A Real Option Approach to Telecommunications Network Optimization

Dohyun Pak and Jussi Keppo

Department of Industrial and Operations Engineering, The University of Michigan

1205 Beal Avenue, Ann Arbor, Michigan 48109-2117, USA

E-mail: pakd@umich.edu and keppo@umich.edu

We optimize network flow by minimizing network blocking and/or delay and by modeling network routing possibilities as real options. The uncertainties in the network are driven by stochastic point-to-point demands and we consider correlations among them in a general network structure. We derive an analytical approximation for the blocking/delay probabilities and solve the optimal network flows by using a global optimization technique. We illustrate the model with examples.

Key words: Optimization, network, basket option.

1. Introduction

The basic goal of network routing is to respond randomly fluctuating network demands by rerouting traffics and reallocating resources. Nowadays many network systems, for instance telecommunications networks, are able to do this so well that in many respects large-scale networks appear as coherent and almost intelligent organisms. However, as network structures become more complicated and new network services and products are developed, the more efficient methodologies for network routing selection and the estimation of network blocking and delay probabilities are needed. The development of such methodologies presents challenges of a mathematical, engineering, and economic nature.

In this paper we optimize network flows by minimizing network blocking/delay in a general network and under stochastic point-to-point demands. Our network optimization method is based on real option modeling. We assume that point-to-point transmission times are independent of network routings and the network capacities are constant. That is, regardless of the selected point-to-point connection, transmission time is constant and we do not optimize the network capacities. This is the case in telecommunications networks and, thus, this paper is a real option application to telecommunications network routing. If the network under consideration is a calling network then our method minimizes the network blocking, and if the network has buffers, such as Internet network, then we minimize the network delays. The network blocking/delay can be approximated analytically by using a financial basket option formula. The network has an option to change its routing and this flexibility is a real option. That is, routing alternatives are real options and, therefore, a routing change corresponds to the exercise of one of these options. Basic real option results imply that the network routers have high value if there is a high uncertainty in the network demands since in this case there is a high probability that network routing is changed. Further, if a network has lot of nodes and routers then the value of the network is high because it has lot of routing options. The network optimization model of this paper maximizes the value of network and, hence, the value of the routing options. Real option theory is summarized, e.g., in Dixit and Pindyck (1994) and telecommunications applications are considered, for instance, in Alleman and Noam (1999) and Keppo (2001, 2003). After the real option modeling we solve optimal network routing by using routing probabilities and the analytical representation of network blocking/delay. The optimization of routing probabilities is a nonconvex optimization problem and we utilize global optimization techniques [for global optimization methods see, e.g., Horst and Pardalos (1995) and Neumaier (2004)]

Many papers have analyzed network traffics. Caceres, Danzig, Jamin, and Mitzel (1991), Leland, Taqqu, Willinger, and Wilson (1994), and Feldmann (1996) have shown that telecommunications traffic is quite complex, exhibiting phenomena such as long-tail probability distributions, long-range dependence, and self-similarity. Therefore, various assumptions on traffic processes are made to simplify the network routing models. For instance, Kelly (1991, 1996) assumes an independent Poisson process for actual network traffic demand and Norros (1994) models traffic with a fractional Brownian motion. In this paper we consider Brownian motion driven network demands and this way we are able to apply real option framework and derive the analytical approximation for network's blockings/delays. A similar demand model is used, e.g., in Zhao and Kockelman (2002), Ryan (2002), and Gune and Keppo (2002). Gune and Keppo show empirically that dial-up demand is usually distributed according to a log-normal distribution. They also show that there are cycles in the demand and hence the parameters of the distribution depend on time. In the present paper we utilize this result and model network point-to-point demands with log normal distributions.

Network routing is carried out along various routes. Routing has been studied extensively over the last few decades. For instance, some telecommunication companies have extended traditional static call routing methods to dynamic strategies. These strategies route calls depending on the given network load and, therefore, they guarantee better quality [see e.g. Ash and Oberer (1989), Ash, Chen, Frey, and Huang (1991), Ash (1998), and Gune and Keppo (2002)]. For instance, DAR (Dynamic Alternate Routing) by British Telecom routes calls along the direct routing between the start and end points as long as there is free capacity and if the direct connection is blocked, the call is routed along an alternative route. Further, if there exist several alternative routing possibilities, the alternative routing is selected by using the historical blocking data. Our model can be seen as an extension to DAR since we use a similar routing strategy. However, our routing selection is based on the future network blocking that depends on the current network demands, their stochastic processes, and the correlations between the demands. For instance, one point-to-point blocking probability might be high even though there have not been any blockings in the history if the point-to-point demand is currently high first time. Further, in contrast to DAR we also consider explicitly the interactions between the point-to-point routing decisions. That is, if one point-to-point routing is changed then it affects the blocking probabilities of the other point-to-point connections and in order to find the optimal routing these effects have to be considered. Mitra, Morrison and Ramakrishnan (1996, 1999) consider network optimization in multi-service broadband networks. They maximize network revenue by using Poisson processes and the corresponding end-to-end loss probabilities for each service and route. They assume no buffer, i.e., if there is insufficient bandwidth on a link an arriving call is blocked and lost. In this paper we also maximize the revenues of the network but we use Brownian motion driven network demands in order to model the correlations between the demands and interactions between the routing decisions.

In addition to calling networks, our network model can also be used in the optimization of Internet's backbone network. Currently the routing of this network is static. However, for instance, Rai and Samaddar (1998) and Bieler and Stevensen (1998) predict that the amount of transmitted data and the number of Internet hosts and connections increase exponentially. This creates network delays in the future and, because routing optimization is a cheap alternative to capacity investments, backbone network optimization becomes important. This paper suggests a framework to the backbone network optimization by minimizing the network delays.

The rest of the paper is divided as follows: Section 2 introduces the underlying models used in the paper and derives a representation for the point-to-point traffic. The stochastic processes for the demands are defined and these processes are then used in the network optimization in Section 3. Section 4 illustrates the model with examples. Section 5 discusses the implementation of the optimization method and finally Section 6 concludes.

2. Network flow representation

Telecommunications and Internet networks are in general modeled as graphs, which have nodes (verticals) and edges. Graphs are a natural choice for telecommunications and Internet networks, because the networks are not fully meshed. Fully or almost fully meshed networks can be modeled with transition matrices as is done, e.g. in Harrison (1988) with processing networks. In the graph, nodes act as endpoints for point-to-point connections. A direct point-to-point connection constitutes an edge also known as a link. Other connections are modeled as a sequence of links, often referred as routes, and correspond to paths in the graph. In this research, we consider blocking probabilities, capacities, and traffic on network links. We make the following assumption on network capacities.

ASSUMPTION 2.1 *Network capacities are constant.*

According to Assumption 2.1 in our model, we are not able to optimize the capacities. An investment model where also capacities are optimized is considered, e.g., in Keppo (2003). The point-to-point demands follow stochastic processes and we specify these processes in Section 3.

2.1 Simple three point-to-point network structure

In this section we consider a simple network of three fixed point-to-point capacities (C) and demands (D). For example, the points could be New York, Los Angeles, and Atlanta, and all the point-to-point capacities are OC-3 (155.52 Mbps). Figure 2.1 illustrates the situation.

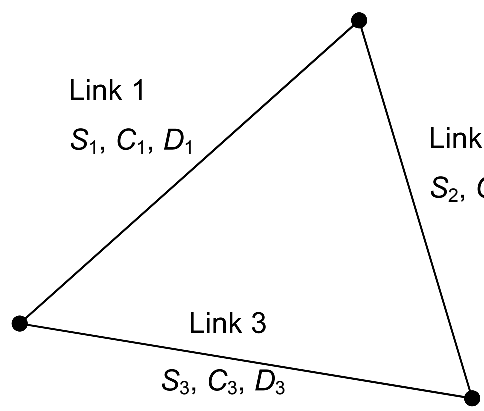

Figure 2.1 *Network costs* (S), *capacities* (C), *and demands* (D).

The S-costs in Figure 2.1 are costs from the blocking/delay of the links. For instance, S_1 is the blocking/delay cost between the up and left points by using the direct routing between them. Note that S_1 does not necessarily equal the realized cost between the up and left points because if we have $S_2 + S_3 \leq S_1$ then it is optimal to use the longer routing and, therefore, the realized cost between the up and left points equals $S_2 + S_3$.

According to Assumption 2.1 each link's capacity is constant. However, the demands are stochastic and, therefore, the network routing might be changing all the time. For instance, part of demand D_1 might be routed via the alternative routing due to the limited capacity on the direct routing. This means that using Figure 2.1 we can represent the routing possibilities with the following table.

Table 2.1 *Routing alternatives. The numbers represent the routing numbers.*

	Link 1	Link 2	Link 3
Demand 1	1	2	2
Demand 2	2	1	2
Demand 3	2	2	1

In Table 2.1, the first row implies that Link 1 is used as D_1's first routing possibility (routing number 1) and links 2 and 3 construct the second routing possibility (routing number 2). In the same way, the second row implies that Link 2 is D_2's first routing possibility and links 1 and 3 are the second routing possibility. On the other hand, the first column implies that the traffic on Link 1 consists of D_1's first routing, D_2's second routing, and D_3's second routing.

We denote $\beta_1^{\,1}$ as the proportion of demand D_1 that is routed through its first routing possibility and, therefore, $(1-\beta_1^{\,1}) = \beta_1^{\,2}$ as the proportion that is routed via the second routing. Thus, demand $\beta_1^{\,1}D_1$ is for the first routing and $\beta_1^{\,2}D_1$ for the second. For demands D_2 and D_3 we have similar representation. Therefore, $\beta_2^{\,1}D_2$ and $\beta_3^{\,1}D_3$ are for their first routings and $\beta_2^{\,2}D_2$ and $\beta_3^{\,2}D_3$ for the second routings. Then we can represent the flow on the links as follows

$$\begin{aligned} F_1(t) &= \beta_1^1 D_1(t) + \beta_2^2 D_2(t) + \beta_3^2 D_3(t) \\ F_2(t) &= \beta_2^1 D_2(t) + \beta_1^2 D_1(t) + \beta_3^2 D_3(t)\,, \\ F_3(t) &= \beta_3^1 D_3(t) + \beta_1^2 D_1(t) + \beta_2^2 D_2(t) \end{aligned} \tag{2.1}$$

where $F_i(t)$ is the flow on Link i at time t and $\sum_{j=1}^{2} \beta_i^j = 1$ for all $i \in \{1,2,3\}$.

According to (2.1) F_1 consists of the first routing part of D_1 and the second routing parts of D_2 and D_3. Thus, we just read the first column of Table 2.1. In the same way, F_2 and F_3 are given by the second and third columns of Table 2.1. Note that D-processes are demands between the start and end nodes over all possible routings and F-processes are total flows on the physical links between the nodes.

2.2 General network structure

With a more complex network structure, the flow representation on each link is similar as in the previous simple model. We first construct the routing table and then from that table we get the flow representation for each link as a function of betas and demands.

We consider a general telecommunications network with n point-to-point connections. Therefore, there exist n point-to-point capacities (C) and demands (D) and the problem is to construct the corresponding flow representation. As expected, this flow representation depends on the structure of the network, i.e., even though there is the same number of point-to-point connections, the representation can be different with different network structures. First, we consider a seven point-to-point connections' network example as an extension to Section 2.1. The network structure for this example is illustrated in Figure 2.2.

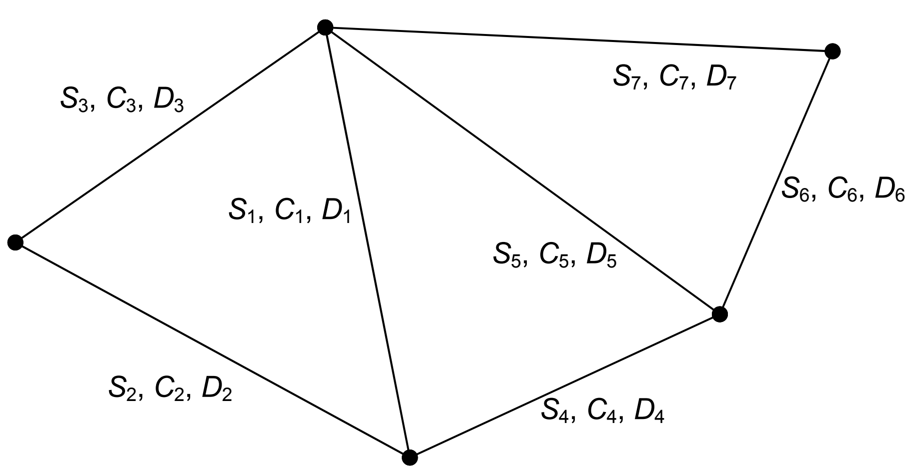

Figure 2.2 *Network costs* (S), *capacities* (C), *and demands* (D).

From Figure 2.2, we can construct the routing table, Table 2.2. We use the following rule in numbering the routings. The less links in the route the smaller is the routing number and the routings with the same number of links have an arbitrary sequence among them. For instance, for D_1, Link 1 gets the routing number 1, Link 2 + Link 3 and Link 4 + Link 5 get the routing number 2 or 3, and Link 4 + Link 6 + Link 7 gets the routing number 4.

Table 2.2 *Routing alternatives. The numbers represent the routing numbers.*

	Link 1	Link 2	Link 3	Link 4	Link 5	Link 6	Link 7
D_1	1	2	2	3, 4	3	4	4
D_2	2	1	2, 3	3, 4	3	4	4
D_3	2	2, 3, 4	1	3, 4	3	4	4
D_4	2, 4	3, 5	3, 5	1	2, 3	4, 5	4, 5
D_5	2	4	4	2, 4	1	3	3
D_6	3	4	4	3, 4	2	1	2, 3, 4
D_7	3	4	4	3, 4	2	2, 3, 4	1

The first row of Table 2.2 indicates that Link 1 is used as the first routing of D_1. Links 2 and 3 are the second routing of D_1, links 4 and 5 are the third, and links 4, 6, and 7 are the fourth routing. The other demands can be analyzed in the same way. On the other hand, the first column of Table 2.2 implies that the flow on Link 1 consists of the D_1's routing 1, D_2's routing 2, D_3's routing 2, D_4's routings 2 and 4, D_5's routing 2, D_6's routing 3, and D_7's routing 3. Thus, the amount of flow on Link 1 can be represented as a linear combination of the demands. The parameters of the linear mapping are our decision parameters, betas. The construction of the routing table is again the first step in the flow representation. This table is created row by row by using the point-to-point demand representations and then the corresponding flows are the columns of that table.

As in Section 2.1, we can create the flow representation from Table 2.2's columns. For instance, the total flow on Link 1 is demand D_1 multiplied by the direct routing probability $\left(\beta_1^1\right)$ plus the flow from other demands, i.e.,

$$F_1(t) = \beta_1^1 D_1(t) + \beta_2^2 D_2(t) + \beta_3^2 D_3(t) + (\beta_4^2 + \beta_4^4) D_4(t) + \beta_5^2 D_5(t) + \beta_6^3 D_6(t) + \beta_7^3 D_7(t)\,, \tag{2.2}$$

where $0 \leq \beta_i^j \leq 1$ for all $i \in \{1, \ldots, 7\}$ and $j \in \{1, \ldots, 4\}$, and $\sum_{j=1}^{4} \beta_i^j = 1$ for all $i \in \{1, \ldots, 7\}$. Similarly with the other links we have from Table 2.2

$$\begin{aligned} F_2(t) &= \beta_2^1 D_2(t) + \beta_1^2 D_1(t) + (\beta_3^2 + \beta_3^3 + \beta_3^4) D_3(t) + (\beta_4^3 + \beta_4^5) D_4(t) + \beta_5^4 D_5(t) + \beta_6^4 D_6(t) + \beta_7^4 D_7(t) \\ &\vdots \\ F_7(t) &= \beta_7^1 D_7(t) + \beta_1^4 D_1(t) + \beta_2^4 D_2(t) + \beta_3^4 D_3(t) + (\beta_4^4 + \beta_4^5) D_4(t) + \beta_5^3 D_5(t) + (\beta_6^2 + \beta_6^3 + \beta_6^4) D_6(t) \end{aligned} \tag{2.3}$$

where $0 \leq \beta_i^j \leq 1$ for all $i \in \{1, \ldots,7\}$ and $j \in \{1, \ldots, m_i\}$, and $\sum_{j=1}^{m_i} \beta_i^j = 1$ for all $i \in \{1, \ldots,7\}$. Variable m_i is the number of possible routings for demand i, and in this case, all the demands except D_4 have four possible routings and D_4 has five, therefore $m_4 = 5$ and $m_i = 4$ for $i \in \{1, \ldots,7\}-\{4\}$ in this example.

For the general network problem, let $I_k^{i,j}$ be a network structure indicator and it is defined as follows

$$I_k^{i,j} = \begin{cases} 1, & \text{if Link } k \text{ is used in } D_i\text{'s } j\text{'th routing} \\ 0, & \text{otherwise} \end{cases}. \tag{2.4}$$

Note that this indicator corresponds to the routing tables, tables 2.1 and 2.2.

Using this notation the total flow on the k'th link in a general network is represented as follows

$$F_k(t) = \sum_{j=1}^{m_1} \beta_1^j I_k^{1,\,j} D_1(t) + \sum_{j=1}^{m_2} \beta_2^j I_k^{2,\,j} D_2(t) + \cdots + \sum_{j=1}^{m_n} \beta_n^j I_k^{n,\,j} D_n(t) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \beta_i^j I_k^{i,\,j} D_i(t)\,, \tag{2.5}$$

where n is the number of links in the network, $0 \leq \beta_i^j \leq 1$ for all $i \in \{1, \ldots,n\}$ and $j \in \{1, \ldots, m_i\}$, and $\sum_{j=1}^{m_i} \beta_i^j = 1$ for all $i \in \{1, \ldots, n\}$.

According to (2.5) the total flow on each link can be represented as a linear combination of the point-to-point demands. In Section 5, we consider $m = \max_i [m_i]$ as a parameter to increase the efficiency of the network optimization.

3. Network Routing Optimization

In this section we optimize the network routing by using the flow representation of Section 2. The objective in the network routing is to maximize the network revenues minus the costs. That is,

$$\text{(3.1)} \qquad \sup \mathrm{E}\left[\int_0^T e^{-rt} \sum_{k=1}^{n} [P_k(t) - S_k(t)] dt\right],$$

where n is the number of links in the network, $[0,T]$ is the planning horizon, r is a constant discount rate, P_k and S_k are the k'th point-to-point connection's price and blocking/delay cost. The price P_k corresponds to the demand D_k and the cost S_k corresponds the network flow F_k. That is, P_k does not depend on the network structure because D_k is the total demand between the start and end points. Therefore, we make the following assumption.

ASSUMPTION 3.1 *The k'th point-to-point price P_k is independent of the network routing for all $k \in \{1, \ldots, n\}$.*

Assumption 3.1 implies that P_k depends on network capacity $C = (C_1, \ldots, C_n)$ that is constant and on demand $D = (D_1, \ldots, D_n)$. Hence, we can write $P_k(t) = P_k(D(t), C)$ and, therefore, the uncertainties in P_k are only from the demand process D. Further, because only the capacities and the demands affect the point-to-point price, the capacity prices are independent of the network routing and they do not affect the routing optimization. Pricing of the network has been studied in many papers. For instance, Kelly (1997) considers the optimal pricing model of a network by maximizing users' aggregated utility and shows a competitive equilibrium. Johari and Tsitsiklis (2004) extend Kelly's research to a game model, where selfish users anticipate the price effects of their actions.

If the network under consideration is a network with buffers, for instance Internet, then the blocked demands are added to the demand of the next period. That is,

$$D(t) = \widehat{D}(t) + B(t-),$$

where $\widehat{D}(t)$ is the new demand at time t and $B(t\text{-})$ is the buffer prior to time t. Mikosch et al. (1991) showed that this kind of cumulative broadband network traffic is well modeled by fractional Brownian motion. If we consider calling network then $B(t\text{-}) = 0$ since there are no buffers. In both cases, we make the following assumption on the stochastic process for demands in order to use the real option framework.

ASSUMPTION 3.2 *The process of the expected i'th demand $D_i(t,T)$ is given by the following Itô stochastic differential equation*

$$dD_i(t,T) = D_i(t,T)\sigma_i(t,T)dB_i(t) \quad \text{for all} \quad i \in \{1,\ldots,n\}, \quad t \in [0,T], \quad T \in [0,\tau] \tag{3.2}$$

where $D_i(t,T) = E\left[D_i(T)\middle|F_t\right]$ is the expected total demand of the i'th point-to-point connection at time T calculated with respect to the information at time t, $\sigma_i(\cdot,T)$ is deterministic and bounded, $B_i(\cdot)$ is the Brownian motion corresponding to the i'th point-to-point connection on the probability space (Ω, F, P) along with the standard filtration $\{F_t\colon t \in [0,\tau]\}$, and we denote by $\rho_{i,z}$ the correlation between the i'th and z'th Brownian motions.

According to equation (3.2), the stochastic processes for the expected point-to-point demand over all possible routings follow an exponential process. The boundedness of the volatility parameter guarantees the existence and uniqueness of the solution to (3.2). Assumption 3.2 is valid, e.g., if we can model the number of network users with a lognormal distribution and assume that each user receives the same amount of capacity. In this case $D_i(T)$ is distributed according to a lognormal distribution with mean $D_i(t,T)$ and variance $D_i^2(t,T)\left[\exp\left(\int_t^T \sigma_i^2(y,T)dy\right) - 1\right]$. Since we model expected value $D_i(t,T)$, the demand process $D_i(t)$ can be e.g. geometric Brownian motion or mean-reverting [see for instance Schwartz (1997)]. Similar demand models are used, e.g., in Zhao and Kockelman (2002), Ryan (2002), and Gune and Keppo (2002). Gune and Keppo show empirically that Assumption 3.2 usually holds with dial-up data that can be seen as a calling point-to-point demand. Therefore, Assumption 3.2 is most convenient for calling networks. However, for simplicity in this paper we use this log-normal assumption for all telecommunications networks. Note that if the above assumption does not hold we can extend our real option modeling to other demands by changing the process in (3.2), e.g., to a Poisson process.

Given Assumption 3.2 we can consider a QoS pricing example that is based on Keppo, Rinaz, and Shah (2002). In equation (3.1) we assumed that $P_k(t)$ is independent of the routing at time *t*. This is the case, e.g., if the point-to-point prices are leasing contract prices (or forward prices), i.e., if the prices are fixed at time 0. Let us first assume that T is small and the network structure is given by Figure 2.1. In this case $\beta_i^1 \approx 1$ if $D_i(T) \leq C_i$ and if $D_i(T) > C_i$ then $\beta_i^1 D_i(T) \approx C_i$. Further, the blocking indicator of demand D_1 is given by

$$\mathbf{1}\left\{D_1(T) \geq C_1\right\}\left(1 - \prod_{k\in\{2,3\}} \mathbf{1}\left\{D_k(T) \leq C_k - \left(D_1(T) - C_1\right)\right\}\right)$$

where, $\mathbf{1}\{X \geq Y\} = \begin{cases} 1, & \text{if } X \geq Y \\ 0, & \text{otherwise} \end{cases}$. Thus, if the above equation is equal to one then part of D_1 is blocked. Based on this blocking equation we get that *T*-maturity price for the first point-to-point connection at time 0 is given by [see details and extensions from Keppo, Rinaz, and Shah (2002)]

$$P_1(0,T) = P_1^0\left[N(d_1) + G\left(-d_1, d_2, d_3, -\rho_{1,2}, -\rho_{1,3}, \rho_{2,3}\right)\right]$$

where P_1^0 is a constant, $N(\cdot)$ is a cumulative standard normal distribution, $G\left(-d_1, d_2, d_3, -\rho_{1,2}, -\rho_{1,3}, \rho_{2,3}\right)$ is the area under a standard trivariate normal distribution function covering the region from -∞ to -d_1, -∞ to d_2, and -∞ to d_3, the three random variables have correlations $-\rho_{1,2}$, $-\rho_{1,3}$, and $\rho_{2,3}$, the variables of the cumulative distributions $d_1 = \dfrac{\ln\left(\frac{C_1}{D_1(0,t)}\right) + \frac{1}{2}\sigma_1^{\,2}(T)T}{\sigma_1(T)\sqrt{T}}$, $d_k = \dfrac{\ln\left(\frac{C_k - r_1(0,T)}{D_k(0,T)}\right) + \frac{1}{2}\sigma_k^{\,2}(T)T}{\sigma_k(T)\sqrt{T}}$ for all $k \in \{2,3\}$, and $r_1(0,T) = D_1(0,T)N\left(-d_1 + \sigma_1(T)\sqrt{T}\right) - C_1 N\left(-d_1\right)$. Note that the other point-to-point prices and other maturities are given in the same way. Thus, we can calculate point-to-point prices for all maturities and, because these prices are fixed at time 0, the future routing decisions do not affect the prices, i.e., Assumption 3.1 holds.

The next assumption gives the blocking/delay cost function.

ASSUMPTION 3.3 *Blocking/Delay cost S_k is given by*

$$S_k(t) = \max[F_k(t) - C_k, 0] \quad \text{for all} \quad t \in [0,T], \qquad k \in \{1,\ldots,n\}.$$

We assume that the investment cost for the fixed capacity is a sunk cost that is already paid. Therefore, our model considers only the costs from the blocking/delay of the network links that are due to the overflow of the links. Assumption 3.3 implies that this blocking/delay depends on the network demands and capacities as well as on the routing probabilities through F_k. Using (2.5) the cost S_k can be represented as follows

$$\begin{aligned} S_k(t,\beta) &= \max\left[F_k(t) - C_k, 0\right] = \max\left[\sum_{i=1}^{n}\sum_{j=1}^{m_i} \beta_i^j I_k^{i,j} D_i(t) - C_k, 0\right] \\ &= \max\left[\sum_{i=1}^{n} w_{k,i}(\beta) D_i(t) - C_k, 0\right], \end{aligned} \tag{3.3}$$

where $w_{k,i}(\beta) = \sum_{j=1}^{m_i} \beta_i^j I_k^{i,j}$, $\beta = \{\beta_1,\ldots,\beta_n\}$, and $\beta_i = \{\beta_i^1,\ldots,\beta_i^{m_i}\}$.

Equation (3.3) is similar as the payoff function of a financial basket call option where the demand processes are considered as the underlying assets, $w_{k,i}$ is the weight of the i'th underlying asset, and C_k is the strike price. By using the basket option pricing model of Gentle (1993) we get

$$E[S_k(T,\beta)] = F_k(\beta)\Big[c_k(\beta) N\big(l_1(T-t,\beta)\big) - \big(\tilde{K}_k(\beta) + c_k(\beta) - 1\big) N\big(l_2(T-t,\beta)\big)\Big], \tag{3.4}$$

where

$$c_k = \exp\left(\left[\frac{1}{2}\sum_{i,j=1}^{n} \rho_{i,j}\widehat{w}_{k,i}\widehat{w}_{k,j}\sigma_i\sigma_j - \sum_{j=1}^{n} \widehat{w}_{k,j}\sigma_j^{\,2}\right](T-t)\right),$$

$$l_{1,2}(t) = \frac{\ln c_k - \ln(\tilde{K}_k + c_k - 1) \pm 0.5 v_k^{\,2} t}{v_k\sqrt{t}},$$

$$v_k^{\,2} = \sum_{i,j=1}^{n} \rho_{i,j}\widehat{w}_{k,i}(\beta)\,\widehat{w}_{k,j}(\beta)\,\sigma_i\sigma_j,$$

$$\widehat{w}_{k,i}(\beta) = \frac{w_{k,i}(\beta)\,S_i}{\sum_{i=1}^{n} w_{k,i}(\beta)\,S_i}, \qquad \tilde{K}_k = \frac{e^{-r(T-t)}C_k}{\sum_{i=1}^{n} w_{k,i}(\beta)\,S_i},$$

and $N(\cdot)$ is a cumulative standard normal distribution.

Next we make the following assumption on the routing probabilities.

ASSUMPTION 3.4 *The routing probabilities are constants over the routing optimization time interval* $[0,T]$.

Assumption 3.4 implies that our routing optimization problem is static on $[0,T]$. This implies that we calculate and change the routing probabilities in discrete time e.g., every hour or every day. This is true in practice since network routing probabilities are changed in discrete time due to, e.g., data collection.

Using assumptions 3.1-3.4 and equation (3.1) we get that the optimization problem can be represented as follows

$$\inf_{\beta} E\left[\sum_{k=1}^{n}\int_{0}^{T} e^{-rt}S_{k}\left(t,\beta\right)dt\right] \tag{3.5}$$

such that $0 \leq \beta_i^j \leq 1$ and $\sum_{j=1}^{m_i}\beta_i^j = 1$ for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m_i\}$.

Equations (3.1) and (3.5) imply that because the point-to-point prices are independent of the network routing, the optimal network routing is solved by minimizing the point-to-point blocking/delay costs. Note that if we optimized each routing as follows

$$\inf_{\beta_i} E\left[\sum_{k=1}^{n}\int_{0}^{T} e^{-rt}S_{k}\left(t,\beta\right)dt\right], \tag{3.6}$$

where $\beta_i = \{\beta_i^1 \ldots \beta_i^{m_i}\}$, then we would get a game equilibrium where each point-to-point demand is optimized according to (3.6) and the optimal network routing would correspond to a mixed strategy equilibrium. Note that we would have a game because the routing decision of a point-to-point demand affects the routing of the other network demands and each point-to-point demand minimizes its own costs. There would exist a mixed strategy equilibrium for this problem [see Nash (1950)] since there is a finite number of demands and possible routings. However, the solution would not necessarily be unique and numerical techniques would be required to solve the game. For the network routing by using a game model see Lambert, Epelman, and Smith (2002).

Since we use (3.5) we do not have a game and we solve the global optimum. This is a nonconvex optimization problem and numerical global optimization techniques have to be used. We use the generalized reduced gradient optimization method developed by Lasdon and Waren (1978). Fylstra, Lasdon, Watson, and Waren (1998) have analyzed this method in more detail.

4. Examples

In this section we illustrate our optimization model with two examples, which are based on the network structures of figures 2.1 and 2.2.

4.1 Three point-to-point network

We consider first the simple three point-to-point network structure of Figure 2.1. The network flows are as follows

$$F_1(t,\beta) = \beta_1^1 D_1(t) + \beta_2^2 D_2(t) + \beta_3^2 D_3(t)$$
$$F_2(t,\beta) = \beta_1^2 D_2(t) + \beta_2^1 D_1(t) + \beta_3^2 D_3(t) \quad (4.1)$$
$$F_3(t,\beta) = \beta_3^1 D_3(t) + \beta_1^2 D_1(t) + \beta_2^2 D_2(t).$$

Using (3.4), we get that the network costs are given by

$$\sum_{k=1}^{3} S_k(t,\beta) = \sum_{k=1}^{3} F_k(\beta)\Big[c_k(\beta)N\left(l_1(T-t,\beta)\right) - \left(\tilde{K}_k(\beta) + c_k(\beta) - 1\right)N\left(l_2(T-t,\beta)\right)\Big] \quad (4.2)$$

where $\sum_{j=1}^{2} \beta_i^j = 1$ for all $i \in \{1,2,3\}$.

Let us assume that we can approximate equation (3.5) as follows

$$\inf_{\beta} E\left[\sum_{k=1}^{n} S_k\left(T,\beta\right)\right] e^{-rT} T\,. \quad (4.3)$$

That is, the cumulative discounted blocking/delay cost on $[0,T]$ is approximated by the time T discounted cost.

In order to solve this optimization problem we assume the following parameter values: time horizon $T = 1$, discount rate $r = 0$, expected demand $D_i(0,T) = 10$, and capacity $C_i = 12$ for all i

$\in \{1,2,3\}$. Note that with these values we have $S_k(0,\beta) = 0$ for all k and, therefore, equation (4.3) is a good approximation if T is small. We analyze the effect of correlations and the volatilities of D_1, D_2 and D_3 on the optimal routing probabilities from (4.1) – (4.3).

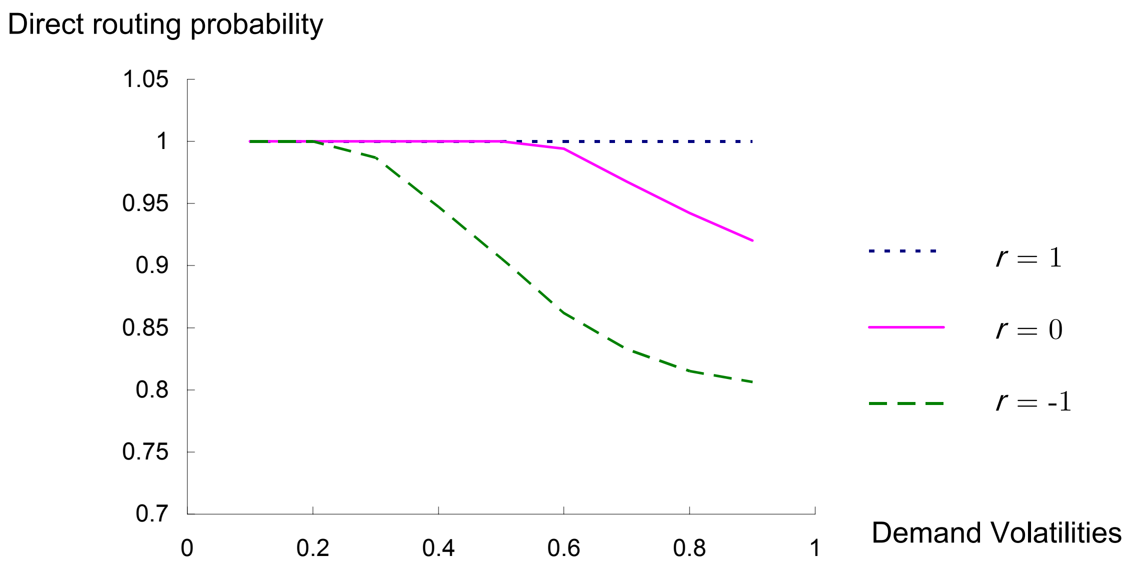

Figure 4.1 *Relationships between demand volatilities* (σ_1=σ_2=σ_3) *and the optimal direct routing probabilities of* D_1 *with different correlations* (*network structure of Figure* 2.1, $r = r_{1,2}, r_{1,3}$). *Parameter values:* $T = 1$, $r = 0$, $D_i(0,T) = 10$, *and* $C_i = 12$ *for all* $i \in \{1,2,3\}$.

Figure 4.1 shows the optimal direct routing probabilities for different demand volatilities (σ_1, σ_2, σ_3) and correlations with demand 2 and 3 ($r_{1,2}$, $r_{1,3}$). We assume that demand 2 and 3 are perfectly correlated ($r_{2,3} = 1$). Note that the lines in Figure 4.1 describe the fractions of D_1 that is routed directly in different cases. We can see that there is a small probability to use the alternative routing when the correlations are positive because in this case the alternative routing and the direct routing are full at the same time. Thus, the lower the correlations and the higher the demand uncertainties the more the alternative routing is used and, therefore, the more valuable the network routing options.

4.2 Seven point-to-point network

Next we consider the seven point-to-point network structure of Figure 2.2. From Table 2.2, the network flows are given as follows

$$F_1(t) = \beta_1^1 D_1(t) + \beta_2^2 D_2(t) + \beta_3^2 D_3(t) + (\beta_4^2 + \beta_4^4) D_4(t) + \beta_5^2 D_5(t) + \beta_6^3 D_6(t) + \beta_7^3 D_7(t)$$

$$F_2(t) = \beta_2^1 D_2(t) + \beta_1^2 D_1(t) + (\beta_3^2 + \beta_3^3 + \beta_3^4) D_3(t) + (\beta_4^3 + \beta_4^5) D_4(t) + \beta_5^4 D_5(t) + \beta_6^4 D_6(t) + \beta_7^4 D_7(t)$$

$$\vdots$$

$$F_7(t) = \beta_7^1 D_7(t) + \beta_1^4 D_1(t) + \beta_2^4 D_2(t) + \beta_3^4 D_3(t) + (\beta_4^4 + \beta_4^5) D_4(t) + \beta_5^3 D_5(t) + (\beta_6^2 + \beta_6^3 + \beta_6^4) D_6(t)$$

where $0 \le \beta_i^j \le 1$ and $\sum_{j=1}^{m_i} \beta_i^j = 1$ for all $i \in \{1, \ldots, 7\}$ and $j \in \{1, \ldots, m_i\}$.

We have 7 demands and 29 betas ($m_4 = 5$ and $m_i = 4$ for $i \in \{1, \ldots, 7\}$-$\{4\}$) in this example. We consider direct routing probabilities for D_1 on Figure 4.2 and assume that D_2 and D_3 are perfectly correlated with each other and D_4, D_6 and D_7 are perfectly correlated with each other. That is, in addition to D_1 we assume three sets of demands $\{D_2, D_3\}$, $\{D_5\}$, and $\{D_4, D_6, D_7\}$, where the demands in a same set are perfectly correlated. Then we consider the effects of the volatilities of D_1, D_2, …, D_7 and the correlations between the demand sets to the optimal direct routing probabilities of D_1.

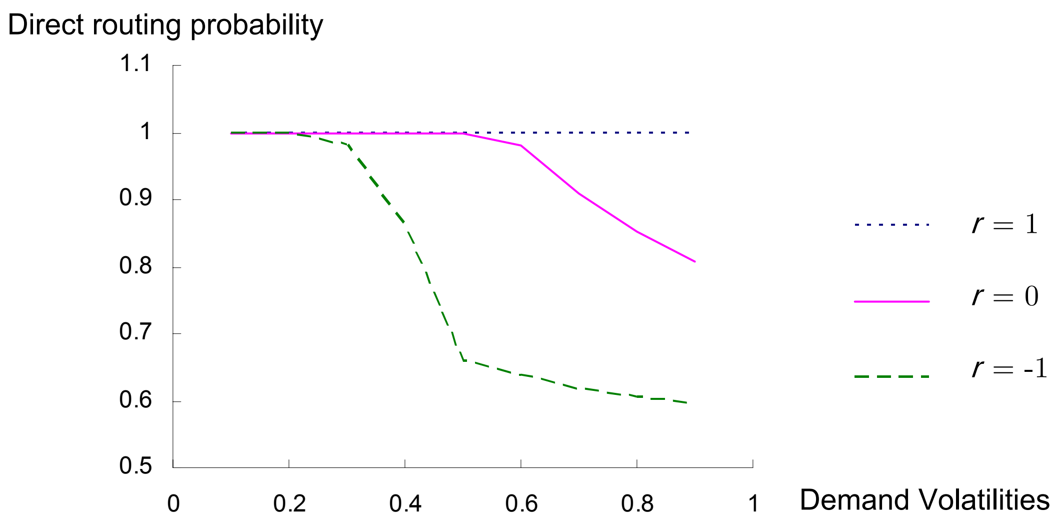

Figure 4.2 Relationship between demand volatilities ($s_1 = s_2 = \ldots = s_7$) and the optimal direct routing probabilities with different correlations (network structure of Figure 2.2, $r = r_{1,2}, r_{1,3}, \ldots, r_{1,7}$). Parameter values: $T = 1$, $r = 0$, $D_i(0,T) = 10$, and $C_i = 12$ for all $i \in \{1, \ldots, 7\}$.

As in Section 4.1, we calculate the optimal routing probabilities for D_1. We assume the same parameter values as in Section 4.1. Figure 4.2 shows the optimal direct routing probability of D_1 with different demand volatilities (s_1, s_2, …, s_7) and correlations. Similarly as with the three point-to-point network, the higher (the lower) the volatilities (the correlations) the lower is the direct routing probability. Comparing figures 4.1 and 4.2 we see that the alternative routings are used more in the complex network. This is simply because there are more routing possibilities in the complex network.

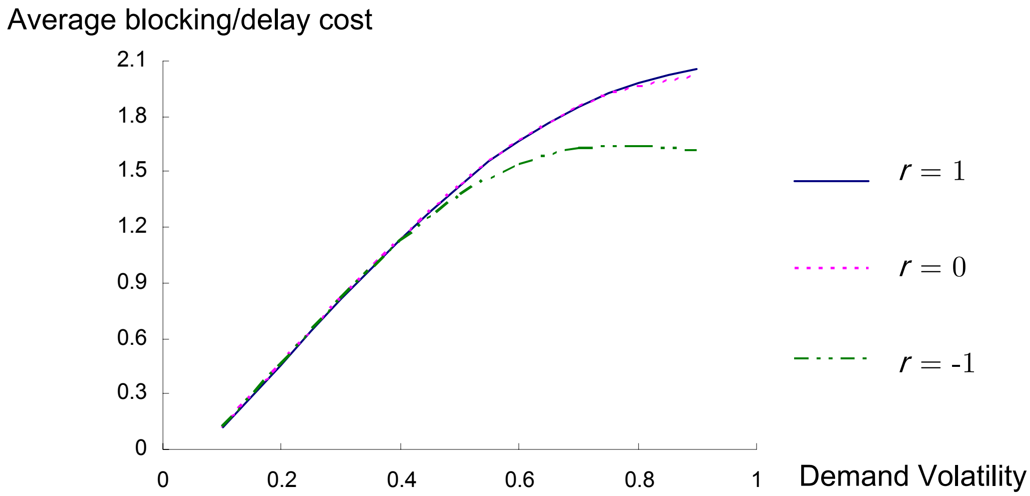

Figure 4.3 Relationship between demand volatilities ($s_1 = s_2 = s_3$) *and the average blocking/delay costs of all the links with different correlation structures* (*network structure of Figure* 2.1, $r = r_{1,2}, r_{1,3}$). *Parameter values:* $T = 1$, $r = 0$, $D_i(0,T) = 10$, *and* $C_i = 12$ *for all* $i \in \{1,2,3\}$.

Let us consider the average one link blocking/delay costs of the two network examples. Figure 4.3 shows the average blocking/delay cost in the three point-to-point network. It implies that the higher the demand uncertainties and the correlations the higher the average blocking cost. Similar results are obtained for the seven point-to-point network in Figure 4.4. However, comparing figures 4.3 and 4.4 the average blocking/delay costs in the seven point-to-point network are lower than those in the three point-to-point case. The average differences corresponding to correlations $r = 1$, $r = 0$ and $r = -1$ are 0.2%, 0.7% and 6.1%. Thus, the value of routing options increases as a function of the network complexity. This is because the network complexity increases the rerouting possibilities and reduces the blocking/delay costs.

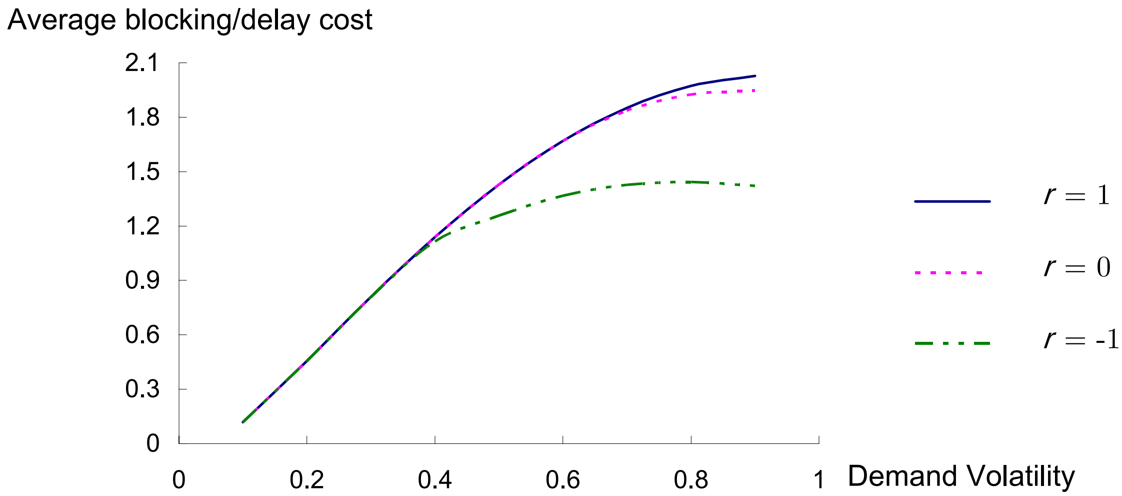

Figure 4.4 Relationship between demand volatilities (s_1 = s_2 = … = s_7) *and the average blocking/delay costs of all the links with different correlations* (*network structure of Figure* 2.2, $r = r_{1,2}, r_{1,3}, \ldots, r_{1,7}$). *Parameter values:* $T = 1$, $r = 0$, $D_i(0,T) = 10$, *and* $C_i = 12$ *for all* $i \in \{1, \ldots, 7\}$.

Note that according to figures 4.3 and 4.4, when r = -1 the cost can decrease when the demand volatility is high enough. The reason is that, in this case if a direct routing is full then the probability that there is free capacity on the alternative routings increases, and with volatilities higher than 0.9 this probability reduces the average blocking/delay costs. This cost decreasing threshold of demand volatility depends on the initial values of demand and capacity, i.e., if initial demand is higher than the capacity in our example then there is blocking/delay cost at time 0 and the threshold is observed with the demand volatility lower than 0.9.

5. Discussions

In the general routing problem, the number of parameters increases as a function of the number of nodes and paths. In Figure 5.1 we have four nodes and five links (demands). There are three betas for each demand and 15 betas total ($m_i = 3$ for $i \in \{1, \ldots, 5\}$). However, if the structure of the network is different then even with the same number of nodes the number of betas may vary a lot. In Figure 5.2 there are four nodes and six demands, all of which have five betas. This gives 30 betas total ($m_i = 5$ for $i \in \{1, \ldots, 6\}$).

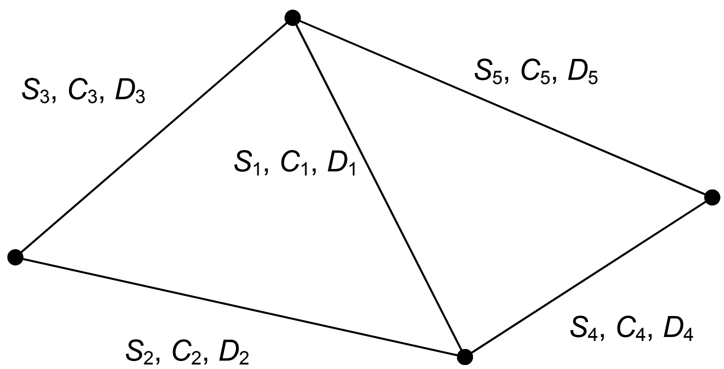

Figure 5.1 *Network blocking costs* (S), *capacities* (C), *and demands* (D).

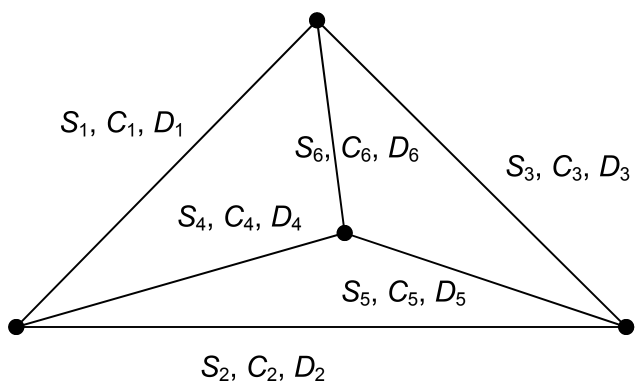

Figure 5.2 *Network blocking costs* (S), *capacities* (C), *and demands* (D).

By using the approach of Section 3 and the routing tables for the network structures, we can calculate the optimal routing probabilities. However, the required computational and data collection time for the whole optimization method increases as a function of the network complexity. Therefore, in these cases it is important to increase the efficiency of our method. This can be done by considering only the most significant routing candidates. In order to show this, let us consider a general network. If there are a direct routing and some alternative routings with one or two intermediate nodes then the probability of using routings that have more than, for example, three intermediate nodes is quite small. This is because if the cost of at least one link goes up then the whole routing cost increases. For example, the probability of using the fifth routing of D_4 in Figure 2.2 (Link 2+Link 3+Link 7+Link 6) is usually small at the network optimum (it is, in fact, zero in all the cases of Section 4.2). Therefore, in order to decrease the computational time in a general model, we can reduce the number of parameters in the network problem and simplify the calculations by reducing the maximum number of possible routings.

Table 5.1 *Loss of accuracy on delay cost by lowering the maximum number of routing possibilities* (m).

σ_1	0.1	0.2	0.3	0.4
Total cost (m=5)	0.8310	3.1945	5.6661	7.8073
Total cost (m=3)	0.8310	3.1945	5.6661	7.8073
Loss of accuracy	0.00%	0.00%	0.00%	0.00%

σ_1	0.5	0.6	0.7	0.8	0.9
Total cost (m=5)	8.8087	9.5661	9.9886	10.1027	9.9496
Total cost (m=3)	8.8317	9.6193	10.0854	10.2521	10.1545
Loss of accuracy	0.261%	0.561%	0.969%	1.479%	2.059%

Table 5.1 shows the loss of accuracy in the seven point-to-point example (Figure 2.2) by lowering $m = \max_i[m_i]$ from 5 to 3 and by assuming negative correlation between D_1 and the other demands (r = -1). All the other parameter values are the same as in Section 4.2. The optimal routing probabilities are the same when m = 4 or 5 since the probability of using the fifth routing of D_4 is zero. The blocking/delay cost in Table 5.1 is the sum of all links' costs. From Table 5.1 we can see that the loss of accuracy is negligible even though there are negative correlations between D_1 and the other demands. These results are, of course, case-specific. We do not define an

efficient value for m since the loss of accuracy depends on the network and the parameter values. However, Table 5.1 indicates that analyzing m is important since it can reduce the computational time required to calculate the optimum.

6. Conclusion

We have suggested a new method for network routing optimization by using real option modeling. Our approach can be divided into three steps. First, we represented the total flow on each link by using point-to-point demands and routing probabilities. Second, we calculated the blocking/delay costs by using the flow representation and a financial basket option model. Finally, we calculated the optimal network routing by using global optimization techniques. The fundamental idea is to use real option concepts in the modeling of network routings. Our approach considers demand correlations and the interactions between the routing decisions. Because we use routing probabilities in the network optimization, this routing strategy is similar to DAR (Dynamic Alternate Routing) and can be viewed as an extension to that.

Numerical examples in Section 4 showed that the higher the demand uncertainties and correlations the higher the network blocking/delay probability. Further, we showed how the complexity of the network lowers the network blocking/delay and, therefore, increases the value of the network.

Acknowledgement

The authors are grateful to conference participants at the INFORMS 2003 Annual Meeting. The authors also thank Mingyan Liu, Tudor Stoenescu, Xu Meng and Sophie Shive for useful discussions.

References

[1] Alleman, J. and E. Noam (Eds.), 1999. *The New Investment Theory of Real Options and its Implications for Telecommunications Economics*. Regulatory Economics Series, Kluwer Academic Publishers, New York.

[2] Ash, G.R. and E. Oberer, 1998. Dynamic routing in the AT&T network-improved service quality at lower cost. Proceedings in IEEE Global Telecommunications Conference, Dallas, TX.

[3] Ash, G.R., J.S. Chen, A.E. Frey, and B.D. Huang, 1991. Real-time network routing in a dynamic class-of-service network. Proceedings of 13th Tele-traffic Congress, Copenhagen, Denmark.

[4] Bieler, D. and I. Stevenson, 1998, Internet Market Forecasts: Global Internet Growth 1998-2005, *Ovum Report*, December.

[5] Black, F. and M. Scholes, 1973. The pricing of options and corporate liabilities. *Journal of Political Economy,* 81, 659-683.

[6] Brockmeyer, E., H.L. Halstrom and A. Jensen, 1948. The life and the works of A. K. Erlang. The Copenhagen Telephone Company, Copenhagen.

[7] Caceres, R., P.G. Danzig, S. Jamin, and D.J. Mitzel, 1991. Charateristics of wide-area TCP/IP conversations. Computer Communications Review, 21, 101 – 112.

[8] Dixit, A.K. and R.S. Pindyck, 1994. *Investment under Uncertainty*, Princeton University Press, Princeton.

[9] Feldmann, A. 1996. Modeling characteristics of TCP connections. Preprint, AT&T Laboratories.

[10] Fylstra, D., Lasdon, L., Watson, J. and Waren, A. 1998, Design and use of the Microsoft Excel Solver, *INTERFACES*, Vol. 28, No. 5, Sept-Oct, pp. 29-55.

[11] Gentle, D., 1993. Basket weaving. *Risk,* 6, 51-52.

[12] Gune, A. and J. Keppo, 2002. Call Routing with Continuous Uncertainties. Preprint, University of Michigan.

[13] Harrison, M., 1988. Brownian Models of Open Queueing Networks with Heterogeneous Customer Populations, IMA Volumes in Mathematics and its Applications, 10, 147-186.

[14] Horst, R. and P.M. Pardalos (eds.), 1995. Handbook of Global Optimization, Kluwer, Dordrecht.

[15] Johari, R. and Tsitsiklis, J., 2004 Efficiency Loss in a Network Resource Allocation Game, working paper, Massachusetts Institute of Technology.

[16] Kelly, F.P., 1997. Charging and rate control for elastic traffic, *European Transactions on Telecommunications,* Vol.8, pp. 33 – 37.

[17] Kelly, F.P., 1991. Network routing. *Philosophical Transactions of the Royal Society London A*, 337, 343 – 367.

[18] Keppo, J., 2001. Pricing of point-to-point bandwidth contracts. Preprint, University of Michigan.

[19] Keppo, J., 2003. Optimality with Telecommunications Network. *IMA Journal of Management Mathematics*, 13, 211-224.

[20] Lambert, T.J., M.A. Epelman, and R.L. Smith, 2002. A Fictitious Play Approach to Large-Scale Optimization, Technical Report 02-01, Department of Industrial and Operations Engineering, University of Michigan.

[21] Lasdon, L.S., A.D. Warren, A. Jain, and M. Ratner, 1978. Design and testing of a generalized reduced gradient code for nonlinear programming, *ACM Trans. Math. Software*, 4, 34--50.

[22] Leland, W.E., M.S. Taqqu, W. Willinger, and D.V. Wilson, 1994. One the self-similar nature of Ethernet traffic. *IEEE/ACM Transactions on Networking*, 2, 1-15.

[23] Mikosch, T., S. Resnick, H. Rootzén, and A. Stegeman, 2002. Is Network Traffic Approximated By Stable Lévy Motion Or Fractional Brownian Motion? *Annals of Applied Probability*, 12, 23–68.

[24] Mitra, D., J. A. Morrison, and K.G. Ramakrishnan, 1996. ATM Network Design and Optimization: A Multirate Loss Network Framework, *IEEE/ACM Transactions on Networking*, 4, 531-543.

[25] Mitra, D., J. A. Morrison, and K.G. Ramakrishnan, K.G., 1999. Optimization and Design of Network Routing using Refined Asymptotic Approximations, Performance Evaluation, 36, 267-288.

[26] Neumaier, A., 2004. Complete Search in Continuous Global Optimization and Constraint Satisfaction, in: Acta Numerica (A. Iserles, ed.) Cambridge University Press.

[27] Norros, I. 1994. A storage model with self-similar input. *Queuing Systems*, 16, 387-396.

[28] Rai, A., T. Ravichandran, and S. Samaddar, 1998, How to anticipate the Internet's global diffusion, *Communications of the ACM*, 41, 97-106.

[29] Ryan, S.M., 2002. Capacity Expansion for Random Exponential Demand Growth with Lead Times. Preprint, Industrial and Manufacturing Systems Engineering, Iowa State University.

[30] Zhao, Y. and K.M. Kockelman, 2002. The Propagation of Uncertainty through Travel Demand Models. *Annals of Regional Science*, 36, 145-163.