Sensible decisions based on QoS*

Erol Gelenbe

Department of Electrical and Electronic Engineering, Imperial College of Science,
Technology and Medicine, Exhibition Road, London, SW7 2AZ, UK (e-mail: erol@cs.ucf.edu)

Abstract. Network Quality of Service (QoS) criteria of interest include conventional metrics such as throughput, delay, loss, and jitter, as well as new QoS criteria based on power utilization, reliability and security. Variable and adaptive routing have again become of interest in networking because of the increasing importance of mobile ad-hoc networks. In this paper we develop a probability model of adaptive routing algorithms which use the expected QoS to select paths in the network. Our objective is not to analyze QoS, but rather to design randomized routing policies which can improve QoS. We define QoS metrics as non-negative random variables associated with network paths which satisfy a sub-additivity condition along each path. We define the QoS of a path, under some routing policy, as the expected value of a non-decreasing measurable function of the QoS metric. We discuss *sensitive* and *insensitive* QoS metrics, the latter being dependent on the routing policy which is used. We describe routing policies simply as probabilistic choices among all possible paths from some source to some given destination. *Incremental* routing policies are defined as those which can be derived from independent decisions taken at certain points (or nodes) along paths. *Sensible* routing policies are then introduced: they take decisions based simply on the QoS of each available path. Sensible policies, which make decisions based on the QoS of the paths, are introduced. We prove that the routing probability of a sensible policy can always be uniquely obtained. A hierarchy of *m-sensible* probabilistic routing policies is then introduced. A $0 - sensible$ policy is simply a random choice of routes with equal probability, while a $1 - sensible$ policy selects a path with a probability which is inversely proportional to the (expected) QoS of the path. We prove that an $m + 1 - sensible$ policy provides better QoS on the average than an $m - sensible$ policy, if the QoS metric is insensitive. We also show that under certain conditions, the same result also holds for sensitive QoS metrics.

Keywords: Mathematics of Networks, Quality of Service, Routing

1 Introduction

Quality of Service (QoS) has now become a central issue in network design, and there is a vast and significant literature on the problem of *estimating certain specific quality of service parameters* (e.g. loss or delay) for *given traffic characteristics* and a given network topology (Gelenbe et al. 1996; Srinivasan et al. 1996). Typically such work has considered single buffer models (finite or infinite), or models of cascaded nodes with or without interfering traffic. There has also been much work on schemes for obtaining better QoS through routing (Chen and Nahrstedt 1998; Yin-Dar Lin et al. 2002; Nelakuditi and 2002), on scheduling techniques in routers to achieve desired QoS objectives (Hao et al. 2002), as well as on the analysis of QoS resulting from the detailed behavior of protocols such as TCP/IP.

The mixed wired and wireless network topologies that are becoming common, including fixed and ad-hoc connections, create the need to rationally exploit dynamically variable routing as a function of network conditions, since the a pplications that use such networks have QoS requirements such as delay, loss or jitter, as well as reliability and low power utilization.

In recent years we have conducted research on *routing algorithms* with two quite different applications in mind:

- *routing in robotic navigation*, and
- routing algorithms in *wired networks and wireless ad-hoc networks*.

The applied research we are currently conducting basically addresses *adaptive routing algorithms in a discrete structure*. In robotic navigation, the discrete structure is a discrete grid of points in a terrain (i.e. a "terrain database"). In a communication network the discrete structure is obviously the set of nodes in a Mobile Ad-hoc Network or in a wired network, and routing decisions are based on optimizing QoS objectives.

The "QoS metrics" used to route robots in a terrain are based on minimizing travel times to destination, minimizing the probability of being destroyed by adversaries (which may be fixed or mobile), minimizing power utilization due to motion, and the obvious need to avoid obstacles. In an ad-hoc network, QoS requirements include the minimization of packet loss, the minimization of end-to-end delay, the minimization of overhead, the minimization of power consumption for routing purposes, or a combination of some of these criteria.

Motivated by our prior work on adaptive network routing algorithms (Gelenbe and Lent 2002; Gelenbe et al. 1999a, 2000, 2001, 2002), in this paper we investigate some basic mathematical problems concerning QoS driven routing. The aim of this work is not to analyze QoS, but rather to show that certain randomized routine policies can improve QoS.

We define QoS metrics as non-negative random variables associated with network paths which satisfy a sub-additivity condition along each path. We then describe routing policies simply as probabilistic choices among all possible paths from some source to some destination. Incremental routing policies are defined as

those which can be derived from independent decisions along each sub-path. We define the QoS of a path, under some routing policy, as the expected value of a measurable function of the QoS metric. We discuss *sensitive* and *insensitive* QoS metrics, the latter being dependent on the routing policy which is used. *Sensible* routing policies are then introduced; these policies take decisions based simply on the QoS of each allowable path.

Finally, a hierarchy of *m-sensible* probabilistic routing algorithms is introduced. The 0-sensible ruting policy is simply a random choice of routes with equal probability, while the 1-sensible policy uses the relative QoS for each alternate route to make select a path. An m-sensible policy uses the $m - th$ power of the QoS for each alternate path, rather than just the $1st$ power. Thus it simply uses the same information in a different manner. It is particularly interesting that we can prove that an $m + 1$-sensible policy provides better resulting average QoS than an m-sensible policy, provided that the QoS metric is insensitive. We also prove that under certain sufficient conditions, the same result holds for senstive QoS metrics.

1.1 Quality of Service (QoS) metrics

A QoS metric relates to some specific *data unit*, the most obvious example being a *packet*. However more broadly, a data unit may be a significant sequence of packets which belong to the same connection. A QoS metric q can be illustrated by the following examples:

- q_D may be the delay D experienced by a packet as it traverses some path in the network, or
- It may be the binary variable $q_r = 1[the - path - is - connected]$, or
- $q_{LR} = L/n$ may be the number of packets lost L divided by the number of packets sent n (i.e. the packet loss rate) for a sequence of packets, or
- q may be the average jitter experienced by n successive packets:

$$q_J = \frac{1}{n - 1} \sum_{l=2}^{n} |(R_l - R_{l-1}) - (S_l - S_{l-1})|, \tag{1}$$

  where S_l is the date at which packet l was sent from the source, and R_l is the time at which packet l arrives to its destination, etc., or
- q may be the number of hops a packet has to traverse, or
- q may be the power expended by the network nodes to service and forward a packet as it travels through a path in the network, or
- q may be the "effective delay" obtained by composing some of these values, such as:

$$q = (1 - q_{LR})q_D + q_{LR}(T_o + q_D),$$
$$= q_D + q_{LR}T_o.$$

where T_o is the (large) time-out delay which triggers a packet's retransmission.

Some paths may not be connected because physical links may not exist. Also since we are dealing with potentially unreliable wired networks as well as ad-hoc networks, some or all links in the path may be connected only with some probability. If a link along a path V is disconnected, we can still compute quality of service metrics, for instance we may then have $q_D(V) = +\infty$, $q_{LR}(V) = 1$, and $q_J(V) = +\infty$.

1.2 Routing policies

Let the nodes in a network be denoted by a fine set of natural numbers $\{0, 1, 2, \dots N\}$.

Definition A *path* in the network starting at node i and ending at node v_d is denoted by $V_i = (i, \dots, v_d)$. It is a sequence of nodes such that the first node is i, the last node is v_d, and no node in the sequence V_i appears more than once. We associate QoS metrics with paths in the network. Let $FV_i(v_d) = \{V_i^1, V_i^2, \dots, V_i^m\}$ be the set of all distinct, but not necessarily disjoint, paths from node i to node v_d in the network.

Definition A *routing policy* for source-destination pair (i, v_d) is a *probability distribution* $\pi^{FV_i(v_d)}$ on the set $FV_i(v_d)$, that selects path $V_i^j \in FV_i(v_d)$ with probability $\pi^{FV_i(v_d)}(V_i^j)$ for each individual data unit which is sent from node i to node v_d.

For any $V_i^j \in FV_i(v_d)$, we may write $V_i^j = (i, \dots, l, n, \dots, v_d)$ as a concatenation of a prefix path and a suffix path: $V_i^j = P_i^p . S_n^s$ where $P_i^p = (i, \dots, l)$, $S_n^s = (n, \dots, v_d)$. Consider now the sets of paths from i to l, $FV_i(l)$ and from n to v_d, $FV_n(v_d)$. We now define a class of routing policies which can be incrementally constructed from partial policies on sub-paths.

Definition A *routing policy* $\pi^{FV_i(v_d)}$ for source-destination pair (i, v_d) is said to be *incremental*, if there exist policies, $\pi^{FV_i(l)}$ on $FV_i(l)$ and $\pi^{FV_n(v_d))}$ on $FV_n(v_d)$ such that for any $V_i^j \in FV_i$ with $V_i^j = P_i^p . S_n^s$:

$$\pi^{FV_i(v_d)}(V_i^j) = \pi^{FV_i(l)}(P_i^p) . \pi^{FV_n(v_d)}(S_n^s). \tag{2}$$

Note that an incremental policy is simply equivalent to one where all decisions can be taken independently at each node. We can also consider policies which are incremental only at certain selected nodes.

Definition Consider a subset of nodes Φ appearing in the set of paths FV_i. We will say that the *routing policy* $\pi^{FV_i(v_d)}$ for source-destination pair (i, v_d) is *incremental on the set* Φ if for each $n \in \Phi$, and $V_i^j = (i, \dots, l, n, \dots, v_d)$, there exist policies $\pi^{FV_i(l)}$ on $FV_i(l)$ and $\pi^{FV_n(v_d))}$ on $FV_n(v_d)$, such that for $V_i^j = P_i^p . S_n^s$:

$$\pi^{FV_i(v_d)}(V_i^j) = \pi^{FV_i(l)}(P_i^p) . \pi^{FV_n(v_d)}(S_n^s). \tag{3}$$

Thus a policy is incremental if it is incremental *for all nodes in the set of paths* FV_i.

Whenever needed, Π will denote the routing policy for the network as a whole, i.e. the set of rules that assign unique paths for each data unit moving from any source node to any destination in the network. FV will denote the set of all paths from all possible source to destination nodes in the network.

2 QoS metrics

Definition A QoS metric for path V is a random variable $q^{\Pi}(V)$ which takes values in $\{0, +\infty\}$, such that for $V = V_1.V_2$ (i.e. V is composed of path V_1 followed by path V_2):

$$q^{\Pi}(V) \leq q^{\Pi}(V_1) + q^{\Pi}(V_2), \quad a.s.$$

Note that the requirement that the QoS metric be *sub-additive* covers many strictly additive metrics of interest such as packet or cell loss rates, delay, path length (number of hops), and power dissipation. Other metrics such as path reliability and available bandwidth are sub-additive, and are also covered by our definition. For a path V composed of two successive sub-paths $V = V_1.V_2$, the following are obviously sub-additive:

$$q_{available-BW}(V) = inf(q_{available-BW}(V_1), q_{available-BW}(V_2)) \ a.s.,$$
$$\leq q_{available-BW}(V_1) + q_{available-BW}(V_2) \ a.s.,$$
$$q_r(V) = [q_r(V_1) \textbf{ and } q_r(V_2)] \ a.s.,$$
$$\leq q_r(V_1) + q_r(V_2) \ a.s.$$

where $q_r(.)$ is treated as a logical binary random value in the third equation, and as a numerical (binary) random value in the last equation.

2.1 QoS metrics and QoS

In the sequel $q^{\pi^{FV_i(v_d)}}(V_i^j)$ will denote the *QoS metric q measured on path* V_i^j, when the policy $\pi^{FV_i(v_d)}$ is applied to data units travelling from node i to v_d using the set of paths $FV_i(v_d)$, $V_i^j \in FV_i(v_d)$.

We sometimes write q with a subscript, e.g. $q_D^{\pi^{FV_i(v_d)}}(V_i^j)$ or $q_{LR}^{\pi^{FV_i(v_d)}}(V_i^j)$ to indicate that it designates some specific metric such as packet delay or packet loss rate.

Definition Let *u be a non-decreasing measurable function* and q be a QoS metric. The *QoS for data units sent on the path* V_i^j *using policy* $\pi^{FV_i(v_d)}$, *from source i to destination* v_d *along the set of paths* $FV_i(v_d)$ is simply the expected value $E[u(q^{\pi^{FV_i(v_d)}}(V_i^j))]$, i.e. the expected value of a measurable function of a QoS metric.

The reason for assuming that u is an increasing function (i.e., non-decreasing) is that we want to the QoS to reflect the trend of the QoS metric. If the QoS metric has a larger value reflecting some degradation in the path, we want the QoS also to reflect this degradation, or at least not to reflect an improvement.

Let us illustrate this with an example: let the QoS metric $q_D^{\pi^{FV_i(v_d)}}(V_i^j))$ be the path delay, and let $u(.) = 1(.)$ be the indicator or characteristic function. The QoS

$$E\left[1\left(q_D^{\pi^{FV_i(v_d)}}(V_i^j)) > T\right)\right] = Prob\left[q_D^{\pi^{FV_i(v_d)}}(V_i^j)) > T\right]$$

is the probability that the path delay is larger than T.

2.2 Sensitive QoS metrics

The value for some path of a *routing sensitive*, or simply *sensitive* QoS metric q *increases* when the probability of directing traffic into that path increases; examples include path delay and path loss ratio. An example of an *insensitive* QoS metric is the number of hops along a path; the power dissipated per data unit on a path may also be insensitive. Even when the probability of sending traffic down a given path is zero, we may assume that the path can be infrequently tested to obtain the value of the QoS metric of that path, or the path QoS may be known via prior information (e.g., available bandwidth, number of hops, or the path's power dissipation).

Definition We will say that the QoS metric q is *sensitive on the set* $FV_i(v_d)$, if for any two routing policies $\pi^{FV_i(v_d)}$ and $\pi'^{FV_i(v_d)}$ and any path $V_i^j \in FV_i(v_d)$:

$$\left\{\pi^{FV_i(v_d)}(V_i^j) > \pi'^{FV_i(v_d)}(V_i^j)\right\} \Rightarrow Prob\left[q^{\pi^{FV_i(v_d)}}(V_i^j) > x\right]$$
$$\geq Prob\left[q^{\pi'^{FV_i(v_d)}}(V_i^j) > x\right], \quad \text{for all } x > 0.$$

We say that q is *insensitive on the set* $FV_i(v_d)$ if for any path V_i^j, and any two routing policies such that $\pi^{FV_i(v_d)}(V_i^j) \neq \pi'^{FV_i(v_d)}(V_i^j)$:

$$Prob\left[q^{\pi^{FV_i(v_d)}}(V_i^j) > x\right] = Prob\left[q^{\pi'^{FV_i(v_d)}}(V_i^j)\right], \quad \text{for all } x > 0.$$

3 Sensible routing policies

A *sensible routing policy* is one which:

 – Selects paths only using the *expected value of the QoS*, i.e. the expected value of a function u of the QoS metric q for each path, as the criterion for selecting the *probability* that a path is chosen,
 – Selects the path for a new data unit independently of the decision taken for the previous data unit.

The practical motivation for considering sensible routing policies is that (1) averages of QoS metrics, or of functions of QoS metrics, are typically easy to estimate, and (2) decisions which are successively independent for successive data units are easier to implement.

Definition Let u be a non-decreasing measurable function. A *sensible routing policy* (SRP) from node i to destination v_d based on the QoS metric q is a probability distribution $\pi^{FV_i(v_d)}$ on the set $FV_i(v_d)$ such that:

$$\pi^{FV_i(v_d)}(V_i^j) = f_i^j(E[u(q^{\pi^{FV_i(v_d)}}(V_i^1)), u(q^{\pi^{FV_i(v_d)}}(V_i^2)),$$
$$\dots, u(q^{\pi^{FV_i(v_d)}}(V^{|FV_i(v_d)|}))]), \tag{4}$$

for a function $f_i^j : R^m \to [0,1]$, for each $V_i^j \in FV_i(v_d)$, such that:

–

$$\sum_{V_i^j \in FV_i(v_d)} \pi^{FV_i(v_d)}(V_i^j) = 1, \tag{5}$$

– and for each path V_i^j, the function $f_i^j(y_1, \dots, y_j, \dots, y_{|FV_i(v_d)|})$ defined in (4) is strictly decreasing in its argument y_j, with

$$\lim_{y_j \to +\infty} f_i^j(y_1, \dots, y_{|FV_i(v_d)|}) = 0.$$

Comment Thus a SRP is a routing policy which decides on routing based only on the QoS of each path, such that whenever the value of the QoS for any path increases then the probability of selecting that path decreases.

Example A simple example of a SRP is the following:

$$\pi^{FV_i(v_d)}(V_i^j) = \frac{\dfrac{1}{E[q_D^{\pi^{FV_i(v_d)}}(V_i^j)]}}{\sum_{all\ s} \dfrac{1}{E[q_D^{\pi^{FV_i(v_d)}}(V_i^s)]}}, \tag{6}$$

which says that packets are directed to the paths with a probability which is inversely proportional to the average delay. In a sensible routing policy, the probability that a specific path is selected will depend on the QoS of that path, which in general depends on the policy itself, i.e. on the probability that the path is selected, unless the policy is insensitive. Thus there is the question of whether we are able to compute the routing probabilities. The following theorem provides sufficient conditions for being able to do this.

Theorem 1 If $\pi^{FV_i(v_d)}$ is a sensible routing policy on $FV_i(v_d)$, then the solution to (4) exists and is unique for each path V_i^j.

Proof. For any path V_i^j, consider:

- the function $f_i^j(y_1, \ldots, y_j, \ldots, y_{|FV_i(v_d)|})$ of equation (4), which (strictly) decreases when y_j increases, and
- the path QoS $y_j(\pi^{FV_i(v_d)}) = E[u(q^{\pi^{FV_i(v_d)}}(V_i^j)]$. Since

$$\left\{ \pi^{FV_i(v_d)}(V_i^j) > \pi'^{FV_i(v_d)}(V_i^j) \right\} \Rightarrow Prob\left[q^{\pi^{FV_i(v_d)}}(V_i^j) > x \right]$$
$$\geq Prob\left[q^{\pi'^{FV_i(v_d)}}(V_i^j) > x \right], \quad (7)$$

for all $x > 0$, the path QoS is an increasing function (not strictly) $y_j(\pi)$ of its argument, the probability π, because of (7), and because u is an increasing function.

Thus the solution of equation (4) for any V_i^j is obtained at the intersection of a non-negative, strictly decreasing function f_i^j of y_j which tends to zero, and an increasing non-negative function y_j of f_i^j.

4 m-Sensible routing policies (m-SRP)

In this section we extend the concept of a sensible policy to more sophisticated usage of QoS to make routing decisions. We construct a hierarchy of *m-sensitive* policies, where the 1-sensitive policy is just the sensitive policy defined earlier, and the 0-sensitive policy is a random uninformed choice between paths with equal probability. What is particularly interesting is that, just by increasing the value of m we are guaranteed to achieve better overall QoS, when the QoS metric is insensitive. The same result can be obtained in the sensitive case as well under certain sufficient conditions.

Definition For a natural number m, an *m-sensible routing policy* ($m-$SRP) from node i to destination v_d based on the QoS metric q is a probability distribution $\pi^{FV_i(v_d)}$ on the set $FV_i(v_d)$ such that:

$$\pi^{FV_i(v_d)}(V_i^j) = \frac{\frac{1}{E[u(q^{\pi^{(FV_i(v_d)}}(V_i^j))^m]}}{\sum_{all\ s} \frac{1}{E[u(q^{\pi^{FV_i(v_d)}}(V_i^s))^m]}}. \quad (8)$$

We will use the notation $\pi^{m-SRP[FV_i(v_d)]}$ to denote the fact that the policy π on the set of paths $FV_i(v_d)$ is $m-sensible$, and the corresponding QoS value will be denoted by $q^{\pi^{m-SRP[FV_i(v_d)]}}(V_i^s)$ for path V_i^j. Note that a $0-$SRP is just a random choice among paths, with equal probability.

4.1 The m-sensible routing theorem when the QoS metric is insensitive

In this section we assume that q is insensitive on the set $FV_i(v_d)$, and consider m-SRP routing policies as defined in (8).

To simplify the notation, let us associate the index j with the path V_i^j and write:

$$W_j(m) = E\left[u\left(q^{\pi^{m-SRP[FV_i(v_d)]}}(V_i^j)\right)\right]. \tag{9}$$

When q is insensitive, we will simply write W_j. Using (8) and (9) we have:

$$Q^{\pi^{m-SRP[VF_i]}} = \frac{\sum_{j=1}^n \frac{W_j}{W_j^m}}{\sum_{j=1}^n \frac{1}{W_j^m}}. \tag{10}$$

We first prove the following simple result.

Lemma 1 For any $W_j \geq 0$, $W_k \geq 0$,

$$\frac{1}{2}(W_j + W_k) \geq \frac{2}{\frac{1}{W_j} + \frac{1}{W_k}}$$

or

$$\frac{W_j}{W_k} + \frac{W_k}{W_j} \geq 2.$$

Proof. Since $(W_j - W_k)^2 \geq 0$, we have $(W_j^2 + W_k^2) \geq 2W_j W_k$, and therefore $(W_j + W_k)^2 \geq 4W_j W_k$, or:

$$\frac{1}{2}(W_j + W_k) \geq \frac{2}{\frac{1}{W_j} + \frac{1}{W_k}}$$

and therefore:

$$(W_j + W_k)\left(\frac{1}{W_j} + \frac{1}{W_k}\right) \geq 4$$

which can be written as:

$$2 + \left(\frac{W_k}{W_j} + \frac{W_j}{W_k}\right) \geq 4$$

completing the proof.

We will call the following result the **m-SRP theorem** (m-sensible routing theorem) for insensitive metrics.

Theorem 2 If q is insensitive on the set $FV_i(v_d)$, the policy $(m+1)$-SRP is better than m-SRP for $m \geq 1$, *i.e.*:

$$Q^{\pi^{m-SRP[VF_i]}} \geq Q^{\pi^{(m+1)-SRP[VF_i]}}.$$

Proof. From Lemma 1, we have that for any $W_j \geq 0$, $W_k \geq 0$,

$$\frac{W_j}{W_k} + \frac{W_k}{W_j} \geq 2,$$

and multiplying both sides by $1/(W_j^m W_k^m)$ we obtain:

$$\frac{1}{W_j^{m-1} W_k^{m+1}} + \frac{1}{W_j^{m+1} W_k^{m-1}} \geq \frac{2}{W_i^m W_j^m}.$$

Summing for $j, k = 1, \ldots, n$ and adding identical terms on both sides, we have:

$$\sum_{j=1}^{n} \frac{1}{(W_j^m)^2} + \sum_{j,k=1; j \neq k}^{n} \left\{ \frac{1}{W_j^{m-1} W_k^{m+1}} + \frac{1}{W_j^{m+1} W_k^{m-1}} \right\} \geq \sum_{j=1}^{n} \frac{1}{(W_j^m)^2}$$

$$+ \sum_{j,k=1; j \neq k}^{n} \frac{2}{W_j^m W_k^m},$$

or

$$\left(\sum_{j=1}^{n} \frac{1}{W_j^{m-1}} \right) \left(\sum_{j=1}^{n} \frac{1}{W_j^{m+1}} \right) \geq \left(\sum_{j=1}^{n} \frac{1}{W_j^m} \right)^2,$$

This can be written as:

$$\frac{\sum_{j=1}^{n} \frac{1}{W_j^{m-1}}}{\sum_{j=1}^{n} \frac{1}{W_j^m}} \geq \frac{\sum_{j=1}^{n} \frac{1}{W_j^m}}{\sum_{j=1}^{n} \frac{1}{W_j^{m+1}}},$$

or in the final form:

$$\frac{\sum_{j=1}^{n} \frac{W_j}{W_j^m}}{\sum_{j=1}^{n} \frac{1}{W_j^m}} \geq \frac{\sum_{j=1}^{n} \frac{W_j}{W_j^{m+1}}}{\sum_{j=1}^{n} \frac{1}{W_j^{m+1}}},$$

which completes the proof. $\square$

It is obvious that for an *insensitive* QoS metric, selecting m to be very large is good, since this will lead to choosing the path with the best QoS if such a path exists. We summarize this point in the following remark. However, if the QoS metric is *sensitive* then the matter is quite different, as will be discussed in the next section.

Remark 3 Suppose that q is *insensitive on the set* $FV_i(v_d)$, and that path V_i^1 is best in the following sense:

$$W_1 < W_2, \ldots W_n.$$

Then $\lim_{m \to \infty} Q^{\pi^{m-SRP[VF_i]}} = W_1$.

Proof. Using:

$$Q^{\pi^{m-SRP[VF_i]}} = \frac{\sum_{j=1}^{n} \frac{W_j}{W_j^m}}{\sum_{j=1}^{n} \frac{1}{W_j^m}}, \tag{11}$$

$$= \frac{W_1 + \sum_{j\neq 1}^{n} W_j \left(\frac{W_1}{W_j}\right)^m}{1 + \sum_{j\neq 1} \left(\frac{W_1}{W_j}\right)^m}, \tag{12}$$

which yields the result when we take $m \to \infty$.   □

4.2 The sensible routing theorem for sensitive QoS metrics

When the QoS metric is sensitive, the QoS varies with the load on the paths. This is of course the most common situation in practice, e.g. for QoS metrics such as delay, packet or cell loss, etc.. Thus we cannot generalize Theorem 2 to the case where the QoS metric is sensitive. However we can provide necessary and sufficient conditions which will yield a similar result.

Let δ_m be defined as follows:

$$\delta_m = \sum_{j=1}^{n} \frac{1}{(W_j(m+1))^m} - \sum_{j=1}^{n} \frac{1}{(W_j(m))^m}. \tag{13}$$

This leads to the **m-SRP Theorem for sensitive QoS metrics**.

Theorem 4 If q is *sensitive on the set $FV_i(v_d)$*, the policy $(m+1)$-SRP is better than m-SRP for $m \geq 1$, *i.e.*:

$$Q^{\pi^{m-SRP[VF_i]}} \geq Q^{\pi^{(m+1)-SRP[VF_i]}},$$

provided that the following condition holds:

$$\delta_{m-1} \sum_{j=1}^{n} \frac{1}{(W_j(m+1))^{m+1}} \leq \delta_m \sum_{j=1}^{n} \frac{1}{(W_j(m+1))^m}. \tag{14}$$

Proof. From Theorem 1, we know that for any $W_j \geq 0$, $j = 1, \dots n$,

$$\frac{\sum_{j=1}^{n} \frac{1}{W_j^{m-1}}}{\sum_{j=1}^{n} \frac{1}{W_j^m}} \geq \frac{\sum_{j=1}^{n} \frac{1}{W_j^m}}{\sum_{j=1}^{n} \frac{1}{W_j^{m+1}}},$$

and in particular this is true if we set $W_j = W_j(m+1)$, so that

$$\frac{\sum_{j=1}^{n} \frac{1}{(W_j(m+1))^{m-1}}}{\sum_{j=1}^{n} \frac{1}{(W_j(m+1))^m}} \geq \frac{\sum_{j=1}^{n} \frac{1}{(W_j(m+1))^m}}{\sum_{j=1}^{n} \frac{1}{(W_j(m+1))^{m+1}}},$$

or:

$$\left[\sum_{j=1}^{n} \frac{1}{(W_j(m+1))^{m-1}} \right] \left[\sum_{j=1}^{n} \frac{1}{(W_j(m+1))^{m+1}} \right]$$
$$\geq \left[\sum_{j=1}^{n} \frac{1}{(W_j(m+1))^{m}} \right] \left[\sum_{j=1}^{n} \frac{1}{(W_j(m+1))^{m}} \right].$$

Using the definition of δ_m we have:

$$\left[\sum_{i=1}^{n} \frac{1}{(W(m))^{m-1}} + \delta_{m-1} \right] \left[\sum_{j=1}^{n} \frac{1}{(W_j(m+1))^{m+1}} \right]$$
$$\geq \left[\sum_{j=1}^{n} \frac{1}{(W_j(m))^{m}} + \delta_m \right] \left[\sum_{j=1}^{n} \frac{1}{(W_j(m+1))^{m}} \right],$$

so that:

$$\left[\sum_{j=1}^{n} \frac{1}{(W_j(m))^{m-1}} \right] \left[\sum_{j=1}^{n} \frac{1}{(W_j(m+1))^{m+1}} \right]$$
$$- \sum_{j=1}^{n} \frac{1}{(W_j(m))^{m}} \right] \left[\sum_{j=1}^{n} \frac{1}{(W_j(m+1))^{m}} \right]$$
$$\geq \delta_m [\sum_{j=1}^{n} \frac{1}{(W_j(m+1))^{m}}] - \delta_{m-1} \left[\sum_{j=1}^{n} \frac{1}{(W_j(m+1))^{m+1}} \right].$$

Condition (14) is therefore sufficient since it implies that: $\delta_m [\sum_{j=1}^{n} \frac{1}{(W_j(m+1))^{m}}] - \delta_{m-1} [\sum_{j=1}^{n} \frac{1}{(W_j(m+1))^{m+1}}] \geq 0.$ □

5 Conclusions

In this paper we suggest a theory of routing based on QoS. We have distinguished between QoS metrics, and QoS. Variable and adaptive routing have again become of interest in networking because of the increasing importance of mobile ad-hoc networks. In this paper we have developed a framework for the study of adaptive routing algorithms which use the expected QoS to select paths to their destination. Our objective is not to analyze QoS, but rather to design randomized routing policies which can improve QoS.

We define QoS metrics as non-negative random variables associated with network paths that satisfy a sub-additivity condition along each path. We define the

QoS of a path as the expected value of a non-decreasing measurable function of the QoS metric. We discuss *sensitive* and *insensitive* QoS metrics, the latter being dependent on the routing policy which is used. An example of an insensitive QoS metric is the number of hops on a path, since it will not change with the fact that this particular path is selected by the route selection.

We describe routing policies as probabilistic choices among all possible paths from some source to some given destination. *Sensible* routing policies are then introduced: they take decisions based simply on the QoS of each possible path. Sensible policies, which make decisions based on the QoS of the paths, are introduced. We prove that the routing probability of a sensible policy can always be uniquely determined. A hierarchy of *m-sensible* probabilistic routing policies is then introduced. A $0 - sensible$ policy is simply a random choice of routes with equal probability, while a $1 - sensible$ policy selects a path with a probability which is inversely proportional to the (expected) QoS of the path. We prove that an $m + 1 - sensible$ policy provides better QoS on the average than an $m - sensible$ policy, if the QoS metric is insensitive. We also show that under certain conditions, the same result also holds for sensitive QoS metrics.

In future work we will consider myopic policies which only examine partial information based on the QoS of initial portions of the possible paths in order to make decisions. We will also consider *Incremental* routing policies which can be derived from independent decisions taken at certain points (or nodes) along paths. Finally, we plan to exploit the sub-additivity of the QoS metrics to prove asymptotic properties for very large networks.

References

Ahmadi H, Guerin R, Nagshineh M (1991) Equivalent capacity and its application to bandwidth allocation in high speed networks. *IEEE J. on Sel. Areas in Comm.* 9: 968–981

Bertsekas D, Gallagher R (1989) *Data Networks*. Prentice Hall, Englewood Cliffs, NJ

Chen S, Nahrstedt K (1998) *An overview of Quality-of-Service routing for the next generation high-speed networks: Problems and Solutions*. Network Magazine, Nov/Dec.

Gelenbe E, Lent R (2002) Mobile Ad-Hoc Cognitive Packet Networks. *Proc. IEEE ASWN*, Paris

Gelenbe E, Mang X, Feng Y (1996) Diffusion cell loss estimates for ATM with multiclass bursty traffic. *Computer Systems–Science and Engineering* (Special Issue: ATM Networks) 11(6): 325–334

Gelenbe E, Seref E, Xu Z (1999a) Towards networks with intelligent packets. *Proc. IEEE-ICTAI Conference on Tools for Artificial Intelligence*. Chicago

Gelenbe E, Mao Zhi-Hong, Da-Li Y (1999b) Function approximation with spiked random networks. *IEEE Trans. on Neural Networks* 10(1): 3–9

Gelenbe E, Lent R, Xu Z (2000) Towards networks with cognitive packets. Keynote Paper, *Proc. IEEE MASCOTS Conference* (ISBN 0-7695-0728-X) pp 3–12

Gelenbe E, Lent R, Xu Z (2001) Measurement and performance of Cognitive Packet Networks. *J. Comp. Networks* 37: 691–701

Gelenbe E, Lent R, Montuori A, Xu Z (2002) Cognitive packet petworks: QoS and performance. Keynote Paper, *Proc. IEEE MASCOTS Conference*, San Antonio, TX

Hao F, Zegura EW, Ammar MH (2002) QoS routing for anycast communications: motivation and an architecture for Diffserv networks. *IEEE Comm. Magazine* 46(2): 48–56

Kodialam M, Lakshman TV (2002) Restorable quality of service routing. *IEEE Comm. Magazine* 46(2): 72–81

Nelakuditi S, Zhang Zhi-Li (2002) A localized adaptive proportioning approach to QoS routing granu-
    larity. *IEEE Comm. Magazine* 46(2): 66 – 71
Srinivasan V, Ghanwani A, Gelenbe E (1996) Block cell loss reduction in ATM systems. *Computer
    Communications* 19: 1077-1091
Ying-Dar Lin, Nai-Bin Hsu, Ren-Hung Hwang (2002) QoS routing granularity in MPLS networks. *IEEE
    Comm. Magazine* 46(2): 58–65