Scalability, Delay Characterization, and Satisfiability of Quality-of-Information in Wireless Networks

***Abstract*—Quality of Information (QoI) provides a context-dependent measure of the utility that a network delivers to its users by incorporating non-traditional information attributes. Quickly and easily predicting performance and limitations of a network using QoI metrics is a valuable tool for network design. Even more useful is an understanding of how network components like topology, bandwidth, protocols, etc. impact these limitations. In this paper, we develop a QoI-based framework that can provide this understanding of limitations and impact by modeling the various contributors to delay in the network, including channel rate and contention, competing traffic flows, and multi-hop propagation effects, and relating them to QoI requirements, especially completeness and timeliness. Analysis shows that large tradeoffs exist between network parameters, such as QoI requirements, topology, and network size. Simulation results also provide evidence that the developed framework can estimate network limits with high accuracy. Finally, this work also introduces *scalably feasible QoI regions*, which provide upper bounds on QoI requirements that can be supported for certain network applications.**

I. Finding Limits and Characterizing Delay

As explained in Section **??**, delay of a flow can be expressed as

$$D = \frac{k_{req} \cdot I_S \cdot CF \cdot TF}{W} + \frac{P_S \cdot DF \cdot (PL-1)}{W} \quad (1)$$

We will make some substitutions to get the following version

$$D = \frac{P_S \cdot CF \cdot P_N \cdot TF}{W} + \frac{P_S \cdot DF \cdot (PL-1)}{W} \quad (2)$$

Building on this equation for delay, we will use the following equation to describe the delay from i given a destination of j:

$$D_{i|j} = \frac{P_S \cdot CD \cdot P_N \cdot TF_{i|j}}{W} + \frac{P_S \cdot DF \cdot (PL(i,j)-1)}{W} \quad (3)$$

Here, we use $PL()$ as a function that provides the path length between i and j. We also assume that P_N is a random variable that describes the number of packets in a given request, capturing both the possible randomness of k_{req} and I_S. Also, recall that TF is a random variable of the flows being forwarded at the bottleneck node along the path of the flow. Let us define two constants to simplify the expression:

$$C_1 = \frac{P_S \cdot CF}{W}$$
$$C_2 = \frac{P_S \cdot DF}{W}$$

Then, we can express the delay as

$$D_{i|j} = C_1 \cdot P_N \cdot TF_{i|j} + C_2 \cdot PL(i,j) \quad (4)$$

We can develop the following expression for a distribution of delay:

$$P(D_{i|j} \le d) = P(C_1 \cdot P_N \cdot TF_{i|j} + C_2 \cdot PL(i,j) \le d)$$
$$P(P_N \cdot TF_{i|j} \le \frac{d - C_2 \cdot PL(i,j)}{C_1})$$
$$\sum_{tf=1}^{tf_{max}} P(P_N \cdot TF \le \frac{d - C_2 \cdot PL(i,j)}{C_1} | TF = tf) \cdot f_{TF_{i|j}}(tf)$$
$$\sum_{tf=1}^{tf_{max}} P(P_N \le \frac{d - C_2 \cdot PL(i,j)}{C_1 \cdot tf}) \cdot f_{TF_{i|j}}(tf)$$
$$F_{D_{i|j}}(d) = \sum_{tf=1}^{tf_{max}} F_{P_N}(\frac{d - C_2 \cdot PL(i,j)}{C_1 \cdot tf}) \cdot f_{TF_{i|j}}(tf)$$

Then, we can generalize the expression to give a distribution for a flow originating in node i with an unknown destination by conditioning over all possible destinations, j.

$$F_{D_i} = \sum_{j \ne i} [\sum_{tf=1}^{tf_{max}} F_{P_N}(\frac{d - C_2 \cdot PL(i,j)}{C_1 \cdot tf}) \cdot f_{TF_{i|j}}(tf)] \cdot p(j) \quad (5)$$

Finally, we can get an average distribution of all flows' delays by summing over all sources and dividing by the the number of sources. This average delay distribtuion is in Equation (5).

A. Minimum Timeliness/Maximum Query Rate

The first useful information that can be gathered from the delay distribution is the maximum expected delay, d_{max}, of a flow in the network, which occurs at the delay d at which F_D reaches its maximum value of 1. If the average rate of queries, λ, is greater than $\frac{1}{d_{max}}$, then the traffic will exceed the network capacity and the number of active queries in the system will grow without bound, causing packets to be dropped and/or delays to grow without bound. Therefore, the maximum query rate is $\lambda_{max} = \frac{1}{d_{max}}$, and, consequently, the minimum timeliness for which *all* flows can be expected to complete before the deadline is d_{max}.

In some applications, having a certain amount of queries not complete by the timeliness requirement may be acceptable. In these situations, more useful information can be extracted from the delay distribution in Equation **??**. Specifically, this delay distribution can be interpreted as the expected percentage of queries that will finish within the timeliness constraint if the timeliness constraint was d. As we will show in Section **??**, this relationship follows a Normal distribution CDF.

$$F_D(d) = \frac{1}{N} \cdot \sum_{i=1}^{N} \sum_{j \neq i} \sum_{tf=1}^{tf_{max}} F_{P_N}\left(\frac{d - C_2 \cdot PL(i,j)}{C_1 \cdot p_N}\right) \cdot f_{TF_{i|j}}(tf) \cdot p(j) \tag{6}$$