Population Distribution Effects in Backbone Network Cost

Dimitris Maniadakis and Dimitris Varoutas, *Member, IEEE*

***Abstract*— Population characteristics such as density or dispersion are heavily used in exploring the viability of infrastructure investments and telecommunication networks' development, but not much attention has been paid to the effects of distribution of population sizes. In this paper, a fast method is presented for estimating the backbone network cost using population size distribution as an input parameter, which is discovered to play a significant role in cost behavior. Combining a variation of the already proposed gravity model for creating realistic complex network topologies with Pareto distribution model for city sizes, allows building simulations, and presenting results for fixed operators' backbone network growth. An application of the methodology is also illustrated, providing estimation and comparison of the network cost for three European countries while increasing their population coverage. The main aim of this work is to provide a simple tool for rapid cost estimation of a countrywide physical backbone network and highlight the population distribution impact, assisting both telecommunication providers and regulators.**

***Index Terms*— backbone network topology, gravity model, population coverage, population distribution, spatial network growth.**

I. Introduction

When an operator is planning to enter or expand in the backbone network market, either as a service provider or a wholesaler, it is crucial to consider the expected cost of investment, among others. It is often spent substantial time and resources to obtain reasonably accurate estimates of deployment costs in particular countries or regions. Long techno-economic analyses specific to each region are a commonplace to determine the feasibility of the project. They usually combine information on telecommunication services demand and costs with geographic and demographic characteristics, and can demonstrate how cost structure is affected while expanding the network [1]. Most times results show that operators should not develop network infrastructure in remote or rural high-cost areas, because investment is not commercially viable [2]. Similar results are expected in the backbone network as well.

In the backbone, the central part of a network that handles the major traffic, what operators have deployed most of the times is microwave radio or satellite links especially when connecting remote and small sites [3]. These solutions are usually capable of delivering speeds that meet the current traffic needs and are relatively cheap solutions. However, their capacity can meet only a small portion of broadband demand for the foreseeable future. On the contrary, wireline solutions such as fiber-optics will always offer more capacity, paying a high price wherever developed. It is generally assumed that fiber-optic infrastructure will eventually dominate in connecting large cities and small towns and replacing most of today's backbone wireless technologies.

That physical backbone network construction does not usually follow a strict topology pattern, but a number of rules-of-thumb, or even better complex rules. This explains why each operator has built a very different backbone network [4-6] and only few contributions have been made on backbone network modeling [5, 6]. Backbones actually connect and serve areas where populations are located and this makes population distributions a presumable factor to consider their influence on the network design [7-9], and cost [9-11]. According to [12] a telecommunications system is not just a technological system but a complex system of people and technology. Also, [13] supports that the interconnectivity of spatially distributed populations is a key to the construction of bottom-up models of techno-social systems. It is therefore of importance to introduce population size distribution in a network cost estimation method and investigate if it has a notable effect on deployment cost, focusing here in terrestrial cable networks i.e. fiber-optics, that need ducts for installation, besides the traditional approach of taking into account only customer distribution and forecasted sales.

This analysis will endeavor to do three things; (1) propose a fast methodology for estimating the backbone cost (2) determine the effect of population distribution in the backbone cost (3) apply the methodology and try to validate it in three European countries of different scale of population distribution.

The rest of this paper is structured as follows. Section 2, contains a review of previous related studies. The methodology, the assumptions and the resulting distributions are presented in Section 3, the dataset and the results of applying the methodology are in Section 4, while Section 5 concludes the analysis.

Manuscript received July 2, 2010; revised August 30, 2010. The authors are with the Department of Informatics and Telecommunications, University of Athens, Panepistimiopolis Ilissia GR-15784, Greece (e-mail: D.Maniadakis@di.uoa.gr; D.Varoutas@di.uoa.gr).

II. Backbone Topology Representation

A. Network Models

Understanding the complex and evolving nature of networks and especially their interactions with locations and populations has garnered an increasing amount of interest [7, 8, 10, 14-16]. In recent years several contributions have been made on modeling network topology of air transportation, electricity, social or telecommunication networks [16-22].

Despite the engineer's dominant role, as in the case of the topology of backbone networks at the physical level, emergent and unplanned topological traits appear in [15, 21, 23]. They rarely can be described by traditional patterns such as star, bus, ring, hierarchical or full mesh topologies. A variety of approaches from complex network theory have been discussed lately on the complex formation for most of these networks. This research area involves networks with non-trivial topological features that are neither purely regular nor purely random. Complex network theory influenced by ideas and tools from graph theory and statistical physics has led to the creation of numerous growth models [17-19, 24]. Some of these models seem to be promising for the representation of realistic, still not cost-optimized, large-scale backbone networks.

The proposed models could be divided into three categories; those that focus on node degree properties and ignore geography, those that involve real space limitations and finally those that consider intrinsic attributes of nodes.

The oldest model in the first category is the Erdős–Rényi random graph model [24], which has proven to be an unrealistic pattern of real-world networks [17]. A widely accepted model in the first category is the Barabási–Albert model [17], in which a new node is more likely to be connected to an already well-connected node, using a preferential attachment mechanism. This concept has been employed to explain the existence of power-laws in several layers of networks' topologies [25]. However, in the case of backbone networks this perspective does not seem to be realistic because geographical distances have been omitted, which are an important cost factor [18]. On the second category, Waxman's [26] and Barthelemy's [18] models are good candidates for applying in backbone networks; however, they are too simplistic [5] and fail to describe any traffic flow effects on topology [5, 22]. Gravity models can be considered to establish a third class. They are mathematical models based on an analogy with Newton's gravitational law, later proposed to be applied to international trade flows [27], and recently to telecommunication traffic [19, 20, 22].

Gravity models use intrinsic attributes of the nodes to represent their fitness to win edges, where fitness could be interpreted as social skills, GDP, business employees, number of Internet hosts, or populations, etc [22, 28]. They combine these intrinsic attributes and the geographical influence in a try to predict the traffic level between nodes, and to construct the network more efficiently [22]. In regards to telecommunication interactions between cities, the gravity-like models relate traffic to population and distance [9, 20, 28]. Specifically, the magnitude of inter-city traffic flows is estimated as an increasing function of the population sizes of communicating partners, and a decreasing function of the distance between them. One of the many variations for traffic between nodes *i* and *j* takes the form [19]:

$$T_{ij} \sim \frac{M_i M_j}{d_{ij}^2} \tag{1}$$

where *T* is the expected traffic, *d* is the distance between the nodes and *M* is usually represented by population sizes for simplicity, as in [19, 20, 22]. The general concept is that after estimating the average traffic volumes (fluctuations are omitted, although there is evidence of their importance [29]) of every node pair between new node and old nodes, they are arranged in descending order, and then these pairs are connected in sequence of *T* until an edge threshold is reached.

B. Population Parameters

As far as population metrics are concerned, a principal measure used in techno-economics literature is population density, which is a measurement of population in a unit area [2, 16, 30]. Another key term is population dispersion which measures how population is spread in an area and is usually associated with the percentage of the landmass inhabited by a certain percentage of population [30]. They either calculate average quantities or just give little knowledge about the relative proportion of population sizes and their spatial distribution. The aforementioned information is necessary for studying the traffic patterns, thus the efficient topology of the backbone network. Unfortunately, both of these metrics are incapable of providing sufficient information for developing a backbone network, but could be valuable for the deployment of the access network [2]. However, connecting cities is different from connecting intra-city locations. An alternative way to examine the effects of geodemography on coverage is by considering the distribution of population sizes within countries. It would be rationale to suggest that a country with the majority of its population clustered in a few geographic areas should be easier to cover with telecommunication networks [30]. This convenience, in turn, could indicate higher potential for serving a big part of the population with low cost than a country with a more uniform population distribution.

The literature on the population distribution of city sizes is extensive, focusing mainly on applying power-law or Pareto distribution [31, 32]. The form of the size distribution of cities as first suggested in [33] takes the following Pareto distribution

$$y = A \cdot x^{-a} \tag{2}$$

where *x* is a particular population size, *y* is the number of cities with populations greater than *x*, and *A* and *α* are constants (*A*, *α*>0). The term *α* is also called "Pareto exponent". The size distribution, when plotted on double logarithmic axes, shows a remarkable linear pattern where the slope of the line is approximately -1, corresponding to the Zipf distribution. The

"Pareto exponent" is a measure of how evenly distributed the population is. The higher the value of α, the more uniform in size the cities are. The empirical validity of the Pareto principle for cities has recently been examined using data on the city populations from 73 countries in [32] and every country was characterized by a different α. The range of values for α is between 0,7 (Guatemala) and 1,7 (Kuwait), with the mean of α approximately 1,1.

III. Methodology

The procedure that is followed is mainly based on the combination of the models described in (1) and (2). Using this methodology, simulations were built to identify the population distribution effects on the cost of expanding the network.

A continuous two-dimensional plane is considered. The square kilometers of the surface are in accordance with the country or region area actually examined. A network of nodes (cities) and links (physical backbone connections) shall be constructed by gradually making it grow. The locations of the cities are randomly chosen and a gravity model, based on that developed in [19] and described by (1), was used to construct realistic backbone topologies. If commercial considerations govern the laying of telecom backbone, it would first be laid on routes yielding adequate returns on investment [3]. Therefore, it was assumed that cities are connected in sequence of population size. As far as the number of links per new node is concerned, it was decided that connecting with two nodes is more realistic than connecting with one, thus increasing the number of rings and independent physical paths and introducing redundancy in the topology. This assumption is in accordance with real-world WAN (wide area networks) measurements found in [4, 5].

Similarly, a population size distribution model is needed in order to distribute the population in each city. Research in this area has shown that population distribution follows a Pareto distribution [31, 32], as discussed earlier in Section 2. The power-law PDF (probability density function) described in (2) can be transformed in:

$$x = A^{\frac{1}{a}} \cdot y^{-\frac{1}{a}} \tag{3}$$

Then, considering x as P (population) and y as *rank*, that expression can be rewritten as a power-law:

$$P_i = A^{\frac{1}{a}} \cdot rank_i^{-\frac{1}{a}} \tag{4}$$

In Fig. 1, a typical population size distribution is plotted and every city's population can be found deploying (4).

Undoubtedly, using the population to denote the fitness of a city is just intuitive; more complex demographic and socio-economic factors are also important for the traffic [28, 34]. The expected traffic can be redefined introducing the population as intrinsic attribute of cities to compete for links:

$$T_{ij} = \frac{P_i \ P_j}{d_{ij}^2} \tag{5}$$

The above form can now be used to come to a combined

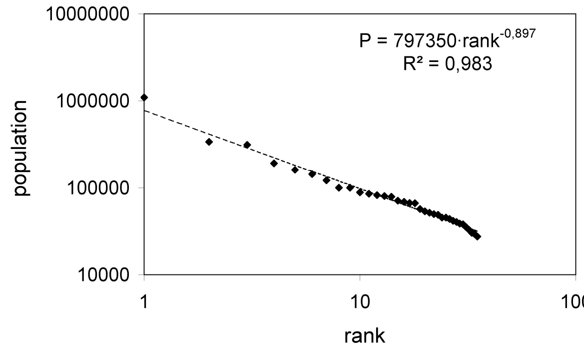


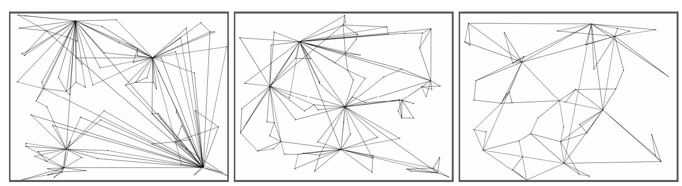
Fig. 1. Typical population distribution is a power-law distribution. The case of Bulgaria (2001).

equation of gravity model and Pareto model described in:

$$T_{ij} = \frac{A^{\frac{2}{a}} \cdot rank_i^{-\frac{1}{a}} \cdot rank_j^{-\frac{1}{a}}}{d_{ij}^2} \tag{6}$$

New nodes are gradually introduced in every discrete time step in the network and connected according to the gravity model concept. The traffic level for every node pair, between the new and the previously inserted nodes, is calculated using (6). Then, the new node is connected with the two nodes indicating the highest values for T. After inserting all N nodes in the network, the procedure ends.

As far as the simulations are concerned, three different population distributions were compared, while keeping the same area and total population. There were constructed networks (Fig. 2) for α=0,7, α=1,1 and α=1,7, which are the lower, the average, and the upper limit of α, respectively [32].



(a) (b) (c)

Fig. 2. Typical networks for the same area (1000x1000), same total population and different population distributions (a) α=0,7 (b) α=1,1 (c) α=1,7.

These assumptions led to different number of cities, less for the highest α. In order to have results that cover the entire population, the percentage population coverage was increased in steps of 0,001, with 1000 runs generated for each step and finally keeping only the average of the length value.

This study focused on the cost of the network, as represented by the total length of all its edges (trunk lines). It is obvious that fiber ducts cannot be installed physically everywhere; normally they must follow the roads infrastructure. Therefore, there is a need to correct the Euclidean distances by multiplying with an adjustment road factor, assuming that on average, real distances (across the roads) are about $\sqrt{2}$ times straight-line distances [35]. For every single topology generated, the total length of the network was calculated using the following equation:

$$TotalLength = \sum_{i<j}^{N} B_{ij} d_{ij} r \tag{7}$$

where B is the adjacency matrix, d is the shortest geographic distance, r is the adjustment road factor and N is the number of cities to be connected.

In Fig. 3 it can be observed how length is increasing while covering more and more population. The population coverage at each level of expenditure is in a linear relationship with the cost for deploying it for high values of α, but this is not true for lower values of α, where a remarkable exponential behavior is seen. It is clear that the cost of deployment is very sensitive to the population distribution in cities. Low values of α lead to very high total cost of deployment when covering almost 100% of the population. However, when covering a smaller part of the population e.g. 50%-90%, the total expense is moderate and lower than that of high α.

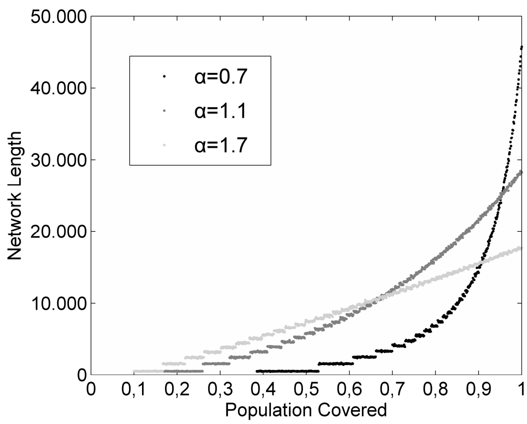


Fig. 3. Relationship between the network length and the cumulative population coverage for different population distributions.

IV. Dataset and Results

As a preliminary validation of whether or not the present methodology produces reasonable results, the described method is applied to three European countries, one with the maximum, one with the minimum, and another with an average value of α. The values of α and A were collected from [32] and the European countries with maximum, average and minimum α, were found to be Belgium, Bulgaria, Belarus, respectively. Census data for the validation of the numbers of cities was obtained from [36] for years 2001 for the first two and 2009 for the third country, and surface area data was collected from [37]. The countries under examination are approximately of same total population size, but quite different area and population size distribution (Table I).

Given that network operators are reluctant to share information about their physical infrastructure, *Google Maps* have been used in order to compute the equivalent network lengths. The connection rules of the simulation methodology of Section 3 were followed, respecting the geographic limitations (i.e. real locations of cities, road restrictions, etc) and assigning real population sizes to cities. The number of connected cities, k, for 50%, 90% and 99% of population coverage was estimated (Table II) using (8). However, the real number of cities and the real required length were calculated only for the 50% of population (Table II and Fig. 6), as a preliminary validation, due to the calculating difficulties in the cases of 90% and 99%, where there are hundreds of cities to connect. The algorithm used is:

TABLE I
Country Characteristics

Country	α	A	Population	Area (km^2)
Belgium	1,5895	803.750.000	10.296.350	30.528
Bulgaria	1,1140	3.753.500	7.932.984	110.912
Belarus	0,8435	206.140	9.648.533	207.600

TABLE II
Number Of Cities Connected

Country	Real number for 50% coverage	Estimation for 50% coverage	Estimation for 90% coverage	Estimation for 99% coverage
Belgium	100	102	418	530
Bulgaria	35	38	426	681
Belarus	14	9	631	3369

$$\max \quad k,$$

$$subject \quad to: \quad \sum_{i=1}^{k} P_i < c \cdot \sum_{j=1}^{N} P_j, \quad where \quad c \in \{50\%, 90\%, 99\%\} \quad (8)$$

There were generated 1000 runs for each of the above percentages of population coverage of each country and finally only the average length values were kept. Results are presented in Fig. 4. The rate of length increase for Belarus is found to be much higher than Belgium and Bulgaria, as already noticed in Fig. 3 for countries with low α. On the other hand, Belgium, with a high α, shows an almost linear relationship between length and coverage expansion. It is noticeable that there is an agreement between real measurements and estimated network lengths (Fig. 5). Also, the agreement in the differences between lengths and number of cities in each country is

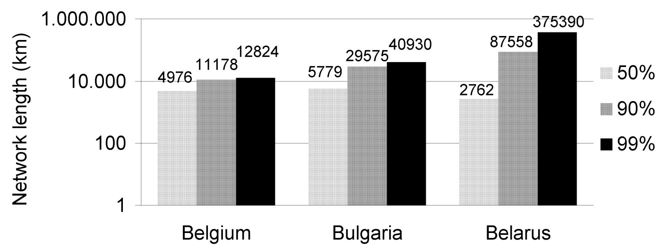


Fig. 4. Estimated network length for different percentages of coverage.

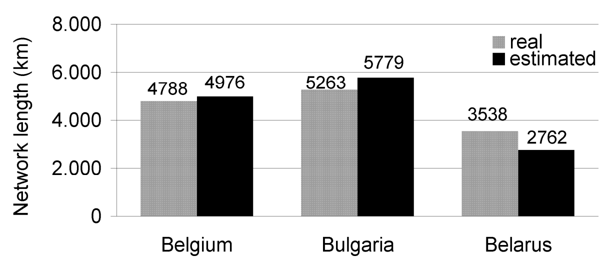


Fig. 5. Real and estimated network length for 50% population coverage.

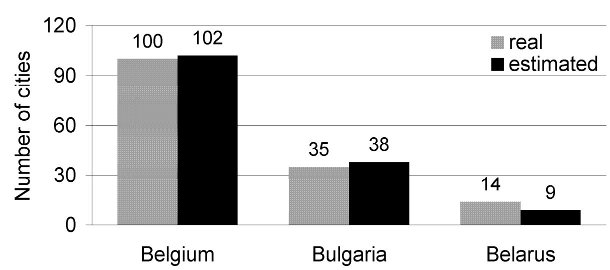


Fig. 6. Real and estimated number of cities for 50% population coverage.

noteworthy and probably the errors in the length estimation are due to the errors in cities' number estimation (Fig. 6).

Civil works account for most of the cost of constructing fiber-optic cable networks [2, 38]. The cost of investment in money was estimated assuming that the cost is approximately 20€/m, including ditch, duct, fiber and labor costs, based on benchmarking from European countries. The resulting costs in millions of Euros are presented in Table III.

TABLE III
ESTIMATED NETWORK COST (M€)

Country	50% coverage	90% coverage	99% coverage
Belgium	66,71	141,14	160,89
Bulgaria	76,35	361,90	498,16
Belarus	40,15	1.057,70	4.511,68

V. CONCLUSION

In this paper, a simple and fast way was provided to estimate the cost level for any percentage of population coverage without the need of detailed data. A variation of the gravity complex network model combined with the Pareto population distribution model was used, as a realistic growing backbone network. After running several simulations and applying the concept to the case of three European countries, the illustrated methodology was found to provide realistic results. Results related to the first 50% of population could be valuable during the early stages of network rollouts, while results concerning the last 10% of population appear to be very sensitive to the population distribution, pointing out that countries with low "Pareto exponent" will face an additional obstacle to bridge that digital divide. The findings of this work could be exploited by telecom managers and engineers, as well as policy makers and regulators.

REFERENCES

[1] S. De Maesschalck, *et al.*, "Pan-European optical transport networks: An availability-based comparison," *Photonic Network Communications,* vol. 5, pp. 203-225, May 2003.

[2] T. Monath, N.K. Elnegaard, P. Cadro, D. Katsianis, and D. Varoutas, "Economics of fixed broadband access network strategies," *IEEE Communications Magazine*, vol. 41, pp. 132-139, Sep 2003.

[3] H.V. Singh and R. Samarajiva, "One Backbone, or Two?," in *ICT Infrastructure in Emerging Asia: Policy and Regulatory Roadblocks*, S. India, ed: IDRC, 2008.

[4] B. Chatelain, M.P. Belanger, C. Tremblay, F. Gagnon, and D.V. Plant, "Topological Wavelength Usage Estimation in Transparent Wide Area Networks," *IEEE/OSA Journal of Optical Communications and Networking,* vol. 1, pp. 196-203, Jun 1 2009.

[5] C. Pavan, R.M. Morais, J.R.F. da Rocha, and A.N. Pinto, "Generating Realistic Optical Transport Network Topologies," *IEEE/OSA Journal of Optical Communications and Networking,* vol. 2, pp. 80-90, Jan 1 2010.

[6] R. Zhang-Shen and N. McKeown, "Designing a predictable internet backbone network," in *Hot Topics in Networks III workshop*, 2004.

[7] S.P. Gorman and R. Kulkarni, "Spatial small worlds: new geographic patterns for an information economy," *Environment and Planning B-Planning & Design,* vol. 31, pp. 273-296, Mar 2004.

[8] E.J. Malecki, "The economic geography of the Internet's infrastructure," *Economic Geography,* vol. 78, pp. 399-424, Oct 2002.

[9] M.C. Sinclair, "Improved Model for European International Telephony Traffic," *Electronics Letters,* vol. 30, pp. 1468-1470, Sep 1 1994.

[10] S. Appleby, "Estimating the Cost of a Telecommunications Network Using the Fractal Structure of the Human-Population Distribution," *IEE Proceedings-Communications,* vol. 142, pp. 172-178, Jun 1995.

[11] M. Naldi and C. Salaris, "Rank-size distribution of teletraffic and customers over a wide area network," *European Transactions on Telecommunications,* vol. 17, pp. 415-421, Jul-Aug 2006.

[12] T.N. Andrew and D. Petkov, "The need for a systems thinking approach to the planning of rural telecommunications infrastructure," *Telecommunications Policy,* vol. 27, pp. 75-93, Feb-Mar 2003.

[13] A. Vespignani, "Predicting the Behavior of Techno-Social Systems," *Science,* vol. 325, pp. 425-428, Jul 24 2009.

[14] A. Lakhina, J.W. Byers, M. Crovella, and I. Matta, "On the geographic location of Internet resources," *IEEE Journal on Selected Areas in Communications,* vol. 21, pp. 934-948, Aug 2003.

[15] L.A. Schintler, *et al.*, "Complex network phenomena in telecommunication systems," *Networks & Spatial Economics,* vol. 5, pp. 351-370, Dec 2005.

[16] S.H. Yook, H.W. Jeong, and A.L. Barabasi, "Modeling the Internet's large-scale topology," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 99, pp. 13382-13386, Oct 15 2002.

[17] A.L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509-512, Oct 15 1999.

[18] M. Barthelemy, "Crossover from scale-free to spatial networks," *Europhysics Letters,* vol. 63, pp. 915-921, Sep 2003.

[19] W.S. Jung, F. Wang, and H.E. Stanley, "Gravity model in the Korean highway," *Epl*, vol. 81, pp. 48005-48005, 2008.

[20] G. Krings, F. Calabrese, C. Ratti, and V.D. Blondel, "Urban gravity: a model for inter-city telecommunication flows," *Journal of Statistical Mechanics-Theory and Experiment,* Jul 2009.

[21] M.L. Mouronte, *et al.*, "Complexity in Spanish optical fiber and SDH transport networks," *Computer Physics Communications,* vol. 180, pp. 523-526, Apr 2009.

[22] J.H. Qian and D.D. Han, "A spatial weighted network model based on optimal expected traffic," *Physica A-Statistical Mechanics and Its Applications,* vol. 388, pp. 4248-4258, Oct 1 2009.

[23] J. Rutherford, A. Gillespie, and R. Richardson, "The territoriality of Pan-European telecommunications backbone networks," *Journal of Urban Technology,* vol. 11, pp. 1-34, Dec 2004.

[24] P. Erdős and A. Rényi, "On random graphs," *Publ. Math. Debrecen,* vol. 6, pp. 290–297, 1959.

[25] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the Internet topology," in *ACM-SIGCOMM*, 1999, pp. 251-262.

[26] B.M. Waxman, "Routing of Multipoint Connections," *IEEE Journal on Selected Areas in Communications,* vol. 6, pp. 1617-1622, Dec 1988.

[27] J. Tinbergen, *Shaping the world economy : suggestions for an international economic policy*. New York: Twentieth Century Fund, 1962.

[28] A. Dwivedi and R.E. Wagner, "Traffic model for USA long-distance optical network," presented at the Optical Fiber Communication Conference, OSA Technical Digest Series, 2000.

[29] D.-H. Kim and A.E. Motter, "Fluctuation-driven capacity distribution in complex networks," *New Journal of Physics,* vol. 10, May 2008.

[30] OECD (2008). *Broadband Growth and Policies in OECD Countries*. Available: www.oecd.org/sti/ict/broadband/growth

[31] K.T. Rosen and M. Resnick, "The Size Distribution of Cities - an Examination of the Pareto Law and Primacy," *Journal of Urban Economics,* vol. 8, pp. 165-186, 1980.

[32] K.T. Soo, "Zipf's Law for cities: a cross-country investigation," *Regional Science and Urban Economics,* vol. 35, pp. 239-263, May 2005.

[33] F. Auerbach, "Das Gesetz der Bevölkerungskonzentration," *Petermann's Geographische Mitteilungen,* vol. 59, pp. 74-76, 1913.

[34] A. Díaz-Pinés, "Indicators of Broadband Coverage," in *OECD Digital Economy Papers*. vol. 165, ed: OECD, Directorate for Science, Technology and Industry, 2009.

[35] C.S. Phibbs and H.S. Luft, "Correlation of travel time on roads versus straight line distance," *Medical Care Research and Review,* vol. 52, pp. 532-542, Dec 1995.

[36] T. Brinkhoff. *City Populations*.Available: http://www.citypopulation.de/

[37] United Nations Statistics Division, "Total surface area as of 19 January 2007," 2007.

[38] M. Hanna and B. Ramarao, "Cost Optimization for Transmission and Backhaul Technologies," *Accenture Communications Solutions Point of View,* 2006.