MULTIPLE INCENTIVE INTERNET PRICING FOR NRNs: A CASE STUDY

Telecommunications Policy

Apostolopoulos Theodore, Athens University of Economics and Business, Department of Informatics, 76, Patission St., 104 34, Athens, Greece, Tel: +301 8203173, Fax: +301 8203158, E-mail: thodoros@aueb.gr.

Courcoubetis Costas, Athens University of Economics and Business, Department of Informatics, 76, Patission St., 104 34, Athens, Greece, Tel: +301 8203653, Fax: +301 8214903, E-mail: courcou@ics.forth.gr.

Cohen Sandra, Athens University of Economics and Business, Department of Business Administration, 76, Patission St., 104 34, Athens, Greece, Tel: +301 8203229, Fax: +301 8214903, E-mail: sandra@aislab.aueb.gr.

Psiakki Xanthi, University of Patras, Department of Computer Engineering and Informatics, University Campus, Patras, Greece, Tel: +301 7485160, Fax: +301 7710693, E-mail: xpsiaki@gsrt.gr.

MULTIPLE INCENTIVE INTERNET PRICING FOR NRNs: A CASE STUDY

Telecommunications Policy

Abstract

In this paper we develop a simple, flexible and adaptable pricing model for NRNs offering best effort services, taking into account the specific characteristics of these networks which differenciate them from those of the commercial ISPs, such as congestion patterns, cost sharing and government subsidization, governmental policy towards equal access and fairness, and over dimensioning of critical resources. The proposed model captures the above multi-objective policy concepts by using a pricing scheme consisting of two parts, a fixed part, that reflects specific policy parameters, and a variable part, that reflects the usage of the congested links. This later part is used in order to reduce congestion and provide the revenue needed in order to upgrade the congested resources. We discuss the case-study of applying the above model for calculating the tariffs for 1999 for the Greek Academic Research Network (GR-NET).

Acknowledgements

We would like to point out the contribution of the Network Operation Centre of the National Technical University of Athens for providing specific technical information on the traffic profile of GR-NET.

MULTIPLE INCENTIVE INTERNET PRICING FOR NRNs: A CASE STUDY

Telecommunications Policy

1 Introduction

1.1 Academic and research networks

The recent developments in telecommunication services and the Internet are extremely important to the academic and research community. Academic and research institutions by their nature, tend to be extremely heavy Internet users, and request advanced network services. In many cases, such institutions tend to be the most demanding and technologically advanced customers of Internet and connectivity providers, and due to their size and traffic, request special care.

The fact that academic and research institutions are such heavy and demanding network users, led to the development of national academic and research networks (NRNs), which are also interconnected through the Internet. NRNs tend to be over-designed, since the experimental nature of their usage requires much larger capacity than traditional ISPs would provide for the same type of services. This is in most cases justified by the advanced multimedia applications that these institutions use for educational purposes, and by the extra bandwidth needed for experimentation. Furthermore, since such institutions are responsible for educating the future members of the Information Society, higher quality in Internet access and information retrieval is justified by its longer lasting effects. Clearly, end-users hosted in institutions accustomed to higher Internet and information service quality will eventually become the new market segment, which will highly value and hence pay for the development and the deployment of new advanced commercial network services. By creating and operating their own networks, academic and research institutions control their service quality and availability.

Consequently, NRNs are characterised at present by two basic features: very high fixed costs due to their over-designed backbones and their costly international links [1], and the lack of congestion in the largest part of the network. Indeed, congestion exists, and is almost always concentrated on the international links, which tend to be the bottlenecks. A major reason is their relatively high cost (compared to national links) which constraints the amount of purchased bandwidth. Also, most interestingly, the above congestion usually occurs in the transatlantic links, in the direction from the USA to Europe. This is due to the fact that most of the current Internet information is hosted in the USA, and hence most WWW accesses are made to servers located in networks on the other side of the Atlantic.

Designing and implementing a pricing policy for NRNs is not a trivial issue, due to the specific characteristics of these networks and the general practice existing so far. From a technical point of view, until recently, NRNs had a backbone topology consisting of 2 to 3 nodes connected via low speed links, connecting a limited number of institutions. From the pricing point of view, such networks were totally subsidised, and hence they either did not charge their customers (institutions) or they used flat rates based on their access speed. Even in the case of flat rates they compensated for a small percentage of their actual cost.

The situation has changed significantly during the last years, where, as mentioned above, such NRNs have developed large national backbones with multiple nodes connected via high-speed links. Moreover, there has been a need for upgrading and expanding the network in the short-term, as new services and applications become available, and the access links to the Internet become congested. In most cases NRNs are still subsidised up to a certain extend but subsidy can not recover the total cost of upgrade by the time this update is required. Hence the network must find a way to recover the cost of the upgrades by the revenues from their customers, and most probably in advance, prior to the upgrade. This situation motivates the use of charging in order to recover the cost of the network, that is not covered by the subsidies. By applying charges NRNs obtain the flexibility to conduct their own policy towards upgrading the network and its services.

The goal of such a charging scheme is quit complex, since it must fulfil a number of constraints in addition to balancing the budget and producing the additional revenue for upgrades. Some of these constraints deal with the issues of fairness and equal access, which are very sensitive issues in the case of academic institutions. Small institutions are very sensitive in subsidising larger ones, and feel that they should be given incentives to increase their network usage even though their current usage pattern does not justify such a need. Nevertheless, defining the size of an institution is not a trivial matter. There are many alternatives for doing that ranging from the number of actual network end- users in an institution to the number of potential users in an institution, or the volume of data transferred. Another issue is the distance of an institution to the closest network node. Institutions that must pay a larger fee for access to the academic network due to the long distance tariffs that apply in their case feel that they should be compensated for this by paying a lower charge to the NRN for Internet access. Also there may be a number of special policy issues that apply to any particular country, which should be reflected in the structure of the tariffs of the NRNs.

1.2 The use of charging

Charging deals also with another important issue, that is the reduction of congestion and the maximisation of social surplus. Since NRNs are by their definition not profit-seeking institutions, their goal is the maximisation of the benefit that the whole academic and research community obtains by using the network. This is achieved by reducing the social cost due to congestion, and by controlling the sharing of the resources of the network among the contending institutions in an economically optimal fashion. In other words, institutions that have a justifiable need for more bandwidth should get it to the expense of other users whose demand is more elastic. Clearly, such a task can be achieved only through the use of prices, which in this case play the role of a control mechanism for resource sharing. Alternatively, the network could employ some hard-coded rules for sharing the above resources. The latter is rather harder to accept in our case. It is easier for each institution to decide the extend of network usage, given a pre-defined set of tariffs. This task will become even more vital in the future, when the set of network services provided will extend to more services than best-effort.

A last but equally important use of pricing is the provisioning of signals for capacity expansion and upgrades. If the usage of the network is not accounted for in the charge then institutions tend to over-utilise the network, and consequently create congestion. In case the prices posted do not induce a rational use of the network, it is very hard for the network manager to deduce which links of the network should be upgraded. With flat rate, congestion cripples in, and remains there for most of the time in the critical links. Recent experiments, such as the Internet Demand Experiment at U.C. Berkeley [2] suggest that network usage is extremely price elastic, and even a low usage-based component in the charge makes people rational, and reduces the excessive traffic that causes congestion. It is a standard result in network economics that in order to increase the overall welfare of the system, one should compare the congestion price of the various links to the marginal cost for expansion, and then proceed accordingly. Hence, prices serve as signals for capacity expansion, and should be part of any well-designed system.

On the other hand, there is a subtle issue about propagating the right incentives back to the end-users within the institutions by the means of pricing. When an institution is charged as a single entity, the above signals do not necessarily propagate back to the individuals (end-users) that generate the traffic. In other words, there is a cross-subsidisation between end-users in the same institution. We believe that in the long term, rationally managed institutions will measure their traffic patterns and associate internal costs to smaller divisions (schools, departments, user groups, projects, etc.) within the institution. This is the only way to prevent cross-subsidisation, and control unnecessary network usage.

The current practice on charging for Internet services is the use of flat rate charging, especially in the case of ISPs. The basic argument is simplicity and fast market penetration. The counter-argument is quality, which due to congestion is in many cases rather low. The low rates charged force the ISPs to highly overbook their network, and generate substantial revenue from advertisement. This pushes quality even lower, and reduces the value of the service to a large class of users that want to use the network for business purposes. The difference regarding NRNs providing Internet services is the fact that such networks are already subsidized by national funds, and hence can afford to keep the service quality at high levels compared to the ISPs. Economic theory suggests that usage-based charging is crucial for increasing social welfare and reducing unnecessary congestion. Since most NRNs were traditionally over designed, congestion was not an issue in most cases, and flat tariffs were used in order to recover parts of the operating costs. When congestion became a critical issue, especially in the transatlantic links, then usage-based charging became a candidate tool for reducing congestion[3]. In the case of JANET in the UK, see [4], users are being charged for the traffic they generate on the transatlantic link, and are given incentives to use the national caches. The basic objective is both to reduce congestion and to collect the 80% of the cost for the upgrade of the congested link, see [5].

The model of New Zealand (1995), which was abandoned due to the complex calculations required, was based mainly on measuring traffic in both directions through the main Internet Gateway for each participating site. The charging was made according to traffic volume [6], and different price bands were used for different volume ranges. The price was high enough to cover actual costs plus a percentage for upgrades. As demand grew the resulting development funds were used to buy more capacity. To provide predictability the notion of "committed traffic volume" per month was introduced. The charging scale had large steps and the price per Megabyte decreased as the traffic volume increased. Each site made an initial choice of its committed volume, and thus monthly charge. The actual traffic was monitored month by month and reported back to all the sites. If a site's traffic fell into a different price band for more than one month, that site's committed volume was changed to the actual rate. This allowed a site to have a single unusual month, and it gave at least a month's warning of a change in the charge.

The pricing model we present in this paper is different from the one used in JANET by attributing to each institution a fixed part besides the variable part that accounts for the usage of the congested links. This fixed part is different for each institution, and implements the multiple policy objectives set at the national level. It approximates the classical two-part tariffs, see [**7**], where the price of the variable part is the marginal cost for expanding the congested resource and the fixed part is computed according to a more equitable manner than just sharing the cost equally. In essence, this part depends on the value the various institutions obtain from the use of the network, and on the other objectives (fairness, access cost of distant institutions, etc.). An important point is that, in contrast to the standard economic model where the number of customers is unknown or customers might decide not to join if the fixed fee is too high, in our case the customers of the network are fixed. The reason is that institutions can not operate without Internet connectivity, and due to the subsidy, the prices they end up paying are much lower than what they would pay for the same service to an ISP. In the case of the Greek NRN, this resulted by carefully selecting the amount of subsidy that would produce the above effect.

This paper is organized as follows. In Section 2 we discuss a generic model for NRNs that highlights some important aspects that will motivate our charging model, which is introduced in detail is Section 3. Then, in Section 4 we describe the particular implementation of the above pricing model in the case of the Greek NRN. In Section 5 we discuss certain interesting issues that we discovered while trying to implement the charging model, and which had to do with the choice of the various parameters. In this section we also analyze the sensitivity of the resulting charges with respect to the parameters of the model, and compare these charges to the ones that would result form other pricing models such as the one of JANET. Finally, in Section 6 we have the conclusions of this case study.

2 Characteristics of NRNs

In this section we discuss the specific requirements of NRNs in terms of resources and pricing policy. Although there is no well-accepted definition for such networks, the current practice in Europe suggests that these tend to evolve in a similar fashion, and hence face reasonably similar problems. In order to facilitate our discussion, we describe a generic model that captures all essential aspects that are relevant to pricing.

NRNs are entities that provide at a national level Internet services to the academic and research institutions. The current trend is that such a network supports a national backbone, which operates at high speeds (34 to 155Mbps), and provides IP connectivity, see Figure 1. The bearer transport services of the backbone are usually ATM over SDH, which interconnect the IP routers. These services are in generally obtained from some other party, usually a wholesale bandwidth provider. The above infrastructure allows the NRNs to provide other services than IP connectivity, such as guaranteed quality bandwidth pipes in the form of ATM Virtual Paths. Such services can be provided on demand, and be used for multimedia, tele-education, and other research activities. There are plans for such guaranteed bandwidth services to be offered by European NRNs in the near future.

Currently, the type of service typically offered by NRNs is best-effort IP service. By best-effort we mean a service where the various IP flows of information share the available bandwidth with no performance guarantees, and each flow gets the "best that can be offered" by the network on an equal basis. Although there are extensions to the IP protocol that allow for the booking of resources and the use of priorities for individual flows, such mechanisms are presently not implemented.

The national backbone is usually connected to the rest of the Internet through one or more high-speed connections. The trend in European NRNs is to have one connection to the European Internet backbone and another one directly to the Internet backbone in the USA. The reason for supporting more than one such connections is that the largest part of the traffic originates from the USA, where most of the Internet information is hosted, and hence one prefers to have some dedicated bandwidth with this part of the Internet for performance reasons. These connections are constructed at the transport level as ATM Virtual Paths of guaranteed bandwidth, which use part of the physical layer bandwidth. For example, the border router of the NRN can use a virtual path of 20Mbps from a 34Mbps SDH connection in order to connect to the Internet backbone router. In this case the NRN pays both for the cost of the 34Mbps SDH connection to the raw bandwidth provider, and a fee for the 20Mbps Internet access capability to the legal entity that manages the Internet backbone. Given the current tariffs for international connectivity, for NRNs that require international links to connect to the nearest Internet node, the price of these international Internet connections (link + access) are the largest part of the cost of the NRN.

The legal entity responsible for the NRN is managing both the national backbone and the Internet connections. The connected institutions are responsible and pay for their direct connections to the national backbone. In most case, there are some value-added services that the NRN provides besides IP connectivity. These currently include multicasting (MBONE service), proxy services, caches, directory services, traffic monitoring and accounting, and the support of alternative routes for reliability. In the near future, more advanced services will be provided such as IP telephony, PBX interconnection, VPN services with advanced security and quality-of-service features, services for the mobile users, etc.

2.1 Differences with ISPs

There are some interesting features that distinguish NRNs from typical ISPs. We start our discussion by examining some of the cost-related issues.

NRNs are over-designed in order to ensure that the services provided by the network are of high quality and that there is enough unused bandwidth for allowing academic institutions (the users) to experiment with new technologies. Furthermore, such networks tend to be more technologically advanced and use state-of-the are technology to a larger extend than commercial ISPs which are extremely cost-conscious. This drives costs to higher levels in comparison to ISP's investment for the same traffic conveyance.

The topology of the network is predefined since the location of the academic and research institutions is known from the beginning. On the other hand, ISP do not know in advance the location of their users (PoPs). Similarly, in the case of NRNs, the number of customers (institutions) and their capacity requirements are both known and more or less stable for a specific time period. On the contrary, ISPs unquestionably face unstable and to a significant extent unpredictable demand. So, NRNs are more confident for the cost they are going to incur for a specific time period and consequently can more easily schedule cost recovery and plan resource allocation.

Furthermore, the customers (institutions) of NRNs are characterised by homogeneity. One can create a rather typical network usage profile that scales with the type and the size of the institution. For example, universities have a more-or-less homogeneous traffic profile, which is different from research centres. Such differences are explained by the different working habits of the end-users in the institutions, the working hours, etc. Of course, there may be also important singularities that are explained by the specifics of the various groups. The above observations suggest that NRNs should be able to better predict their usage patterns in contrast to typical ISPs. On the other hand, institutions could prove to be more unpredictable than commercial users, which follow standard market trends.

2.2 Pricing policies in NRNs

In the case of pricing policies, there are again striking differences with the case of ISPs. The most important feature of NRNs is that they are not profit seeking organisations in contrast to the ISPs. Their main goal is cost recovery that ensures that there is an adequate surplus for upgrading the network. On the other hand there are many constraints on how to recover the above amount due to policy goals that are established at the national level.

The pricing strategy of NRNs has to be stable for a lengthy time period (mainly a year) and has to be known in advance so that the proper provisions are incorporated into the rigid budgets of the academic institutions. Frequent changes in tariffs (even due to the provision of new services) cannot be dealt with due to the above rigidity in the budget planning of the NRN customers. On the contrary, the prices and the tariffs set by ISPs are flexible as they are driven by the market forces and are frequently updated in order to implement aggressive marketing strategies due to the introduction of new competitive services.

The prices set by NRNs usually reflect the governmental strategy towards the expansion of the Information Society, since the current members of the academic and research institutions will be the main force to drive such an expansion. So, the prices may also help towards improving social welfare by being non-discriminatory and by ensuring equal and fair access to all users. The pricing structure of ISPs does not reflect any such social and/or governmental considerations.

Finally, NRNs are usually to a large extend financed by subsidies granted by the government or by other public institutions (i.e. projects financed by the European Commission). Consequently, the total revenues of NRNs come both from the customers (institutions) and the subsidies. This fact makes it possible for institutions not to pay the true cost of the network and hence the NRN can offer extremely competitive prices to its customers (taking into account the quality of the services). ISPs are very rarely granted subsidies, and hence can not compete with NRNs. This makes the customers of the NRN loyal and predictable.

The differences stated above between commercial ISPs and NRNs reflect a necessity for a customised pricing model for academic networks. This model should be simple enough to implement (as NRN customers are not accustomed to even being charged!), flexible to incorporate different goals that may evolve through time without radical changes in the pricing philosophy, and adaptable to changes in technology and the introduction of new services.

3 The pricing model

This paper has been motivated by the work involved in the construction of the tariffs for the Greek NRN (GR-NET). In this section we describe the underlying approach in its generality, since it can provide some interesting guidelines for creating tariffs for other NRNs. The key idea is to use usage-based prices in order to control the optimal sharing of the scarce resources, which in our case correspond to the bandwidth of the incoming link that connects the NRN to the rest of the Internet (in particular, to the Internet backbone in the US).

The major concerns were three-fold: i) the control of congestion by the propagation of charging signals back to the institutions for the rational use of the expensive resources, ii) the sharing of the fixed cost of the infrastructure on a fair fashion that reflects a number of policy objectives, and iii) the effective use of the subsidies.

An important assumption is that the cost of the network and of the proposed upgrades are known in advance for the charging period (one year) and the major constraint is the recovery of at least the above cost reduced by the government subsidies. In this case, the problem reduces to the calculation of the optimal prices which maximise the economic efficiency of the system under certain constrains. These constraints are to balance the budget by following certain fairness and policy constraints for sharing the common network cost among the different institutions.

3.1 The utility function

A measure of the economic efficiency is social welfare, that is defined as the value generated to the users by the use of the network, minus the corresponding cost for supplying the above network service. Institutions derive value from the network that is related to the amount and the type of the information they receive. Since the network transport service is not responsible for the content of the information being transported, a simple assumption for defining the value of such a service to a user is to assume that the above information is of the same "average" type. This allows the value of the service to be directly measured in terms of the quantity of the information being transported. This motivates a rather simple definition of the utility function of a network user. We model the above utility function as an increasing function of the input flow of information, averaged over some time period. Clearly, there is a different utility for receiving information during different periods of the day, and hence one should be careful in defining the appropriate time intervals over which the model remains accurate. In the approach described in this paper we did not address the above time-of-day aspects, and we used instead the simplest model of an "average" utility function, that depends on the average flow of input data during the complete charging period. Extending this model to the multiple time-of-day intervals is straightforward.

In the case that a user receives input flows from different information sources, the utility could be written as the sum of the utilities for each of the incoming flows. Consider the model of the network in Figure 1.

(Insert Figure 1 here)

3.2 Maximising the economic efficiency

In Figure 1, we have n users connected to the national backbone, which is connected to the rest of the Internet (denoted as user a). The corresponding access link bandwidths are B_i for each user i (two-way links). The details of the backbone network are omitted for the moment, and as we will argue in the sequel, these are not important for determining the usage part of the charge.

Since the utility of each user is an increasing function of the input flows from the other users connected to the above network (including the rest of the Internet), the socially optimal decision is to allow each user to request the maximum amount of information from the information sources available. The problem is that capacity is finite and hence after a certain point, increasing the information flow provided to one of the users implies the decrease of the similar flows to the other users. The goal of the social planer is to maximise the overall value that the network creates to all of its users by allocating the capacity among the various flows in a way that reflects the utility of the users to the particular flows. Hence, the social welfare maximisation problem becomes

$$\max_{\{x_{ji}\}} \sum_{i,j \neq i} U_{ij}(x_{ji}) \quad \text{such that the set of flows } \{x_{ij}\} \text{ is feasible.} \tag{3.1}$$

In the above equation, $U_{ij}(x_{ji})$ is the utility rate obtained by user i when he receives an average rate of information x_{ji} from user j. Current practice suggests that the utility functions are increasing and concave, and it is reasonable to assume that after a certain point they become almost flat. This models the fact that there is a saturation phenomenon, since more information from the same information source is not of any actual value to the user receiving the information. Clearly, larger institutions have larger saturation points.

In most pragmatic NRN situations, the optimal flow assignment in (3.1) will leave all links of the access part of the network and of the backbone partially empty, while filling completely the Internet link in the incoming direction. This is due to the fact that the utility of the institutions for information that originates from local sites of the NRN is easily saturated for small flow values relative to the available bandwidth. On the other hand, most of the utility is derived form information that comes from the Internet node, and the bandwidth of the incoming Internet link does not suffice to saturate the above information needs. Actual measurements suggest that the internal flows of the NRN are rather negligible compared to the Internet in-flows. Hence the equation to solve simplifies and becomes

$$\max_{\{x_{ai}\}} \sum_i U_{ia}(x_{ai}) \text{ such that } \sum_i x_{ai} \le B_a \text{, and } x_{ai} \le B_i . \tag{3.2}$$

This is since we can always construct the complete solution of (3.1) by superimposing the internal NRN flows at their maximum saturating values with the solution of (3.2), without violating any constraints, due to the large amounts of excess capacity available internally to the NRN.

We can simplify (3.2) even more by observing that we can "internalise" the access constraints into the definition of the utility function of the institutions. We redefine the utility function $U_{ia}(x)$ to be constant for $x \ge B_i$, and omit the corresponding network constraint. Intuitively, it does not pay to increase the rate to an institution at a level higher than his access speed. Using these redefined utility functions, the optimisation problem in (3.2) is equivalent to

$$\max_{\{x_{ai}\}} \sum_i U_i(x_{ai}) \text{ such that } \sum_i x_{ai} \le B_a , \tag{3.3}$$

where for simplicity we have dropped the subscript a from the utility functions.

3.3 The pricing mechanism

One way to solve the above optimisation problem is by the use of prices which correspond to the Lagrange multiplier of the bandwidth constraint. The social planner must find a price p^* per unit of flow on the Internet link, which, if posted, the rational flow choices on behalf of the users will generate the optimal amount of flow for the complete system. In other words, we need to find a

p^* such that, if x_{ai} is the solution of $\max_x U_i(x) - p^* x$ (the rational flow choice of user i) for each i, then the set $\{x_{ai}\}$ satisfies the constraint $\sum_i x_{ai} \le B_a$ with equality.

A way to construct such an optimal price is by a simple market mechanism where the social planner posts a price p, users solve their local optimisation problems and generate some amount of input flow form the Internet. If the total demand is in excess of the available bandwidth B_a, the price is increased, else it is decreased. A practical way to implement that is, after posting the price, to observe the congestion level at the bottleneck link. If the congestion persists, this translates in demand being more than capacity. In best-effort Internet links, congestion can be measured by either measuring packet losses or by observing the link utilisation level. Congestion levels become unacceptable when utilisation exceeds a certain level, which is easy to establish in practice.

An important corollary is that capacity should be expanded if its marginal cost is less than the optimal price that solves (3.3). This is rather trivial to see, since the optimal price equals the marginal utility of the users of the network. This suggests the following steps: use as a price the current value of the marginal cost for expanding the capacity of the congested link. If the resulting demand is less than the existing capacity, then there is no need to perform upgrades, and one should rather decrease the capacity of the link to save money. If this is not the case, then one should buy more capacity in order to increase social efficiency.

It is not always straightforward to translate the above simple economic theory results into a pricing model for the NRNs. One main reason is convergence. In order to converge to the right price utilities must remain constant while price convergence takes place. Hence, if prices are not allowed to change rapidly (in the case of NRNs, prices can change in a yearly basis), utilities must remain constant over time periods of many years. Clearly, this is not the case in the current Internet, where demand for new services expands at exponential rates. On the other hand, capacity must also remain constant during the same time. But current practice suggests that capacity is expanded on an almost yearly basis. It is plausible that, due to the dramatic drop in capacity prices, the same price that was charged per kilobyte the year before, must be charged per megabyte the year after. Despite all these we should keep in mind that charging for volume transported on the congested link is definitely in the right direction for improving social welfare.

The above discussion suggests a more pragmatic approach for choosing the above price. The operator of the NRN must decide what percentage of the revenue should be based on the usage, and then choose a price that will produce the above usage-based revenue. The only concern should be that the above price should be low enough as no to permit any irrational reactions.

A simple relation that describes this approach is

$$C = W - U = F + G, \tag{3.4}$$

where C is the total cost to be recovered through charging, W is the actual cost of the network and its upgrades, U is the amount of the subsidy, G is the amount to be recovered through usage-based pricing, and F is the fixed part of the charge to be recovered. This suggests that we first determine C and then decide how to split it into a fixed and a variable part (function of usage). The next issue is to determine how to split the fixed part F of the charge among the institutions. By doing so, each institution i will be charged by a two-part tariff

$$z_i(V_i) = F_i + pV_i, \tag{3.5}$$

where p is the price per unit volume determined before, $\sum_i F_i = F$, and V_i is the volume produced on the congested link [8] that is attributed to the above institution. It is clear that we can not use a standard Coase type tariff where $F_i = F / n$ (n is the number of participants), since the above amount will be excessive for the smaller institutions and some of them might choose to leave the consortium. Furthermore, such a split does no convey any policy objectives and is rather unfair since different institutions have different marginal utilities of income. We have adopted the following simple but general methodology to share F among the institutions.

We assumed that institutions can be characterised by a finite number of attributes $A = \{1, \ldots, M\}$, each taking values in the interval $[0, L]$. We denote by $y_{ij} \in [0, L]$ the value of the j th attribute of institution i. Such attributes can model the size of the institution, its network access cost for reaching the nearest node of the NRN, the access speed to the NRN, etc. A reasonable way to split the fixed charge is

$$F_i = \frac{\sum_j w_j y_{ij}}{\sum_k \sum_j w_j y_{kj}} F , \qquad (3.6)$$

where w_j is a non-negative weight attributed to attribute j, with $\sum_j w_j = 1$. The definition of the set of attributes and their corresponding weights expresses the policy of the NRN towards the fair allocation of the costs and the propagation of other incentives besides usage. For instance, the incentives for reducing access speed or for choosing a particular access technology can be captured by such a formula.

In the following section we describe a case study where we use the above charging model in order to derive the tariffs used in the GR-NET. There are some interesting observations regarding the application of the above ideas in an actual real life situation, including the definition of the appropriate usage price p, and the definition of the weights $\{w_i\}$.

4 The case study of GR-NET

In this part we describe the technical features of GR-NET as well as the calculation of the network cost and the methodology followed for the implementation of the pragmatic charging model introduced in the previous section. The charging scheme was designed for 1999, using the information available from 1998. We provide a brief description of the state of GR-NET when this study started

(Insert Figure 2 here)

4.1 GR-NET

GR-NET is the Greek National Research & Technology Network, providing Internet services to the academic and research institutions in Greece. In 1998, the GR-NET topology consisted of network nodes in 6 major Greek cities, that is, Athens, Thessaloniki, Patras, Ioannina, Xanthi and Heraklion. At the beginning of 1999 the backbone operated at 2/34 Mbps interconnecting 121.000 end-users (hosted in institutions) out of the 289.000 total Internet end- users in Greece. GR-NET was connected to the European Internet Research backbone (TEN-155) network via a 10Mbps ATM link connecting Athens to Milan [9]. Traffic congestion was constantly observed in the inbound part of the international link during the whole period of 1998. The total incoming international traffic increased significantly in 1998, i.e. started from 6,000,000 Mbits in January and resulted in 11,000,000 Mbits in December 1998 [10]. During 1999 (after this study took place), the international link was upgraded to 34Mbps using subsidies from the European project TEN155/QUANTUM, and it will be further upgraded to 155Mbps in June 2000.

4.2 The charging scheme

The calculation of the total cost of GR-NET for 1999 was based on a projection of the costs (both network and operating costs) expected to incur during that period. This estimation was made on the basis of signed contracts and agreements for line rentals, depreciation of existing equipment and scheduled upgrading, pragmatic needs for technical and administrative assistance and a reasonable set of operating cost elements. Hence, the cost that is going to be covered from charging is based on budgeted figures since the actual costs will be only known at the end of the corresponding period.

As mentioned in section 3, the charge z_i of institution i consists of two parts, a fixed charge F_i and a variable part pV_i, where V_i is the volume carried on behalf of institution i on the inbound part of the international connection. The approach taken consists of the following steps:

1. determine the price p of the usage component of the charge,
2. estimate the amount of traffic on which the usage charge will be applied, i.e., the projected amount $V = \sum_i V_i$ for 1999,
3. determine the leftover part of the cost that must be recovered through the fixed parts of the charges.

This later part is $F = W - U - p\sum_i V_i$, where W is the total projected cost of the network for 1999, U is the total amount of subsidy to be subtracted from the network cost for the same period of time, and $p\sum_i V_i$ is the amount recovered through usage charges.

4.2.1 Calculation of p

In our case study approach, the value of p is defined as the average incremental cost per Mbps of the expansion of the international link. This is considered to be the best approximation of the marginal cost mentioned in section 3, since this international link can be expanded by fixed amounts, determined by the raw bandwidth providers.

The above definition implies that if the international link is expanded from the current configuration a to configuration b for some extra cost $W_b - W_a$, and the effective transfer capability for the given time period (one year) in Mbytes is increased from Q_a to Q_b, then

$$p = \frac{W_b - W_a}{Q_b - Q_a} = \frac{\Delta W}{\Delta Q}. \quad (4.1)$$

In the case of Internet traffic, the effective transfer capability of a link is not equal to the bandwidth of the link times the time period over which data transfer takes place. The reason is that the Internet protocols, mainly TCP, can not load fully the link due to the way these operate (the slow start mechanism of TCP). Experience shows that TCP operates reasonably well when the load (average data rate divided by the link capacity) does not exceed the 70% of the capacity of the link.

What is not clear is whether the right time period to define the data transfer capability is the period during which the link is congested or the whole duration of the charging period. The right interpretation of (3.3) implies that the congestion price should be posted only when demand might exceed capacity, and hence the cost of the upgrade should be shared among the customers that compete for bandwidth during the peak period of operation. In this case, the term ΔQ in (4.1) should be computed over the interval of time during which congested operation occurs.

Interestingly enough, our observations of the traffic patterns indicate that the incoming international link remains fully loaded for most of the time [10]. The above observation implies that the term ΔQ in (4.1) should be computed over the complete period of operations of the link (a year). This leads to the following formula for computing the price p in (4.1)

$$p = \frac{W_b - W_a}{0.7(B_b - B_a)T} = \frac{\Delta W}{0.7\Delta BT} \quad , \tag{4.2}$$

where the above quantity is selected to be the smallest among all possible expansions B_b. This is because the incremental cost of the expansion ΔW depends on the network design and the intended policy for the upgrade of the congested link, and there may be many such alternatives. In our case, the most cost-effective upgrade solution was to upgrade the bandwidth by 24Mbps, for a cost of 171 million GRDs (approximately 0.52 MEuro). Hence

$$p = \frac{171mGRD}{24Mbps \times 0.7 \times 31.536.000\sec} = 0.32GRD/\ Mbit.$$

For reference reasons it is worth mentioning that for the solution selected the incremental cost of the expansion refers to the 19% of the total cost of the international link.

4.2.2 *Projection of the carried Volume V for 1999*

Since the institutions could not accurately predict in advance the incoming international traffic they were going to incur, an estimation of the total incoming international traffic for 1999 had to be done on the basis of projections. These projections were mainly based on the assumption that the expansion of the bandwidth of the international link in 1999 would result to an increase in the incoming international traffic proportional to the upgrade of the bandwidth minus 20%. The usage charge pV_i for institution i for the year 1999 was calculated by assuming that the percentage of the traffic on the international link corresponding to the above institution would be the same as in 1998.

The amount that would be covered by the variable part of the charging scheme is the product $p \sum_i Vi$ where $\sum_i Vi$ is the volume produced on the congested international link by all users.

4.2.3 *Calculation of the subsidies*

GR-NET is granted subsidies by both EU and the Greek State. GR-NET's policy is to transfer part of the yearly subsidies to the institutions, and to keep the rest for coping with unpredictable situations. The selection of the corresponding percentage is based on considerations about the competitive state of the market. The goal is to keep charges slightly lower than what these would be for competitive services of similar quality, speed and reliance offered by regular ISPs. This is reasonable in order to keep all participating institutions in the GR-NET consortium. The subsidy transferred to the institutions in 1999 was the 23% of the total yearly cost of GR-NET. In order to come up with this figure, we had to go through a number of iterations where we compared the resulting charges to the Internet access prices offered in the Greek market. In many cases this was not a simple task since most institutions, due to their size, have requirements that can not be met by most of the Greek ISPs, and hence the existing tariffs could not be used as a point of reference. We could also make the remark that in order to obtain such large customers, most ISPs should redesign their network and hence obtain larger economies of scale. Furthermore, the quality offered by today's ISPs is by far inferior to the one demanded by an NRN, adds even more difficulties in determining the above competitive prices.

4.2.4 Calculation of the fixed part of the charges

The fixed part of the charge is shared among the various institutions by taking into account several attributes that describe the profile of each institution, as described in equation (3.6). There are five such attributes that GR-NET has defined in order to exercise its policy towards the fair allocation of costs and the propagation of important incentives to the participating institutions. These attributes are: A1- access speed to GR-NET, A2 - size of the user, A3 - GR-NET shareholder category, A4 - local or long-distance access to GR-NET, and A5 - access port technology.

Each of the above attributes takes several values that depend on the particular attribute. For instance, the attribute A1 – access speed to GR-NET takes the possible values 64Kbps, 128 Kbps, 256Kbps, 512Kbps, 2048 Kbps, etc. To each such attribute-specific value (the "relative" value of the attribute) we associate an absolute value in a scale from 1 to 10 (1 = less important, 10 = more important), except for attribute A2, which has a different range of values. These absolute values correspond to the coefficients y_{ij} in equation (3.6). The attribute-specific as well as the absolute values of the four attributes (excluding A2 - size of the user) is presented in table 1. The mapping of relative to absolute attribute values is part of the policy decisions of GR-NET.

(Insert Table 1 here)

In order to analyse and rate A2-size we measured each institution's total incoming traffic X_i during the charging period, and we assigned a value y_{i2} according to the formula :

$$y_{i2} = 10 \frac{X_i}{\max_k X_k}. \qquad (4.3)$$

We have observed that in many cases institutions with high values of the A1-access speed attribute were having relatively low values for the A2- user size attribute. This is illustrated in Table 2, where we have selected seven indicative institutions.

(Insert Table 2 here)

5 Results and Discussion

Since the calculation of the F_is depends on the choice of the weights w_j for the various attributes, an important issue was the investigation of the sensitivity of the resulting F_is with respect to the choice of the above weights. Since policy objectives are encoded in the choices of the weights, an important related question is the robustness of the charges for small deviations of the policy objectives. An obvious thing to avoid is small changes in policy objectives (which are not so rigid anyway) to imply great variations in the charges.

The sensitivity analysis was based on a set of five scenarios that were considered adequate to span some important policy objectives, in which emphasis was given to either the network access and user size related attributes (A1 and A2) or the other policy – incentive related attributes (A3 to A5).

The first three scenarios emphasized the importance of user size and access related attributes by allocating to them a cumulative weight of 80% and by leaving the policy incentive related attributes with a total weight of 20%. In particular, in scenario 1 w_1 is set equal to 80% and w_2 is set equal to 0%, thus in this scenario the attribute of the user size is totally ignored and all the emphasis is placed on the type of the access link. Scenario 3 is exactly the opposite (w_1 = 0% and

w_2 = 80%), while scenario 2 lies between the previous two extreme cases (w_1 = 40% and w_2 = 40%).

The last two scenarios are more hypothetical as they place a rather small, less than 50%, weight to the user size and access related attributes, and investigates the effects of the policy-incentive attributes. The weights of each attribute for any of the five scenarios is presented analytically in Table 3.

(Insert Table 3 here)

The F_i s calculated for the various scenarios show considerable differences. In order to visualise the impact of the different weight choices we have selected seven indicative institutions, which correspond to the profiles introduced in Table 2. The following two Figures compare the total charges, i.e., $F_i + pV_i$, resulting from the above scenarios for each of the seven institution types.

(Insert Figure 3 here)

(Insert Figure 4 here)

The above Figures suggest the following:

1) Low access and low volume institutions ranging from 64 Kbps to 256 Kbps do not present considerable differences among different scenarios, while high access (2Mbps) but low volume users present a slight decrease in their corresponding total charge with scenarios 3 and 4 compared to scenarios 1, 2 and 5. Therefore, for this institution profile the selection of the scenario does not have a considerable effect on the resulting total charge.

2) High volume institutions incur considerable differences among different scenarios, namely, prices from scenarios 2 and 3 are higher compared to the respective prices from scenarios 1, 4 and 5. Hence large universities and research institutions are very sensitive to scenario selection since they may end up paying for a considerable part of the total cost.

The decision has been to select scenario 2. The reason is that it places a reasonably strong emphasis on institution size as this is reflected through its overall network usage (40% weight on

A2-usage size attribute) and on access capability (40% weight on A1-access speed). It places a lesser emphasis on issues such the subsidisation of the access line cost (only 8% on A4-access distance) and the fact that the institution did not choose to be a shareholder of GR-NET. Also the cost of the access port did not influence much (only 4% weight).

This decision is in contrast to current practice where institutions are charged solely on access speed on a flat basis (no usage based component). We did not consider this scenario (scenario 1) since we believe that access capability is not playing a crucial role in the economics of our resource sharing problem. Charging according to access speed for institution i would make sense if changing the above parameter would influence the sharing of the bandwidth on the congested link. From our measurements it became obvious that during the peak hours when the international link was congested, the average bandwidth that any institution received was well below its access capability. Indeed, access to GR-NET has been over-designed due to subsidisation from other sources that paid for the above access services. The reason that we took access into account (40% weight) is merely for historic reasons, and to discourage unjustified requests for high-speed access. Interestingly enough, the above scenario produces charges that are close to those advertised by commercial ISPs for the low to medium size institutions, taking into consideration quality and bandwidth parameters (see Figure 5). We must stress that we could not infer the ISP charges for the large institutions in the actual market, since these institutions are not typical ISP customers for which tariffs were available. Since most of them would generate high values of congestion on the typical ISP backbone, an ISP should expand his network in order to provide high quality service to such customers. On the other hand, without the above re-dimensioning of the ISPs network, the perceived performance would be greatly reduced from the performance offered currently in GR-NET, and hence, although the charge from scenario 2 is much higher than any probable ISP charge, the difference of service quality would discourage such institutions to switch providers. Of course, this is not the case for smaller institutions that are more unstable in that respect. But scenario 2 results in competitive prices for these institutions.

Looking back at the rest of the scenarios, we can make the following remarks. The selection of scenario 1 and 5 would be most preferable for institution profile D, that is the big universities and research institutions that consume the greatest amount of bandwidth and that are indirectly responsible for the size of the network. The policy objectives of GR-NET are not to favour this type of minority users against the majority of smaller institutions. On the other hand the selection of scenarios 3 and 4 would favour high access but low volume users (user profile E). GR-NET would rather penalise this type of users that give wrong signals for the dimensioning of the network since they ask for a high access link that is eventually poorly utilised.

(Insert Figure 5 here)

5.1 Comparison with the Janet charging model

It is interesting to compare the charges resulting from scenario 2 with these that would be produced if we would use the charging model of JANET, the UK research network, which has also implemented a usage based pricing model. In contrast to our model, the pricing scheme of JANET comprises only a variable part, corresponding to the 80% of the cost of the upgrade for the international link. There is no fixed part in the above pricing scheme. Specifically, the JANET charging model

a) is based on measuring the incoming traffic on the international link for each institution,

b) the value of p in JANET is defined as 1.25GRD/Mbit which is higher than the 0.32 GRD/Mbit of GR-NET.

(Insert Figure 6 here)

By comparing the total charges that the seven types of institutions would pay under scenario 2 of GR-NET and the charging scheme of JANET, some interesting remarks can be drawn. In order to do that, we have computed for each GR-NET institution the equivalent charge per Mbit that would produce the same total charge that results from scenario 2, assuming that each such institution produces its projected traffic for 1999, and is charged based only on volume. In other words, we divided the total charge of each user under scenario 2 by the projected volume of the incoming international traffic for 1999. The results are shown in Figure 6.

- High volume users pay less under the GR-NET pricing scheme, since the total charge is not solely based on the traffic volume.
- Low volume and low access users are charged more, but not excessively, under the GR-NET pricing scheme since they pay a fixed charge regardless of their traffic.
- High access but low volume users are charged well above the price that they would pay under the JANET formula, since their fixed part of the invoice is the bulk of their total charge.

6 Conclusion

In this paper we presented a pricing scheme based on a flexible structure customised for academic-research networks. The model consists of a predefined basic structure (fixed part + traffic sensitive part) taking into account a range of parameters, but focusing mainly on the usage of the network. The model is easily adaptable to a rich set of policy goals and network features (access technologies, identity of congested resources, etc.), by adding or eliminating attributes and by using different weighting factors. In addition it is independent of the underlying physical network and can be easily adapted to various technologies such as Frame Relay, ISDN, ATM, etc. The pricing model used in this paper can be easily extended to handle time-of-day pricing where different prices are used during the peak and the off-peak period, in order to provide incentives to reduce the load during the peak hours of operation. Traffic statistics show that there are still periods when the incoming link is not congested, and we would like to give incentives for scheduling delay-insensitive traffic during these periods.

An area of on-going research is the charging of differential quality of service. Modern Internet technology such as IPv6 and the Differentiated Services Architecture provide the capability for allocating resources to traffic aggregates and hence guaranteeing performance in addition to providing the traditional best-effort services. Important research is conducted in constructing charging models for Service Level Agreements that specify the service contracts between the user and the network. Our plan is to be able to implement such charging schemes when GR-NET will deploy these new service architectures. Related issues are the capability to auction bandwidth and hence provide the incentives to institutions to acquire the network resources they truly need, and release resources they do not need. Minimising such asocial waste is an important task that we need to handle.

We should finally comment on the way one should handle the subsidy of the network cost. In this paper we have directly subtracted the amount of subsidy from the cost of GR-NET, and we have divided the remaining cost among the participants. An other option for handling the above subsidy is to decide on the percentage of the subsidy that will be transferred directly to the institutions in order to make GR-NET prices competitive, since such subsidy would be given to an institution only if GR-NET is chosen as its ISP. This could be more important in the cases of small institutions since these are more price than quality sensitive. Also, having this subsidy being added as part of the institution's budget for telecommunications, it provides the right incentives for not over-spending and for finding ways to reduce overall network cost, in order to use the amount saved for other purposes.

ATTRIBUTE VALUES FOR THE CALCULATION OF Fi	
A1 – Access speed to GR-NET	**Value**
64kbps	1
128 kbps	1
256 kbps	2
512 kbps	3
2048 kbps	5
4 Mbps	6
6 Mbps	7
8 Mbps	8
10 Mbps	9
12 Mbps	10
A3- GR-NET Shareholder category	**Value**
No shareholder	8
Shareholder	4
A4-Geografic distribution of the physical link	**Value**
Local link	8
Local distance link	2
A5-Access port technology	**Value**
Framerelay	2
PCM/Hellascom	6
ATM	2
ISDN	2

Table 1: Attribute values for the calculation of F_i

Institution	Access link	User size (yearly total traffic)	Value
Institution A	64Kbps	VLV (57.500 Mbytes)	0,18
Institution B	128Kbps	VLV (57.500 Mbytes)	0,18
Institution C	256Kbps	VLV (57.500 Mbytes)	0,18
Institution D	2Mbps	VHV (2.545.000 Mbytes)	7,95
Institution E	2Mbpss	VLV (57.500 Mbytes)	0,18
Institution F	2Mbps	MV (903.100 Mbytes)	2,82
Institution G	2Mbps+64kbps backup	HV (1.888.335 Mbytes)	5,90

VLV=Very Low Volume, VHV=Very High Volume, MV=Medium Volume, HV=High volume

Table 2: Indicative values of attribute A2 – user size

Telecommunications Policy

Scenario	Access link	User size	Shareholder category	Local/long distance physical link	Access port technology
Scenario 1	80%	0%	8%	8%	4%
Scenario 2	40%	40%	8%	8%	4%
Scenario 3	0%	80%	8%	8%	4%
Scenario 4	20%	20%	15%	25%	20%
Scenario 5	30%	20%	30%	10%	10%

Table 3: Attribute weights corresponding to alternative scenarios for the calculation of F_i

REFERENCES

[1] MacKie-Mason, J. K. and Varian H. R., Economic FAQs about the Internet. *Journal of Economic Perspectives,* Summer 1994, Volume 8, Number 3, p.p. 75-96.

[2] Edell, J. R., McKeown N. and Varaiya P. P.**,** *Billing Users and Pricing for TCP*. INDEX Project Report #98-004P, April 15, 1995 (Synopsis updated April 27, 1999). Available at: http://www.index.berkeley.edu/public/index.phtml

[3] Kelly, P. F., Charging and Accounting for Bursty Connections. Presented at MIT Workshop on Internet Economics March 1995. Available at http://www.press.umich.edu/jep/works/KellyCharg.html

[4] JANET traffic accounting site. URL: http://bill.ja.net

[5] "Usage charging for transatlantic network traffic", University of Cambridge, UK. URL: http://www.cam.ac.uk/CS/NetDiv/uscharge.html

[6] Brownlee, N., New Zealand Experiences with Network Traffic Charging. *ConneXions,* December 1994, vol. 8, no. 12 , http://www.press.umich.edu/jep/works/BrownNewZe.html.

[7] MacKie-Mason, J. K. and Varian H. R., Pricing the Internet. *Public Access to the Internet,* editors Kajin B. and Keller J., Prentice Hall, Englewood Cliffs, New Jersey, 1994.

[8] Kelly, F. P., Charging and rate control for elastic traffic. *European Transactions on Telecommunications,* 1997, volume 8, p.p 33-37, updated version available at: http://www.statslab.cam.ac.uk/~frank/elastic.html

[9] http://www.dante.net/quantum.html.

[10] Monthly traffic statistics for GR-NET, *Network Operation Centre, National Technical University of Athens, Issues January 1998- December 1998*. Available at: http://www.grnet.gr/reports

MULTIPLE INCENTIVE INTERNET PRICING FOR NRNs: A CASE STUDY

Telecommunications Policy

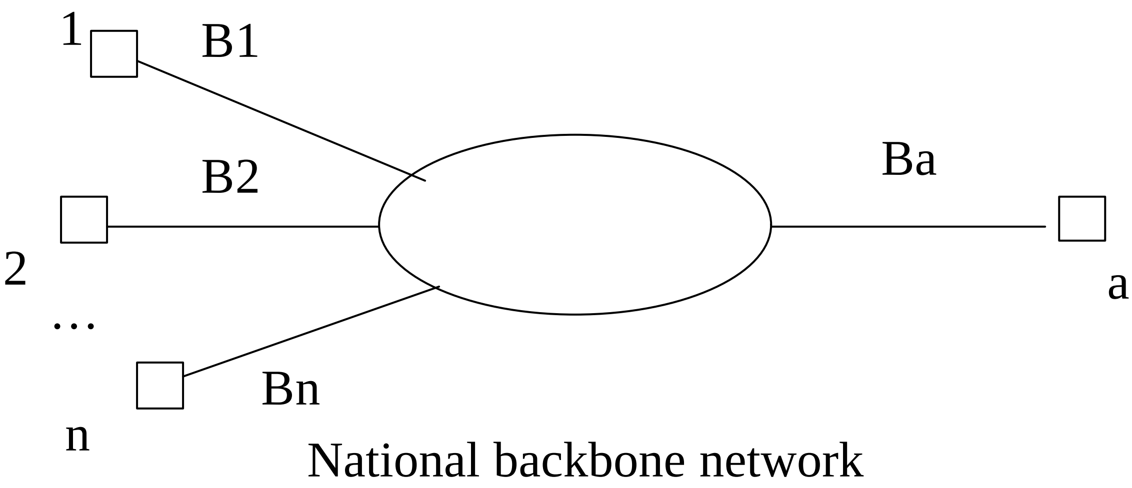

Figure 1

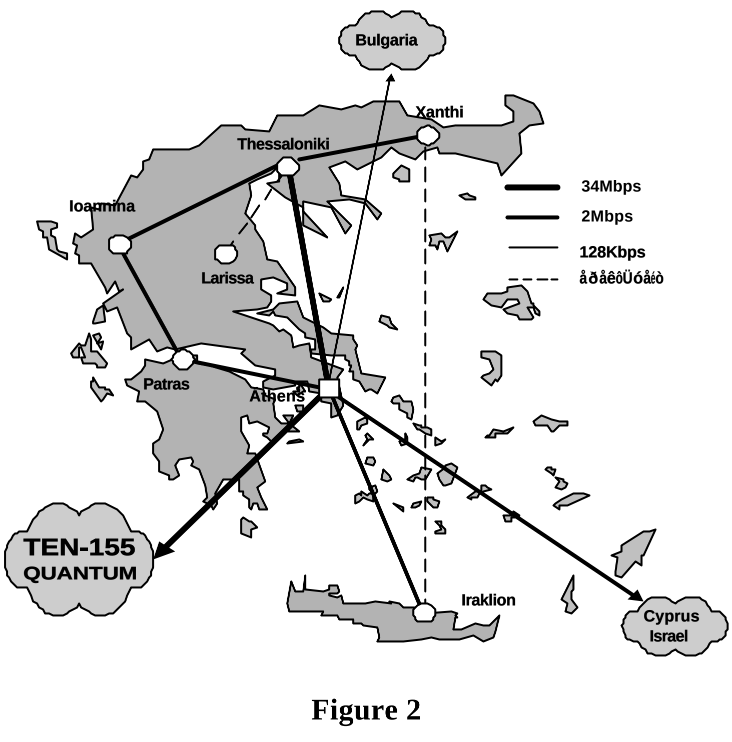

Figure 2

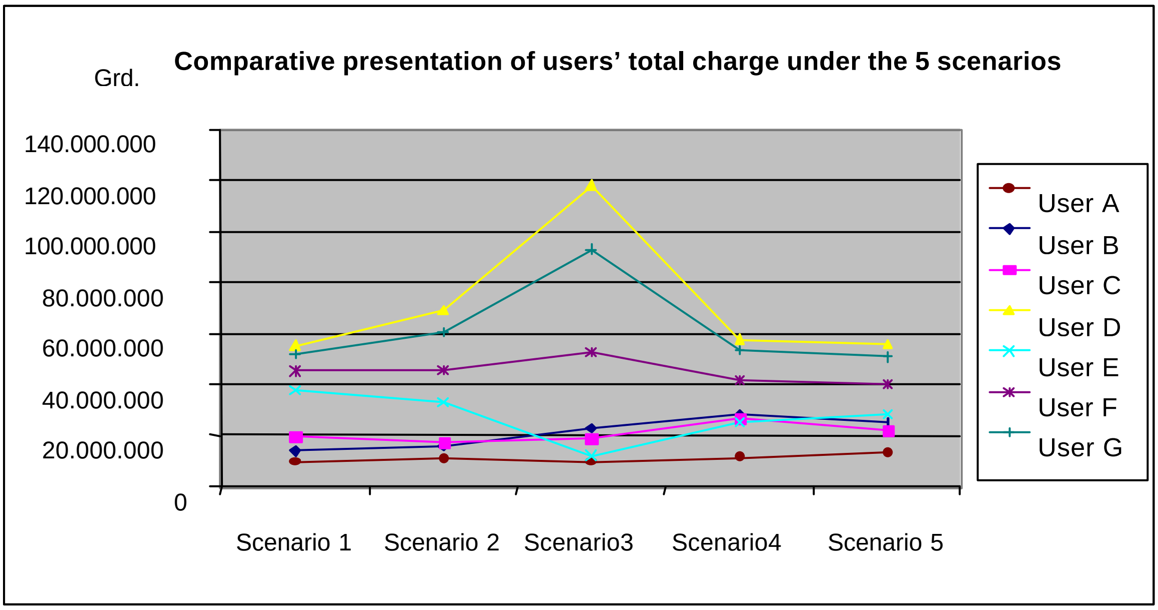

Figure 3

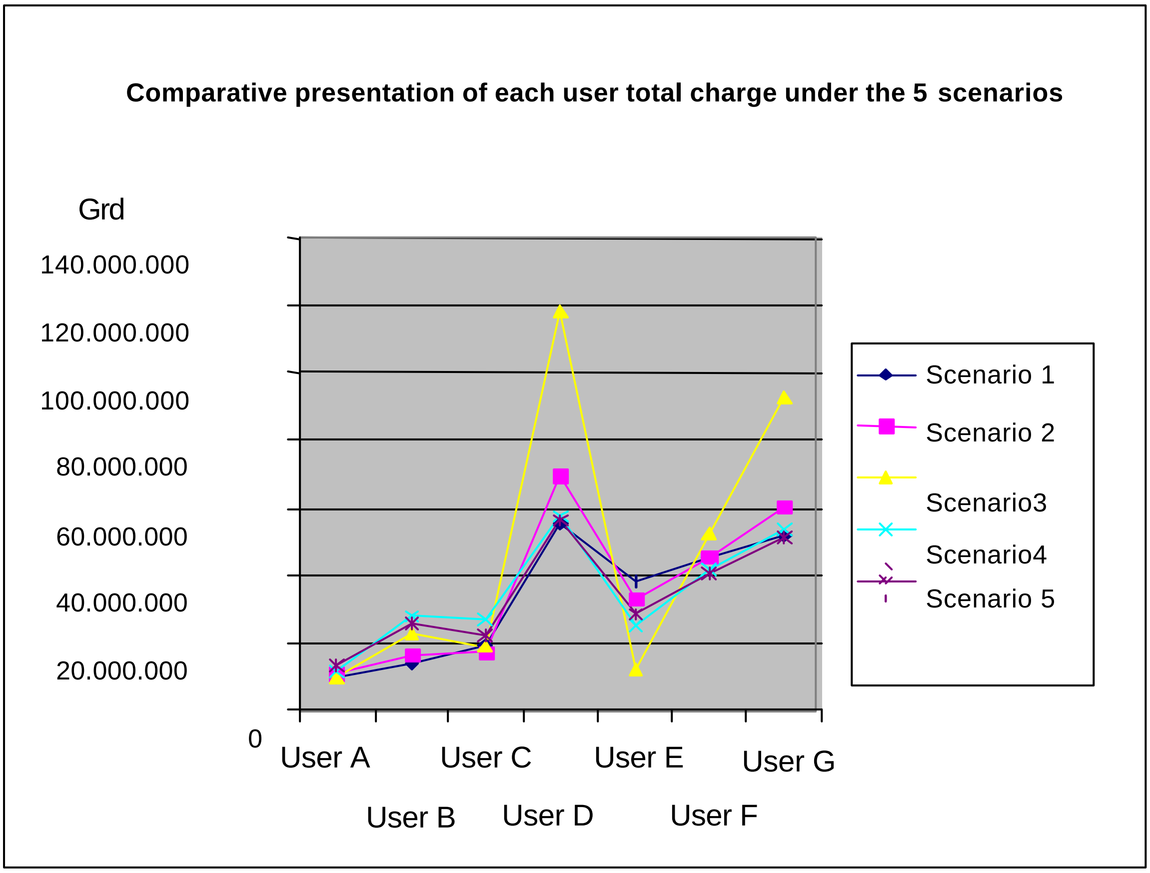

Figure 4

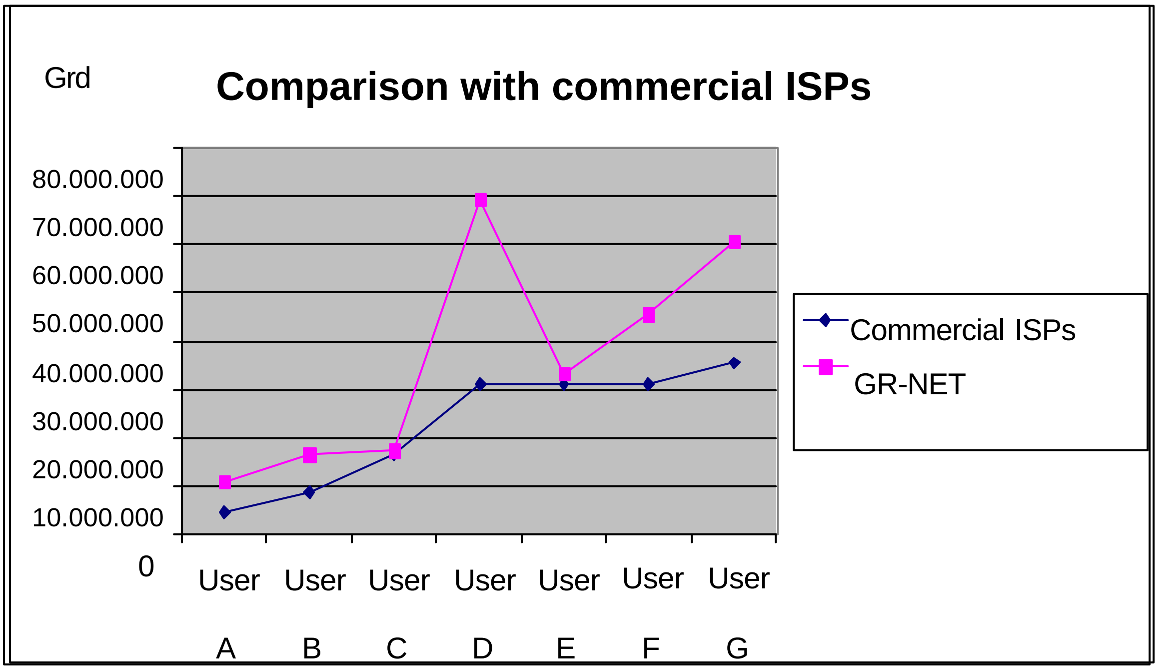

Figure 5

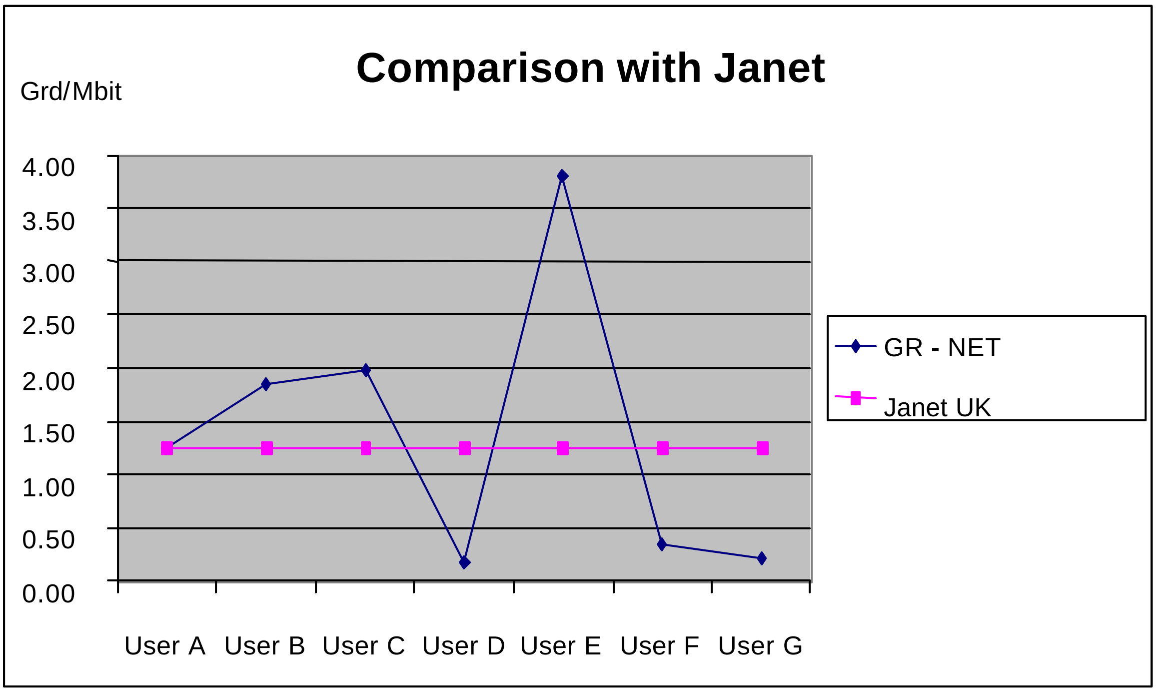

Figure 6