Jointly Optimal Congestion and Contention Control Based on Network Utility Maximization

Jang-Won Lee, *Member, IEEE*, Mung Chiang, *Member, IEEE*, and A. Robert Calderbank, *Fellow, IEEE*

***Abstract*— We study joint end-to-end congestion control and per-link medium access control (MAC) in ad-hoc networks. We use a network utility maximization formulation, in which by adjusting the types of utility functions, we can accommodate multi-class services as well as exploit the tradeoff between efficiency and fairness of resource allocation. Despite the inherent difficulties of non-convexity and non-separability of the optimization problem, we show that, with readily-verifiable sufficient conditions, we can develop a distributed algorithm that converges to the globally and jointly optimal rate allocation and persistence probabilities.**

***Index Terms*— Congestion control, medium access control, ad-hoc wireless network, network utility maximization, optimization.**

I. INTRODUCTION AND RELATED WORK

SINCE the publication of the seminal paper [1], the network utility maximization (NUM) framework has found applications in network rate allocation through congestion control protocols. In the NUM framework, each end-user (or source) has its utility function and link bandwidths are allocated so that network utility (*i.e.*, the sum of all users' utilities) is maximized. A utility function can be interpreted as the level of satisfaction attained by a user as a function of resource allocation. Efficiency of resource allocation algorithms can thus be measured by the achieved network utility. Utility functions can also be interpreted as the 'knobs' to control the tradeoff between efficiency and fairness. Different shapes of utility functions lead to different types of fairness. For example, the following family of utility functions, parameterized by $\alpha \geq 0$, is proposed in [2]:

$$U^{\alpha}(x) = \begin{cases} (1-\alpha)^{-1}x^{1-\alpha}, & \text{if } \alpha \neq 1 \\ \log x, & \text{otherwise} \end{cases} . \quad (1)$$

If we set $\alpha = 0$, NUM reduces to system throughput maximization. If $\alpha = 1$, proportional fairness among competing users is attained; if $\alpha = 2$, then harmonic mean fairness; and if $\alpha \rightarrow \infty$, then max-min fairness. To accommodate multi-class services and attain the desired tradeoff between efficiency and fairness, it is important that the NUM framework can handle general types of concave utility functions.

Manuscript received September 10, 2005. The associate editor coordinating the review of this letter and approving it for publication was Prof. Iakovos Venieris. This work was supported by Yonsei University research fund of 2005 and NSF Grants CCF-0440443, CNS-0417607, CCF-0448012, and CNS-0427677.

J.-W. Lee is with CITY-Center for Information Technology of Yonsei University, Dept. of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea (e-mail: jangwon@yonsei.ac.kr).

M. Chiang and A. R. Calderbank are with the Dept. of Electrical Engineering, Princeton University, Princeton, NJ USA (e-mail: {chiangm, calderbk}@princeton.edu).

Digital Object Identifier 10.1109/LCOMM.2006.03021.

Motivated by unfairness in existing medium access control (MAC) protocols especially in conjunction with end-to-end congestion control protocols, the NUM framework has very recently been extended for the design of contention-based MAC protocols (*e.g.*, [3] and references therein) and joint congestion control and MAC protocols (*e.g.*, [4], [5]) in ad-hoc networks. In [4], a deterministic approximation approach is used. However, we have shown [3] that the deterministic approximation approach cannot accurately model random-access-based MAC, which are generally used for the MAC protocol in ad-hoc wireless networks. In [3], [5], a probabilistic model is used, which formulates the optimization problem directly in terms of persistence probabilities of each link and node. However, the main difficulty in the probabilistic approach is that in general, the problems formulated turn out to be non-convex and non-separable optimization, which is difficult to solve, especially in a distributed way. For example, only the restrictive case of log utility (*i.e.*, $\alpha = 1$ in (1) and the associated proportional fairness) has been studied [5] in the probabilistic approach for TCP/MAC joint design. In this letter, we provide the methodology to solve the joint congestion and medium access control problem optimally and distributively for general classes of concave utility functions.

II. SYSTEM MODEL AND NOTATION

We consider an ad hoc wireless network represented by a directed graph $G(V, E)$, where V is the set of nodes and E is the set of logical links. Each source node s has its utility function $U_s(x_s)$, which is a function of its end-to-end data rate x_s and uses a subset $L(s)$ of links for its communication. We define $S(l)$ as the subset of sources that are traversing link l. Let $L_{out}(n)$ denote the set of outgoing links from node n, $L_{in}(n)$ the set of incoming links to node n, t_l the transmitter node of link l, and r_l the receiver node of link l. Node n is said to transmit data if one of its outgoing links transmits data. Now define $N^I_{to}(l)$ as the set of nodes whose transmissions cause interference to the receiver of link l, excluding the transmitter node of link l, (*i.e.*, t_l), and $L^I_{from}(n)$ as the set of links whose transmissions get interfered from the transmission of node n, excluding outgoing links from node n (*i.e.*, $l \in L_{out}(n)$). Hence, if the transmitter of link l and a node in set $N^I_{to}(l)$ transmit data simultaneously, the transmission of link l fails. If node n and the transmitter of a link l in set $L^I_{from}(n)$ transmit data simultaneously, the transmission of link l also fails.

Each node and each link has a contention resolution protocol based on the transmission persistence probability. Each node n transmits data with a probability P^n. When it determines to transmit data, it chooses one of its outgoing links with

1089-7798/06$20.00 © 2006 IEEE

a probability q_l, $\forall l \in L_{out}(n)$, such that $\sum_{l \in L_{out}(n)} q_l = 1$, and transmits data only on the chosen link. Hence, there is no collision among links that have the same transmitter node. Consequently, link l, $l \in L_{out}(n)$, transmits data with a probability $p_l = P^n q_l$ such that $\sum_{l \in L_{out}(n)} p_l = P^n$, $\forall n$. We call q_l and p_l the conditional persistence probability and the persistence probability of link l, respectively. Based on a random access algorithm, which will be studied in next section, each link adjusts its persistence probability.

III. Joint Congestion and Contention Control

Consider the following utility maximization problem subject to constraints at both transport and MAC layers, with variables $\mathbf{x}$ (end-to-end rates controlled by TCP) and $\mathbf{p}$ (per link persistence probability controlled by contention-based MAC).

$$\begin{array}{ll} \max & \sum_s U_s(x_s) \\ \text{s. t.} & \sum_{s \in S(l)} x_s \le c_l p_l \prod_{k \in N_{to}^I(l)} (1 - \sum_{m \in L_{out}(k)} p_m), \forall l \\ & \sum_{m \in L_{out}(n)} p_m \le 1, \ \ \forall n \\ & x_s^{min} \le x_s \le x_s^{max}, \ \ \forall s \\ & 0 \le p_l \le 1, \ \ \forall l, \end{array} \quad (2)$$

where c_l is a fixed data rate of link l, and x_s^{min} and x_s^{max} are the minimum and maximum data rates of source s, respectively. The first constraint states that the traffic load on each link must be less than or equal to the average data rate that is offered by that link. The second constraint states that the persistence probability of each node n, P^n, must be less than or equal to one.

Due to the first constraint, problem (2) is in general a non-convex and non-separable optimization problem, which is difficult to solve, especially in a distributed way. For any solution algorithm, convexity is the key for its global optimality and separability for its distributed nature. Nonetheless, we will show with appropriate problem transformation and under readily-verifiable conditions, a distributed algorithm can be developed to solve for the globally optimal transport layer rates and MAC layer persistence probabilities.

To this end, we first take the log at both sides of the first constraint in problem (2) and a log change of variables and constants: $x'_s = \log x_s$, $x_s'^{max} = \log x_s^{max}$, $x_s'^{min} = \log x_s^{min}$, $U'_s(x'_s) = U_s(e^{x'_s})$, and $c'_l = \log c_l$. This reformulation turns the problem into:

$$\begin{array}{ll} \max & \sum_s U'_s(x'_s) \\ \text{s. t.} & \log(\sum_{s \in S(l)} e^{x'_s}) - c'_l - \log p_l \\ & - \sum_{k \in N_{to}^I(l)} \log(1 - \sum_{m \in L_{out}(k)} p_m) \le 0, \forall l \\ & \sum_{m \in L_{out}(n)} p_m \le 1, \ \ \forall n \\ & x_s'^{min} \le x'_s \le x_s'^{max}, \ \ \forall s \\ & 0 \le p_l \le 1, \ \ \forall l. \end{array} \quad (3)$$

It can be verified, *e.g.*, through the second derivative test, that $\log(\sum_{s \in S(l)} e^{x'_s})$ is a convex function in $\mathbf{x}'$. However, for (3) to be a convex optimization problem, we also need the concavity of the objective function, which may not be true for any $U'_s(x'_s)$ even when $U_s(x_s)$ is concave. Now consider

$$g_s(x_s) = \frac{d^2 U_s(x_s)}{dx_s^2} x_s + \frac{dU_s(x_s)}{dx_s}.$$

Then, we have the following lemma.

Lemma 1: [3] If $g_s(x_s) \le 0$, $U'_s(x'_s)$ is a concave function of x'_s.

For example, considering the utility functions (1) parameterized by α, we can easily show that if $\alpha \ge 1$, $U'_l(x'_s)$ is a concave function. Throughout this paper, we will assume that the condition in Lemma 1 is satisfied and thus problem (3) is now convex optimization. To solve problem (3), we use similar approaches as those used in [5]. Due to the page limit, in this letter we only consider a primal-based algorithm, *i.e.*, a penalty function approach, rather than the alternative of dual-based algorithm.

We first define $h_l(\mathbf{p}, \mathbf{x}') = \log(\sum_{s \in S(l)} e^{x'_s}) - c'_l - \log p_l - \sum_{k \in N_{to}^I(l)} \log(1 - \sum_{m \in L_{out}(k)} p_m)$ and $w_n(\mathbf{p}) = \sum_{m \in L_{out}(n)} p_m - 1$. Then, problem (3) can be rewritten as

$$\begin{array}{ll} \max & \sum_s U'_s(x'_s) \\ \text{s. t.} & h_l(\mathbf{p}, \mathbf{x}') \le 0, \ \ \forall l \\ & w_n(\mathbf{p}) \le 0, \ \ \forall n \\ & x_s'^{min} \le x'_s \le x_s'^{max}, \ \ \forall s \\ & 0 \le p_l \le 1, \ \ \forall l. \end{array} \quad (4)$$

Instead of solving problem (4) directly, we apply the penalty function method and consider the following problem:

$$\begin{array}{ll} \max & V(\mathbf{p}, \mathbf{x}') \\ \text{s. t.} & x_s'^{min} \le x'_s \le x_s'^{max}, \ \ \forall s \\ & 0 \le p_l \le 1, \ \ \forall l, \end{array} \quad (5)$$

where $V(\mathbf{p}, \mathbf{x}') = U'_s(x'_s) - \kappa \sum_l \max\{0, h_l(\mathbf{p}, \mathbf{x}')\} - \kappa \sum_n \max\{0, w_n(\mathbf{p})\}$ and κ is a positive constant.

Since the objective function of problem (5) is concave, problem (5) is convex optimization with simple, decoupled constraints, which can be solved by using a subgradient projection algorithm. We can easily show that

$$\frac{\partial V(\mathbf{p}, \mathbf{x}')}{\partial p_l} = \kappa \left(\frac{\epsilon_l}{p_l} - \frac{\sum_{k \in L_{from}^I(t_l)} \epsilon_k}{1 - \sum_{m \in L_{out}(t_l)} p_m} - \delta_{t_l} \right) \quad (6)$$

and

$$\frac{\partial V(\mathbf{p}, \mathbf{x}')}{\partial x'_s} = \frac{\partial U'_s(x'_s)}{\partial x'_s} - \kappa e^{x'_s} \sum_{l \in L(s)} \frac{\epsilon_l}{\sum_{k \in S(l)} e^{x'_k}}, \quad (7)$$

where

$$\epsilon_l = \begin{cases} 0, & \text{if } \sum_{n \in S(l)} e^{x'_n} \le c_l p_l \prod_{k \in N_{to}^I(l)} (1 - \sum_{m \in L_{out}(k)} p_m) \\ 1, & \text{otherwise} \end{cases}$$

and

$$\delta_n = \begin{cases} 0, & \text{if } \sum_{m \in L_{out}(n)} p_m \le 1 \\ 1, & \text{otherwise} \end{cases}.$$

Then, the subgradient projection algorithm that solves problem (5) is obtained as follows. On each logical link l, transmission is decided to take place with probability

$$p_l(t+1) = \left[p_l(t) + \beta(t) \frac{\partial V(\mathbf{p}, \mathbf{x}')}{\partial p_l} |_{\mathbf{p}=\mathbf{p}(t), \mathbf{x}'=\mathbf{x}'(t)} \right]_0^1, \quad (8)$$

and concurrently at each source s, end-to-end rate is adjusted:

$$x'_s(t+1) = \left[x'_s(t) + \beta(t) \frac{\partial V(\mathbf{p}, \mathbf{x}')}{\partial x'_s} |_{\mathbf{p}=\mathbf{p}(t), \mathbf{x}'=\mathbf{x}'(t)} \right]_{x_s'^{min}}^{x_s'^{max}}, \quad (9)$$

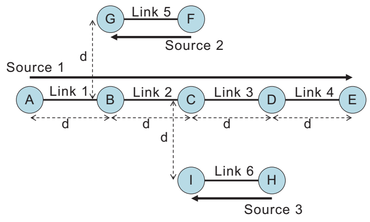

Fig. 1. Physical and logical topologies for a numerical example.

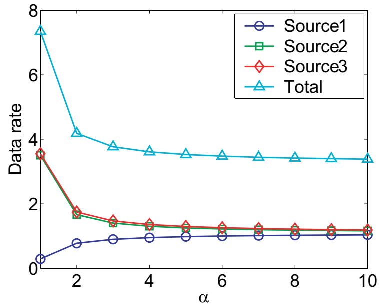

Fig. 2. Optimized source rates as fairness index α changes.

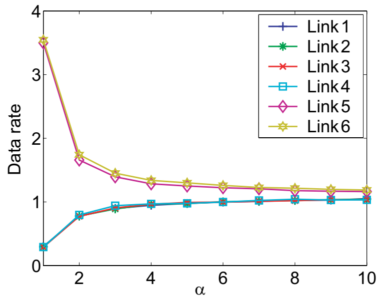

Fig. 3. Average link rates obtained by optimized persistence probabilities as fairness index α changes.

where $[a]_c^b = \max\{\min\{a, b\}, c\}$ and $\beta(t)$ is step-size.

The joint control algorithm in (8) and (9) can be implemented as follows. Each link l (or its transmission node t_l) updates its persistence probability $p_l(t)$ using (8), and concurrently, each source updates its data rate $x_s(t)$ using (9). To calculate the subgradient in (6), each link needs information only from link k, $k \in L^I_{from}(t_l)$, *i.e.*, from links whose transmissions are interfered from the transmission of link l, and those links are in the neighborhood of link l. To calculate the subgradient in (7), each source needs information only from link l, $l \in L(s)$, *i.e.*, from links on its routing path. Hence, to perform the algorithm, each source and link need only local information through limited message passing and the algorithm can be implemented in a distributed way. In particular, note that δ_n is calculated at the transmitter node of each link to update the persistence probability of that link, and does not need to be passed among the nodes. There is no need to explicitly pass around the values of persistence probabilities, since their effects are included in $\{\epsilon_l\}$. And quantities such as $\sum_{m \in L_{out}(t_l)} p_m$ and $\sum_{k \in S(l)} \exp(x_k^{'})$ can be measured locally by each node and each link.

To show convergence and optimality properties of the algorithm in (8) and (9), we first introduce a more compact notation. Let $\mathbf{q}(t) = (\mathbf{p}(t), \mathbf{x}'(t))$, where $\mathbf{p}(t)$ and $\mathbf{x}'(t)$ are obtained by solving (8) and (9), respectively, and $\mathbf{Q}^*$ be the set of the optimal solutions to problem (3). We further define $d(\mathbf{q}, \mathbf{Q}) = \min_{\mathbf{q}' \in \mathbf{Q}} ||\mathbf{q} - \mathbf{q}'||$ and $N_r(\mathbf{Q}) = \{\mathbf{q} \mid d(\mathbf{q}, \mathbf{Q}) \leq r\}$. We have the following results readily derived based on the convergence properties of a subgradient algorithm [5], [6]:

Theorem 1: If $\beta(t) \to 0$, as $t \to \infty$ and $\sum_{t=0}^{\infty} \beta(t) = \infty$, then there exists a $K < \infty$ such that for all $\kappa > K$,

$$\lim_{t \to \infty} d(\mathbf{q}(t), \mathbf{Q}^*) = 0.$$

Theorem 2: For all $r > 0$, there exist $K < \infty$ and $A > 0$ such that if $\beta(t) = \beta < A$ for all t and $\kappa > K$, then

$$\lim_{t \to \infty} d(\mathbf{q}(t), N_r(\mathbf{Q}^*)) = 0.$$

Theorem 1 states that with a diminishing step size, *e.g.*, $\beta(t) = 1/t$, the subgradient projection algorithm in (8) and (9) converges to the optimal solution to problem (3) (*i.e.*, problem (2)). However, in practice, the constant step size can more efficiently track system variations and more practical for implementation than a diminishing step size. Theorem 2 states that with a constant step size, the subgradient projection algorithm in (8) and (9) converges to the neighborhood of the optimal solution to problem (3).

An illustrative numerical example is summarized below for the network shown in Fig. 1. Each source has its utility function parameterized by α as in (1). We assume that if the distance between the transmitter node and the receiver node is less than $2d$, the receiver node gets interfered from the transmitter node. Figs. 2 and 3 show the *optimized* source rates and link rates, respectively, with each data point being the result of distributively solving (5) for a given fairness parameter α. Since source 1 traverses more heavily interfered links, at the optimal rate allocation that maximizes the network utility, it is allocated the lowest data rate. However, as the value of α increases, the gap among the sources decreases, improving fairness among sources. This improvement is achieved by sacrificing the network efficiency, since as the value of α increases, the total rate decreases. This illustrates that, by using different types of utility functions (*e.g.*, different values of α in (1)), we can control the tradeoff between efficiency and fairness of resource allocation.

IV. Conclusion

We formulate a joint congestion control and MAC problem in this letter by using the NUM framework. We propose the methodologies to tackle coupling and non-convexity of the optimization problem and develop a joint congestion control and MAC protocol for ad-hoc wireless networks, which provides a globally optimal solution in a distributed way and accommodates a wide variety of utility functions.

References

[1] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *J. Optical Research Society*, vol. 49, pp. 237–252, Mar. 1998.

[2] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. Networking*, vol. 8, pp. 556–567, Oct. 2000.

[3] J.-W. Lee, M. Chiang, and R. Calderbank, "Optimal MAC design based on network utility maximzation: reverse and forward engineering," in *Proc. IEEE INFOCOM*, 2006.

[4] L. Chen, S. Low, and J. Doyle, "Joint congestion control and media access control design for ad hoc wireless networks," in *Proc. IEEE INFOCOM 2005*, pp. 2212-2222.

[5] X. Wang and K. Kar, "Cross-layer rate control for end-to-end proportional fairness in wireless networks with random access," in *Proc. ACM Mobihoc 2005*, pp. 157-168.

[6] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*. Springer-Verlag, 1985.