10-4-2009

Shared Versus Separate Networks - The Impact of Reprovisioning

Soumya Sen
University of Pennsylvania, soumyas@princeton.edu

Roch A. Guérin
University of Pennsylvania, guerin@acm.org

Kartik Hosanagar
University of Pennsylvania, kartikh@wharton.seas.upenn.edu

Paper presented at:
ACM ReArch'09 Workshop, Rome, Italy, December 2009.
URL: http://conferences.sigcomm.org/co-next/2009/workshops/rearch/

Shared Versus Separate Networks - The Impact of Reprovisioning

Abstract

As networks improve and new services emerge, questions arise that affect service deployments and network choices. The Internet is arguably a successful example of a network shared by many services. However, combining heterogeneous services on the same network need not always be the right answer, and technologies such as virtualization make deploying new services on separate networks increasingly more viable. So, which is the right option? The question is not unique to networks, and there is a large body of work in the manufacturing systems literature that explores the trade-off between flexible and dedicated plants. This paper highlights an important feature missing from these earlier works, namely, the ability to ``reprovision'' resources in response to changes in demand. It demonstrates that this feature alone can affect the choice of network solutions, and argues for models that incorporate it.

Keywords

Networks, Economics, Theory

Disciplines

OS and Networks | Systems and Communications

Comments

Paper presented at:

ACM ReArch'09 Workshop, Rome, Italy, December 2009.

URL: http://conferences.sigcomm.org/co-next/2009/workshops/rearch/

Shared Versus Separate Networks The Impact of Reprovisioning *

Soumya Sen
ESE, U.Pennsylvania
Philadelphia, PA 19104, USA
ssoumya@seas.upenn.edu

Roch Guérin
ESE, U.Pennsylvania
Philadelphia, PA 19104, USA
guerin@ee.upenn.edu

Kartik Hosanagar
Wharton, U.Pennsylvania
Philadelphia, PA 19104, USA
kartikh@wharton.upenn.edu

ABSTRACT

As networks improve and new services emerge, questions arise that affect service deployments and network choices. The Internet is arguably a successful example of a network shared by many services. However, combining heterogeneous services on the same network need not always be the right answer, and technologies such as virtualization make deploying new services on separate networks increasingly more viable. So, which is the right option? The question is not unique to networks, and there is a large body of work in the manufacturing systems literature that explores the trade-off between flexible and dedicated plants. This paper highlights an important feature missing from these earlier works, namely, the ability to "reprovision" resources in response to changes in demand. It demonstrates that this feature alone can affect the choice of network solutions, and argues for models that incorporate it.

Categories and Subject Descriptors

H.1.0 [**Information Systems**]: Models and Principles—*General*

General Terms

Economics, Theory

Keywords

Network Services, Virtualization, Resource Allocation

1. INTRODUCTION

The ubiquity and capabilities of the Internet have led to an "explosion" of 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 benefits of a shared infrastructure notwithstanding, combining services with disparate requirements onto a single network has a cost. 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 or the need for tracking and trouble-shooting problems of previously little consequences, *e.g.,* minor routing instabilities don't affect most data services but can severely degrade voice or video quality. Assessing the benefits of sharing a network across services calls, therefore, for understanding the trade-off between the economies of scale and scope it allows, and the diseconomies of scope it gives rise to.

Developing models that explore this trade-off is the initial motivation for this paper. Models should capture the costs of the different components involved in deploying and operating networks, how these costs are affected by the needs of different services, and allow meaningful comparisons between shared and separate network solutions. We note that networks are not the first to face such a question. There is a long tradition in the manufacturing sector of models aimed at gauging the benefits of flexible but more expensive manufacturing plants, versus those of dedicated plants. We review this parallel in Section 2, but next we point to what we believe is an important difference; one that is at the core of this paper.

Specifically, the time-lag involved in building a new manufacturing plant is such that once made, decisions are difficult if not impossible to revisit. This implies that if the production capacity of a new plant is insufficient to meet the realized demand for its product, the excess demand is typically lost[1]. In contrast, networks are becoming more akin to services, and adjusting network capacity in response to an unexpected increase in demand can often be realized relatively quickly[2]. Furthermore, the emergence of network virtualization technology [6, 7] is likely to make this even more common place, and makes the question of whether to add a new service on an existing network or on a new network "slice" a more realistic one.

The main purpose of this paper is to demonstrate that unlike the more traditional manufacturing setting, comparing

*Supported by NSF grant CNS-0721610.

[1]This is slowly changing as "on-demand" manufacturing facilities become available, but remains a significant hurdle at least in the heavy manufacturing sector.
[2]Not so long ago the provisioning of a T1 connection required several months, while "dialing-up" an additional Gigabit/sec of bandwidth is now commonly available.

the relative merits of shared and separate network solutions calls for models that incorporate a "reprovisioning" phase. The paper's main contributions are to demonstrate how reprovisioning can influence which network solution is more effective, and provide insight into when and why this happens. This establishes the pertinence of the models proposed in the paper, and paves the way for a more systematic investigation of shared versus separate network solutions, which is the topic of ongoing work.

The rest of the paper is structured as follows. Section 2 reviews previous works from the manufacturing literature, and articulates their relevance. Section 3 introduces our model and its cost factors. Section 4 presents the optimization framework used to solve the model, and explores the impact of reprovisioning. Section 5 summarizes the paper's findings and outlines our ongoing investigations in applying its model.

2. RELATED LITERATURE

There is a long tradition of investigating the trade-off between flexible (but more expensive) and dedicated resources in the manufacturing systems literature. These works target various related managerial decision problems. For example, the Manufacturing Process Flexibility literature has focused on efficient plant-product assignments [3, 4] (how to best allocate product demand to manufacturing plants), and the effect of process flexibility in handling demand variability [1]. The stream of works most relevant to our present discussion is one that addresses optimal resource planning and allocation in the presence of demand uncertainty [2, 5].

In these models, investment decisions in manufacturing plants of given capacity have to be made before the actual product demands are realized. 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 plants and how much to build into dedicated plants. Fine and Freund [2] develop a *two-stage model* to analyze this trade-off. Plant capacity investment decisions are made in the first stage, when product demand is still uncertain. Production decisions are implemented in the second stage (after product demands have been realized), given the decisions of the first-stage. The authors set up an optimization problem to establish the firm's optimal investments in flexible and dedicated resources, and the optimal production levels. A similar setting was considered by Van Mieghem [5], with however an emphasis on the role of price margin and cost mix differentials. It showed that investment in flexible resources can be beneficial even with perfectly positively correlated product demands, *i.e.,* because a flexible plant can shift production towards the product with a higher profit margin.

Our model shares basic structural properties with these works. Choosing between shared and separate 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 model and these earlier works. First, rather than simply explore the benefits of a flexible (shared) plant (network) in dealing with uncertain demand (correlated or not), our focus is on investigating the impact of various economies and diseconomies of scope in the underlying cost factors. A second and more important difference is that 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[3]. As a result, even if some excess demand is ultimately lost, *i.e.,* adjusting provisioning decisions may incur a penalty, networks can recover from insufficient provisioning. This affects not only (optimal) resource provisioning decisions, but as we shall see can also influence the choice of network solutions, *i.e.,* shared or separate.

3. MODEL FORMULATION

We develop a model aimed at exploring when multiple services are best deployed over shared or separate networks. Without loss of generality, we limit the discussion to the case of two services. Furthermore, for simplicity we assume that the first service has already been deployed and runs on an existing network. As a result, demand uncertainty is present only for the second service. This is one of the most basic settings in which the question of network sharing arises, and we use it to illustrate the importance of certain features in any model seeking to explore these issues.

3.1 Model Parameters

Service 1 is the existing service and has a stable demand D_1, with a corresponding provisioning level that can support K_1 users. For simplicity, we assume $K_1 = D_1$. The new service, Service 2, has uncertainty in its demand D_2 for which only the distribution f_{D_2} is known. The provisioning level (number of users to be supported) for Service 2 corresponds to a decision variable denoted by K_{s2} or K_2 for shared and separate networks, respectively[4].

Each network solution involves cost and revenue components. To facilitate comparisons, we follow standard accounting principles, *e.g.,* [2, 5], and normalize up-front and future investments as well as recurring revenues and expenses to their value over a single period.

Specifically, service prices/revenues are assumed exogenously driven by market forces, and are denoted by p_1 and p_2 that correspond to the discounted value of all present and future earnings apportioned over a single period. Similarly, costs are categorized into fixed and variable costs, with the latter consisting of two components. One that grows with the *realized demand* for the service and the other that grows with the *level of provisioning* in anticipation of a certain realization of the demand. Note that the latter is incurred irrespective of the actual realized demand (this is the price of uncertainty). Normalized fixed costs are denoted as c_s in a shared network, and as $c_i,\ i = \{1, 2\}$ in separate networks. We assume that a shared network affords economies of scope in fixed costs so that $c_s < c_1 + c_2$.

The quantities v_{s1} and v_{s2} correspond to the variable costs that scale with realized service demands D_1 and D_2 in a shared network, while v_1 and v_2 denote corresponding quantities for separate networks. Similarly, the variables a_{s2}, a_2, correspond to the cost components that depend on the levels of provisioning K_{s2} and K_2 for Service 2, and a_{s1}, a_1 for pro-

[3]As mentioned in the previous section, the advent of virtualization technology will contribute further to this ability.

[4]We initially ignore economies of scale in resources, so that the total amount of resources provisioned in a shared network can support $K_1 + K_{s2}$ users.

visioning level $K_1 (= D_1)$ for Service 1, in shared and separate networks, respectively. Both types of variable costs can exhibit either economies or diseconomies of scope depending on assumptions on how potential savings associated with sharing of equipment or personnel compare to cost increases that arise from more sophisticated equipment or greater operational complexity in shared networks. For example, when network sharing is by way of an overlay, $v_{s1} = v_1$ and $a_{s1} = a_1$, while in a truly integrated network $a_{s1} > a_1$ since more expensive equipment is usually required but $v_{s1} \leq v_1$ since various facilities are shared across services. All the above cost parameters take positive values only. The model detailed in the next sub-sections can accommodate all possible combinations of economies and diseconomies of scope in the cost components.

The last parameter of the model, α, denotes the extent to which it is possible to capture realized demand in excess of what the network was originally provisioned for (see Section 3.3 for details on the provisioning procedure). When $\alpha = 0$ any excess demand is lost, while $\alpha = 1$ corresponds to a scenario where network provisioning can be adjusted without penalty to accommodate the full demand. In other words, when $\alpha = 1$, there is no need for a "provisioning phase," since resources can be secured on-the-fly. By varying α, we account for different levels of flexibility in the allocation of network resources, *e.g.,* as afforded by different types of virtualization technology. Of interest, as discussed in Section 4, is the fact that different values of α can translate into different answers regarding whether shared or separate networks are more effective.

3.2 Network Costs and Revenues

This section details cost and revenue models for the shared and separate network solutions based on the parameters introduced in the previous section.

3.2.1 Separate Networks

For Service 1, the provider incurs a fixed cost of c_1, a variable (operational) cost of v_1 per customer and a variable cost of a_1 for the resources needed to support it, thus giving a total cost of $c_1 + v_1 D_1 + a_1 D_1$. The Profit for Service 1 is therefore given by

$$\Pi_1 = (p_1 - v_1 - a_1) D_1 - c_1 \tag{1}$$

For Service 2, a fixed cost of c_2, a variable (deployment and operational) costs of v_2 per customer and a variable cost of a_2 for the provisioned resources are incurred. The profit depends on whether the realized demand, D_2, is greater than or less than the resources, K_2, provisioned for it. When the realized demand D_2 is less than K_2, the total cost is $c_2 + v_2 D_2 + a_2 K_2$, and the profit from Service 2 is

$$R_2(D_2 < K_2) = (p_2 - v_2) D_2 - a_2 K_2 - c_2 \tag{2}$$

When the realized demand D_2 exceeds K_2, the network provisioning needs to be adjusted upward[5] to account for the excess demand, of which a fraction α can then be accommodated, *i.e.,* resources are increased to $K_2 + \alpha(D_2 - K_2)$, which correspond to a total cost of $c_2 + (v_2 + a_2)(K_2 + \alpha(D_2 - K_2))$. Thus profit from Service 2 in this scenario will be

$$R_2(D_2 > K_2) = (p_2 - v_2 - a_2)(K_2 + \alpha(D_2 - K_2)) - c_2 \tag{3}$$

3.2.2 Shared Networks

In a shared network, a fixed cost of c_s is jointly borne by the two services. The provider incurs a cost of $(v_{s1} + a_{s1}) D_1$ for Service 1, where both v_{s1} and a_{s1} can differ from their corresponding quantities in a dedicated network. Service 2 costs depend on how its realized demand, D_2, compares to the level of provisioning, K_{s2}.

When $D_2 < K_{s2}$, the network operates at less than full capacity and the cost incurred from Service 2 is $v_{s2} D_2 + a_{s2} K_{s2}$, thus giving a net profit from the two services equal to

$$\begin{aligned} R_s(D_2 < K_{s2}) = \quad & (p_2 - v_{s2}) D_2 - a_{s2} K_{s2} \\ & + (p_1 - v_{s1} - a_{s1}) D_1 - c_s \end{aligned} \tag{4}$$

When $D_2 > K_{s2}$, additional resources are secured to ultimately accommodate a fraction α of the excess demand, *i.e.,* resources are increased to $K_{s2} + \alpha(D_2 - K_{s2})$. The profit from Service 2 is then $(p_2 - v_{s2} - a_{s2})(K_{s2} + \alpha(D_2 - K_{s2}))$, and thus the total profit from the two services is

$$\begin{aligned} R_s(D_2 > K_{s2}) \quad = \quad & (p_2 - v_{s2} - a_{s2})(K_{s2} + \alpha(D_2 - K_{s2})) \\ & + (p_1 - v_{s1} - a_{s1}) D_1 - c_s \end{aligned} \tag{5}$$

3.3 Three Stage Model

The presence of uncertainty in the demand for Service 2 is the sole unknown in determining how to provision shared or separate networks, and consequently which one is more cost effective. In the absence of demand uncertainty, the "optimal" provisioning of either network solution is deterministic, *i.e.,* as given by setting $D_2 = K_2$ in eqs. (2-3) or $D_2 = K_{s2}$ in eqs. (4-5). As a result, identifying which is more effective is immediate once the respective economies and diseconomies of scope of each approach have been specified. This sub-section introduces the solution method used to compute optimal network provisioning levels in the presence of demand uncertainty for Service 2. For simplicity, the description given is for a dedicated network for Service 2, but a similar approach applies for a shared network.

The solution method consists of three logical phases. Phase 1 is the provisioning phase in anticipation of the demand for Service 2 based on its distribution f_{D_2}. Phase 2 is elementary and maps the realized demand onto the resources provisioned in Phase 1. Phase 3 accounts for the fact that a fraction α of any excess demand not accommodated in Phase 2 can eventually be captured. Under this model, the expected revenue R_2 *given* a provisioning level K_2 can be expressed as

$$\begin{aligned} \mathbf{E}(R_2 | K_2) \quad & = \int_0^{K_2} R_2(D_2 < K_2 | K_2) f'_{D_2} d(D_2) \\ & + \int_{K_2}^{D_2^{\max}} R_2(D_2 > K_2 | K_2) f'_{D_2} d(D_2) \end{aligned} \tag{6}$$

where $R_2(D_2 < K_2 | K_2)$ and $R_2(D_2 > K_2 | K_2)$ are given in eqs. (2) and (3), respectively, and f'_{D_2} is the density function of the demand for Service 2. For analytical tractability, D_2 is assumed uniformly distributed in $[0, D_2^{\max}]$. This choice magnifies the impact of uncertainty by making all possible levels of demand equally likely. However, it does not affect

[5]Note that we assume that resources can not be revised downward when $D_2 < K_2$, *e.g.,* because of contractual constraints.

findings regarding the influence of α in deciding the best network solution.

Based on eq. (6), the optimal provisioning level K_2^* is obtained from comparing profit when K_2 is such that $\frac{\partial \mathbf{E}(R_2|K_2)}{\partial K_2} = 0$ to (boundary) profits when $K_2 = 0$, and $D_2^{\max}$ (see Section 4.1.1 for details).

4. ANALYSIS

This section introduces the solution to the optimal resource allocation problem, and investigates the impact on the choice of network solution (shared or separate) of the parameter α that captures the ability to "re-provision" to accommodate excess demand.

4.1 Optimal Resource Allocations & Profits

As mentioned earlier, optimal resource allocation is relevant only for Service 2 that exhibits uncertainty in its demand. The optimal provisioning level maximizes eq. (6) in the case of separate networks, and a similar expression for shared networks. In this section, we derive expressions for these quantities under the assumption that Service 2 is profitable.

4.1.1 Separate Networks

Service 1 has a stable demand equal to D_1, so that $K_1 = D_1$ and the corresponding profit Π_1 earned from Service 1 is as given in eq. (1). As stated in Section 3.3, the optimal amount of resources for Service 2, K_2^*, (typically) satisfies $\frac{\partial \mathbf{E}(R_2|K_2)}{\partial K_2} = 0$ in eq. (6). This gives

$$K_2^* = \frac{(1-\alpha)(p_2 - v_2 - a_2)D_2^{\max}}{(1-\alpha)(p_2 - v_2) + \alpha a_2} \tag{7}$$

As expected, eq. (7) shows that when $\alpha = 1$, $K_2^* = 0$, *i.e.,* the ability to reprovision without penalty obviates the need for provisioning. On the other hand, when $\alpha = 0$, K_2^* is maximum, *i.e.,* the required provisioning is the highest to account for the fact that any excess demand is lost. More specifically, we have:

PROPOSITION 1. *Assuming that offering Service* 2 *is profitable,* i.e., $p_2 - v_2 - a_2 > 0$, *we have*

$$\frac{\partial K_2^*}{\partial \alpha} = \frac{-a_2(p_2-v_2-a_2)}{[(1-\alpha)(p_2-v_2)+\alpha a_2]^2} < 0$$
$$\frac{\partial K_2^*}{\partial a_2} = \frac{-(1-\alpha)(p_2-v_2)}{[(1-\alpha)(p_2-v_2)+\alpha a_2]^2} < 0$$
$$\frac{\partial K_2^*}{\partial v_2} = \frac{-(1-\alpha)a_2}{[(1-\alpha)(p_2-v_2)+\alpha a_2]^2} < 0$$

Optimal provisioning for Service 2*,* K_2^**, decreases as* α *increases, because of the greater ability to accommodate excess demand by upgrading resources. Similarly, increases in* v_2 *(the cost incurred per unit of demand), or* a_2 *(the cost per unit of provisioning) lower the profit margin* $p_2 - v_2 - a_2$ *per unit of demand, and so the optimal provisioning level is also lowered.*

Substituting K_2^* in $\mathbf{E}(R_2|K_2)$, gives the expected profit for Service 2 under optimal provisioning:

$$\Pi_2 = \frac{(p_2 - v_2 - a_2)D_2^{\max}}{2}\left(1 - \frac{(1-\alpha)a_2}{(1-\alpha)(p_2 - v_2) + \alpha a_2}\right) - c_2 \tag{8}$$

The total Profit from the two separate networks for Services 1 and 2 can be written as $\Pi_{sep} = \Pi_1 + \Pi_2$:

$$\begin{aligned}\Pi_{sep} = \;& \frac{(p_2 - v_2 - a_2)D_2^{\max}}{2}\left(1 - \frac{(1-\alpha)a_2}{(1-\alpha)(p_2 - v_2) + \alpha a_2}\right) \\ & + (p_1 - v_1 - a_1)D_1 - (c_1 + c_2)\end{aligned} \tag{9}$$

4.1.2 Shared Networks

In a shared network, Service 1 users are again allocated $K_1 = D_1$, which gives profit of $(p_1 - v_{s1} - a_{s1})D_1$. For Service 2, the expected profit for uniform demand distribution in $[0, D_2^{\max}]$ is computed from eqs. (4-5).

$$\begin{aligned}\mathbf{E}(R_s|K_{s2}) = \;& \int_0^{K_{s2}} R_2(D_2 < K_{s2}|K_{s2}) f'_{D_2} d(D_2) \\ & + \int_{K_{s2}}^{D_2^{\max}} R_2(D_2 > K_{s2}|K_{s2}) f'_{D_2} d(D_2)\end{aligned} \tag{10}$$

The optimal provisioning level K_{s2}^* is then given by

$$K_{s2}^* = \frac{(1-\alpha)(p_2 - v_{s2} - a_{s2})D_2^{\max}}{(1-\alpha)(p_2 - v_{s2}) + \alpha a_{s2}} \tag{11}$$

By similarity with eq. (7), we have

PROPOSITION 2. *The value of* K_{s2}^* *decreases with* v_{s2}*,* a_{s2} *and* α*,* i.e., $\frac{\partial K_{s2}^*}{\partial v_{s2}} < 0$, $\frac{\partial K_{s2}^*}{\partial a_{s2}} < 0$ *and* $\frac{\partial K_{s2}^*}{\partial \alpha} < 0$.

The corresponding optimal expected profit Π_{shr} is

$$\begin{aligned}\Pi_{shr} = \;& \frac{(p_2 - v_{s2} - a_{s2})D_2^{\max}}{2}\left(1 - \frac{(1-\alpha)a_{s2}}{(1-\alpha)(p_2 - v_{s2}) + \alpha a_{s2}}\right) \\ & + (p_1 - v_{s1} - a_{s1})D_1 - c_s\end{aligned} \tag{12}$$

4.2 The Impact of α on Network Choices

We focus on scenarios with $\Pi_2 > 0$, where the choice is between shared and separate networks. Inserting the expressions for K_2^* and K_{s2}^* of eqs. (7) and (11) in eqs. (9) and (12) gives the following relation for preferring shared over separate networks, *i.e.,* $\Pi_{shr} > \Pi_{sep}$

$$a_2 K_2^* - a_{s2} K_{s2}^* > 2\gamma, \tag{13}$$

where

$$\begin{aligned}\gamma = \;& \left[\left((v_{s2} + a_{s2})\frac{D^{\max}}{2} + (v_{s1} + a_{s1})D_1 + c_s\right)\right. \\ & \left. -\left((v_2 + a_2)\frac{D^{\max}}{2} + (v_1 + a_1)D_1 + (c_1 + c_2)\right)\right]\end{aligned} \tag{14}$$

The parameter γ captures the difference in the expected costs of shared and separate networks in the absence of any impact from provisioning decisions, *i.e.,* assuming the network is perfectly provisioned to accommodate the realized demand as would be the case when $\alpha = 1$. As a result, γ is independent of α.

On the other hand, the term $a_2 K_2^* - a_{s2} K_{s2}^*$ in eq. (13) depends on α, so that varying α can affect whether or not the inequality in eq. (13) holds. Hence, a different α can change network preference from shared to separate (or vice versa). We explore this next.

At $\alpha = 1$, the left hand side of eq.(13) is zero (as $K_2^* = K_{s2}^* = 0$ since provisioning is not needed). Therefore, a shared network is preferred when $\gamma < 0$, and a separate is otherwise. The effect of a decrease in α on the inequality of eq. (13) will then depend on (i) the magnitude of γ

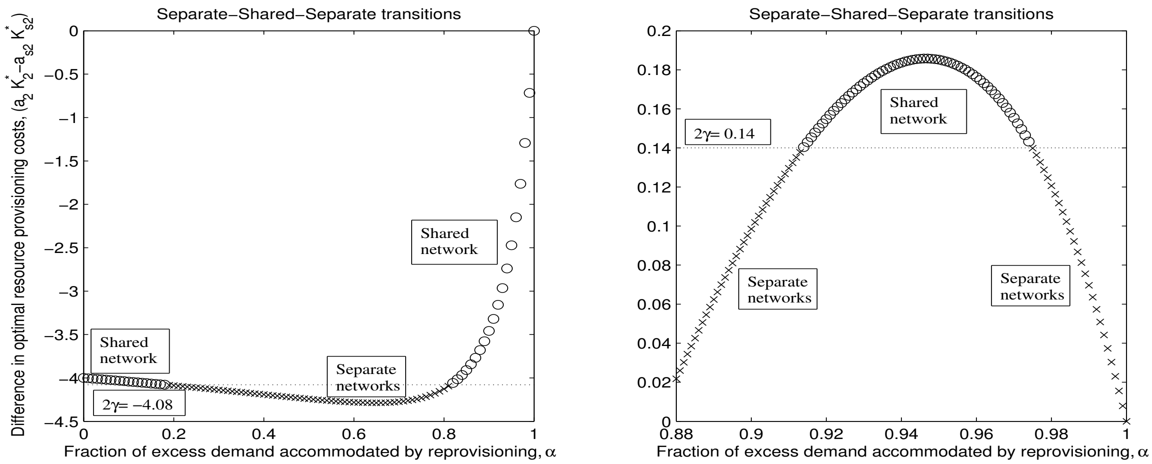

Figure 1: Impact of α on the choice of network solution

(how far it is from zero), and (ii) the sign of the derivative[6] $\frac{\partial(a_2 K_2^* - a_{s2} K_{s2}^*)}{\partial \alpha}$ at $\alpha = 1$.

Next, we provide conditions for a decrease in α to effect changes in the inequality of eq. (13).

Case 1: Shared to Separate

At $\alpha = 1$, a shared network is preferred if $\gamma < 0$. As α decreases from 1, a transition from preferring shared to separate can occur if the left hand side of eq. (13) also decreases with α to a value less than γ[7]. This requires $\frac{\partial(a_2 K_2^* - a_{s2} K_{s2}^*)}{\partial \alpha}\Big|_{\alpha=1} > 0$, which gives the condition $p_2 - v_{s2} - a_{s2} > p_2 - v_2 - a_2$, *i.e.,* the profit margin for a shared network should be higher than for separate networks. Intuitively, this implies that the loss of some excess demand (from decreasing α) results in a higher marginal loss on a shared network. This translates into higher provisioning levels in a shared network, thus making $a_2 K_2^* - a_{s2} K_{s2}^*$ more negative.

Proposition 3. *If at $\alpha = 1$, a shared network is preferred, then as α decreases so that some of the excess demand is lost, a transition to separate networks can occur if (i) $p_2 - v_{s2} - a_{s2} > p_2 - v_2 - a_2$ and (ii) $0 > \gamma > \min_\alpha(a_2 K_2^*(\alpha) - a_{s2} K_{s2}^*(\alpha))$.*

Case 2: Separate to Shared

A separate network is preferred at $\alpha = 1$ if $\gamma > 0$. A transition to using separate networks occurs when decreasing α, if $\frac{\partial(a_2 K_2^* - a_{s2} K_{s2}^*)}{\partial \alpha}\Big|_{\alpha=1} < 0$, *i.e.,* the left hand side of eq. (13) increases and eventually exceeds γ. This corresponds to a symmetric condition and interpretation as that of Proposition 3, namely,

Proposition 4. *If at $\alpha = 1$ a separate network is preferred, then as α decreases and some of the excess demand is lost, a transition to a shared network can occur if (i) $p_2 - v_{s2} - a_{s2} < p_2 - v_2 - a_2$ and (ii) $0 < \gamma < \max_\alpha(a_2 K_2^*(\alpha) - a_{s2} K_{s2}^*(\alpha))$.*

The above discussion demonstrates that the ability to reprovision a network to accommodate unexpected excess demand, as captured by α, can affect the choice of network solution. In the remainder of this section, we illustrate that α can have even more far-reaching effects, and for example result in multiple transitions from, say, ‘shared to separate to shared’ as it varies.

This is illustrated in the left hand-side of Figure 1, with the right hand-side displaying a symmetric behavior starting from 'separate'. The choice of parameters for the left hand-side of Figure 1 is ($D_1 = D_2^{\max} = 10$, $p_1 = 6$, $p_2 = 20$, $c_1 = 15$, $c_2 = 10$, $c_s = 15$, $v_1 = 2$, $a_1 = 2$, $v_{s1} = 2$, $a_{s1} = 4.796$, $v_2 = 15$, $a_2 = 1$, $v_{s2} = 20/3$, $a_{s2} = 4/3$). This corresponds to a scenario where a shared network exhibits economies of scope in its fixed costs and in the deployment costs of Service 2. However, diseconomies of scope arise in the operational costs of both Services 1 and 2 in the shared network. Under those conditions, we see that a shared network is preferred when $\alpha = 1$ as well as when $\alpha = 0$, with an intermediate region where separate networks are preferred. The situation for $\alpha = 1$ is as predicted by Proposition 3, but the double transition (to separate and back to shared) as α decreases from 1 to 0 calls for additional conditions. Specifically, it can be shown that this double transition requires $\frac{p_2 - v_2 - a_2}{a_2} < \frac{p_2 - v_{s2} - a_{s2}}{a_{s2}}$, *i.e.,* the ratio of profit to cost of provisioning per user needs to be higher in the shared than in the separate networks.

Conversely, in the right hand-side of Figure 1, parameters are chosen to correspond to an overlay network scenario for Service 2 as follows: ($D_1 = D_2^{\max} = 10$, $p_1 = 6$, $p_2 = 20$, $c_1 = 10$, $c_2 = 10$, $c_s = 16.07$, $v_1 = 2$, $a_1 = 2$, $v_{s1} = 2$, $a_{s1} = 2$, $v_2 = 15$, $a_2 = 1$, $v_{s2} = 14.8$, $a_{s2} = 2$). As a result, Service 1 is essentially unaffected by the use of a shared network, but Service 2 sees limited economies of scope in its deployment and still experiences diseconomies of scope in its operation, *e.g.,* because of possible interactions in using a shared infrastructure. Under this scenario, the conditions of Proposition 4 predict the preference for shared networks when $\alpha = 1$, but the presence of a double transition first to separate and then back to shared when $\alpha = 0$ calls

[6]The rate of change of the difference in provisioning costs w.r.t. α, *i.e.,* $\frac{\partial(a_2 K_2^* - a_{s2} K_{s2}^*)}{\partial \alpha}$, is given by

$$\frac{-{a_2}^2(p_2 - v_2 - a_2)}{[(1-\alpha)(p_2 - v_2) + \alpha a_2]^2} + \frac{{a_{s2}}^2(p_2 - v_{s2} - a_{s2})}{[(1-\alpha)(p_2 - v_{s2}) + \alpha a_{s2}]^2}$$

[7]Obviously, this also requires γ to be greater than the minimum possible value of $a_2 K_2^* - a_{s2} K_{s2}^*$.

again for additional conditions. Specifically, this requires a symmetric condition to that of the left hand-side of Figure 1, *i.e.,* $\frac{p_2 - v_2 - a_2}{a_2} > \frac{p_2 - v_{s2} - a_{s2}}{a_{s2}}$.

5. CONCLUSION & FUTURE WORK

This paper sought to investigate when shared or separate networks offer a more effective solutions to the deployment of a new service. The focus was on highlighting that the increased flexibility available in allocating network resources, *e.g.,* through technologies such as virtualization, calls for models that incorporate a reprovisioning phase. The paper established the impact such a capability can have on the choice of network solutions. It represents a first step towards a full-fledged investigation. Using the models outlined in the paper, we are currenly exploring what factors and service features influence the trade-off between the economies of scope and scale of a shared network and the diseconomies of scope that interactions between services can create.

6. ACKNOWLEDGEMENTS

The authors would like to acknowledge Kristin Yamauchi for help with simulations that provided useful insight, and Profs. A. Odlyzko and Z.-L. Zhang for their comments as the work evolved.

7. REFERENCES

[1] E. K. Bish, A. Muriel, and S. Biller. Managing flexible capacity in a make-to-order environment. *Management Science*, 51(2):167–180, 2005.

[2] C. H. Fine and R. M. Freund. Optimal investment in product-flexible manufacturing capacity. *Management Science*, 36(4):449–465, 1990.

[3] S. C. Graves and B. T. Tomlin. Process flexibility in supply chains. *Management Science*, 49(7):907–919, 2003.

[4] W. C. Jordan and S. C. Graves. Principles of the benefits of manufacturing process flexibility. *Management Science*, 41(4):577–594, 1995.

[5] J. A. Van Mieghem. Investment strategies for flexible resources. *Management Science*, 44(8):1071–1078, 1998.

[6] L. Peterson, S. Shenker, and J. Turner. Overcoming the Internet impasse through virtualization. In *Proc. ACM HotNets-III*, 2004.

[7] J. Touch, Y. Wang, L. Eggert, and G. Finn. A virtual Internet architecture. In *Proc. ACM FDNA'03*, Karlsruhe, Germany, 2003.