economides and J. Tag. Net neutrality on the internet: A two-sided market analysis. *Working Paper no. 07-14, NET Institute, New York University*, 2007.

[21] S. Verbrugge *et al.* Modeling operational expenditures for telecom operators. In *Proc. of Conference on Optical Network Design and Modeling*, pages 455–466, 2005.

[22] S. J. Erlebacher. Optimal and heuristic solutions for the multi-item newsvendor problem with a single capacity constraint. *Production and Operations Management*, 9(3):303–318, 2000.

[23] J. Farrell and G. Saloner. Installed base and compatibility: Innovation, product preannouncements, and predation. *American Economic Review*, 76:940–955, 1986.

[24] J. Farrell and G. Saloner. Converters, compatibility, and the control of interfaces. *The Journal of Industrial Economics*, 40(1):9–35, 1992.

[25] C. H. Fine and R. M. Freund. Optimal investment in product-flexible manufacturing capacity. *Management Science*, 36(4):449–465, 1990.

[26] K. Flanigan. $1.5 million ad revenue suggests free apps could be the future. *http://148apps.biz/1-5-million-ad-revenue-suggests-free-apps-could-be-the-future/*, 2010.

[27] L. A. Fourt and J. W. Woodlock. Early prediction of market success for grocery products. *Journal of Marketing*, 25:31–38, 1960.

[28] D. Fudenberg and J. Tirole. *Game Theory*. MIT Press,, Cambridge, MA, 1991.

[29] M. Garnatt. AWS by the numbers. *http://www.elastician.com/2010/06/aws-by-numbers.html*, 2010.

[30] GENI. *www.geni.net*.

[31] R. F. Göx. Capacity planning and pricing under uncertainty. *Journal of Management Accounting Research*, 14(1):59–78, 2002.

[32] S. C. Graves and B. T. Tomlin. Process flexibility in supply chains. *Management Science*, 49(7):907–919, 2003.

[33] P. E. Green. Hybrid models for conjoint analysis: An expository review. *Journal of Marketing Research*, 21(2):155–169, 1984.

[34] P. E. Green, A. Krieger, and Y.Wind. Thirty years of conjoint analysis: Reflections and prospects. *Interfaces*, 31(3):S56–S73, 2001.

[35] The Hammersmith Group. The internet of things- networked objects & smart devices. *Research Report*, February 2010.

[36] A. Hagiu. Pricing and commitment by two-sided platforms. *The RAND Journal of Economics*, 37(3):720–737, 2006.

[37] A. Hagiu. Multi-Sided Platforms: From Microfoundations to Design and Expansion Strategies. *Harvard Business School Strategy Unit Working Paper No. 09-115*, 2009.

[38] C. Holahan. iLike: a model for the recording industry. *http://www.businessweek.com/the_thread/techbeat/archives/2008/07/ ilike_a_model_for_the_recording_industry.html*, 2008.

[39] D. Horsky. A diffusion model incorporating product benefits, price, income and information. *Marketing Science*, 9(4):342–365, 1990.

[40] NSF NeTS FIND Initiative. *http://www.nets-find.net/*.

[41] Y. Jin, S. Sen, R. Guerin, K. Hosanagar, and Z.-L. Zhang. Dynamics of competition between incumbent and emerging network technologies. *Proc. of ACM NetEcon'08*, 2008.

[42] W. C. Jordan and S. C. Graves. Principles of the benefits of manufacturing process flexibility. *Management Science*, 41(4):577–594, 1995.

[43] D. Joseph, N. Shetty, J. Chuang, and I. Stoica. Modeling the adoption of new network architectures. In *CoNEXT'07 Conference*. ACM, 2007.

[44] D. Joseph, N. Shetty, J. Chuang, and I. Stoica. Modeling the adoption of new network architectures. In *Proc. ACM CoNEXT'07*, 2007.

[45] JSR. IMS Communication Enablers (ICE). http://jcp.org/en/jsr/detail?id=325, 2010.

[46] J. P. Kassirer. The next transformation in the delivery of health care. *New England Journal of Medicine*, 332(1):52–54, 1995.

[47] M. L. Katz and C. Shapiro. Network externalities, competition, and compatibility. *The American Economic Review*, 75(3):424–440, 1985.

[48] M. L. Katz and C. Shapiro. Systems competition and network effects. *The Journal of Economic Perspectives*, 8(2):pp. 93–115, 1994.

[49] H. K. Khalil. *Nonlinear Systems (Second Edition)*. Prentice Hall, 1996.

[50] M. Khouja. The single-period (news-vendor) problem: literature review and suggestions for future research. *Omega*, 27(5):537 – 553, 1999.

[51] P. H. Koebbe. A case for separate networks. Communications News, http://www.comnews.com/features/2007_november/1107a_case.aspx, 1 November 2007.

[52] K. Kogan and V. Portougal. Multi-period aggregate production planning in a newsvendor framework. *Journal of the Operational Research Society*, 57:423–433, 2006.

[53] N. Laoutaris, G. Smaragdakis, P. Rodriguez, and R. Sundaram. Delay tolerant bulk data transfers on the internet. In *SIGMETRICS '09: Proceedings of the eleventh international joint conference on Measurement and modeling of computer systems*, pages 229–238, New York, NY, USA, 2009. ACM.

[54] H. S. Lau. The newsboy problem under alternative optimization objectives. *The Journal of the Operational Research Society*, 31(6):pp. 525–535, 1980.

[55] H. S. Lau and A. H. L. Lau. The newsstand problem: A capacitated multiple-product single-period inventory problem. *European Journal of Operational Research*, 94(1):29 – 42, 1996.

[56] V. Mahajan, E. Muller, and Y. Wind, editors. *New-Product Diffusion Models*. Springer, 2000.

[57] M. Maisto. Mobile app revenue soaring, although most are free. *http://www.eweek.com/index2.php?option=content&do_pdf=1&id=58530*, 2010.

[58] K. Martin and I. Todorov. How Will Digital Platforms Be Harnessed in 2010, and How Will They Change the Way People Interact with Brands? *Journal of Interactive Advertising*, 10(2):61–66, 2010.

[59] B. Martorelli. Internet service providers: Addressing the customer service challenge. White paper, Hurwitz Group Inc. Framingham, MA., 2000.

[60] J. A. Van Mieghem. Investment strategies for flexible resources. *Management Science*, 44(8):1071–1078, 1998.

[61] P. Mileff and K. Nehez. Solving capacity constraint problems in a multi-item, multi-period newsvendor model. *Proc. of microCAD 2007, 21th International Scientific Conference*, pages 169–176, 2007.

[62] J. Musacchio, G. Schwartz, and J. Walrand. A two-sided market analysis of provider investment incentives with an application to the net-neutrality issue. *Review of Network Economics*, 8(1):22–39, 2009.

[63] J. A. Norton and F. M. Bass. A diffusion theory model of adoption and substitution for successive generations of high-technology products. *Management Science*, 33(9):1069–1086, 1987.

[64] U. Olsson and M. Stille. Communication services - the key to IMS service growth. *Ericsson Review*, 2008.

[65] S. S. Oren and S. A. Smith. Critical mass and tariff structure in electronic communications markets. *The Bell Journal of Economics*, 12(2):467–487, 1981.

[66] J.C. Panzar and R.D. Willig. Economies of scope. *The American Economic Review*, 71(2):268–272, 1981.

[67] G. G. Parker and M. W. Van Alstyne. Two-sided network effects: A theory of information product design. *Management Science*, 51(10):1494–1504, 2005.

[68] H. Peterson, S. Sen, J. Chandrashekar, L. Gao, R. Guerin, and Z. L. Zhang. Message-efficient dissemination for loop-free centralized routing. *Computer Communications Review*, 38(3):63–74, 2008.

[69] L. Peterson, S. Shenker, and J. Turner. Overcoming the Internet impasse through virtualization. In *Proc. ACM HotNets-III*, 2004.

[70] N. C. Petruzzi and M. Dada. Pricing and the Newsvendor Problem: A Review with Extensions. *Operations Research*, 47(2):183–194, 1999.

[71] Cisco's Featured Routers Products. http://www.cisco.com/en/us/products/hw/routers/.

[72] R. Reed. Facebook places: The de facto platform for lbs. *http://momentfeed.com/2010/11/facebook-places-platform-lbs/*, 2010.

[73] Frost & Sullivan Press Release. Growth in green buildings market spurred by rising interests and government support. *www.frost.com/prod/servlet/press-release.pag?docid=189796979*, January 2010.

[74] J. C. Rochet and J. Tirole. Two-sided markets: a progress report. *The RAND Journal of Economics*, 37(3):645667, 2006.

[75] S. Sastry. *Nonlinear Systems: Analysis, Stability and Control*. Springer, 1999.

[76] E. I. Schwartz. Behind the surge in mobile advertising. *http://www.technologyreview.com/business/26648/*, 2010.

[77] S. Sen, R. Guerin, and K. Hosanagar. Shared versus separate networks: The impact of re-provisioning. In *Proc. of ReArch'09*, Rome, Italy, 2009.

[78] S. Sen, R. Guerin, and K. Hosanagar. Functionality-rich versus minimalist platforms: A two-sided market analysis, 2011. under submission.

[79] S. Sen, Y. Jin, R. Guérin, and K. Hosanagar. Modeling the dynamics of network technology adoption and the role of converters. Technical report, University of Pennsylvania, June 2009. Available at http://repository.upenn.edu/ese_papers/496/.

[80] S. Sen, Y. Jin, R. Guerin, and K. Hosanagar. Modeling the dynamics of network technology adoption and the role of converters. *IEEE/ACM Transactions on Networking*, 18(6):1793–1805, 2010.

[81] S. Sen, K. Yamauchi, R. Guerin, and K. Hosanagar. The impact of reprovisioning on the choice of shared versus dedicated networks. In *Proc. of the Ninth Workshop on E-Business (WEB 2010),*, St. Louis, Missouri, 2010.

[82] O. Shy. Technology revolutions in the presence of network externalities. *International Journal of Industrial Organization*, 14(6):485–800, 1996.

[83] J. Touch, Y. Wang, L. Eggert, and G. Finn. A virtual Internet architecture. In *Proc. ACM FDNA'03*, Karlsruhe, Germany, 2003.

[84] Internet World Statistics: Usage and Population Statistics. Top 20 countries with the highest number of internet users. *www.internetworldstats.com/top20.htm*, 2010.

[85] I. Wilkes. Cloud platform choices: a developer's-eye view. *arstechnica.com/business/news/2010/02/cloud-basics.ars*, 2010.

[86] WPI. Location based applications are driving innovation in wireless communication industry. http://www.wpi.edu/news/20101/2010wireless.html, 2010.

[87] B. Yoo, V. CHoudhary, and T. Mukhopadhyay. A model of neutral b2b intermediaries. *Journal of Management Information Systems*, 19(3):43–68, 2002.

[88] B. Zhang and S. Du. Multi-product newsboy problem with limited capacity and outsourcing. *European Journal of Operational Research*, 202(1):107 – 113, 2010.