A Knowledge Plane as a Pricing Mechanism for Aggregate, User-Centric Utility Maximization

Vladimir Marbukh
National Institute of Standards and Technology
100 Bureau Drive, Stop 8920, Gaithersburg,
MD 20899-8920, USA
Tel: 1 301 975 2235
E-mail: marbukh@nist.gov

ABSTRACT

This paper proposes pricing user centric requirements as a potential role for the Knowledge Plane. Assuming elastic users capable of modifying their behavior in response to the pricing signals, this approach may result in optimal resource allocation without necessity for the users to acquire detailed information on the network state as well as advanced knowledge of the user requirements by the network.

Categories and Subject Descriptors

C.4 **[Performance of Systems]**: – *modeling techniques, performance attributes, reliability, availability, and serviceability.*

General Terms

Algorithms, Management, Performance, Design, Theory.

Keywords

Network, elastic users, utility, performance, pricing.

1. INTRODUCTION

Since typically network performance is characterized by multiple competing criteria, the network management requires resolving the corresponding trade-offs. The problems include finding the Pareto optimal frontier in the space of these criteria and selecting the desired operating point on this frontier. Mathematically, finding the Pareto optimal frontier can be framed as a constrained optimization problem. Solution to this problem can be expressed in terms of the Lagrange multipliers, which characterize the sensitivity or "the price" of one competing criterion or constraint with respect to another. Once these prices are flooded throughout the network, the agents have the necessary and sufficient information for making the optimal decisions on the amount and mixture of the resources to request. These decisions may relate to flow control, routing, etc. Finding the optimal operating point on the corresponding Pareto optimal frontier in a distributed environment is equivalent to aggregate utility maximization, where the aggregate utility is the sum of individual utilities of all the elastic users.

The concept of an elastic user capable of adjusting its rate and QoS requirements in response to the external stimuli by maximizing its individual net utility has been articulated and the corresponding aggregate utility maximization framework has been proposed in [1]. The pricing based approach to the distributed maximization of the aggregate utility has been developed in [2] for a case when utilities are expressed in terms of the single network resource (bandwidth). This approach assumes that users are aware of the mapping between the user centric requirements and network resources allocated to the user. Users may become aware of this mapping through probing or with help of the Knowledge Plane (*KP*) proposed in [3].

Since architecture of the current Internet assumes that all the intelligence is concentrated on the edges, *KP* may help users (applications) to make informed decisions on the amount of the network resources to request. Due to distributed nature of the network it is desirable that signals provided by *KP* contain all sufficient and only necessary information needed for the users to make the optimal decisions. Such necessary and sufficient information can be provided to the users in a form of pricing of the user level requirements. Assuming that elastic users modify their behavior according to the received pricing signals it is possible for the system to achieve the optimal balance among competing user requirements.

This short paper briefly discusses the possibility of pricing user centric requirements for the purpose of aggregate utility maximization. The approach is illustrated on an example of balancing throughput with reliability for each user as well as across different users in a case of unreliable network [4]-[6]. Providing reliable networking services under adverse conditions requires additional resources, e.g., link bandwidth, storage memory, transmission power, etc., as compared to the best effort services. Adverse conditions may be a result of limited reliability of the network elements, such as network links and/or nodes, as well as result of malicious attempts to disrupt the network services by adversary/adversaries. Additional network resources are required to accommodate redundancy needed to counteract the adverse conditions. Conventional research on protection and restoration is usually concentrated on evaluation of various survivability metrics of certain protection and restoration schemes leaving aside the problem of balancing survivability and economic efficiency for each user as well as across various users. Proposed approach has potential of achieving optimal balance among these competing user requirements.

2. AGGREGATE UTILITY MAXIMATION

Consider a network of unidirectional links l shared by users $s \in S$. Let R_s be the set of feasible routes for a user s. User s has utility $u_s(\mu)$ of transmitting at rate μ, where function $u_s(\mu)$ is strictly concave, increasing and differentiable in

$\mu \geq 0$. The purpose of the flow control and routing strategies is to maximize the social welfare

$$W(x) = \sum_{s \in S} u_s \left(\sum_{r \in R_s} x_{sr} \right) - \sum_{l \in L} f_l \left(\sum_{s \in S} \sum_{r: l \in r} x_{sr} \right) \tag{1}$$

$$\max_{x_{sr} \geq 0} W(x) \tag{2}$$

where penalty function $f_l(y_l)$ is associated with congestion on a link l and assumed to be non-negative, strictly increasing and convex. For appropriately selected penalty functions solution to optimization problem (1)-(2) satisfies the capacity constraints.

Due to our assumptions optimization problem (1)-(2) has unique solution, which can be effectively calculated by a distributed algorithm based on the congestion pricing. Define the cost of a route r to be the sum of the costs of links comprising this route:

$$d_r = \sum\nolimits_{l \in r} d_l \tag{3}$$

where the cost of a link l is

$$d_l(y) = df_l(y)/dy \tag{4}$$

Given link costs $d = (d_l)$, introduce the minimum cost routes $r \in R_s^*$ for a user $s \in S$: $\breve{d}_s \overset{def}{=} \max_{r \in R_s^*} d_r = \min_{r \in R_s} d_r$, where the set of feasible routes for user s is R_s. It can be shown that the necessary and sufficient conditions for the flow vector $x^* = (x_r^*)$ to solve optimization problem (1)-(2) are as follows: (a) load is carried only on the minimum cost routes and (b) user $s \in S$ transmission rate minimizes the individual net utility:

$$\max_{\mu \geq 0} \{ u_s(\mu) - \breve{d}_s \mu \}. \tag{5}$$

Note that optimization problem (1)-(2) can be solved by proposed in [2] distributed algorithm.

Making network management decisions by maximizing the aggregate, user-centric utility has been proposed in [1]. Under certain assumptions this idea can be formalized as a following optimization problem:

$$W = \sum_{s \in S} u_s (\pi_{s1}, .., \pi_{sk}) - \sum_{l \in L} f_l \left(\sum_{s \in S} \sum_{r: l \in r} x_{sr} \right) \tag{6}$$

where user centric parameters are

$$\pi_{si} = \varphi_{si} (x_{sr}, r \in R_s) \tag{7}$$

Substituting (7) into (6) we obtain the following optimization problem:

$$\max_{x_{sr} \geq 0} \left\{ \sum_{s \in S} u_s [\varphi_{si} (x_{sr})] - \sum_{l \in L} f_l \left(\sum_{s \in S} \sum_{r: l \in r} x_{sr} \right) \right\} \tag{8}$$

If user $s \in S$ is aware of mapping (7), this user may request bandwidth x_r on feasible routes $r \in R_s$ by solving his individual optimization problem

$$\max_{x_{sr} \geq 0} \left\{ u_s [\varphi_{si} (x_{sr})] - \sum_{r \in R_s} d_r x_{sr} \right\} \tag{9}$$

Solution to the individual optimization problem (9) also solves the global optimization problem (6) if the prices d_r are "right".

Alternative approach to global optimization (6) in a case when users are unaware of mapping (7) may be based on direct pricing of user-centric parameters $\pi_s = (\pi_{si}, i = 1,..,k)$. If *KP* is aware of mapping (7) and is capable of implementing user-level requirements $\pi_s = (\pi_{si}, i = 1,..,k)$:

$$x_{sr} = \psi_{sr}(\pi_{si}), \ r \in R_s, \tag{10}$$

then, the global optimization problem (6) becomes

$$\max_{\pi_{sr} \geq 0} \left\{ \sum_{s \in S} u_s (\pi_{si}) - \sum_{l \in L} f_l \left(\sum_{s \in S} \sum_{r: l \in r} \psi_{sr} (\pi_{si}) \right) \right\} \tag{11}$$

Given implementation (10), user $s \in S$ individual optimization problem is

$$\max_{\pi_{sr} \geq 0} \left\{ \sum_{s \in S} u_s (\pi_{si}) - \sum_{i} g_{si} \pi_{si} \right\} \tag{12}$$

where the prices of fulfilling user level requirements are

$$g_{si} = \sum_{l} \left[d_l \left(\sum_{s' \in S} \sum_{r': l \in r} \psi_{s'r'} \right) \frac{\partial \psi_{s'i'}}{\partial \pi_{si}} \right] \tag{13}$$

Prices (13), of course, depend on the implementation (10). Given user-centric requirements $\pi_s = (\pi_{si}, i = 1,..,k)$, the optimal implementation (10) is determined by solution to optimization problem

$$\min_{x_{sr} \geq 0} \sum_{l \in L} f_l \left(\sum_{s \in S} \sum_{r: l \in r} x_{sr} \right) \tag{14}$$

subject to constraints (7). Thus maximization (8) can be decomposed into individual optimization by the elastic users (12), based on the pricing and implementation (13)-(14) by the *KP*.

In general, pricing user level requirements and pricing bandwidth result in different total costs to the users: $\sum\nolimits_i g_{si} \pi_{si} \neq \sum\nolimits_r d_{sr} x_{sr}$. This discrepancy is undesirable due to unfairness, price instability in competitive environment, etc. A natural approach to keeping pricing user level requirements consistent with cost of the resources is adding fixed component to the price of user level requirements:

$$P_s^{opt} = P_s^0 + \sum_{i} g_{si}^* \pi_{si} \tag{15}$$

where asterisk indicates the operating point, and

$$P_s^0 = \sum_r d_{sr}^* x_{sr}^* - \sum_i g_{si}^* \pi_{si}^* \tag{16}$$

3. EXAMPLE: PRICE OF SURVIVABILITY

Consider a network where link capacities are subject to fluctuations due to fading, mobility, link failures, etc, and these fluctuations occur on much faster time scale than the network management decisions are made. We model this situation by assuming that each user s may be assigned only "average" bandwidth $\widetilde{x}_{sl}$ on link $l \in r \subseteq R_s$. However at any given moment this user transmission rate on link $l \in r \subseteq R_s$ cannot exceed

$$x_{sl} = (1 - \xi_l)\widetilde{x}_{sl} \tag{17}$$

where $\xi_l \leq 1$ are jointly statistically independent random variables for all links in the network, and $E[\xi_l] = 0$. The maximum "expected" aggregate rate for user s subject to assigned link bandwidths $\widetilde{x}_{sl}$ can be easily calculated. However, since actual link bandwidths are random variables, the actual maximum aggregate rate x_s is also a random variable.

It is known that under optimal coding scheme user s data rate μ_s can be recovered at the destination if μ_s does not exceed the aggregate transmission rate from this user x_s. The following exponent characterizes the reliability of providing data rate μ_s for user s:

$$\gamma_s = -\log P\{x_s \geq \mu_s\} \tag{18}$$

Users can combat the adverse conditions by a combination of route diversity and redundant transmissions creating trade-offs among needs for data transmission and reliability for each user as well as across different users.

Assuming that user $s \in S$ utility u_s is a function of its data (payload) transmission rate μ_s and reliability parameter (18):. $u_s = u_s(\mu_s, \gamma_s)$, it is natural to balance these competing requirements by maximizing the following criterion:

$$W = \sum_{s \in S} u_s(\mu_s, \gamma_s) - \sum_{l \in L} f_l\left(\sum_{r:l \in r} x_r\right) \tag{19}$$

where the penalty associated with link loads approaching link capacities is represented by penalty functions $f_l(y)$. The main difficulty in maximization (19) lies in difficulty of calculating of the parameter (18) for large-scale networks. However, in some cases, including a case of routes without overlapping links and a case of small variations in the link capacities, calculations (18) significantly simplify.

Consider a particular case of balancing throughput with survivability in a case of link failures assuming that each user $s \in S$ is capable of redundant transmitting it data on all possible subsets $R \subseteq R_s$ of the set of feasible routes R_s, e.g.,

$$\widetilde{x}_r = \begin{cases} \mu & if \quad r \in R \\ 0 & if \quad r \notin R' \end{cases} \tag{20}$$

where μ is the corresponding data rate. Given user $s \in S$, consider reliability exponents

$$\gamma_R = -\log P\left\{\sum_{r \in R} \min_{l \in r}(1 - \xi_l) \geq 1\right\} \tag{21}$$

and the corresponding price rates

$$d_R = \sum\nolimits_{r \in R} d_r \tag{22}$$

for all possible $R \subseteq R_s$: $\mathrm{T}_s = \{(\gamma_R, d_R) : \forall R \subseteq R_s\}$, where the route costs d_r are given by (3)-(4).

Let $\mathrm{T}_s^* \subseteq \mathrm{T}_s$ be the corresponding Pareto optimal (east-south frontier) service offering. The *KP* takes the link costs d_r and link failure probabilities as the input, and presents each user $s \in S$ with "service offering" T_s^*. Each user $s \in S$ selects a set of routes $R = R_s^* \subseteq R_s$ and data rate $\mu = \mu_s^*$ in an attempt to maximize its net utility:

$$\max_{\mu \geq 0} \max_{(\gamma, d) \in \mathrm{T}_s^*} \{u_s(\mu, \gamma) - d\mu\} \tag{23}$$

Then, the *KP* implements the requested service, recalculates link costs, and the process repeats.

4. CONCLUSION

This paper proposes pricing user centric requirements as a potential role for a Knowledge Plane. Future efforts should be directed towards developing adaptive algorithms leading to the optimal resource allocation resulting in maximization of the aggregate utility.

5. REFERENCES

[1] S. Shenker, "Fundamental design issues for the future Internet," *IEEE JSAC*, 13 (1995) 1176-1188.

[2] F.P. Kelly, A.K. Maulloo, and D.H.K. Tan, "The rate control for communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, pp. 237-252, vol. 409, 1998.

[3] D. Clark, C. Partridge, C. Ramming, and J. Wroclawski, "A Knowledge Plane for the Internet," SIGCOMM'03, August 25-29, 2003, Karlsruhe, Germany.

[4] J. MacKie-Mason, "What does economics have to do with survivability?" InfoSurv. Tutorial, June 1997. available at http://www-personal.umich.edu/~jmm/presentations/darpa-econ-tutorial.pdf.

[5] V. Marbukh, "Towards market approach to providing survivable services," *Proc. Conf. on Info. Sciences and Systems*, Princeton Univ., Mar. 17-19, 2004

[6] V. Marbukh, "On aggregate Utility Maximization Based Network Management: Challenges and Possible Approaches," *IEEE International Communications Conference (ICC 2004)*, Paris, France, 2004.