An Analysis of Incentives for Network Infrastructure Investment under Different Pricing Strategies

ONLINE APPENDIX

Appendix A.1 Direction of Network Capacity Expansion θ*

Here, we will provide the discussion of second order conditions for the expression (4) for the normalized vector of the optimal direction of increase of the capacity of our theoretical network. We start with our maximization problem as follows:

$$\max_{\theta} B(K_0 + b\theta) - cb(\theta \underline{e}) \qquad \text{such that} \quad \|\theta\| = 1 \tag{A1}$$

Our Lagrangian is:

$$L(\theta,\lambda) = B(K_0 + b\theta) - cb(\theta \underline{e}) - \lambda(\|\theta\|^2 - 1) \qquad \text{(A2)[1]}$$

Taking first derivative with respect to each θ_m , *(m = 1,..,M)* and setting it to equal zero, we will get the optimal direction of capacity adjustment. The first order conditions are:

$$\frac{\partial L}{\partial \theta_m} = b(\frac{\partial B}{\partial K_m} - c) - 2\lambda\theta_m = 0 \qquad m=1,\dots,M \tag{A3}$$

$$\theta_m^* = \frac{b}{2\lambda}(\frac{\partial B}{\partial K_m} - c) \qquad m=1,\dots,M \tag{A4}$$

The entire optimal direction vector becomes:

$$\theta^* = \frac{b}{2\lambda}(\nabla B - c\underline{e}) \qquad \underline{e} = (1,\dots,1) \in R^M \tag{A5}$$

where: $\nabla B \equiv \frac{\partial B}{\partial K_m}$, $m=1,\dots,M$

---

[1] Our constraint $\|\theta\| = 1$ is expressed as $\|\theta\|^2 = 1^2 = 1$, providing a clearer way of presenting first order conditions.

The value of a scalar b/2λ can be retrieved using the constraint from (A1):

$$\left\|\theta^*\right\| = \frac{b}{2\lambda}\left\|\nabla B - c\underline{e}\right\| = 1 \tag{A6}$$

$$\frac{b}{2\lambda} = \frac{1}{\left\|\nabla B - c\underline{e}\right\|} \tag{A7}$$

Therefore our direction vector of capacity expansion can be written as:

$$\theta^* = \frac{\nabla B - c\underline{e}}{\left\|\nabla B - c\underline{e}\right\|} \tag{A8}$$

while:

$$\lambda = 0.5b\left\|\nabla B - c\underline{e}\right\| \tag{A9}$$

We start the second order condition analysis by differentiating first order conditions, as shown in A3, with respect to θ_j:

$$\frac{\partial^2 L}{\partial \theta_j^2} = b^2\left(\frac{\partial^2 B}{\partial K_j \partial K_m}\right) - 2\lambda\delta_{jm} = 0 \qquad (j=1,..,M),\ (m=1,..,M) \tag{A10}$$

where $\delta_{jj} = 1$ and $\delta_{jm} = 0$ for $j \neq m$.

Substituting for A9 into A10 our second order conditions are:

$$\frac{\partial^2 L}{\partial \theta_j^2} = b^2\left(\frac{\partial^2 B}{\partial K_j \partial K_m}\right) - b\left\|\nabla B - c\underline{e}\right\|\delta_{jm} = 0 \qquad (j=1,..,M),\ (m=1,..,M) \tag{A11}$$

Note that the diagonal of the Hessian matrix is proportional to:

$$b\left(\frac{\partial^2 B}{\partial K_j^2}\right) - \left\|\nabla B - c\underline{e}\right\| = 0 \qquad (j=1,\ldots,M) \tag{A12}$$

As long the step size b is kept sufficiently small, each component j of A12 is approximately equal to:

$$-\left\|\nabla B - c\underline{e}\right\| \tag{A13}$$

A13 is strictly negative as long as $\left\|\nabla B - c\underline{e}\right\|$ is not equal to zero. Also note that off-diagonal elements of our Hessian matrix A11 are proportional to:

$$b\left(\frac{\partial^2 B}{\partial K_j \partial K_m}\right) \quad (j=1,..,M),\ (m=1,..,M),\ j \neq m \tag{A14}$$

Again, note that A14 can be made arbitrarily small, by choosing a small step size *b*. The analysis provided above shows that the Hessian matrix A11 can be made arbitrarily proportional to -I (where I is the unit diagonal matrix), which implies that the second order conditions are satisfied. In other words, if $\left\|\nabla B - c\underline{e}\right\|$ is not equal to zero, our solution is satisfactory for sufficiently small step sizes. In practice, all one needs to do is check that the objective function actually increases with each step. If a decrease is observed, the step size should be reduced. This recursive procedure will eventually lead to the point where $\left\|\nabla B - c\underline{e}\right\|$ is equal to zero, implying that, as long as the exogenous arrival rate does not change, capacity of the system should remain unchanged.

Appendix A2. Assessing the impact of ignoring the budget constraint during the capacity expansion process

In order to address this impact, we have conducted an additional set of simulations implementing the server capacity reallocation in the following manner: The capacity of the set of servers was increased incrementally in the direction of ∇B, adjusting the step size so $\nabla B\underline{e}$ is equal to a fixed small increment. The iterations were then performed of adjusting capacity in the direction of $\nabla B - \left[\nabla B\underline{e}\right]\left[\underline{e}\frac{1}{n}\right]$, corresponding to the zero increase in aggregate system capacity. In essence, this procedure reallocates the capacity of the servers without changing the aggregate system capacity and budget involved (under assumption of zero reallocation costs). The above-average performing servers, in terms of marginal benefits gained, will have their capacity increased proportionally to their contribution, while those performing below average will have their capacities proportionally reduced. This re-allocation at a given budget level is then repeated until this adjustment process became negligible in term of incremental benefits gained. This process is conducted at all budget increments. We present the set of the results, comparing the benefits under the old gradient method and the revised method that includes then capacity reallocation. The comparison table shows that at lower levels of aggregate capacity, the reallocation process produced more noticeable changes, with more visible improvements in benefits generated.

Capacity Level	Benefits Generated under old gradient method	Maximum Benefits Generated when capacity reallocation is conducted	Percentage change
12.80	0.025061	0.025798	2.94%
19.20	0.031137	0.032023	2.85%
25.60	0.036258	0.037611	3.73%
32.00	0.040773	0.042288	3.72%
38.40	0.044642	0.04576	2.50%
44.80	0.047733	0.048464	1.53%
51.20	0.050212	0.050632	0.84%
57.60	0.052358	0.052728	0.71%

Table 1. Result comparison of the gradient method and the gradient method with budget-constrained capacity reallocation. Capacity level in the last row corresponds to K^* resulting in the highest level of net benefits for arrival rate x=100 requests/sec.

Appendix A.3 Expansion of Network's Capacity under Different Exogenous Arrival Rates: Benefits Generated

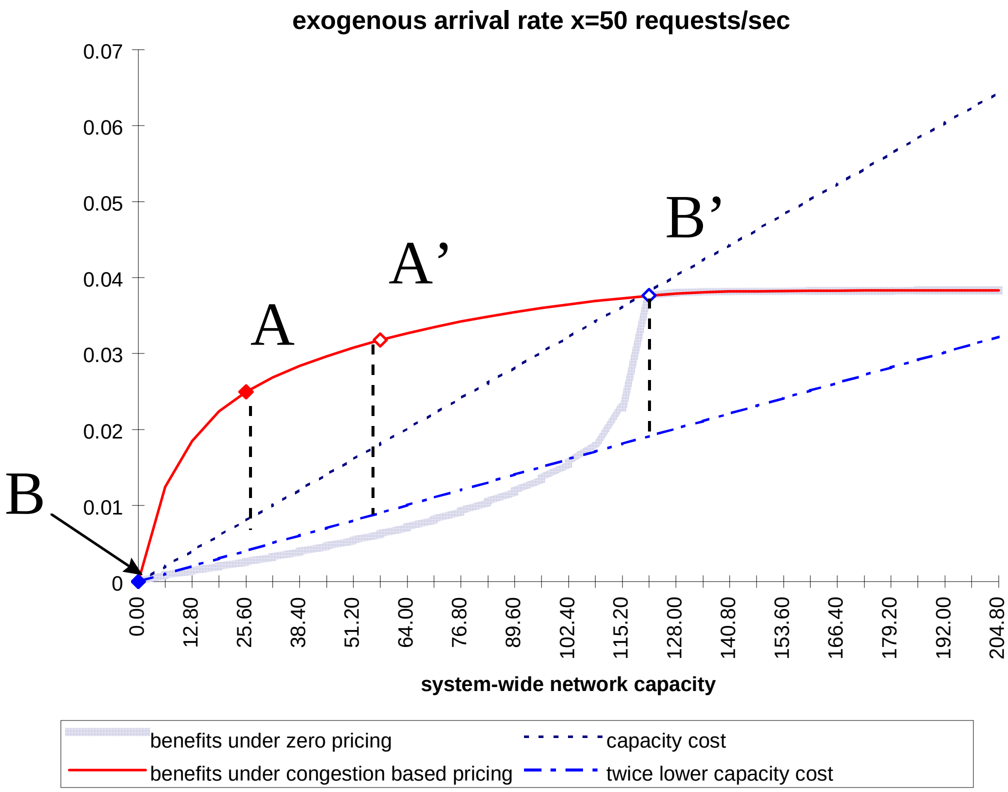


Figure A3.1: Benefits Generated under Both Pricing Schemes for the Exogenous Arrival Rate of 50 Requests/Second

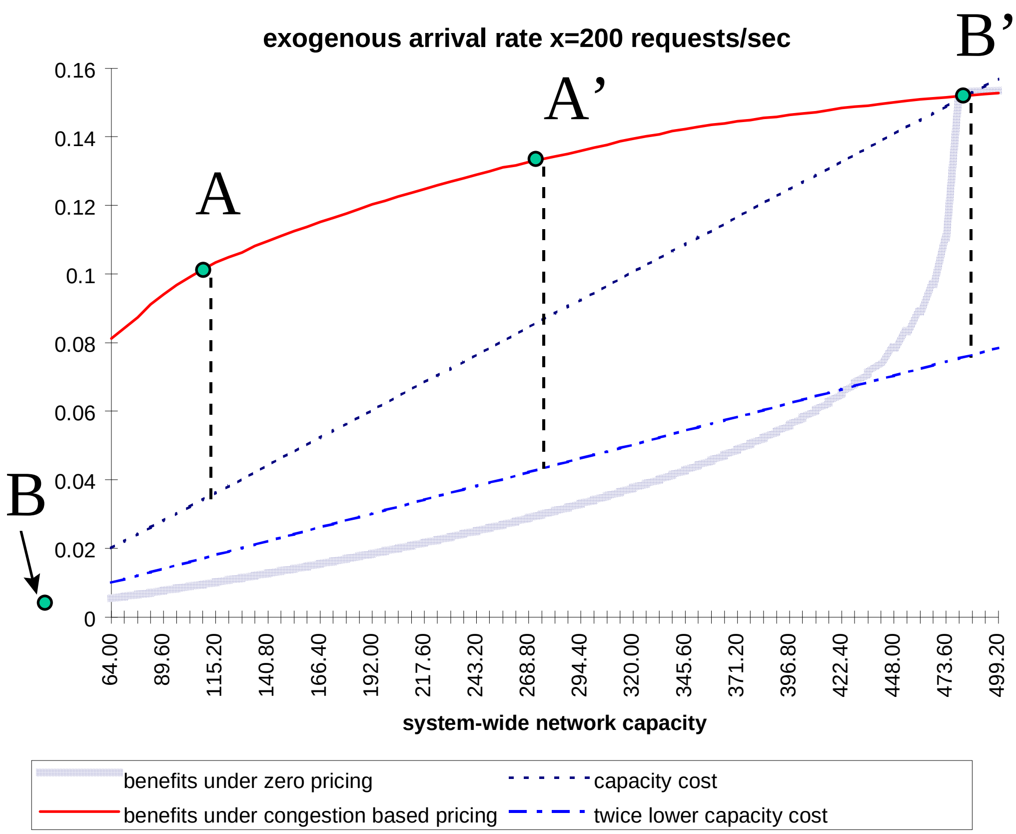


Figure A3.2: Benefits Generated under Both Pricing Schemes for the Exogenous Arrival Rate of 200 Requests/Second

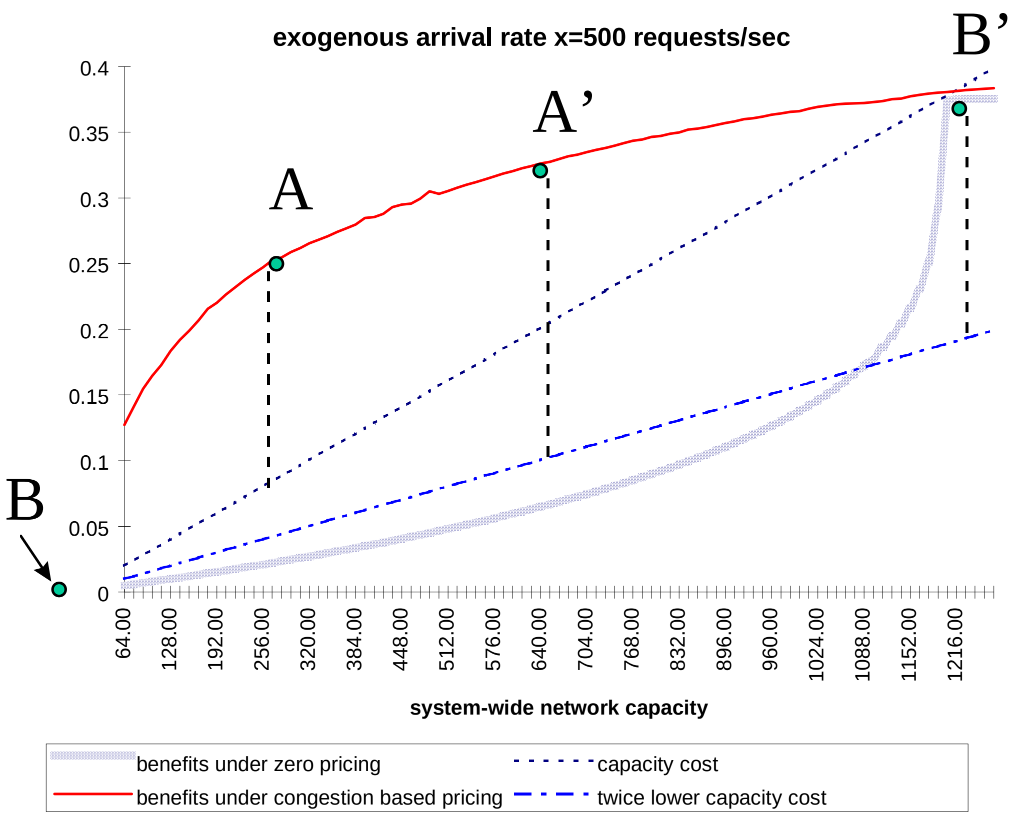


Figure A3.3: Benefits Generated under Both Pricing Schemes for the Exogenous Arrival Rate of 500 Requests/Second:

Appendix A.4 Expansion of Network's Capacity under Different Exogenous Arrival Rates: Profits Generated

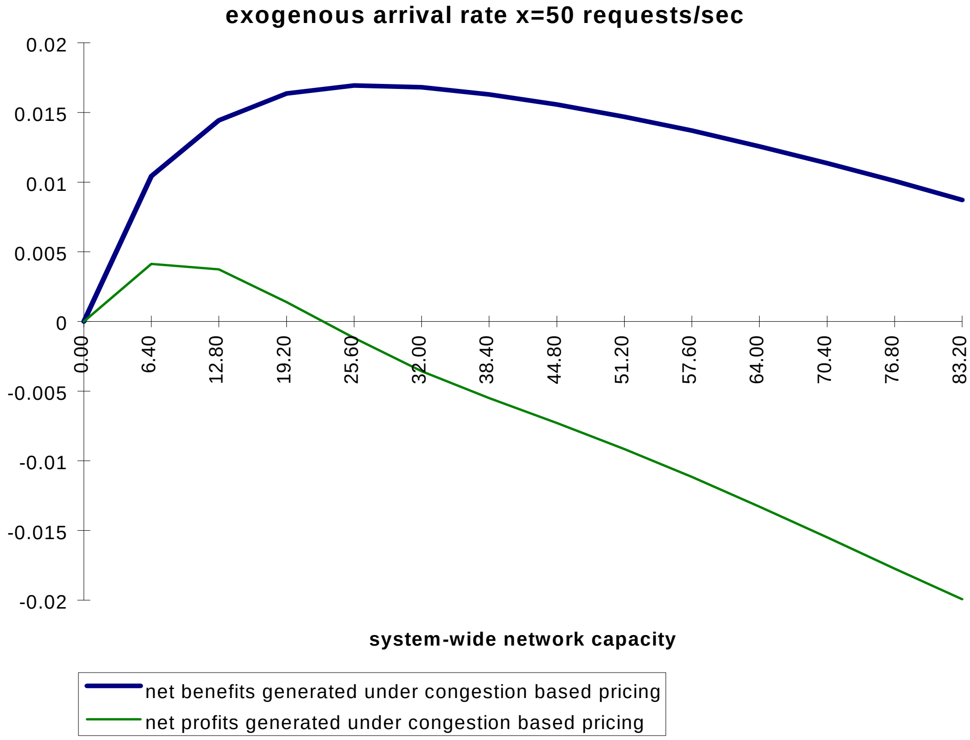


Figure A4.1: Net Benefits and Net Profits under Congestion-based Pricing and Original per Unit Capacity Cost for the Exogenous Arrival Rate of 50 Requests/Second

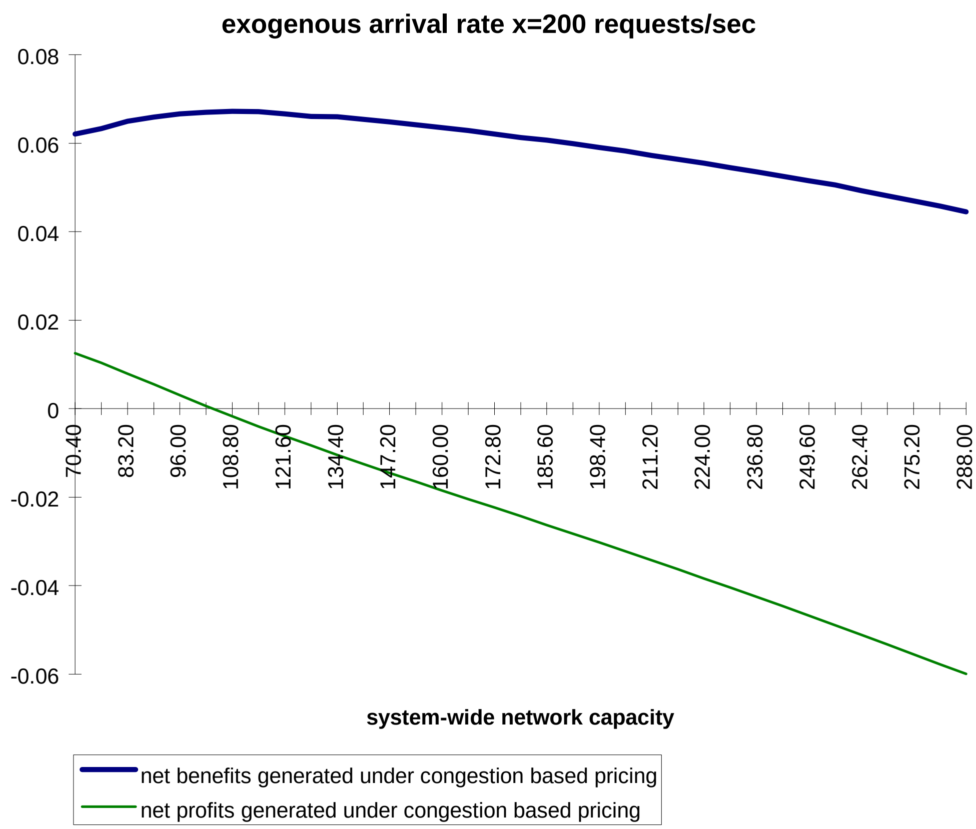


Figure A4.2: Net Benefits and Net Profits under Congestion-based Pricing and Original Per Unit Capacity Cost for the Exogenous Arrival Rate of 200 Requests/Second

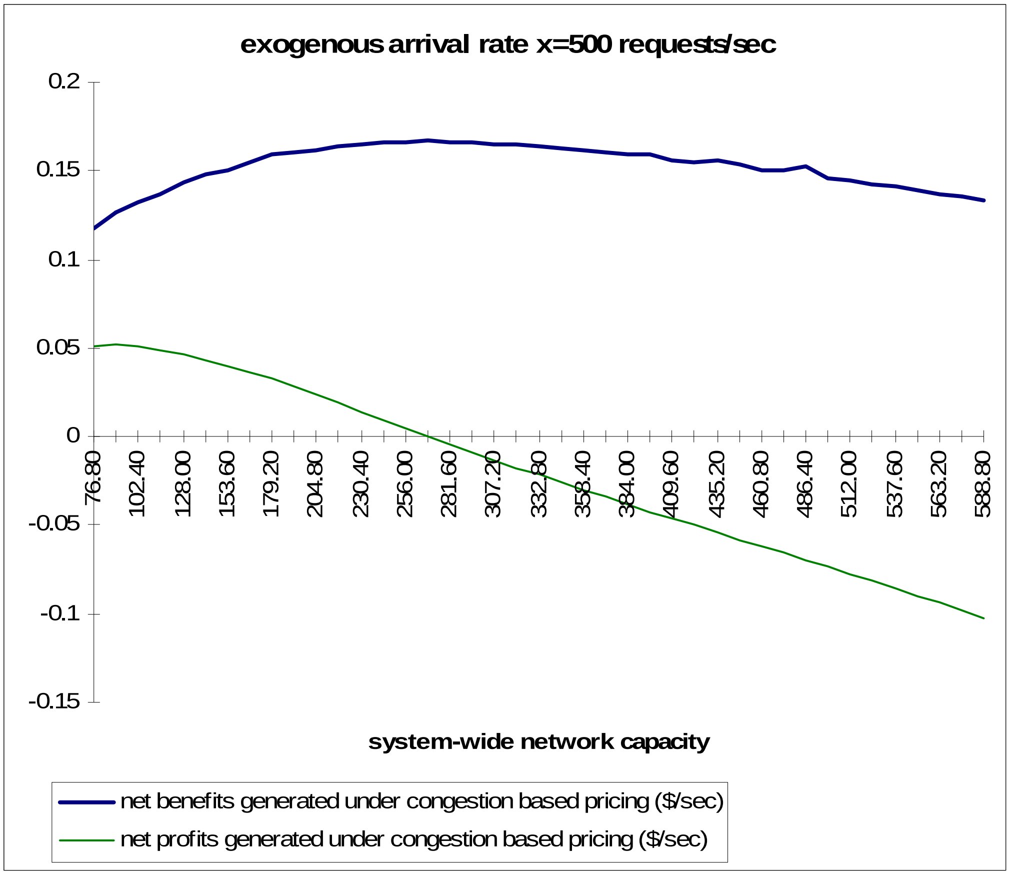


Figure A4.3: Net Benefits and Net Profits under Congestion-based Pricing and Original per Unit Capacity Cost for the Exogenous Arrival Rate of 500 Requests/Second

Appendix A.5 Comparison of Expansion of Network's Capacity under fixed capacity increment and fixed Euclidean step length.

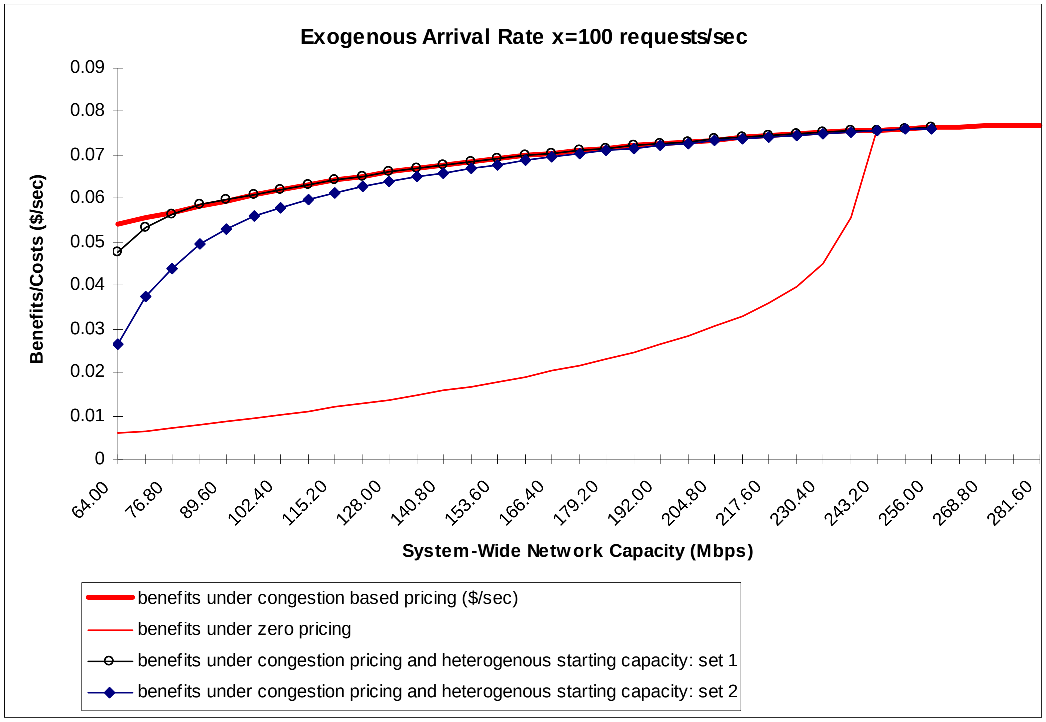
Arrival rate x=100 per sec

	Fixed Capacity increment		**Fixed Euclidean Step Length**	
Extension stage	System Capacity	Net Benefits Generated	System Capacity	Net Benefits Generated
0	6.400000	0.017004	6.4	0.017004
1	12.799994	0.024828	13.16433	0.025391
2	19.199991	0.031284	19.95532	0.031567
3	25.599990	0.036271	26.73633	0.036977
4	31.999985	0.040489	33.28058	0.041483
5	38.400001	0.044649	40.0062	0.045488
6	44.799998	0.047794	46.70478	0.048078
7	51.200016	0.050303	53.366	0.05094
8	57.600008	0.051698	60.02475	0.052972
9	63.999970	0.054074	66.62443	0.054491
10	70.399967	0.055561	72.09554	0.055926
11	76.799961	0.056878	78.69605	0.057291
12	83.199957	0.058107	85.22067	0.058442
13	89.599971	0.059469	91.79234	0.060076

continues through ….

34	224.000002	0.074722	222.8003	0.074638
35	230.399986	0.075086	228.9118	0.075017
36	236.799986	0.075426	235.1197	0.075338
37	243.199992	0.075740	241.2305	0.075635
38	249.600006	0.076005	247.3273	0.075943
39	255.999976	0.076247	253.3022	0.076178

Appendix A.6 Assessing the impact of non-homogenous capacity distribution at the starting point.



The figure above shows the results of an experiment in which the initial distribution of the available service capacity is apportioned in a heterogeneous fashion. Two additional sets of results are described. In the comparison set one, 98 percent of starting capacity is apportioned to four servers that have had the lowest amount of capacity at the end of the original expansion experiment. The remaining two percent of system-wide capacity is equally distributed across the remaining 46 servers. The comparison set two shows the result of a more drastic apportioning of initial capacity in which one server is given 98 percent of he starting capacity and the remaining two percent are apportioned across other 49 servers. While the system-wide capacity increments in each expansion set are again 6.4 Mbps, the aggregate level of initial capacity is set at 64 Mbps, ten times higher than in original experiment. This is done in order to have each increment be a less than a very large percentage of current capacity, further accentuating effects of initial heterogeneous allocation. The results show that heterogeneity of initial distribution has impact early in the expansion process, especially in the second set. Eventually however, the levels of benefits identical to those realized in the original experiment were reached.