Delay Constrained Optimal Resource Utilization of Wireless Networks for Distributed Control Systems

Muhammad Tahir, *Student Member, IEEE,* Sudip K. Mazumder, *Senior Member, IEEE*

Abstract— An optimal resource utilization framework to analyze the impact of end-to-end delay thresholds on the wireless network throughput is developed. In contrast to conventional networks where lumped transmission and queuing delays are used to model the end-to-end delay, the proposed framework models these delay components separately to analyze their relative contribution to the end-to-end delay when stringent delay thresholds are used. A 'log' utility maximization problem is formulated and its distributed implementation using dual decomposition is proposed. Our optimization results show that decreasing the delay threshold leads to larger contribution from the transmission delay to the end-to-end delay.

Index Terms— Delay, optimization, network capacity, distributed control.

I. INTRODUCTION

USING wireless networks for information exchange in distributed control environment require stringent end-to-end delay thresholds to be met [1]. Highly delay dependent performance of a distributed control system poses new challenges to the optimal utilization of the underlying communication network. The network induced delay is widely modelled either as sum of the queuing delay (equal to the waiting time in the queue) and an associated transmission delay [2] or as a measure of network congestion [3]. The end-to-end delay model [2] based on lumped queuing and transmission delays does not provide an insight about the relative significance of the individual delay components (transmission and queuing). From network design perspective, when stringent delay thresholds need to be satisfied as is the case of a distributed control system, it is important to know how the variation in the delay threshold affects each of these delay components. To study this we have developed an optimal resource utilization framework which models the transmission and queuing delays separately in the end-to-end delay model.

In contrast to a conventional network where queuing delay dominates the transmission delay, the two delay components become comparable in case of a distributed control system due to the reason explained in the following by an example. Given a queue buffer of size B and the transmission delay $d^{(t)}$, the average queuing delay $d^{(q)}$ (assuming the link is available for transmission) is given by $d^{(q)} = d^{(t)}E[k], 0 \leq k \leq B$, where k is a random variable and $E[k]$ is the mean number of packets in the buffer. It can be seen that queuing delay, in

general, is large compared to transmission delay for moderate buffer size. This does not hold for small end-to-end delay requirements and as a result requires reduction in the buffer size B. This is the case in a distributed control environment, where stringent delay thresholds require small buffers and as a result the transmission delay becomes comparable to the queuing delay.

II. NETWORK MODEL

We model the wireless network as a directed graph $G(N, L)$, where N is the set of nodes and L is the set of links. Each transmission session $s_i \in S$ (S is the set of all active network sessions and s_i will be replaced by s to simplify the notation), representing an ongoing transmission between a source-destination pair through intermediate nodes, has an associated end-to-end session rate $r_s \in \mathbf{r}$, a shortest path [4] consisting of a subset of links $L(s) \subseteq L$, an end-to-end delay threshold $D'(s)$ due to distributed control system application layer [1] and the minimum rate threshold $R'(s)$. We define *transmission cycle* to be the set H of *transmission schedules* where each transmission schedule $h_i \in H$ has an associated subset of simultaneously transmitting links $L(h_i) \subseteq L$. We consider an interference limited wireless network where the network has more than one transmission in each transmission schedule h_i. Simultaneous transmissions are allowed when distance between the transmitter of link m and the receiver of link l is greater than $1.5t$, where t is the distance between the transmitter and receiver of link l.

1) Link Capacity Model: For any link l representing an active transmitter receiver pair in a transmission schedule, we model the channel behavior using the signal to interference and noise ratio $SINR_l(\mathbf{P})$ [3] as

$$SINR_l(\mathbf{P}) = \frac{\gamma_{ll}P_l}{n_l + \sum_{m \neq l} \gamma_{lm}P_m}. \tag{1}$$

In (1) γ_{lm} is the channel gain from transmitter of link m to the receiver of link l, $P_l \in \mathbf{P}$ and $P_l \leq P_{max} \forall l$ is the transmitter power for link l, P_{max} is the maximum transmit power level and n_l is the additive noise. The accumulated data rate assigned to link l requires $\sum_{s:l \in L(s)} r_s \leq c_l(P)$, where $c_l(\mathbf{P}) = W \log(1 + SINR_l(\mathbf{P}))$ [3] is link capacity and W is the bandwidth allocated to the channel. Without loss of generality we assume that a session s is scheduled only once at link l in a transmission cycle. This gives an average link capacity of $c_l(\mathbf{P})/|H|$ and average session rate of $r_s/|H|$, where $|H|$ is the number of transmission schedules in one transmission cycle. To incorporate the effect of scheduling in the optimization framework we scale the rate and delay thresholds as $R_{min}(s) = R'(s)|H|$ and $D_{max}(s) = D'(s)/|H|$ respectively.

2) End-to-End Delay Model: Using the M/D/1 [4] queuing model the average delay at link l for mean packet length μ is given by $(\mu/c_l(\mathbf{P}) + \mu\rho/(c_l(\mathbf{P}) - \sum_{s:l\in L(s)} r_s))$, where the second term represents the queuing delay and $\rho = \sum_{s:l\in L(s)} r_s/c_l(\mathbf{P})$ is the load factor. The above expression is simplified to $0.5\mu(1/c_l(\mathbf{P}) + 1/(c_l(\mathbf{P}) - \sum_{s:l\in L(s)} r_s))$ and we approximate the queuing delay by $0.5\mu/(c_l(\mathbf{P}) - \sum_{s:l\in L(s)} r_s))$. The approximation increases the actual queuing delay by a factor of $0.5\mu/c_l(\mathbf{P})$ which is negligible compared to the actual queuing delay when $\sum_{s:l\in L(s)} r_s$ is close to $c_l(\mathbf{P})$. The end-to-end delay for each session s, obtained by accumulating the transmission and queuing delays at each intermediate link $l \in L(s)$, is upper bounded by the delay threshold $D_{max}(s)$ and is given by

$$\sum_{l\in L(s)} \left(\frac{\mu}{c_l(\mathbf{P})} + \frac{0.5\mu}{c_l(\mathbf{P}) - \sum_{s:l\in L(s)} r_s} \right) < D_{max}(s). \quad (2)$$

III. CROSS-LAYER OPTIMIZATION PROBLEM FORMULATION

Using the auxiliary variables $d_l^{(t)} \in \mathbf{d}^{(t)}$ and $d_l^{(q)} \in \mathbf{d}^{(q)}$ corresponding to the transmission and queuing delays respectively at link l in (2), we formulate the network utility maximization problem as follow:

$$\text{maximize} \quad \sum_s U(s)$$

$$\text{subject to} \quad \sum_{l\in L(s)} \left(d_l^{(t)} + d_l^{(q)} \right) < D_{max}(s) \quad \forall s,$$

$$\frac{0.5\mu}{c_l(\mathbf{P}) - \sum_{s:l\in L(s)} r_s} \le d_l^{(q)} \quad \forall l,$$

$$\frac{\mu}{c_l(\mathbf{P})} \le d_l^{(t)} \quad \forall l,$$

$$R_{min}(s) < r_s \quad \forall s, \quad 0 < P_l < P_{max} \quad \forall l. \quad (3)$$

In the above expression $U(r_s)$ is the network utility function and is formulated as $\log r_s$. Motivated by Kelly's seminal work [5] we decompose the original problem by associating Lagrange multipliers $\lambda_l \in \Lambda$ and $\psi_l \in \Psi$ to each of the queuing and transmission delay constraints respectively in (3) and construct the partial Lagrangian (after rearranging the terms) as

$$L(\mathbf{r}, \mathbf{P}, \mathbf{d}^{(q)}, \mathbf{d}^{(t)}, \Lambda, \Psi) = \sum_s \left(U(r_s - \sum_{l\in L(s)} \lambda_l r_s) \right)$$
$$+ \sum_l (\lambda_l + \psi_l) c_l(\mathbf{P}) - \sum_l \mu \left(\frac{\lambda_l}{2d_l^{(q)}} + \frac{\psi_l}{d_l^{(t)}} \right). \quad (4)$$

Let $g(\Lambda, \Psi)$ denotes the optimal value of $L(\mathbf{r}, \mathbf{P}, \mathbf{d}^{(q)}, \mathbf{d}^{(t)}, \Lambda, \Psi)$ in variables $\mathbf{r}$, $\mathbf{P}$, $\mathbf{d}^{(q)}$ and $\mathbf{d}^{(t)}$, then the associated dual problem is

$$\text{minimize}_{\lambda_l, \psi_l \ge 0 \ \forall l} \quad g(\Lambda, \Psi). \quad (5)$$

Block diagram in Fig. 1 shows the implementation of the rate-delay-power allocation algorithm using the optimization sub-problems in (4). Both the rate and delay allocation sub-problems are convex in r_s and $d_l^{(t)}$ and $d_q^{(t)}$ respectively. The
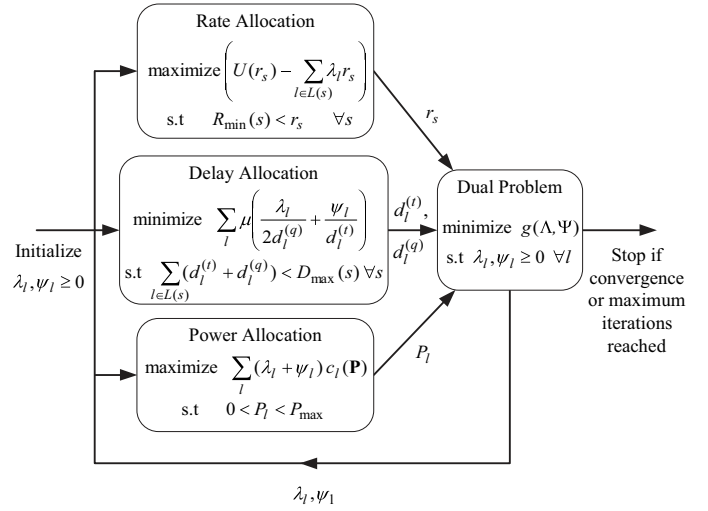

Fig. 1. Block diagram for the distributed implementation of the rate-delay-power allocation algorithm.

power allocation sub-problem is not convex but can be transformed into a convex problem by using 'log' transformation, and then can be efficiently solved by power-update algorithm [3] given by

$$P_l(k+1) = P_l(k) + \beta(k) \left(\frac{(\lambda_l + \psi_l)}{P_l(k)} - \right.$$
$$\left. \sum_{j\neq l} \frac{\gamma_{jl}(\lambda_j + \psi_j)}{\sum_{m\neq j} \gamma_{jm} P_m(k) + n_j} \right). \quad (6)$$

In (6) $\beta(k)$ is the variable step size chosen as $1/\sqrt{k}$, where k is the time index. For the dual problem we observe that $(c_l(\mathbf{P}) - \mu/d_l^{(t)})$ and $(c_l(\mathbf{P}) - \sum_{s:l\in L(s)} r_s - 0.5\mu/d_l^{(q)})$ are subgradients of $g(\Lambda, \Psi)$ with respect to ψ_l and λ_l respectively. As a result solving dual problem in (5) is equivalent to updating dual variables for each link l by

$$\lambda_l(k+1) = \left[\lambda_l(k) - \beta(k) \left(c_l(\mathbf{P}) - \sum_{s:l\in L(s)} r_s - \frac{\mu}{2d_l^{(q)}} \right) \right]^+, \quad (7)$$

$$\psi_l(k+1) = \left[\psi_l(k) - \beta(k) \left(c_l(\mathbf{P}) - \frac{\mu}{d_l^{(t)}} \right) \right]^+. \quad (8)$$

In (7) and (8) $[x]^+$ is defined as $max\{0, x\}$. The communication overhead of the distributed implementation involves exchanging the primal (r_s, P_l, $d_l^{(q)}$ and $d_l^{(t)}$) as well as dual (ψ_l and λ_l) variables among different nodes. A large contribution to the overhead comes from P_l due to $\sum_{m\neq j} \gamma_{jm} P_m(k)$ term in (6) since there can be multiple transmitters active simultaneously. By limiting the contribution to only those transmitters which lie within certain hop distance, we can reduce the communication overhead significantly.

IV. RESULTS

To study the delay and throughput characteristics of the distributed optimization algorithm we use the network shown in Fig. 2(a). One transmission cycle H ensuring interference
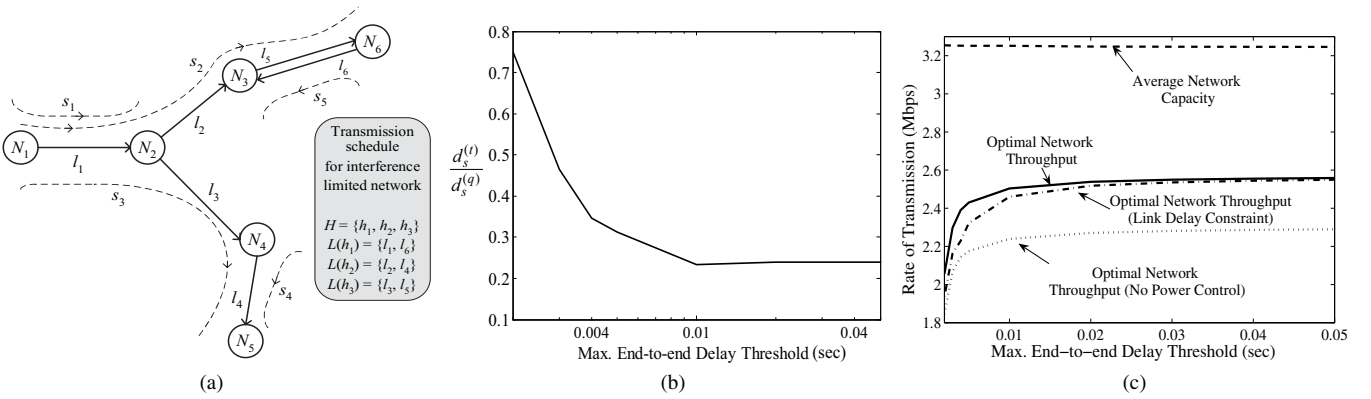
Fig. 2.   (a) Example network with multiple sessions; (b) variation of $d_s^{(t)}/d_s^{(q)}$ for different $D_{max}(s)$; (c) effect of maximum end-to-end delay threshold on the optimal network throughput and network capacity.

limited wireless communication in each transmission schedule is also shown in Fig. 2(a). Constant packet size of 50 bytes, $R_{min}(s)$ of 150 Kbps and P_{max} of 10 dBm is used.

To quantify the performance measure in an interference limited network we define *Average Network Capacity* (C_{net}) and *Optimal Network Throughput* (T_{net}) by summing, the optimal link capacities (c_l^*) and optimal session rates accumulated at each link ($\sum_{s:l\in L(s)} r_s^*$) respectively, in each transmission schedule h_i and averaging over the transmission cycle H, and are given by

$$C_{net} = \frac{1}{|H|} \sum_{h_i \in H} \sum_{l \in L(h_i)} c_l^*,$$

$$T_{net} = \frac{1}{|H|} \sum_{h_i \in H} \sum_{l \in L(h_i)} \sum_{s:l\in L(s)} r_s^*. \qquad (9)$$

The definition of C_{net} is inline with the network capacity measure in [6], since we aim to have maximum number of simultaneous transmissions in each transmission schedule while meeting the distance threshold.

To quantify the contribution of transmission delay in the end-to-end delay we define the end-to-end transmission and queuing delays for each session s by $d_s^{(t)} = \sum_{l\in L(s)} d_l^{(t)}$ and $d_s^{(q)} = \sum_{l\in L(s)} d_l^{(q)}$ respectively. Fig. 2(b) shows the ratio of the transmission to queuing delay for session s_2. For small delay thresholds the contribution of the transmission delay in the end-to-end delay becomes dominant emphasizing the scenario of a distributed control system. As the delay threshold is increased, the optimization algorithm tries to maximize the session rates which in turn increases the queuing delay leading to sharp decrease in the $d_s^{(t)}/d_s^{(q)}$.

Fig. 2(c) shows T_{net} as a function of delay threshold $D_{max}(s)$. We observe that a higher network throughput can be obtained for larger $D_{max}(s)$, but the increase in the throughput becomes trivial for larger $D_{max}(s)$. As expected the plot shows that C_{net} remains unchanged with an increase in $D_{max}(s)$. Fig. 2(c) also shows the plots for two special cases: 1) no power control with the transmitter power fixed at a constant power level of $P_{max}/2$ and 2) the end-to-end delay constraint decomposed into link delay constraints by $D_{max}(l) = min\{D_{max}(s)/|L(s)|$, s.t. $l \in L(s)\}$. The throughput performance of both the special cases is comparable for small delay thresholds. In case of large delay thresholds the throughput performance while meeting $D_{max}(l), l \in L(s)$ is almost equivalent to that of meeting $D_{max}(s)$. For the case of fixed transmitter powers the link capacities are constant and an increase in the delay threshold $D_{max}(s)$ is not exploited leading to the reduction in the network throughput by more than 10% compared to the case when transmitter power control is employed.

V. CONCLUSION

A framework to analyze the impact of end-to-end delay thresholds on the wireless network throughput is developed where transmission and queuing delays are modelled separately. The approximation used in modelling the queuing delay results in higher end-to-end delay estimates. But doing so gives an implicit delay margin to the distributed control system and a stable operation within delay bounds is guaranteed.

REFERENCES

[1] K. Acharya, M. Tahir, and S. K. Mazumder, "Communication fault-tolerant wireless network control of a load-sharing multiphase interactive power network," in *Proc. IEEE Power Electronics Specialists Conf.*, pp. 1167-1174, 2006.

[2] M. Saad, A. Leon-Garcia, and W. Yu, "Optimal network rate allocation under end-to-end quality-of-service requirements," accepted in *IEEE Trans. Network and Service Management*, 2007.

[3] M. Chiang, "Balancing transport and physical layers in wireless multihop networks: jointly optimal congestion control and power control," *IEEE J. Select. Areas Commun.*, vol. 23, no. 1, 2005.

[4] D. Bertsekas and R. G. Gallager, *Data Networks*. Prentice Hall, 1991.

[5] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *J. Operations Research Society*, vol. 49, no. 3, pp. 237-252, 1998.

[6] S. Toumpis and A. J. Goldsmith, "New media access protocols for wireless ad hoc networks based on cross-layer principles," *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, pp. 2228-2241, 2006.