RZ 3358 (# 93404) 08/13/01
Computer Science   20 pages

Research Report

Understanding Point-to-Point Price Development and Telecom Geographics

Giorgos Cheliotis and Chris Kenyon

IBM Research
Zurich Research Laboratory
8803 Rüschlikon
Switzerland
gic@zurich.ibm.com

Research
Almaden · Austin · Beijing · Delhi · Haifa · T.J. Watson · Tokyo · Zurich

Understanding Point-to-Point Price Development and Telecom Geographics

Giorgos Cheliotis and Chris Kenyon

IBM Research, Zurich Research Laboratory, 8803 Rüschlikon, Switzerland

Abstract

Bandwidth commodity markets are an obvious opportunity to build the mechanisms and liquidity necessary to turn any amount of unused bandwidth into a cash flow today. Our approach is rather proactive, as it assumes that some of the current hurdles, such as trading infrastructure integration and contract standardization are solved. Then we ask the question: What will this market look like, how will prices behave, in a global scale? We provide an overview of several related research results. Among these, we explain the basic notion of geographical arbitrage, how QoS considerations affect the choice of provisioning paths and most importantly how spot price dynamics depend on network topology. Data on bandwidth trades is scarce today, but we use empirical observations, some trading data and global network connectivity information to support our analytical and simulation results.

1 Introduction

The dazzling rate of new telecommunications technology adoption, unprecedented capital expenditure on infrastructure and market liberalization have had their toll on the whole industry. The question today is not whether money can be made with telecom services, money *has* to be made to appease creditors and sustain the economies of scale that this business relies on. The burning issue, as we see it, is how to make the most out of existing and planned network infrastructure investments while focusing on network-attached value-add services for future revenue growth. The status quo of telecommunications, with incumbents forming a powerful oligopoly for international traffic and occasionally local price-setting monopolies, is being fundamentally shaken by the credit crisis and competition from new entrants. Even if business models may still significantly differ in the new telecom landscape, the possibility to buy bandwidth on the spot when it's needed or turn unused bandwidth into cash at any moment in time is arguably an attractive proposition to anyone in the business.

Bandwidth commodity markets are an obvious opportunity to build the mechanisms and liquidity necessary to turn any amount of unused bandwidth into a cash flow today. There are certainly some concerns regarding the future of bandwidth commodities and the impact of an efficient global market on profit margins. No matter which side the reader is on, whether the reader would rather see the market flourish or feel threatened by its existence, we suggest a more realistic approach. The market is being built this very moment, it is a reality and those who understand the market's workings better will be able to use it to their advantage. It is our wish that this chapter helps in acquiring a better understanding of these developments. Our approach is being rather proactive in assuming that some of the current hurdles, such as trading infrastructure integration and contract standardization are solved. How will this market then behave, in a global scale? Data on bandwidth trades is scarce today, but we use empirical observations, some trading data and global network connectivity information to support our analytical and simulation results.

1.1 Bandwidth flavors

When talking about bandwidth as a commodity, it is first necessary to define the underlying asset, since bandwidth comes in several flavors.

1.1.1 Connectivity type

It is fundamentally a data transport service, so it is often defined as a point-to-point offering, not unlike cargo transport. Data is allowed to travel at the defined rate (speed) from point A to point B, perhaps with some quality parameters guaranteed between the two points (delay, availability, etc). We can call this the point-to-point paradigm because, providing QoS and privacy are guaranteed to the user, it is equivalent to leasing a separate point-to-point line for the data (even if in reality the connection may be built on top of a shared network). There is also, however, access bandwidth, which is defined as spare capacity at an ingress/egress point of a network and peering bandwidth which is free capacity at an interconnection point between two providers' networks. Both of these are generally not defined in a point-to-point fashion, in fact they refer to a point and a network of accessible destinations. Data is allowed to travel at the defined rate through that point, and into the network, perhaps with some statistical quality guarantees for all destinations directly attached to the network. Some ISP service level agreements define connectivity in this way. Conceivably commodity markets could be built for all these connectivity types.

1.1.2 Technology

Ideally a commodity trader does not care much which technology is used to provision a certain good or service, even when interested in actual delivery. Providing that traded contracts are well-defined, what matters is price, transfer rate, quality guarantees and compensation for outage incidents (e.g. liquidated damages). However, in modeling price dynamics and gaining insight into the internals of the market, understanding the underlying technology is important, as shown later and throughout the chapter. Most of the currently posted offerings are based on SDH/SONET and WDM technology. Since optical fiber has become the standard physical medium for backbone connectivity and all-optical networking is promising substantial performance gains, we focus on including the economics of the fiber market in our modeling. A popular staged investment approach in fiber has been to start early with dark fiber and light it when demand catches up. Dark fiber is cable without the equipment needed to transfer data and deserves special attention, so we explain how to employ a real-options approach to determine dark fiber value in another chapter. Lighting comprises buying and installing networking equipment (e.g. add-drop multiplexers) to operate the fiber connection. Only then is fiber turned into usable bandwidth. With the addition of network management software and intelligent switches/routers managed bandwidth services can be offered, that is finely grained bandwidth ready to be used by any application supporting standard networking protocols, like TCP/IP. Backup services often complement such offerings to create a complete reliable data transfer solution.

1.1.3 Market segmentation

A distinction is usually made between long-distance (backbone) and metropolitan (local-loop) bandwidth. Even though there is a convergence in technology, with broadband metro fiber being deployed in several cities, the economics of metro networks are not necessarily the same as in long-distance. Furthermore they may vary from city to city depending on the amount of local competition and the regulatory framework in place. The focus of this chapter is on inter-city bandwidth, i.e. the backbone market. Traffic termination in the local loop is often not included in long-distance offerings, but some intermediaries are starting to offer bundles, or at least assist in finding a matching local loop contract to complement backbone deals.

This chapter is devoted to the trading of point-to-point long-distance managed bandwidth services over fiber connections, although some concepts introduced later, such as that of network effects on bandwidth prices could be applied to any point-to-point offering.

1.2 Commoditization

There are arguably several more or less valid views of the term "commodity" and varied opinions as to whether it is applicable to bandwidth. Some telecom incumbents dislike the term because they view the network as their core business, their very raison d'être. Even if they would be willing to conduct a few trades, understandably they are not fond of the idea that bandwidth may be regarded as just another commodity, of which they are but one of several indistinguishable providers. Frequently heard arguments against commoditization are that managed bandwidth is too complex and needs are too diverse to introduce a generally accepted standard contract. On the other side there is a handful of incumbents, but mostly new entrants who do accept or actively promote bandwidth commoditization. Our view is that the debate on commoditization has had its rightful share of attention but there is more benefit in a discussion that is not inhibited by today's conflicting views and interests. We rather aim for a forward-looking analysis, which starts with the realization that there *is* today a global market for trading managed bandwidth services and attempts to answer some fundamental questions: How shall we structure commodity contracts to satisfy end-to-end demand and QoS? Moreover, how much will these contracts cost?

The rest of the chapter is organized as follows: we first discuss those features which are important for modeling a bandwidth commodity market in Section 2 and explain how point-to-point contracts are constructed in Section 3. In Section 4 we present a stochastic model for a spot market and an application with simulation results for network NPV in Section 5. Section 6 presents global telecom geographics from the point-of-view of commodity trading and in Section 7 we show some issues of concern to network providers.

2 Bandwidth commodity features

Here we list and discuss factors that may affect bandwidth commodity pricing making contrasts with other commodities, mostly electricity. For a more extensive list see [8].

geographical arbitrage This means that, given the same QoS, the cheapest of all available routes will set the end to end price in a competitive (liquid) market. This is due to the fact that the actual route is irrelevant with respect to data transport, as long as certain QoS requirements are met. The set of routes connecting two given geographical locations at the same QoS level are perfect substitutes.

non-storability Inventories act to smooth variations in supply and demand. When no inventories exist prices can jump if supply or demand change suddenly. Prices can also change suddenly when the perception or expectation of supply or demand status suddenly changes. Bandwidth is non-storable so price jumps and spikes are to be expected. This is a determining factor in electricity price modeling. Jumps and especially spikes are observed often deriving from weather events (e.g. summer 1998, 1999 in Texas, Australia, California) sometimes in combination with equipment failures. In fact even in commodities where storage is possible, like oil, large-scale political events can still cause jumps and spikes in prices.

trading and settlement timescale In some electricity markets 30 minute and even 10 minute blocks are priced, traded and settled. To date bandwidth contracts have had the character of the regulated electricity industry in that month or year contracts are the norm. With universal trading contracts (i.e. bandwidth deregulation and contact feature standardization) and pooling points we expect this to change and approximate the electricity market much more closely.

liquidity Currently the bandwidth market is less liquid than the electricity or indeed any other commodity market. Trading volumes are picking up and ongoing deregulation of the industry along with universal trading contracts will assist in reaching higher levels of liquidity. In any case, we do not expect all traded locations to be equally liquid in the future.

demand in-elasticity Provided that most consumers do not react on the timescale of market trading and settlement this will be a feature of total point-to-point demand on that timescale. However, on an individual link basis demand will be elastic thanks to automation technologies, such as software agents, electronic auctions (e.g. [10]) and least-cost routing which allow for fast switching between substitutes (alternative paths). This elasticity will exploit market liquidity and contribute to it. Inelastic demand is a current feature of short-term (days to weeks) electricity markets.

growth The internet, and network bandwidth available, has had periods where growth was 100% every 3 to 4 months. In the last few years this has slowed to only 100% per year. In addition enormous amounts of dark fiber are being laid to take advantage of and push further growth. The energy industry is growing at a much slower rate. In fact energy growth in the First World is barely 5% per year, if that. This is partly because of increasing efficiency.

technological development A dominant factor in bandwidth development. There are actually two related areas: bandwidth and switching/routers. Bandwidth increases with both the increase in packing down a single wavelength and also with the increase in the number of wavelengths that can go down a single fiber (Dense Wavelength Division Multiplexing, DWDM). Switches are inherently parallel devices so they also exhibit this combination improvement: improve as chips improve; and improve with packing multiple switching units. There is also an expected transition to all optical switching within the next 5 to 10 years. The combined effect of technological development and competition is a continuous drop in the cost of transporting a Megabit per second per mile. Technological development is today almost irrelevant in electricity markets. The only significant new factor has been the development of small gas turbines which have short response times (seconds to minutes). However these are expensive and have mostly been built as peaking units.

supply elasticity Suppliers of bandwidth may have significant flexibility in assigning network resources towards the fulfillment of different end-to-end contracts. This is due to routing and bandwidth management tools which allow for a number of different allocations, depending on which contracts the supplier wishes to offer. The amount of flexibility depends on the design of the underlying network. Generally, a network with more switching points will provide more flexibility (at the expense of QoS, because more contention/failure points are added). In a similar manner, a power plant's resources can be assigned to different markets, but the allocation problem is different since it occurs only at one place, the plant, not on the distribution network.

term structure of volatility In commodity markets in general one of the stylized facts is that there is more volatility in the near-term forward market than in the long-term forward market. It is also usual that for non-investment commodities this volatility does not asymptote to zero. In the bandwidth market whilst in some very long term there may be stability we consider it more likely that in any reasonable planning horizon, say up to 5 years, there are substantial levels of uncertainty. So far prices are continuously falling (with rare exceptions), but even if we were to assume that this trend will continue, there is significant uncertainty concerning the rate of decline (both on a global and local level).

positive prices We assume that prices for bandwidth are always positive. In energy, zero real prices are observed. This is caused by large thermal or nuclear plants that cannot ramp down as quickly as demand can drop. Thus when demand is very low and drops suddenly these plants may give away power for free as this is their only means of disposal. That is, in electricity, there is no free disposal.

3 Satisfying point-to-point demand

Demand in the commodity market is for point-to-point bandwidth. Supply is limited by the location of a seller's resources and a single network does not reach all destinations. In the following paragraphs we will explain how the bandwidth market can be represented with a graph of traded contracts and how demand is in general satisfied with a combination of multiple contracts, so as to connect two points at the best market price. The fact that there may be several alternative routes of equivalent QoS between two points but with different prices is often described by the term "geographical arbitrage", which is used herein as well.

3.1 Contract definitions

The set of all traded point-to-point bandwidth contracts can be represented as a *contract graph* $\mathcal{G}(V, E)$, where vertices represent traffic exchange locations and edges represent traded bandwidth. A bandwidth segment s is a pair of exchange locations (a, b) corresponding to neighboring vertices of the contract graph (i.e. a link). We define a contract $\mathcal{C}$ for segment s as follows: $\mathcal{C}_s = (\beta, G(\mathbf{q}), p)$, where β is the bandwidth offered, $G(\mathbf{q})$ is a set of guarantees (i.e. constraints) on a QoS vector $\mathbf{q}$ and p is the price of the contract. A path π in the contract graph is a sequence of exchange locations $\pi =< a_1, a_2, ..., a_n >, \quad a_i \in \mathcal{G}$.[1]

3.2 Segment and path properties

A segment contract is defined as the triad $(\beta, G(\mathbf{q}), p)$. Let us take a set of commonly used parameters, often referred to in service-level agreements, as our QoS vector $\mathbf{q} =< d, \Delta d, L, R >^T$, where d is delay, Δd is delay variation (jitter), L is the loss ratio (traffic lost over traffic sent) and R stands for reliability (1 minus probability of failure). A QoS vector can characterize a segment or a whole path. When computing the elements of $\mathbf{q}$ for a path we should bear in mind that delay and jitter are additive metrics, reliability is multiplicative [12, 6] and loss ratio is quasi-multiplicative, as shown below. Bandwidth β is a concave metric [12] and price p is additive.

The guarantees of a segment contract $G(\mathbf{q})$ can be expressed in terms of the following relations: $d \leq M_d$, $\Delta d \leq M_{\Delta d}$, $L \leq M_L$ and $R = F_R$, where M_x denotes an upper bound

---

[1] What does the contract graph look like? In early stages of bandwidth trading it is bound to be limited in size, but when the market matures and expands, the graph will grow and the connectivity degree will increase. For more information see the section on telecom geographics.

of property x and F_x is the value of an exact guarantee. For simplicity we assume that the bandwidth property has an exact value guaranteed (i.e. in our notation $\beta = F_\beta$) which reads as: maximum guaranteed bandwidth is F_β. Note that such a bandwidth guarantee can be provided on any network supporting properly engineered resource reservations. It is the responsibility of the connectivity provider to employ a reservation scheme that is suitable for his network (per-flow reservations, relative differentiation of flow aggregates, etc).

We can represent a customer's end-to-end quality requirements $R(\mathbf{q}(\pi))$ with the following bounds on values of the path's QoS vector : $M_{d(\pi)}$, $M_{\Delta d(\pi)}$, $M_{L(\pi)}$, $m_{R(\pi)}$, where m_x is a lower bound on property x. Now, as shown in more detail in [3], in the case of standard commodity contracts being traded for multiple links, a path satisfying the quality requirements and consisting of multiple contracts should satisfy the following constraints:

$$\begin{aligned} nM_d &\leq M_{d(\pi)} && (1) \\ nM_{\Delta d} &\leq M_{\Delta d(\pi)} && (2) \\ {M_S}^n &\geq 1 - M_{L(\pi)} && (3) \\ {F_R}^n &\geq m_{R(\pi)} && (4) \end{aligned}$$

Let $\mathcal{S}$ be the solution space of the system of inequalities (1) to (4). By choosing $\hat{n} = \max(\mathcal{S})$, customer quality requirements can be met by a path consisting of $\hat{n}$ segment contracts (or less). Thus we have shown that the entire quality vector can be reduced to an upper bound for the number of contracts in a path, assuming identical contracts. The number of contracts n can be thought of as the "hop-count" in the contract graph.

The above discussion of path properties shows that under the assumption of identical segment contracts being traded on all segments (or on selected segments indicative of the pricing in an area) geographical arbitrage can be detected by solving the following least-cost hop-constrained path optimization problem (LCHC): *find a path in the contract graph that minimizes cost subject to a hop-count constraint.* If the resulting path's cost is lower than the customer's price limit, then there is a solution. Alternatively, we could try to minimize hop-count under a cost constraint. In part it depends on which constraints a customer has given.

3.3 Computing the optimal path

The LCHC problem is a path-constrained path optimization problem with two additive metrics, which is NP-complete in general [12, 7]. It can be solved using the approach in [5], where one of the additive metrics is discretized and an $O(x^2V^2)$ algorithm is given for minimizing the other metric subject to the constraint on the discrete one. In that algorithm, x is an approximation parameter which at higher values increases the success of the algorithm at an increased complexity cost. Space complexity is $O(xV)$.

Chen and Nahrstedt suggest in [5] the use of an approximation technique, which, in fact, is not necessary for geographical arbitrage detection because the hop-count is constant at every edge of the graph and therefore does not need to be discretized. Without the approximation step, the problem can be solved in $O(\hat{n}^2V^2)$ using the extended Dijkstra shortest path algorithm[2] in [5], which is guaranteed to find a solution, if there is one. Space complexity in this

---

[2] The extended Bellman-Ford algorithm can also be used.

case is $O(\hat{n}V)$.

Having shown how a composite path contract can be built out of several segment contracts and how to efficiently detect geographical arbitrage, we will now move on to a modeling approach for a bandwidth spot market whereby detection and exploitation of geographical arbitrage opportunities are modeled explicitly and influence the stochastic behaviour of prices.

4 A spot market model

In the previous section we described factors that are important for the development of bandwidth prices in general. In this section we will consider how these factors combine to give specific models for traded commodities. The basic method is shown in Figure 1. The inputs to price development are: the graph of traded contracts G; the initial link (contract) prices G_{init}; the network function f which expresses how arbitrage opportunities are removed by market forces and the time constant τ for their removal, i.e. the market liquidity (see Section 4.1.4); and the stochastic process models for each link (see Section 4.1.1).

The price development that Figure 1 describes is that observed prices (the Market Observed Outcome, G_{arb}) may contain arbitrage opportunities. Market forces act both to remove the existence of arbitrage and to disturb the prices via normal and unusual information and changes in supply and demand. These two processes in general combine to produce the next set of observed prices. The arbitrage removal occurs relative to the observed state of contract prices. However, as we just said, other market forces are still acting and this occurs at the same time. Thus the new state may not be arbitrage free, even with high liquidity, because there is always the potential for new arbitrage opportunities to be created. In practice we may model the two forces (geographical arbitrage removal and "the rest") separately or in combination. For simplicity we will here describe them together and then separate the effects.

When we consider bandwidth prices it is vital to be precise on what exactly are the underlying traded commodities. We model price development at the link level. We consider links to be indivisible contracts offered between pooling points (referred to as segment contracts earlier). A pooling point is a facility for the exchange of traffic at a particular geographical point among trading partners in the commodity market. Any party may combine link contracts to form end to end contracts, in the manner that we have already described. Equally any party buying such an end to end contract may be able to split it according to the pooling points along the route to create new (link or multi-link) contracts. Thus prices between any pair of pooling points may be observed on the market but these are formed from link prices. We do not model multi-link price processes directly.

We model link price development as a combination of three factors: link price changes; geographical arbitrage; and liquidity. If the market is completely illiquid then if geographical arbitrage appears there will be no action by market participants affecting prices to move the market in a direction to remove these opportunities. That is, traffic will not take the cheapest available route if the market is illiquid. On the other hand in a completely competitive, i.e. liquid, market arbitrage opportunities will only last until the next trade (at most).

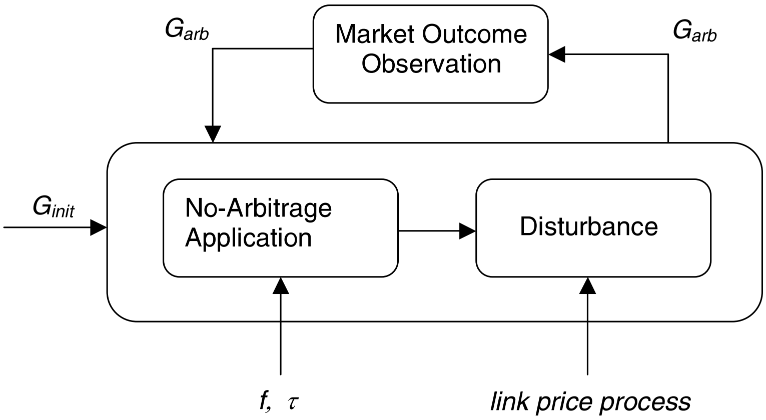

Figure 1: Telecom commodity price development. Inputs are: an initial network graph G_{init} of traded contracts and their prices; the network function f describing geographical arbitrage removal by load balancing; the speed of geographical arbitrage removal (market liquidity) τ; and the stochastic processes for individual links (contracts). Arbitrage removal may be combined with individual link price process. The prices observed in the market G_{arb} may contain geographical arbitrage opportunities.

4.1 Combined price process

We use the following stochastic process (stochastic differential equation) which combines the effects seen on individual links using geographical arbitrage, dA, according to a liquidity factor α. The geographical arbitrage term includes load balancing across different QoS-equivalent routes. The term αdA embodies the network function f described in section 4.1.4 and α plays the role of the previous liquidity term τ in the same section. dA describes a stochastic process that is zero whenever no geographical arbitrage effects are present. When geographical arbitrage effects are present the dA gives the appropriate correction to remove arbitrage from the previously observed prices in the contract network.

$$\begin{aligned} X &= \log(S) \\ dX &= \eta(\overline{X} + GU - X)dt + \sigma dW + GdU + HdV \\ &\quad +\alpha dA \\ d\overline{X} &= -\nu dt + \rho dZ \end{aligned}$$

The above equations express the fact that market forces are acting at the same time as arbitrageurs acting to remove arbitrage opportunities (by profiting from them). Indeed they are part of the market, just modeled explicitly in this case.

4.1.1 Independent link prices: SRR model

The individual link prices (the equations without the αdA term) are generated for each link based on an Orstein-Uhlenbeck process with the addition of a process for the long term mean, spike and jump terms. We also incorporate limited regime switching induced by the spike terms on the effective mean. We term this process a Shock-Regime-Reverting or SRR process.

We make the hypothesis that if there were no network then each link price S would develop as follows

$$\begin{aligned} X &= \log(S) \\ dX &= \eta(\overline{X} + GU - X)dt + \sigma dW + GdU + HdV \\ d\overline{X} &= -\nu dt + \rho dZ \end{aligned}$$

U is a two state $\{0, 1\}$ semi-Markov process, where we make an identification between the states and the numbers zero and unity, with rate parameters λ_U, μ_U and

$$G = \begin{cases} \text{Gamma}(g_U, \alpha_U) & \text{if } dU = +1 \\ G & \text{otherwise} \end{cases}$$

Clearly Gamma stands for a Gamma distribution with scale parameter g and shape parameter α thus Gamma(g,α) has mean $g\alpha$ and variance $g^2\alpha$. When U jumps from state $\{0\}$ to state $\{1\}$ then X increases by G and the mean to which the process is reverting also increases by G. When U jumps from state $\{1\}$ to state $\{0\}$ then X decreases by the same amount that it previously increased by and the extra term in the mean is dropped. Thus a price spike is created. The price will stay in its current state for an exponentially distributed amount of time as given by the rate parameters.

V is a Poisson process with rate parameter λ_V and

$$H = \begin{cases} \text{Gamma}(g_{\text{up}}, \alpha_{\text{up}}) & \text{if jump up} \\ -\text{Gamma}(g_{\text{down}}, \alpha_{\text{down}}) & \text{if jump down} \end{cases}$$

Jumps may be equally probable in both directions or not as determined by the probability of an up jump p_{up}. These jumps act additively on the logarithm of the spot price so represent a multiplication of the current spot price.

η is the speed of price reversion to the average price $\overline{X}$; σ is the scale of the driving Brownian motion of price change increments dW; ν is the (positive) instantaneous rate of average price decrease and there is an uncertainty about this rate of size ρ (recall that $X = \log S$ so we are hypothesising log-normal changes in S); dZ is a Brownian motion uncorrelated with dW. Note that $\overline{X}$ is just called an average price in that it is the price towards which S reverts. It is not actually an arithmetic average in that sense.

Note that since there is no storage there is no requirement that the process be a Martingale under any particular risk-neutral measure, although this could be arranged.

4.1.2 Link prices: Rationale

We will now describe the price process in more detail. First note that the process is a semi-Markov jump diffusion with regime switching, thus the present determines the future. We assume, as in the original paper by Merton [9], that ordinary market news moves the price continuously and is responsible for the basic driving Brownian motion dW. Note that non-Markov descriptions of spikes have also been used when supported by empirical evidence [4].

Commodity markets often show reversion to some long run mean and we expect telecom to be no exception. Hence the use of an Orstein-Uhlenbeck process. However, in a significant departure from non-manufactured commodities, e.g. oil, wheat, etc, there are clear — non-zero — expectations on the speed of telecom capacity development. We expect the long run mean, $\overline{X}$, to mimic the technological development of communications capacity with its exponential improvement (ν). The degree of the exponential improvement is not known but we may use an estimate, for example from the IBM General Technology Outlook. Whilst single-mode fiber capacity has shown exponential growth the development of multi-mode (DWDM) transmission has radically improved capacity. Other such disruptive improvements are possible, e.g. long distance transmission with no repeaters, all optical switching, etc. We model this uncertainty with the scale, ρ, of the driving Brownian motion dZ.

Spikes in prices have been observed in electricity prices and are the result of demand being very close to available supply followed by, say, some equipment failure. Congestion is observed on telecom networks and equipment failures leading to outages have been observed. Thus we include spikes as a feature of the link price proposal processes. We define a spike to be a sudden increase in price followed quickly by a similar decrease in price. During the spike the mean for reversion is altered to include the magnitude of the spike. This change in regime is reversed when the spike ends. Spike sizes are modeled with a Gamma distribution. This particular form is not important — what is important is that spikes generate reversible step changes in price.

Price jumps are observed in oil prices, largely as a result of the perceived status of OPEC. Given that the owners of long distance networks also form an oiligopoly there is the potential for price jumps. These may be local to a single link or more general. We only deal with the simplest case as independent effects on different links. Jump occurrences are modeled with a Poisson process with a given rate that describes how many jumps are expected per unit time. These jumps may be positive or negative and again are modeled with Gamma distributions.

4.1.3 Dependent link prices

The simplest form of dependence between different link prices is to introduce correlations in the driving Brownian motions for the long and short-term price variations, i.e. dZ and dW, between different links. We do not go into this in depth here because we want to highlight a much more network-specific form of dependence: geographical arbitrage. Note however that the introduction of a correlation structure across a node allows us to trivially model the introduction of a pooling point on a previously undivided arc. If the new pooling point is actually redundant and all demand and supply actually crosses it we can use perfect correlations in the driving processes on both sides of it. This does still permit separation of rare events on either side of the pooling point.

4.1.4 Geographical arbitrage and liquidity

Geographical arbitrage is the term that has been used to describe the existence of at least two different end to end prices between two pooling points that are joined by a single link at a given end to end QoS. These two end to end prices may each be formed from one or more links but will both provide a required QoS level. Clearly in a liquid market this situation will not persist, all other factors being equal. More precisely we offer the following definition.

Definition 1 *A* **simple geographical arbitrage** *opportunity exists when multiple links can be substituted for a single link and when the total price of the substituted links is less than that* *of the single link. We assume of course that the QoS is equivalent between the single link and the end to end QoS of the substituting links.*

Recall that a link represents an indivisible contract. Now not all contracts offered on the market may be indivisible in general, thus we have the following specification of geographical arbitrage.

Definition 2 *A* **geographical arbitrage** *opportunity exists when multiple contracts can be substituted for a single contract and when the total price of the substituted contracts is less than that of the single contract. We assume of course that the QoS is equivalent between the single contract and the end to end QoS of the substituting contracts.*

Simple geographical arbitrage is important because it provides an immediate downward pressure on the price of the single link. How fast this pressure acts depends on two factors: how easy the substitute is to identify; and how liquid the market is. A decrease in the price of a single link implies that some part of total end-to-end demand has shifted from that link to the cheaper alternative path. The increase in demand in the alternative path should result in a price increase on all links of that path. The question arising here is how to quantify the effect of movement of demand from a link to an alternative path at the price level.

Now we will put all these ideas together. Let (a, b) be a given link, i.e. an arc on which indivisible contracts are available in the market. Let Λ_{abq} be the set of all paths between a and b which provide at least a QoS q. Now we set

$$p_\Lambda = \{p'_k | k \in \Lambda_{abq}\}$$

Where p' are formed from the observed prices for links. With respect to geographical arbitrage the next price observed in the market for the link (a, b), at a QoS q is p_{abq} and is thus given by (we will combine this with the stochastic process for the link in Section 4.1):

$$p_{abq} = (1 + e^{-\tau\Delta t} f(p_\Lambda, p'_{abq}))p'_{abq}$$

Where p'_{abq} was the previous observed price on the link (a, b) at QoS q. Here $e^{-\tau}$ describes how fast the no-arbitrage correction $f(p_\Lambda, p'_{abq})$ takes effect. The relaxation constant τ is the quantification of the system liquidity. The function $f(p_\Lambda, p'_{abq})$ encapsulates the degree of the arbitrage opportunities available relative to the observed price for the link (a, b) and the appropriate correction to the link price. The no-arbitrage correction function f also embodies the speed and extent to which applications and electronic agents can re-balance flow in the network on the timescale of network price development. Note that the prices on the alternative paths will also be affected.

If the observed link price p'_{abq} is the cheapest alternative out of p_Λ then there is no arbitrage opportunity and no correction takes place. We call f the *network function.*

Let's examine now the case where there is just one cheaper alternative path (i.e. one simple geographical arbitrage opportunity). The correction on the direct link (labeled d here) from p_d to p'_d is the effect of a left shift of the demand curve by x for that link. Remember that total end-to-end demand is inelastic in the timescale we examine (short i.e. here, one day), so this same demand is directed to the alternative path and added to the demand of each link in that path, resulting in shifts to the right by the same amount x, as shown in figure 2.

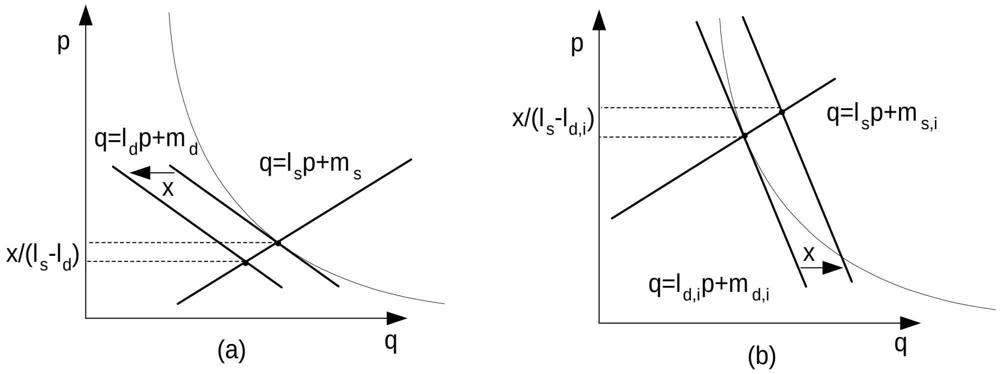

Figure 2: Load Balancing

By extending this to multiple alternative routes, we can resolve a simple arbitrage case by treating demand as a network flow which is conserved point-to-point, while allowed to shift from the direct link to alternative paths so as to achieve a load-balancing effect. Demand shifts on a single link i translate to changes on the price p_i of that link by multiplication with a term a_i which depends on price elasticity of supply and demand. It was shown in [8] that:

$$a_i = (E_s + E_d p_i^{-(E_d+1)})^{-1}$$

where E_s is the constant elasticity (slope) of a linear supply curve and E_d is the (also constant) price elasticity of demand, both assumed uniform across all links in the market (for justification and more details see again [8]). The general problem of load-balancing between several routes can be formulated with a linear system of equations and an additional constraint on demand flows. This is a linearly constrained linear optimization problem, in the form $Ax = b, \quad x \geq 0$, which is the form of a standard Linear Programming (optimization) problem. Linear Programming problems are typically solved by either interior point (polynomial worst case) or simplex (exponential worst case but very good in practice) methods. However, instead of the Linear Programming approach, a simpler and efficient iterative load-balancing algorithm was developed and is being used in our bandwidth market simulation framework [4].

5 Investment: Network NPV

We will now present an example of an application in network investment using a simulation framework that we developed based on the model we described. We consider an undirected triangular network where two of the sides start at the same price and the starting price of the third side is varied over different simulations. This is the most basic market setup where geographical arbitrage can occur. This gives an isosceles triangle with respect to prices. We vary the ratio of one side to the other two between 10% and 190% so the total price to go around the triangle (at the start) runs from 2.1 to 3.9 in some arbitrary price units. We also set the QoS offered on each side to unity and set the allowed QoS to two. Thus there are always two alternative paths between any pair of distinct nodes. The parameters used for these price simulations are given in Table 1. In this simulation we consider a highly liquid market so arbitrage opportunities only last for the time-step on which they are observed. New

parameter	symbol	value
number of simulations		1024
simulation length	t_{max}	1 year (=252 trading days)
simulation granularity	Δt	1 day
price trend (time to halve)	$-\nu$	1.25 years
trend uncertainty	ρ	0% to 40% per year
short-term price volatility	σ	0% to 200% per year
price reversion to trend (time to halve)	η	3 months
price jumps	G	none
price spikes	H	none
liquidity	α	1

Table 1: Parameters for Simple Network Experiments. Liquidity refers to how long arbitrage opportunities last before market forces (arbitrageurs, etc) remove them. In this simulation we consider a highly liquid market. Note that common random numbers were used across different parameter value combinations for variance reduction.

geographical arbitrage opportunities may arise each time-step but load balancing (demand shifts) act to fully remove them on each subsequent time-step together with the usual price drivers embodied in the stochastic processes for the links. A triangular network may appear too simple to observe anything of interest but we show here that the network demonstrates a rich set of features.

Perhaps the most important parameter is the short-term volatility which we examine over a range 10% to 200%. Spot price volatility (day-ahead volatility) in at least seven current markets for other non-storable commodities (i.e. electricity) has been observed at more than 150% and over 200% for five of them (CINERGY INTO, ENTERGY, MAIN, PALO VERDE, PJM, data source *Power Markets Week, 1999*). One month forwards were also observed at 45% to 95%. Thus, for a non-storable commodity, the high side is perhaps more realistic than the low side of the range we are considering.

Figure 3 shows the effect of geographical arbitrage liquidity on the NPV of a triangular network. This NPV is the sum of the NPV for each edge of the triangle. Triangle NPV changes (increases) with increasing short-term volatility (right top panel). The difference between different networks with different triangle ratios is less significant that the change in NPV relative to zero volatility which is nearly 140% of zero volatility NPV. The spread on triangle shape is less than 10%. There is even less difference for triangle shape for the SD of the network NPV in terms of the mean NPV for the same short-term volatility.

The lower panels of figure 3 show the additional effect of geographical arbitrage on network NPV. For acute triangle networks there is a decrease in mean network NPV of up to 10% (left lower panel). This gradually changes with increasing obtuseness to an increase of more than 2% of mean network NPV maximising at around equalateral configurations then gradually decreasing. Changes only become significant for short-term volatilities of 50% and higher. There is always a decrease in the SD of the network NPV (right lower panel). This decrease starts at around a 20% decrease for acute trangles at 200% short-term volatility then becomes less pronounced at a side ratio of 10:10:10 before becoming more pronounced again to a decrease of 12% for very obtuse triangles.

Experiments on price development on individual links showed that topology greatly influ-

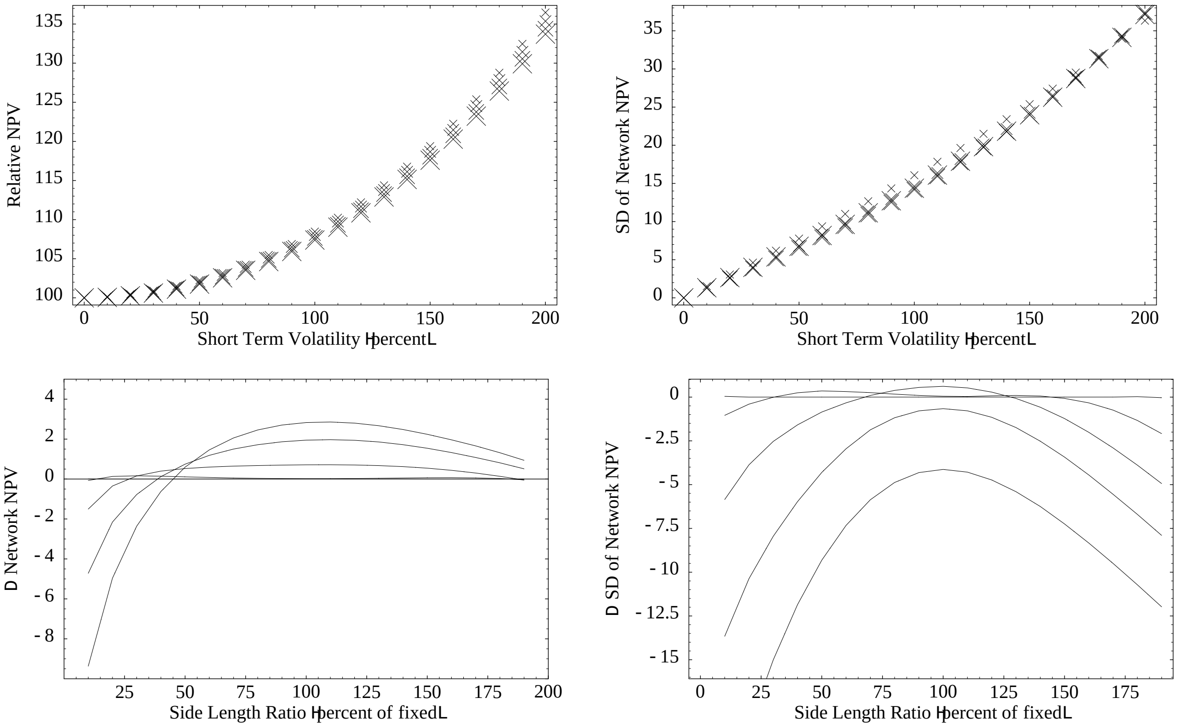

Figure 3: Network effects on network NPV. *top panels:* show the mean NPV and the standard deviation of NPV (SD) for baseline, isolated link triangle. Mean NPV is shown as a percentage of mean NPV with zero volatility. SD of NPV is shown as a percentage of the mean NPV that corresponds to the same short-term volatility level. Increasing sized diagonal crosses show increasingly obtuse triangle networks from {1 *smallest* — 5 — 10 —15 —19 *biggest* }:10:10. *lower panels:* show the additional effect of network effects (geographical arbitrage and load balancing) on network mean NPV and SD of network NPV, i.e. the changes ("Δ"'s). The most acute angled triangles show the most extreme effects (different from zero).

ences price behaviour [8]. Network effects affect the total NPV around the triangle much less than the prices of the individual links. The difference in the percent mean price change for an individual link is up to 250% both up and down, depending on the side you consider, whilst that effect for the total is always less than 10%. In some sense the triangular network is, at least in the mean, a self-hedging instrument whereas individual links are certainly not. However, the result shown here clearly suggests that the rate of decrease in network NPV does depend on topology, an effect which our model is able to capture and which is of importance to anyone owning networking infrastructure. The decrease in the SD of the network NPV is easy to understand as whenever arbitrage is observed prices are moved closer together to some extent depending on liquidity. In these experiments we used a sufficiently high liquidity to eliminate arbitrage on the next time-step after it was observed. There is clearly a trade-off between the value of the arbitrage opportunities and the extend to which no-arbitrage affect network price development. A lower liquidity would have increased the value of the arbitrage opportunities by having them last longer but would have reduced the other effects.

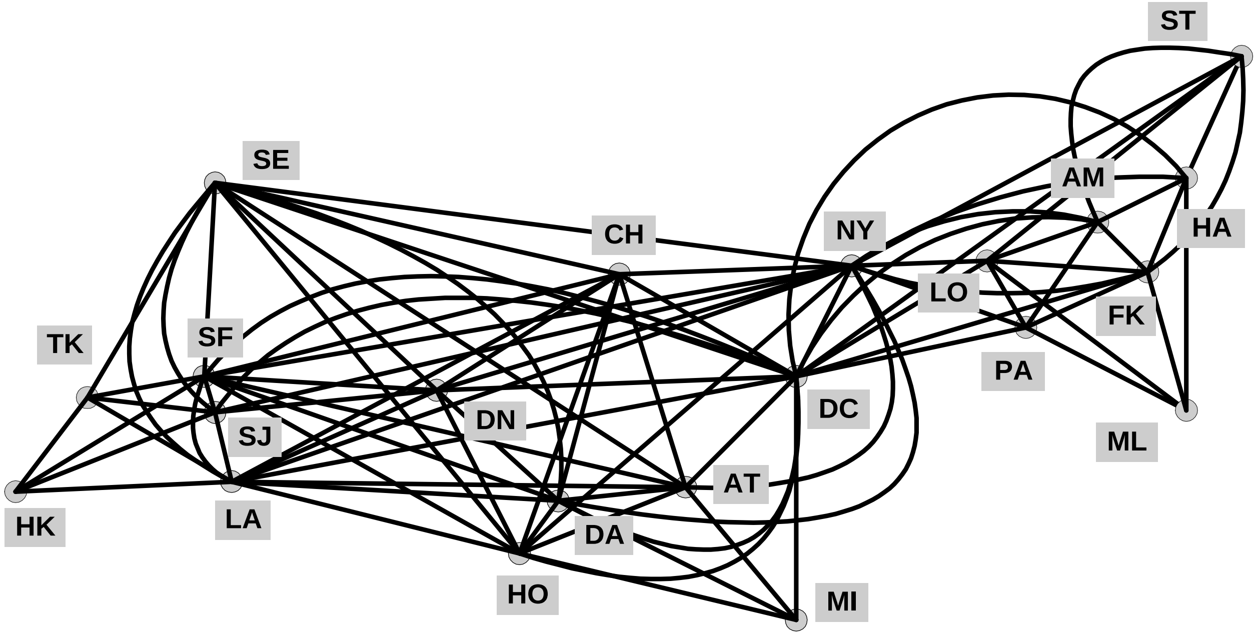

Figure 4: Combined contract network topology for Asia, USA and Europe. Two letter labels give city names, e.g. NY=New York, HK=Hong-Kong. For construction see 6.1

6 Global telecom geographics

Alongside the investigation of dynamical properties of basic triangular networks, we naturally wished to apply the analysis to larger and more realistic topologies. However, there was no such thing available as a global market topology, clearly because the market itself is changing and growing. So we had to construct a realistic international topology (see Figure 4) to analyze the applicability of our results. Specifically, we quantified the potential presence of triangles of link contracts and their side ratio distribution.

6.1 Contract graph construction

Since the market is today at a stage of expansion and transformation there is no single data source at present, certainly nothing like a global map of all underlying link contracts. We are, however, at a position where we can construct a plausible future topology using information on commercial optical fiber backbone maps, pooling point operator deployment plans and trades on the OTC bandwidth marketplaces. The result is a map of what *could* be offered by providers in the bandwidth market as *indivisible* contracts.

We constructed the contract map starting by overlaying several carrier backbone maps. We then create an edge (or link) between two pooling point locations A and B in our map if there is at least one backbone which could connect A and B without crossing any other pooling point. If all possible network-level paths would cross other pooling point locations, A-to-B cannot be offered as an indivisible contract. An example of this is Dallas-London. All network paths would cross the New York or Washington DC pooling points, Dallas-London is therefore not an edge in our graph.

The backbone maps we included in the set of potential providers are those of: Williams Communications Global Network, UUNET Internet Network, AT&T United States IP Backbone, Global Crossing Network, Level 3 Network, Enron Boradband Services United States Network, Genuity United States Fiber Optic Network, Qwest Americas IP Backbone, Cable & Wireless Global N^3 IP Backbone and European Ipergy Network, TeliaNet European IP Backbone, AboveNet Global IP Network, Onyx Networks Global IP Infrastructure and KPNQwest EuroRings. We believe this to be a fairly representative mix of national and global players, incumbents and new entrants.

The set of 21 pooling points is a subset of currently operating and planned locations, communicated to the authors by pooling point developers and operators (Enron, Lightrade) combined with destinations which have been popular in the OTC market ([2]). This gives: Tokyo, Hong-Kong, Seattle, San Francisco, San Jose, Los Angeles, Denver, Dallas, Houston, Chicago, Atlanta, New York, Washington DC, Miami, London, Paris, Milan, Frankfurt, Hamburg, Amsterdam and Stockholm. Even though the map would look different for subsets or supersets of the chosen pooling points, the current version provides a useful view of the topology of the underlying good. As a rough approximation of an edge's length we compute the geographical distance between two locations using the Haversine formula [1]. This method will generally underestimate the actual length of a network connection.

6.2 Observations

A simple observation we can make is that US and Europe are about 70% meshed. What is not immediately visible, but relevant to our analysis, as we show later, is the fact that many pooling points are clustered in relatively small densely connected regions (Asia, US West Coast, US East Coast, Texas, Europe), with a few really long edges from one local cluster to another.

Using this contract map we analyzed the topology for the presence of triangular subnetworks. Recall that we found network effects to be most pronounced with acute angle triangles where the side ratio was 1:10:10. We found 187 triangular subnetworks. We found that if we allowed the two "fixed" sides to differ by only 10% 86 triangles remained (46% of the total). Loosening this requirement to 20% included 129 triangles (69%) and to 30% included 175 triangles (94%). These ratios were constructed from geographical distance ratios not observed prices, since there was not enough pricing information available for many routes. However, at least some network costs are roughly linear with distance and these will dominate the mean trend (ν) for long distance routes, other things being the same (e.g. market and regulatory environment, etc).

The most immediate observation from this subnetwork classification is that there is an overwhelming number of acute angled triangles (1–5:10:10) and that this does not change with a loosening of the requirement to be included. As this requirement is loosened there is the suggestion of a second peak at very obtuse angles. There are very few triangles with a side ratio between 5:10:10 and 15:10:10. The reason for such a high proportion of acute angles triangles in the potential contract network is simple. There are concentrations of high-tech and population on the East and West coasts of the USA with a few mid-Western additions. Thus you get short sides within such concentrations and much longer ones between different concentrations. The Atlantic and Pacific Oceans only serve to reinforce this trend, whatever the difference in cost between land and undersea transmission cables.

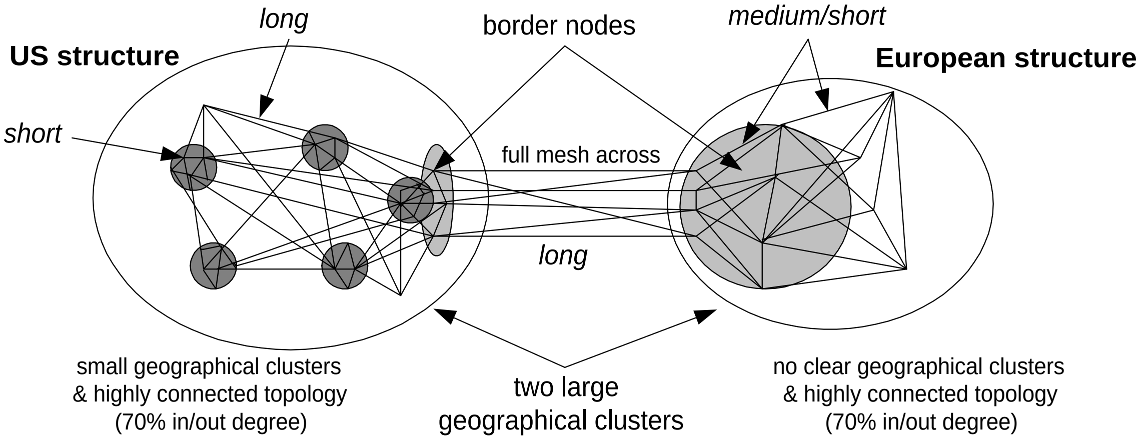

Figure 5: Features of the global wide-area bandwidth market topology . It consists mainly of two large clusters, one of which contains smaller clusters as well. This asymmetry is caused by the difference in distances among telecom hub cities in the US and in Europe. Each cluster is internally highly connected, at about the same degree. Certain nodes of one cluster have links to nodes of the other cluster. We call these "border" nodes. and they are fully meshed across the two clusters. Asymmetry is present again in that US border nodes are fewer than their European counterparts. Relative geographical distances are shown in italics.

We thus conclude that the presence of a high proportion of acute angled triangles in the network is a simple result of geography and expect this to be a very robust result. This implies that the network effects from geographical arbitrage, that are present whatever the ratio, will be very pronounced because these are most extreme for such topologies.

The underlying theme of this work is that the topology, or to be more precise, the geography of long-haul networks which underlies the bandwidth market influences price development of all traded contracts. To prove this statement simple triangular networks and a projection of future topology were used. Since we cannot be sure today of what the market will look like in some years' time, we summarize the most prominent topological features of the global bandwidth market in a general, abstract manner in Fig.5. In the last section we will briefly illustrate one of the issues a network provider is faced with in exploiting the opportunities of a commodity market.

7 Physical-to-contract bandwidth mapping

One interesting question arising from the discussion of telecom geographics and commodity trading is how a provider should allocate unused bandwidth to form commodity contract offerings. To explain the problem, we will use an example, illustrated in Figure 6. A small part of a carrier's network with spare capacity is shown in the star topology and the question is which is the best way to allocate bandwidth so that point-to-point contracts can be offered to the market and revenue is maximized. Three of several alternative allocation schemes are shown in the same figure. The result of the bottom-right picture would be a typical outcome of a *least-restrictive allocation* algorithm, i.e. an algorithm which for every new demand makes

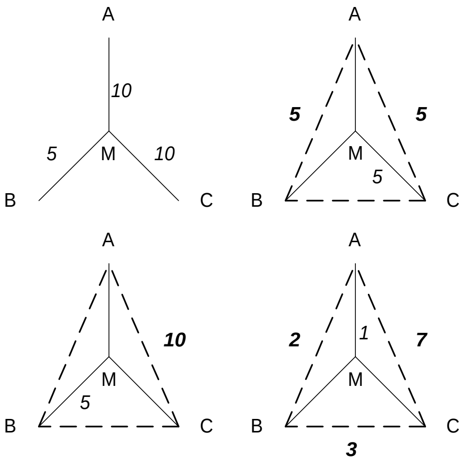

Figure 6: Assume that the top-left star topology is part of a provider's network and that numbers represent spare capacity units. Assume also that point M is not a pooling point, i.e. there is no commodity market for bandwidth originating or terminating at M. Then there is several ways of allocating the available bandwidth, resulting in different point-to-point contracts between A,B and C to be supplied to the marketplace. These are shown in the top-right and bottom drawings. Bold numbers (outside the triangle shape) represent capacity allocated for commodity contract fulfillment and plain numbers capacity not utilized for those contracts.

an allocation which maximizes the probability that future allocations between any two points can be satisfied. This seems to be a good approach since it results in a total of 12 units, as compared to 10 with the other two schemes (bottom-left and top-right).

However, when the allocation is made with trading in mind, what matters is the total amount of expected revenue from these contracts and it could very well be that the bottom-right scheme is not the most profitable, if, for example, links AB and AC are both much more expensive than BC. This would mean that ABC is an acute triangle, which, as shown in the previous section, is all too common. We could imagine the development of new bandwidth allocation algorithms for optimal decision-making in a commodity market environment.

Market conditions are subject to constant change so in any case it is becoming clear that fine-grained control of bandwidth is needed in this context (for example OSPF [11] would not solve the problem) and moreover the carrier with the most flexible network will have a significant advantage in the new market. Flexibility and allocation speed depend on physical layer technologies (e.g. all-optical networking is promising ultra-fast switching and reconfiguration of wavelengths) as well as on network management software. The latter should, among other things, present a network operator with a concise view of allocated resources, current and expected revenue that is generated by these and potential revenue from unused resources under different future allocation scenarios.

8 Discussion

Bandwidth trading is today characterized by a fair amount of uncertainty but at the same time the idea of bandwidth commodities is gaining more acceptance among carriers. Some are cautious about entering the market but even they are certainly following closely the developments, as can be deduced by the fact that several carriers have set up small teams to explore opportunities in this space. Everyone will agree that in light of this uncertainty insight into the workings of this market is key. Our hope is that we have contributed in this respect with focusing on exactly those aspects of bandwidth which set it apart from other commodities. Certainly many issues are still open for further investigation.

Anyone interested in building this market, creating liquidity or valuing the underlying assets clearly needs to have an understanding of the peculiarities of bandwidth. Market-makers and pooling point developers need to plan the deployment of pooling points and the routes on which to support the creation of a liquid market. Trading and decision support tool developers will need reliable stochastic models of prices on a multitude of inter-dependent routes.

On the other hand, arguably a trader should only need to quote a price for a fully standardized commodity, all other variables being already clearly defined in the contract. So is the networked nature of bandwidth, what we call network effects – or geographical arbitrage – irrelevant to a trader? Certainly not, since we have demonstrated that this type of arbitrage can occur very frequently, depends directly on the structure of the underlying (technology, QoS, topology) and will influence price dynamics on all routes.

We have presented the basic setup of this market and an approach to modeling prices taking into account network effects. Like with any other model, we can further refine and extend our approach, while success will ultimately depend on the ability to accurately model the market in the future. At this stage we mostly use the simulation framework based on the concepts described herein to understand the basic dynamics of the market. This is not to say that such modeling is not useful to market participants today. On the contrary, in a market that's not yet well understood analytical methods and simulations can provide not only valuable input to pricing decisions, but also the technological edge for successful planning in the new telecom landscape.

References

[1] B. Chamberlain. What is the best way to calculate the distance between two points? US Census Bureau, GIS FAQ Q5.1, http://www.census.gov/cgi-bin/geo/gisfaq?5.1, 2000.

[2] G. Cheliotis. Bandwidth trading in the real world: Findings and implications for commodities brokerage. 3rd Berlin Internet Economics Workshop, 26-27 May 2000, Berlin.

[3] G. Cheliotis. Qos in bandwidth commodity markets. In *forthcoming, Internet Services: The Economics of Quality of Service in Networked Markets.* MIT Press, Cambridge MA, 2001.

[4] G. Cheliotis and C. Kenyon. Failure contagion and qos elasticity in bandwidth markets. IBM Research Report RZ3345, http://domino.watson.ibm.com/library /cyberdig.nsf/Home . Accepted in IEEE Conference on Communications and Computer Networks 2001, October, 2001.

[5] S. Chen and K. Nahrstedt. On finding multi-constrained paths. In *IEEE International Conference on Communications (ICC 98)*, 1998.

[6] S. Chen and K. Nahrstedt. An overview of quality of service routing for next-generation high-speed networks: Problems and solutions. *IEEE Network*, pages 64–79, November/December 1998.

[7] M.R. Garey and D.S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W.H. Freeman and Co, San Francisco, 1979.

[8] C. Kenyon and G. Cheliotis. Stochastic models for telecom commodity prices. *Computer Networks*, 36(5-6):533–555, 2001. Special Issue on Network Economics.

[9] R. Merton. Option pricing when underlying stock returns are discontinuous. *J. Financial Economics, 3, Jan-Mar*, pages 125–144, 1976.

[10] B. Rupp and V. Grimm. A proposal for an automated bandwidth exchange. Presented at the 3rd Berlin Internet Economics Workshop, 26-27 May 1999, Berlin.

[11] M. Steenstrup. *Routing in Communications Networks.* Prentrice Hall, New Jersey, USA, 1995.

[12] Z. Wang and J. Crowcroft. Quality-of-service routing for supporting multimedia applications. *IEEE Journal on Selected Areas In Communications*, pages 1228–1234, September 1996.